Marginal estimation of aberrations and image restoration by phase diversity

Amandine BLANC

Office National d'Études et de Recherches Aérospatiales

Optics Department BP 72, F-92322 Châtillon cedex, France

and Laboratoire des Signaux et Systèmes

École Supérieure d'Électricité

Plateau de Moulon, 91192 Gif-sur-Yvette, France

Laurent M. MUGNIER

Office National d'Études et de Recherches Aérospatiales

Optics Department BP 72, F-92322 Châtillon cedex, France

Jérôme IDIER

Laboratoire des Signaux et Systèmes

École Supérieure d'Électricité

Plateau de Moulon, 91192 Gif-sur-Yvette, France.

We propose a novel method, called marginal estimator, for estimating the aberrations and the object from phase diversity data. The conventional estimator found in the literature concerning the technique first proposed by Gonsalves is based on a joint estimation of the aberrated phase and the observed object. By means of simulations, we study the behavior of this estimator which is interpretable as a joint Maximum A Posteriori approach and show, in particular, that it has bad asymptotic properties and does not allow an optimal joint estimation of the object and the aberrated phase. We propose a novel marginal estimator of the sole phase by Maximum A Posteriori. It is obtained by integrating the observed object out of the problem. This reduces drastically the number of unknowns, allows the unsupervised estimation of the regularization parameters and provides better asymptotic properties. We show that the marginal method is also appropriate for the restoration of the object. This estimator is implemented and its properties are validated by simulations. The performance of the joint method and the marginal one is compared on both simulated and experimental data in the case of Earth observation. For the studied object, the comparison of the quality of the phase restoration shows that the performance of the marginal approach is higher under high noise level conditions.

1. Introduction

The images recorded by a telescope are often degraded by aberrations. These aberrations can be due to atmospheric turbulence and to imperfections of the optical system. Whatever their origins, they lead to phase variations in the pupil plane, which severely reduce the optical transfer function. The problem is to correct the images from these distortions. Many wavefront sensing techniques have been proposed but few can be used with both a point source and an extended scene. One of them is phase diversity. This technique, first proposed by Gonsalves,[1] has been significantly developed during the last ten years. It has been used successfully by many authors to determine aberrations[2–4] and also to improve the quality of images, as in solar imaging.[5,6] This technique uses a low-cost, optically simple wavefront sensor but requires numerical processing to estimate the unknowns from the images. The conventional processing scheme found in the literature is based on the joint estimation of the aberrations and of the observed object. The reconstruction of these parameters is an ill-posed inverse problem and thus must be regularized. Several methods have been proposed to this end. Concerning the aberrations, implicit regularization is achieved by expanding the phase on a finite linear combination of basis functions. Additionally, in the case of imaging through turbulence, a statistical prior on the turbulent phase is available according to Kolmogorov model.[7] In order to regularize the object, various approaches have been proposed: some authors use a low-pass filter[3,5] , some impose a sieve on the object[6] and more recently a Tikhonov regularization on the object was proposed.[8] All the regularization methods concerning the object require the tuning of regularization parameters. In a joint estimation, these parameters must be adjusted by hand. Fur-

thermore, in the joint restoration framework, the ratio of the number of unknowns (aberrations + object) to the number of data does not tend towards zero when the size of the data set tends to infinity. As a consequence, the joint method does not have good asymptotic properties. The aim of this article is to solve these problems by proposing a robust estimator derived in a Bayesian framework that restores the sole aberrations. It is obtained by integrating the observed object out of the problem. This novel method is called marginal estimator and is based on a Maximum A Posteriori approach. This estimator has good asymptotic properties and allows the unsupervised estimation of the noise variance and of the regularization parameters of the object. In addition, the marginal estimator also provides a simple way to estimate the object once the aberrations and the regularization parameters are estimated: the restoration is given by a Wiener filter as with the joint estimator. The two methods differ in their tuning of the regularization parameters and in the choice of aberration estimators. The outline of the paper is as follows: we begin with the mathematical formulation of the problem. In Section 3 we recall the usual joint estimation scheme and we present the marginal estimator. In Section 4, we present the comparison of the two estimators; we compare, by means of simulations, their asymptotic properties, the influence of the hyperparameters on the quality of the restoration and finally their performances on both simulated and experimental data. Section 5 summarizes the results presented herein.

2. Phase diversity principle and imaging model

The idea of phase diversity is to collect a first image in the focal plane and one (or more) additional image(s) which differ(s) from the focused one by a known phase variation. A simple implementation of this technique is to take the second image in a defocused plane (with a known defocus distance). Figure 1 shows a simplified diagram of the phase diversity setup.

In the isoplanatic patch of the telescope, the image is the noisy sampled convolution of the point spread function h in the observation plane with the object o:

$$i(\boldsymbol{r}) = (h * o)(\boldsymbol{r}) + n(\boldsymbol{r}) \tag{1}$$

where $\boldsymbol{r}$ is a two-dimensional vector in the image plane and n is an additive noise. Under the near field approximation, the point spread function associated with the focused image is given by:

$$h_1(\boldsymbol{r}) = |FT^{-1}\{P(\boldsymbol{u}).\exp[j\phi(\boldsymbol{u})]\}|^2 \tag{2}$$

where $\boldsymbol{u}$ is a two-dimensional vector in the pupil plane, ϕ is the unknown aberrated phase function, P is the binary aperture function and FT^{-1} denotes the inverse Fourier transform. In the defocused plane:

$$h_2(\boldsymbol{r}) = |FT^{-1}\{P(\boldsymbol{u}).\exp[j(\phi(\boldsymbol{u}) + \phi_d(\boldsymbol{u}))]\}|^2 \tag{3}$$

where ϕ_d is the known diversity phase function.

In this paper, the aberrated phase function is expanded on a finite set of Zernike polynomials:[9]

$$\phi(\boldsymbol{u}) = \sum_{i=4}^{k} a_i Z_i(\boldsymbol{u}) \tag{4}$$

5

Note that the coefficients a_{1-3} have not been introduced: the piston coefficient a_1 is a constant added to the phase and has no influence on the point spread function, and the tilt coefficients a_2, a_3 introduce a shift in the image that is of no importance for extended objects. In the following, we will note $\boldsymbol{a} = (a_4, ..., a_k)^t$, the $\{k - 3\}$-dimensional vector which contains the aberration coefficients.

In practice, data are discrete arrays because of the spatial sampling of the images and Equation 1 takes the form:

$$\boldsymbol{i} = H\boldsymbol{o} + \boldsymbol{n} \qquad (5)$$

where H is the Toeplitz-block-Toeplitz matrix that corresponds to the convolution by $\boldsymbol{h}$, and $\boldsymbol{i}$, $\boldsymbol{o}$, $\boldsymbol{h}$ and $\boldsymbol{n}$ are the discrete forms of the previous variables.

The problem is to estimate the unknown parameters (the object $\boldsymbol{o}$ and the aberrations $\boldsymbol{a}$) from the data (focused $\boldsymbol{i}_1$ and defocused $\boldsymbol{i}_2$ images) and the defocus distance. We begin by recalling the conventional method based on the joint estimation of the unknowns.

3. Theory

A. Joint estimator

1. Criterion

The classical method is based on joint estimation of the object and the aberrations. The Bayesian interpretation of such an approach consists in computing the Joint Maximum A Posteriori (JMAP) estimator:

$$(\hat{\boldsymbol{o}}, \hat{\boldsymbol{a}})_{\text{JMAP}} = \arg\max_{\boldsymbol{o},\boldsymbol{a}} f(\boldsymbol{i}_1, \boldsymbol{i}_2, \boldsymbol{o}, \boldsymbol{a}; \boldsymbol{\theta})$$

$$= \arg\max_{\boldsymbol{o},\boldsymbol{a}} f(\boldsymbol{i}_1|\boldsymbol{o}, \boldsymbol{a}; \boldsymbol{\theta}) f(\boldsymbol{i}_2|\boldsymbol{o}, \boldsymbol{a}; \boldsymbol{\theta}) f(\boldsymbol{o}; \boldsymbol{\theta}) f(\boldsymbol{a}; \boldsymbol{\theta}) \qquad (6)$$

where $f(\boldsymbol{i}_1, \boldsymbol{i}_2, \boldsymbol{o}, \boldsymbol{a}; \boldsymbol{\theta})$ is the joint probability density function of the data $(\boldsymbol{i}_1, \boldsymbol{i}_2)$, of the object $\boldsymbol{o}$ and of the aberrations $\boldsymbol{a}$. It may also depend on a set of hyperparameters or regularization parameters $\boldsymbol{\theta}$. $f(\boldsymbol{i}_1|\boldsymbol{o}, \boldsymbol{a}; \boldsymbol{\theta})$ and $f(\boldsymbol{i}_2|\boldsymbol{o}, \boldsymbol{a}; \boldsymbol{\theta})$ denote the likelihood of the data $\boldsymbol{i}_1$ and $\boldsymbol{i}_2$, $f(\boldsymbol{o}; \boldsymbol{\theta})$ and $f(\boldsymbol{a}; \boldsymbol{\theta})$ are the *a priori* probability density function of $\boldsymbol{o}$ and $\boldsymbol{a}$. We assume that the noise is stationary white Gaussian with a variance σ^2 (the same for the two images). We choose a Gaussian prior probability distribution for the object with a mean $\boldsymbol{o}_m$ and a covariance matrix R_o, and also a Gaussian distribution for the aberrations, with a null mean and a covariance matrix R_a. Under these assumptions, the set of the hyperparameters is $\boldsymbol{\theta} = (\boldsymbol{\theta_n}, \boldsymbol{\theta_o}, \boldsymbol{\theta_a})$ where $\boldsymbol{\theta_n} = \sigma^2$, $\boldsymbol{\theta_o} = (R_o, \boldsymbol{o}_m)$ and $\boldsymbol{\theta_n} = R_a$. Hence, we have

$$
\begin{aligned}
f(\boldsymbol{i}_1, \boldsymbol{i}_2, \boldsymbol{o}, \boldsymbol{a}; \boldsymbol{\theta}) &= \frac{1}{(2\pi)^{N^2/2}\sigma^{N^2}} \exp\left[-\frac{1}{2\sigma^2}(\boldsymbol{i}_1 - H_1\boldsymbol{o})^t(\boldsymbol{i}_1 - H_1\boldsymbol{o})\right] \\
&\times \frac{1}{(2\pi)^{N^2/2}\sigma^{N^2}} \exp\left[-\frac{1}{2\sigma^2}(\boldsymbol{i}_2 - H_2\boldsymbol{o})^t(\boldsymbol{i}_2 - H_2\boldsymbol{o})\right] \\
&\times \frac{1}{(2\pi)^{N^2/2}\det(R_o)^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{o} - \boldsymbol{o}_m)^t R_o^{-1}(\boldsymbol{o} - \boldsymbol{o}_m)\right] \\
&\times \frac{1}{(2\pi)^{(k-3)/2}\det(R_a)^{1/2}} \exp\left(-\frac{1}{2}\boldsymbol{a}^t R_a^{-1}\boldsymbol{a}\right) \qquad (7)
\end{aligned}
$$

where $\det(x)$ denotes the determinant of x, N^2 is the number of pixels in the image. The criterion to minimize is then

$$L_{\text{JMAP}}(\boldsymbol{o}, \boldsymbol{a}, \boldsymbol{\theta}) = -\ln f(\boldsymbol{i}_1, \boldsymbol{i}_2, \boldsymbol{o}, \boldsymbol{a}; \boldsymbol{\theta})$$

$$= N^2 \ln \sigma^2 + \frac{1}{2} \ln \det(R_o) + \frac{1}{2} \ln \det(R_a)$$

$$+ \frac{1}{2\sigma^2}(\boldsymbol{i}_1 - H_1\boldsymbol{o})^t(\boldsymbol{i}_1 - H_1\boldsymbol{o}) + \frac{1}{2\sigma^2}(\boldsymbol{i}_2 - H_2\boldsymbol{o})^t(\boldsymbol{i}_2 - H_2\boldsymbol{o})$$

$$+ \frac{1}{2}(\boldsymbol{o} - \boldsymbol{o}_m)^t R_o^{-1}(\boldsymbol{o} - \boldsymbol{o}_m) + \frac{1}{2}\boldsymbol{a}^t R_a^{-1}\boldsymbol{a} + A \qquad (8)$$

where A is a constant. Canceling its derivative with respect to the object gives[1,10] a closed-form expression for the object $\hat{\boldsymbol{o}}(\boldsymbol{a}, \boldsymbol{\theta})$ that minimizes the criterion for given $(\boldsymbol{a}, \boldsymbol{\theta})$:

$$\hat{\boldsymbol{o}}(\boldsymbol{a}, \boldsymbol{\theta}) = (H_1^t H_1 + H_2^t H_2 + \sigma^2 R_o^{-1})^{-1}(H_1^t \boldsymbol{i}_1 + H_2^t \boldsymbol{i}_2 + \sigma^2 R_o^{-1}\boldsymbol{o}_m) \qquad (9)$$

Furthermore, the criterion $L_{\text{JMAP}}(\boldsymbol{o}, \boldsymbol{a}, \boldsymbol{\theta})$ and thus the closed-form expression $\hat{\boldsymbol{o}}(\boldsymbol{a}, \boldsymbol{\theta})$ can be written in the discrete Fourier domain with a circulant approximation (see Appendix A):

$$L_{\text{JMAP}}(\boldsymbol{o}, \boldsymbol{a}, \boldsymbol{\theta}) = N^2 \ln \sigma^2 + \frac{1}{2} \sum_v \ln S_o(v) + \frac{1}{2} \ln \det(R_a)$$

$$+ \sum_v \frac{1}{2\sigma^2}|\tilde{i}_1(v) - \tilde{h}_1(\boldsymbol{a}, v)\tilde{o}(v)|^2 + \sum_v \frac{1}{2\sigma^2}|\tilde{i}_2(v) - \tilde{h}_2(\boldsymbol{a}, v)\tilde{o}(v)|^2$$

$$+ \sum_v \frac{|\tilde{o}(v) - \tilde{o}_m(v)|^2}{2S_o(v)} + \frac{1}{2}\boldsymbol{a}^t R_a^{-1}\boldsymbol{a} + A \qquad (10)$$

$$\text{and} \quad \hat{\tilde{o}}(\boldsymbol{a}, \boldsymbol{\theta}, v) = \frac{\tilde{h}_1^*(\boldsymbol{a}, v)\tilde{i}_1 + \tilde{h}_2^*(\boldsymbol{a}, v)\tilde{i}_2 + \frac{\sigma^2 \tilde{o}_m(v)}{S_o(v)}}{|\tilde{h}_1(\boldsymbol{a}, v)|^2 + |\tilde{h}_2(\boldsymbol{a}, v)|^2 + \frac{\sigma^2}{S_o(v)}}. \qquad (11)$$

where $\tilde{x}$ denotes the two-dimensional Fourier transform of x, v is the spatial frequency and S_o is the power spectral density of the object. Substituting $\hat{o}(\boldsymbol{a}, \boldsymbol{\theta})$ into the criterion yields a "new" criterion that does not explicitly depend on the object:

$$L'_{\mathrm{JMAP}}(\boldsymbol{a}, \boldsymbol{\theta}) = L_{\mathrm{JMAP}}(\hat{o}(\boldsymbol{a}, \boldsymbol{\theta}), \boldsymbol{a}, \boldsymbol{\theta})$$

$$= N^2 \ln \sigma^2 + \frac{1}{2}\sum_v \ln S_o(v) + \frac{1}{2}\sum_v \frac{\left|\tilde{i}_1(v)\tilde{h}_2(\boldsymbol{a}, v) - \tilde{i}_2(v)\tilde{h}_1(\boldsymbol{a}, v)\right|^2}{\sigma^2 \left(|\tilde{h}_1(\boldsymbol{a}, v)|^2 + |\tilde{h}_2(\boldsymbol{a}, v)|^2 + \frac{\sigma^2}{S_o(v)}\right)}$$

$$+ \frac{1}{2}\sum_v \frac{\left|\tilde{h}_1(\boldsymbol{a}, v)\tilde{o}_m(v) - \tilde{i}_1(v)\right|^2 + \left|\tilde{h}_2(\boldsymbol{a}, v)\tilde{o}_m(v) - \tilde{i}_2(v)\right|^2}{S_o(v)\left(|\tilde{h}_1(\boldsymbol{a}, v)|^2 + |\tilde{h}_2(\boldsymbol{a}, v)|^2 + \frac{\sigma^2}{S_o(v)}\right)}$$

$$+ \frac{1}{2}\ln\det(R_a) + \frac{1}{2}\boldsymbol{a}^t R_a^{-1}\boldsymbol{a} + A \tag{12}$$

2. A priori information

The *a priori* information on the aberrations $\boldsymbol{a}$ is the covariance matrix R_a. In the case of aberrations induced by turbulence, R_a is deduced from Kolmogorov statistics.[9] If the aberrations are due to the imperfections of the system (misalignments, thermal effects, etc.) that are essentially low frequency, a few Zernike coefficients (typically the first twenty) are usually enough to describe all the aberrations. Notice that for high frequency imperfections, such as polishing errors on optics, this small number of coefficients is not enough. In the case of low frequency imperfections, regularization due to the truncated expansion of the phase is sufficient and the additional penalization term $\ln\det(R_a) + \frac{1}{2}\boldsymbol{a}^t R_a^{-1}\boldsymbol{a}$ can be omitted.

The *a priori* information required on the object consists of the choice of the object power spectral density model. We choose the following one:

$$S_o(v) \triangleq E[|\tilde{o}(v) - \tilde{o}_m(v)|^2] = k/[v_o^p + v^p] - |\tilde{o}_m(v)|^2 \tag{13}$$

where E stands for the mathematical expectation. This heuristic model and similar ones have been quite widely used.[11] Notice that this model introduces four hyperparameters $\boldsymbol{\theta}_o = (k, v_o, p, o_m)$. The tuning of the hyperparameters of the object $\boldsymbol{\theta}_o$ is not easy: they depend on the structure of the object. Thus, they must be estimated for each object. Unfortunately, with a joint method, the hyperparameters can not be jointly estimated with $\boldsymbol{o}$ and $\boldsymbol{a}$. Indeed, the criterion (Equation 8) degenerates when, for example, one seeks $\boldsymbol{\theta}_o$ together with $\boldsymbol{o}$ and $\boldsymbol{a}$; in particular for the pair $\{\hat{\boldsymbol{\theta}}_o = (k = 0, v_o, p, o_m), \hat{o} = o_m\}$, which does not depend on the data, the criterion tends to minus infinity. So before minimizing L_{JMAP}, these hyperparameters must be chosen empirically by the user. Moreover, concerning asymptotic behavior (i.e., when the observed images incorporate a growing number of pixels), such joint estimations have not good asymptotic properties, as empirically checked in Subsection B. In a very general setting, Little and Rubin[12] provide additional insight to understand the behavior of joint estimators. This leads us to propose the marginal estimation approach.

B.   Marginal estimation

In this method, the phase and the hyperparameters linked to the noise and the object $(\boldsymbol{\theta}_n, \boldsymbol{\theta}_o)$ are first estimated. Then if the parameter of interest is the object, it is restored, in a second step, by Wiener filtering using the previous aberrations and hyperparameters estimates. Thus the object is restored the same way as with the joint method except that, in the marginal method, the aberrations used are estimated by the minimization of a different criterion and that the tuning of the hyperparameters

can be done automatically, as described below.

1. Criterion

The marginal estimator restores the sole aberrations by integrating the object out of the problem (in the vocabulary of probabilities, to integrate out (i.e. to marginalize) a quantity means to compute a marginal probability law by summing over all possible values of the quantity). It is a Maximum A Posteriori estimator for $\boldsymbol{a}$, obtained by integrating the joint probability density function.

$$\hat{a}_{\text{MAP}} = \arg\max_{\boldsymbol{a}} f(\boldsymbol{i}_1, \boldsymbol{i}_2, \boldsymbol{a}; \boldsymbol{\theta}) = \arg\max_{\boldsymbol{a}} \int f(\boldsymbol{i}_1, \boldsymbol{i}_2, \boldsymbol{o}, \boldsymbol{a}; \boldsymbol{\theta}) d\boldsymbol{o} \tag{14}$$

$$= \arg\max_{\boldsymbol{a}} \int f(\boldsymbol{i}_1|\boldsymbol{a}, \boldsymbol{o}; \boldsymbol{\theta}) f(\boldsymbol{i}_2|\boldsymbol{a}, \boldsymbol{o}; \boldsymbol{\theta}) f(\boldsymbol{a}; \boldsymbol{\theta}) f(\boldsymbol{o}; \boldsymbol{\theta}) d\boldsymbol{o}$$

Let $\boldsymbol{I} = (\boldsymbol{i}_1 \ \boldsymbol{i}_2)^T$ denotes the vector which concatenates the data. As a linear combination of jointly Gaussian variables ($\boldsymbol{o}$ and $\boldsymbol{n}$), $\boldsymbol{I}$ is a Gaussian vector. Maximizing $f(\boldsymbol{i}_1, \boldsymbol{i}_2, \boldsymbol{a}; \boldsymbol{\theta}) = f(\boldsymbol{I}, \boldsymbol{a}; \boldsymbol{\theta})$ is thus equivalent to minimizing the following criterion:

$$L_{\text{MAP}}(\boldsymbol{a}, \boldsymbol{\theta}) = \frac{1}{2}\ln\det(R_I) + \frac{1}{2}(\boldsymbol{I} - \boldsymbol{m}_I)^T R_I^{-1}(\boldsymbol{I} - \boldsymbol{m}_I)$$

$$+ \frac{1}{2}\ln\det(R_a) + \frac{1}{2}\boldsymbol{a}^t R_a^{-1}\boldsymbol{a} + B \tag{15}$$

where B is a constant, $\boldsymbol{m}_I = (H_1\boldsymbol{o}_m \ H_2\boldsymbol{o}_m)^T$ and R_I is the covariance matrix of $\boldsymbol{I}$, by definition $R_I \triangleq E[\boldsymbol{I}\boldsymbol{I}^t] - E[\boldsymbol{I}]E[\boldsymbol{I}]^t$. The latter has the following expression :

$$R_I = \begin{bmatrix} H_1 R_o H_1^t + \sigma^2 I_d & H_1 R_o H_2^t \\ H_2 R_o H_1^t & H_2 R_o H_2^t + \sigma^2 I_d \end{bmatrix} \tag{16}$$

where I_d is the identity matrix.

2. Relationship between the joint and the marginal criteria

The calculation of $(\boldsymbol{I} - \boldsymbol{m}_I)^T R_I^{-1}(\boldsymbol{I} - \boldsymbol{m}_I)$ (see Appendix B) leads to the following expression for L_{MAP}:

$$L_{\text{MAP}}(\boldsymbol{a}, \boldsymbol{\theta}) = \frac{1}{2} \ln \det(R_I) + \frac{1}{2} \sum_v \frac{\left| \tilde{\imath}_1(v)\tilde{h}_2(\boldsymbol{a}, v) - \tilde{\imath}_2(v)\tilde{h}_1(\boldsymbol{a}, v) \right|^2}{\sigma^2 \left(|\tilde{h}_1(\boldsymbol{a}, v)|^2 + |\tilde{h}_2(\boldsymbol{a}, v)|^2 + \frac{\sigma^2}{S_o(v)} \right)}$$

$$+ \frac{1}{2} \sum_v \frac{\left| \tilde{h}_1(\boldsymbol{a}, v)\tilde{o}_m(v) - \tilde{\imath}_1(v) \right|^2 + \left| \tilde{h}_2(\boldsymbol{a}, v)\tilde{o}_m(v) - \tilde{\imath}_2(v) \right|^2}{S_o(v) \left(|\tilde{h}_1(\boldsymbol{a}, v)|^2 + |\tilde{h}_2(\boldsymbol{a}, v)|^2 + \frac{\sigma^2}{S_o(v)} \right)}$$

$$+ \frac{1}{2} \ln \det(R_a) + \frac{1}{2}\boldsymbol{a}^t R_a^{-1}\boldsymbol{a} + B \tag{17}$$

The comparison of the expression of the criterion L_{MAP} (see Equation 17) and of the criterion L'_{JMAP} (see Equation 12) shows that the two criteria are related by the following relationship:

$$L_{\text{MAP}}(\boldsymbol{a}, \boldsymbol{\theta}) = \frac{1}{2} \ln \det(R_I) - N^2 \ln \sigma^2 - \frac{1}{2} \sum_v \ln S_o(v) + L'_{\text{JMAP}}(\boldsymbol{a}, \boldsymbol{\theta}) + C \tag{18}$$

where C is a constant. If we focus only on the terms depending on the phase (i.e. suppose that the hyperparameters are known), relationship 18 can be summarized:[13]

$$L_{\text{MAP}}(\boldsymbol{a}) = \frac{1}{2} \ln \det(R_I) + L'_{\text{JMAP}}(\boldsymbol{a}) \tag{19}$$

Thus, the difference between the marginal and the joint estimator consists of a single additional term dependent on the phase, which in the discrete Fourier domain is given by (see Appendix C):

$$\ln \det(R_I) = \sum_v \ln S_o(v) + N^2 \ln \sigma^2 + \sum_v \ln \left(|\tilde{h}_1(\boldsymbol{a}, v)|^2 + |\tilde{h}_2(\boldsymbol{a}, v)|^2 + \frac{\sigma^2}{S_o(v)} \right) \tag{20}$$

Although the two estimators differ only by one term, we shall see in Subsection 4.B that their properties differ considerably. Before that, let us see how the hyperparameters can be estimated.

3. Estimation of the hyperparameters

Both estimators call for the choice of the values of the hyperparameters $\boldsymbol{\theta}$. For the marginal estimator, the estimation of $\boldsymbol{\theta}_o = (k, v_o, p, o_m)$ and $\boldsymbol{\theta}_n = \sigma^2$ can be tackled jointly with the aberrations, according to

$$(\hat{\boldsymbol{a}}, \hat{\boldsymbol{\theta}}_o, \hat{\boldsymbol{\theta}}_n) = \underset{\boldsymbol{a}, \boldsymbol{\theta}_o, \boldsymbol{\theta}_n}{\arg \max} \, f(\boldsymbol{i}_1, \boldsymbol{i}_2, \boldsymbol{a}; \boldsymbol{\theta}) \tag{21}$$

The criterion $L_{\mathrm{MAP}}(\boldsymbol{a})$ (Equation 15) becomes $L_{\mathrm{MAP}}(\boldsymbol{a}, \sigma^2, k, v_o, p, o_m, \boldsymbol{\theta}_a)$. It must be minimized with respect to the aberrations $\boldsymbol{a}$ and the five hyperparameters (σ^2, k, v_o, p, o_m). If we adopt the change of variable $\mu = \sigma^2/k$, the cancellation of the derivative of the criterion with respect to k gives a closed-form expression $\hat{k}(\boldsymbol{a}, \mu, v_o, p, o_m, \boldsymbol{\theta}_a)$ which minimizes the criterion for given values of the other parameters. Injecting $\hat{k}$ into L_{MAP} yields $L_{\mathrm{MAP}}(\boldsymbol{a}, \mu, v_o, p, o_m, \boldsymbol{\theta}_a)$. There is no closed-form expression for $\hat{\mu}$, $\hat{v}_o$, $\hat{p}$ and $\hat{o}_m$ but it is easy to calculate the analytical expression of the gradients of the criterion with respect to these hyperparameters and then to use gradient-based methods for the minimization of the criterion.

4. Restoration of the object

When phase diversity is used as a wavefront sensor, estimation of the aberrations is the unique goal. On the contrary, in the case of image restoration, the object is the parameter of interest. In the case of the joint method, the object is estimated jointly with the aberrations, but in the case of the marginal estimator, only the aberrations and the hyperparameters are directly provided by the minimization of the criterion. However, the marginal method provides a simple way to restore the object. The idea is to calculate $\hat{\boldsymbol{o}}$ once the aberrations and the hyperparameters are estimated,

as $\hat{\boldsymbol{o}}_{\mathrm{MAP}} = \arg\max_{\boldsymbol{o}} f(\boldsymbol{i}_1, \boldsymbol{i}_2, \hat{\boldsymbol{a}}_{\mathrm{MAP}}; \hat{\boldsymbol{\theta}}_{\mathrm{MAP}})$. The object $\hat{\boldsymbol{o}}_{\mathrm{MAP}}$ is the bi-frame Wiener filter associated with the aberration and hyperparameter estimates $\hat{\boldsymbol{a}}_{\mathrm{MAP}}$ and $\hat{\boldsymbol{\theta}}_{\mathrm{MAP}}$. Additionally, although formally the restoration of the object is done in a second step, in practice, its restoration is in fact performed at each computation of the marginal criterion (see Equations 18 and 11), as in the joint method, so it is available at convergence.

4.   Comparison

In this section, we compare the two estimators by means of simulations and experimental data.

A.   Image simulation

The simulations have been obtained in the following way: our object is an Earth view. The aberrations are due to the imperfections of the optical system. The phase is a linear combination of the first 21 Zernike polynomials, with coefficients listed in Table 1; the estimated phase will be expanded on the same polynomials. The defocused amplitude for the second observation plane is 2π radians, peak-to-valley. The simulated images are monochromatic and are sampled at the Shannon rate. They have been obtained by convolution between the point spread function and the object, computed in the Fourier domain using FFT. The result is corrupted by a stationary white Gaussian noise (see Figure 2). The images generated in this way are periodic. This is an artificial situation, under which R_o is truly circulant block circulant, more simple than the real cases.

B. Asymptotic properties of the two estimators for known hyperparameters

We first begin with a comparison of the asymptotic properties of the two estimators on the ground of simulations. For the time being, we consider that the hyperparameters are the "true" ones, i.e. we fit the power spectral density of the object using the true object, and we assume that σ^2 is known (note that the mean objet is set to zero). On the other hand, as in all the following, no regularization on the aberrations is introduced, save the fact that only the Zernike coefficients a_4 to a_{21} are estimated. Consequently, the marginal estimation which was based on a Maximum A Posteriori approach now corresponds to Maximum Likelihood estimation and the joint approach corresponds to Generalized Maximum Likelihood estimation (GML). The minimized criteria will then be denoted L_{ML} and L_{GML}, respectively. Figure 3 shows the root mean square error (RMSE) on the phase as a function of the noise level for three image sizes (128×128, 64×64 and 32×32 pixels). Results are shown for the joint method (left diagram) and for the marginal one (right diagram). The RMSE on the phase is defined as $\left[\sum_{i=4}^{21}\langle(\hat{a}_i - a_i^{true})^2\rangle\right]^{1/2}$, the empirical average is done on 50 different noise realizations. Furthermore, for the image sizes of 32×32 and 64×64 pixels, the RMSE is averaged on all the sub-images of 32×32 pixels (respectively 64×64) contained in the image of 128×128 pixels. For joint estimation, the RMSE does not vary significantly when the number of data increases. This pathological behavior meets several statistical studies:[12,14] bias and variance are not expected to asymptotically vanish, i.e. the estimate does not converge towards the true value as the size of the data set tends to infinity. An intuitive explanation of this phenomenon is that, if a larger image is used to estimate the aberrations, the

size of the object, which is jointly reconstructed, increases also, so that the ratio of the number of unknowns to the number of data does not tend towards zero. On the contrary, for marginal estimation, the ratio of unknowns to data tends towards zero because the number of unknowns stays the same whatever the size of the data set. In this case, the RMSE of the phase decreases when the number of data increases (see Figure 3 right). Indeed, under broad conditions, the marginal estimator is expected to converge, since it is a true maximum likelihood estimator.[15,16]

Furthermore, we have empirically checked that the joint criterion presents local minima whatever the size of the dataset, whereas the marginal criterion tends to be asymptotically more and more regular. The latter observation is in agreement with the asymptotic Gaussianity of the likelihood, which is expected under suitable statistical conditions.

C. Influence of the hyperparameters

An important problem for the estimation of the object and the aberrations is the tuning of the hyperparameters. For the joint estimator, we have pointed out that they must be adjusted by hand. Particularly important is the global hyperparameter μ (see Subsection 3) i.e. the one that quantifies the trade-off between goodness of fit to the data and to the prior. First, let us study its influence on the joint method. Figure 4 shows the RMSE on the phase estimates and on the object estimate as a function of the value of this hyperparameter (its true value is $\mu = 1$). The RMSE on the object is defined as $\left[\sum_{\boldsymbol{r}} \langle (\hat{o}(\boldsymbol{r}) - o^{true}(\boldsymbol{r}))^2 \rangle \right]^{1/2} / \left[\sum_{\boldsymbol{r}} \hat{o}(\boldsymbol{r})^2 \right]^{1/2}$. We see that the best value of this hyperparameter (i.e. the one that gives the lower error on

the estimate) is not the same for the object and for the phase. It means that the object and the phase can not be jointly optimally restored. Note that the optimal hyperparameter value for the object coincides with the true value $\mu = 1$. If the parameter of interest is the phase, the object must be under-regularized to have a better estimation of the aberrations. The behavior of RMSE on the phase strongly depends on the noise level: for a high noise level (14% here), there is an optimal basin but for lower noise levels (4% and below), any value under 1, including a near null regularization ("near null regularization" means that the parameter μ is not equal to zero but to a small arbitrary constant (10^{-16} in our case) in order to avoid numerical problems due to computer precision), is almost optimal with respect to estimation of the aberrations even though the jointly estimated object becomes of the poorest quality. This observation sheds some light on the fact that, when the parameters of interest are the aberrations and when the noise level is low, estimation without object regularization can be successfully used as the literature testifies it.[7,17–19] This empirical observation has also led us to study the asymptotic behavior of the joint estimator with near null regularization. Figure 5 shows the results. In this case, when the number of data increases, the RMSE on the aberrations estimates decreases. Although, the number of data samples over the number of unknowns is the same as that of the estimation with the true hyperparameters, the estimator behaves as if the object were not being estimated. From a statistical viewpoint, the surprising behavior of joint aberrations estimates when the regularization parameter μ vanishes remains to be understood.

Finally, we have analyzed the influence of the global hyperparameter on the

RMSE of marginal estimates (Figure 6) and we can see that there is a unique optimal hyperparameter both for the object and the aberrations (we have drawn the curves only for a noise level of 4% because the general behavior of this estimator is the same for all the noise levels).

D. Unsupervised estimation

We have shown that the marginal estimator has a better coherence than the joint one: the optimal hyperparameters are the same for the aberrations and for the object. We still have to show that the unsupervised estimation of the hyperparameters (i.e. when the hyperparameters are estimated jointly with the aberrations) gives good aberration estimates. To this end, we compare the quality of the aberrations reconstruction obtained either by minimizing $L_{\mathrm{ML}}(\boldsymbol{a})$ with the true hyperparameters (which are the best hyperparameters for the restoration of the aberrations, see Figure 6) or by minimizing $L_{\mathrm{ML}}(\boldsymbol{a}, \mu, v_o, p)$, for several image sizes. From Figure 7, we see that for low noise levels, the unsupervised restoration is very good (the maximum difference is less than 5%). For 128×128 pixels, it is quite good (the maximum difference is less than 15%) for any noise level. Only for 32×32 pixels and high noise levels, the reconstruction is seriously degraded because of the lack of information contained in the noisy data.

E. Performance comparison

1. On simulated data

First, we present the performance comparison of the joint and the marginal estimators, on simulated data. In order to compare these estimators in a realistic way,

18

we use the joint estimator with a near null regularization (which gives good results for the estimation of the aberrations as seen in Subsection C) and the unsupervised marginal estimator, described in Subsection 3. First, let us see their performance on the restoration of the phase. We compared the RMSE of the phase estimates as a function of noise level for two image sizes (32×32 pixels and 128×128 pixels). The results for the two estimators are plotted in Figure 8. We can see two different domains: when the signal to noise ratio (SNR) is high (noise$< 5\%$), the two estimators approximately give the same results. At lower SNR ($5\% <$noise$< 20\%$), marginal estimation is significantly better. Concerning the restoration of the object, Figure 9 shows the results obtained by the two methods (left figure is for the joint estimator, right for the marginal one) for a noise level of 4% and an image size of 128×128 pixels. As we see, when there is no regularization on the object, at the minimum of the joint criterion, the object estimate is completely burial in noise. A solution to avoid this very bad restoration is to stop arbitrarily the iterative minimization before the end (which is an ad-hoc regularization[7,20]). We show this object estimate to illustrate the surprising behavior of the joint estimator which is able to restore the aberrations well even if the jointly estimated object is very degraded.

2. On experimental data

Finally we compare, on experimental data, both phase diversity estimators applied to wavefront sensing on an extended scene. We use a plate to generate known aberrations on the optical system. These aberrations (mainly defocus and astigmatism) depend on the angle between the plate and the optical axis.[21] The optical setup is depicted in Figure 10. Note that this optical setup has been already used in a previous paper.[17]

A slide illuminated by a projector and located in the focal plane of the lens L_1 simulates the extended scene. The slide represents the object used in the simulations. The translation of the lens L_2 allows the successive recording of the focal and the out-of-focus images (with a defocus amplitude peak-to-valley of λ). The data are monochromatic images (λ=633 nm, 10 nm width) and are Shannon-sampled. In order to process experimental data, we have to take into account the problem of edge effects due to the fact that the criterion is expressed in the Fourier domain (the convolutions are made by using FFT and produce severe wrap-around effects on extended scenes). To solve this problem, two solutions have been proposed in the phase diversity literature: the apodization technique[5] or the guard-band technique.[6] In the following, we have chosen the second method. The application of the guard-band technique to our restoration methods is explained in Appendix D. The image size used for the estimation is 64×64 pixels (one focused experimental image is shown Figure 10) and the guard band width used at each edge is equal to 32 pixels. The measurements of five pairs of focused and defocused images are made for an angular position of the plate of 45° and different noise levels. The theoretical values of the aberrations generated are $a_4 = 0.383$ radian (defocus) and $a_6 = 0.442$ radian (astigmatism).

Figure 11 shows the comparison of the RMSE on the phase estimates obtained by the unsupervised marginal estimator and by the joint estimator with a null regularization on the object, as a function of noise level. The curves are the results of the average of the 25 estimates obtained by all the combinations of the five pairs of focused and defocused images. For a noise level of 2%, the two estimators give very close phase estimates. For lower signal to noise ratios, the marginal method behaves

significantly better. This behavior is comparable with the one obtained in simulations (see Figure 8) and thus validates on experimental data the superiority of the marginal estimator for high noise level.

5. Conclusion

We have shown that the conventional method used in phase diversity has bad asymptotic properties (the estimates do not converge towards the true values as the data size tends to infinity and it usually presents local minima), needs the manual tuning of the regularization parameters and does not allow an optimally joint estimation of the object and of the aberrations. We have proposed a novel marginal estimator to solve these problems. We have shown its good asymptotic properties. We have proposed and validated an unsupervised estimation which solves the problem of hyperparameter tuning. We have shown that the aberrations and the hyperparameters estimated by the marginal method could be easily used to restore the object. Finally we have compared the performance of the joint method and the marginal one on simulated and on experimental data. This study has shown that, for the studied object, the two estimators give a very similar performance on the phase estimate for low noise levels but that the marginal method leads to better phase estimates for high noise levels. In this paper, we have used one particular Earth view. Future work should include a more comprehensive comparison of the two methods.

A. Circulant approximation

We assume that $(o - o_m)$ is stationary, so that the covariance matrix R_o is Toeplitz-block-Toeplitz. These matrices can be approximated by circulant block circulant

21

matrices with the approximation corresponding to a periodization. Under this assumption, the covariance matrix R_o and the convolution matrices H_1 and H_2 are diagonalized by a discrete Fourier transform (DFT). We can write

$$R_o = F^{-1}\text{diag}[S_o]F$$

$$H_1 = F^{-1}\text{diag}[\tilde{h}_1]F$$

$$H_2 = F^{-1}\text{diag}[\tilde{h}_2]F$$

where F is the two-dimensional DFT matrix, $diag[x]$ denotes a diagonal matrix having x on its diagonal, tilde denotes the two-dimensional DFT and S_o is the power spectral density of the object. Thus the criterion L_{JMAP} can be written in the Fourier domain.

B. Calculation of $(\boldsymbol{I} - \boldsymbol{m}_I)^T R_I^{-1}(\boldsymbol{I} - \boldsymbol{m}_I)$

The expression of R_I^{-1} is obtained by the block matrix inversion lemma[22]

$$R_I^{-1} = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} \tag{22}$$

with,

$$Q_{11} = \left[(H_1 R_o H_1^t + \sigma^2 I_d) - H_1 R_o H_2^t (H_2 R_o H_2^t + \sigma^2 I_d)^{-1} H_2 R_o H_1^t\right]^{-1}$$

$$Q_{12} = -Q_{11}(H_1 R_o H_2^t)(H_2 R_o H_2^t + \sigma^2 I_d)^{-1}$$

$$Q_{21} = -(H_2 R_o H_2^t + \sigma^2 I_d)^{-1}(H_2 R_o H_1^t)Q_{11}$$

$$Q_{22} = \left[(H_2 R_o H_2^t + \sigma^2 I_d) - H_2 R_o H_1^t (H_1 R_o H_1^t + \sigma^2 I_d)^{-1} H_1 R_o H_2^t\right]^{-1} \tag{23}$$

Substituting the expression of m_I (see Subsection 3.B.1) and of R_I^{-1} (Equation 22) into $(\boldsymbol{I} - \boldsymbol{m}_I)^T R_I^{-1}(\boldsymbol{I} - \boldsymbol{m}_I)$, yields:

$$(\boldsymbol{I} - \boldsymbol{m}_I)^T R_I^{-1}(\boldsymbol{I} - \boldsymbol{m}_I) = (\boldsymbol{i}_1 - H_1\boldsymbol{o}_m)^t Q_{11}(\boldsymbol{i}_1 - H_1\boldsymbol{o}_m) + (\boldsymbol{i}_1 - H_1\boldsymbol{o}_m)^t Q_{12}(\boldsymbol{i}_2 - H_2\boldsymbol{o}_m)$$

$$+ (\boldsymbol{i}_2 - H_2\boldsymbol{o}_m)^t Q_{21}(\boldsymbol{i}_1 - H_1\boldsymbol{o}_m) + (\boldsymbol{i}_2 - H_2\boldsymbol{o}_m)^t Q_{22}(\boldsymbol{i}_2 - H_2\boldsymbol{o}_m)$$

Using the circulant approximation (see Appendix A) for H_1, H_2 and R_o, the previous equation is limited to the sum of diagonal terms. After simplification, $(\boldsymbol{I} - \boldsymbol{m}_I)^T R_I^{-1}(\boldsymbol{I} - \boldsymbol{m}_I)$ has the following expression:

$$(\boldsymbol{I} - \boldsymbol{m}_I)^T R_I^{-1}(\boldsymbol{I} - \boldsymbol{m}_I) = \sum_v \frac{\left|\tilde{i}_1(v)\tilde{h}_2(\boldsymbol{a}, v) - \tilde{i}_2(v)\tilde{h}_1(\boldsymbol{a}, v)\right|^2}{\sigma^2 \left(|\tilde{h}_1(\boldsymbol{a}, v)|^2 + |\tilde{h}_2(\boldsymbol{a}, v)|^2 + \frac{\sigma^2}{S_o(v)}\right)}$$
$$+ \sum_v \frac{\left|\tilde{h}_1(\boldsymbol{a}, v)\tilde{o}_m(v) - \tilde{i}_1(v)\right|^2 + \left|\tilde{h}_2(\boldsymbol{a}, v)\tilde{o}_m(v) - \tilde{i}_2(v)\right|^2}{S_o(v) \left(|\tilde{h}_1(\boldsymbol{a}, v)|^2 + |\tilde{h}_2(\boldsymbol{a}, v)|^2 + \frac{\sigma^2}{S_o(v)}\right)} \quad (24)$$

C.   Determinant of R_I

Let Δ be a matrix that reads

$$\Delta = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

in a block form. Its determinant is given by

$$\det(\Delta) = \det(A)\det(D - CA^{-1}B)$$

Using this formula, it is easy to calculate the determinant of R_I:

$$\det(R_I) = \det(H_1 R_o H_1^t + \sigma^2 I_d)$$

$$\times \det\left[H_2 R_o H_2^t + \sigma^2 I_d - H_2 R_o H_1^t (H_1 R_o H_1^t + \sigma^2 I_d)^{-1} H_1 R_o H_2^t\right] \quad (25)$$

23

(25) can be rewritten in the Fourier domain (see appendix A) as

$$\det(R_I) = \prod_v \sigma^2 I_d \times \prod_v S_o(v) \times \prod_v \left(|\tilde{h}_1(\boldsymbol{a}, v)|^2 + |\tilde{h}_2(\boldsymbol{a}, v)|^2 + \frac{\sigma^2}{S_o(v)} \right) \qquad (26)$$

D. Application of the guard-band technique

To apply the guard-band technique to the joint estimator, in the case of Gaussian noise model, we use the criterion $L_{\mathrm{JMAP}}(\boldsymbol{o}, \boldsymbol{a}, \boldsymbol{\theta})$ (Equation 10).

For the marginal criterion $L_{\mathrm{MAP}}(\boldsymbol{a}, \boldsymbol{\theta})$, a new algorithm for this criterion $L_{\mathrm{MAP}}^{\mathrm{alt}}(\boldsymbol{o}, \boldsymbol{a}, \boldsymbol{\theta})$ called the "alternating" marginal criterion is used. The relationship between the joint criterion and the marginal one (see Equation 18) can be summarized by $L_{\mathrm{MAP}}(\boldsymbol{a}, \boldsymbol{\theta}) = L'_{\mathrm{JMAP}}(\boldsymbol{a}, \boldsymbol{\theta}) + \varepsilon(\boldsymbol{a}, \boldsymbol{\theta})$. The alternating marginal criterion is then defined by:
$L_{\mathrm{MAP}}^{\mathrm{alt}}(\boldsymbol{o}, \boldsymbol{a}, \boldsymbol{\theta}) = L_{\mathrm{JMAP}}(\boldsymbol{o}, \boldsymbol{a}, \boldsymbol{\theta}) + \varepsilon(\boldsymbol{a}, \boldsymbol{\theta})$. And:

$$
\begin{aligned}
\arg\min_{\boldsymbol{o}, \boldsymbol{a}, \boldsymbol{\theta}} L_{\mathrm{MAP}}^{\mathrm{alt}}(\boldsymbol{o}, \boldsymbol{a}, \boldsymbol{\theta}) &= \arg\min_{\boldsymbol{a}, \boldsymbol{\theta}} \left\{ \arg\min_{\boldsymbol{o}} L_{\mathrm{MAP}}^{\mathrm{alt}}(\boldsymbol{o}, \boldsymbol{a}, \boldsymbol{\theta}) \right\} \\
&= \arg\min_{\boldsymbol{a}, \boldsymbol{\theta}} \left\{ \arg\min_{\boldsymbol{o}} \left[L_{\mathrm{MAPJ}}(\boldsymbol{o}, \boldsymbol{a}, \boldsymbol{\theta}) \right] + \varepsilon(\boldsymbol{a}, \boldsymbol{\theta}) \right\} \\
&= \arg\min_{\boldsymbol{a}, \boldsymbol{\theta}} \left[L'_{\mathrm{MAPJ}}(\boldsymbol{a}, \boldsymbol{\theta}) + \varepsilon(\boldsymbol{a}, \boldsymbol{\theta}) \right] \\
&= \arg\min_{\boldsymbol{a}, \boldsymbol{\theta}} L_{\mathrm{MAP}}(\boldsymbol{a}, \boldsymbol{\theta}) \qquad (27)
\end{aligned}
$$

The minimization of $L_{\mathrm{MAP}}^{\mathrm{alt}}(\boldsymbol{o}, \boldsymbol{a}, \boldsymbol{\theta})$ with respect to $\boldsymbol{o}, \boldsymbol{a}$ and $\boldsymbol{\theta}$ is therefore equivalent to the minimization of $L_{\mathrm{MAP}}(\boldsymbol{a}, \boldsymbol{\theta})$ with respect to the sole $\boldsymbol{a}$ and $\boldsymbol{\theta}$. The guard-band can then be applied to the criterion $L_{\mathrm{MAP}}^{\mathrm{alt}}(\boldsymbol{o}, \boldsymbol{a}, \boldsymbol{\theta})$.

Jérôme Idier is now with Institut de Recherche en Communications et Cybernétique de Nantes, École Centrale de Nantes, 1 rue de la Noë, B.P. 92101, 44321 Nantes Cedex 3, France.

Corresponding author Laurent Mugnier's e-mail is `Laurent.Mugnier@onera.fr`.

References

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

2. R. L. Kendrick, D. S. Acton, and A. L. Duncan, "Phase-diversity wave-front sensor for imaging systems," Appl. Opt. **33,** 6533–6546 (1994).

3. D. J. Lee, M. C. Roggemann, B. M. Welsh, and E. R. Crosby, "Evaluation of least-squares phase-diversity technique for space telescope wave-front sensing," Appl. Opt., **36,** 9186–9197 (1997).

4. M. G. Löfdahl and A. L. Duncan, "Fast phase diversity wavefront sensor for mirror control," in *Adaptative Optical System Technologies*, D. Bonaccini and R. K. Tyson, eds., **3353,** 952–963, Proc. SPIE (1998).

5. M. G. Löfdahl and G. B. Scharmer, "Wavefront sensing and image restoration from focused and defocused solar images," Astron. Astrophys. **107,** 243–264 (1994).

6. J. H. Seldin and R. G. Paxman, "Phase-diverse speckle reconstruction of solar data," in *Image Reconstruction and Restoration*, T. J. Schulz and D. L. Snyder, eds., **2302,** 268–280, Proc. SPIE (1994).

7. B. J. Thelen, R. G. Paxman, D. A. Carrara, and J. H. Seldin, "Maximum a posteriori estimation of fixed aberrations, dynamic aberrations, and the object from phase-diverse speckle data," J. Opt. Soc. Am. A **16,** 1016–1025 (1999).

8. O. M. Bucci, A. Capozzoli, and G. D'Elia, "Regularizing strategy for image restoration and wave-front sensing by phase diversity," J. Opt. Soc. Am. A **16,** 1759–1768 (1999).

9. R. J. Noll, "Zernike polynomials and atmospheric turbulence," J. Opt. Soc. Am. **66,** 207–211 (1976).

10. R. G. Paxman, T. J. Schulz, and J. R. Fienup, "Joint estimation of object and aberrations by using phase diversity," J. Opt. Soc. Am. A **9,** 1072–1085 (1992).

11. A. P. Kattnig and J. Primot, "Model of the second-order statistic of the radiance field of natural scenes, adapted to system conceiving," in *Visual information processing VI*, S. K. Park and R. D. Juday, eds., **3074**, 132-141, Proc. SPIE (1997).

12. R. J. A. Little and D. B. Rubin, "On Jointly Estimating Parameters and Missing Data by Maximizing the Complete-Data Likelihood," The American Statistician **37,** 218–220 (1983).

13. Y. Goussard, G. Demoment, and J. Idier, "A new algorithm for iterative deconvolution of sparse spike," in *Proc. IEEE ICASSP*, Piscataway, New Jersey, 1990, pp. 1547–1550.

14. F. Champagnat and J. Idier, "An alternative to standard maximum likelihood for Gaussian mixtures," in *Proc. IEEE ICASSP*, Piscataway, New Jersey, 1995, pp. 2020–2023

15. E. Lehmann, *Theory of point estimation*, Wiley, New York (1983).

16. E. D. Carvalho and D. Slock, "Maximum-likelihood blind FIR multi-channel estimation with Gaussian prior for the symbols," in *Proc. IEEE ICASSP*, Piscataway, New Jersey, 1997, pp. 3593–3596.

17. L. Meynadier, V. Michau, M.-T. Velluet, J.-M. Conan, L. M. Mugnier, and G. Rousset, "Noise propagation in wave-front sensing with phase diversity," Appl. Opt. **38,** 4967–4979 (1999).

18. J. Seldin and R. Paxman, "Closed-loop wavefront sensing for a sparse-aperture, phased-array telescope using broadband phase diversity," in *Imaging Technology and Telescopes*, J. B. Breckinridge and al., eds., **4091,** 48–63, Proc. SPIE (2000).

19. D. A. Carrara, B. J. Thelen and R. G. Paxman, "Aberration correction of segmented-aperture telescopes by using phase diversity," in *Image reconstruction from incomplete data*, M. A. Fiddy and R. P. Millane, eds., **4123,** 56–63, Proc. SPIE (2000).

20. O. N. Strand, "Theory and methods related to the singular-function expansion and landwebers's iteration for integral equations of the first kind," SIAM J. Numer. Anal. **11,** 798–825 (1974).

21. M. Born and E. Wolf, *Principles of Optics*, Pergamon Press, Oxford (1983).

22. F. R. Gantmacher, *Théorie des Matrices Tome I*, Dunod, Paris (1966).
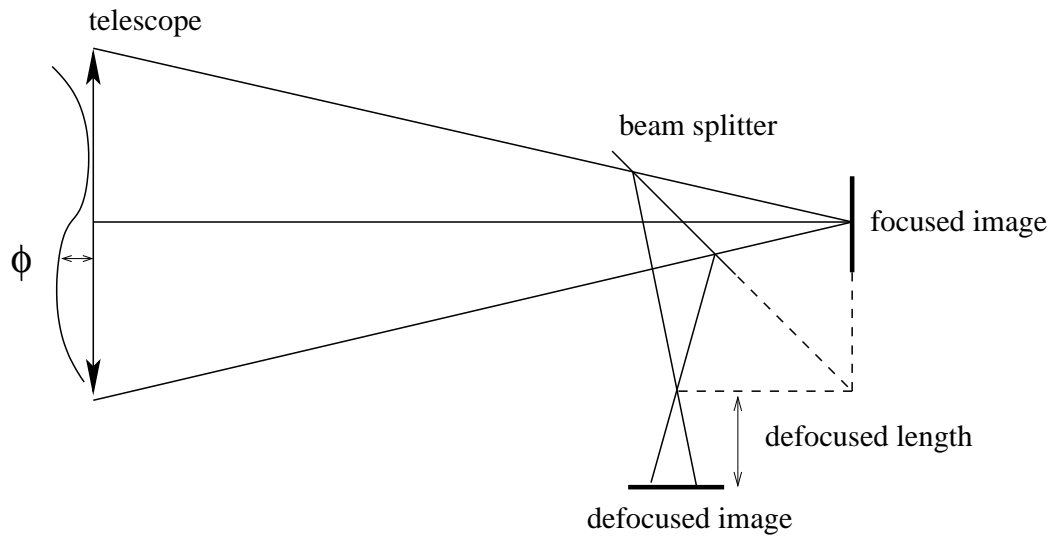
Fig. 1. Phase diversity principle
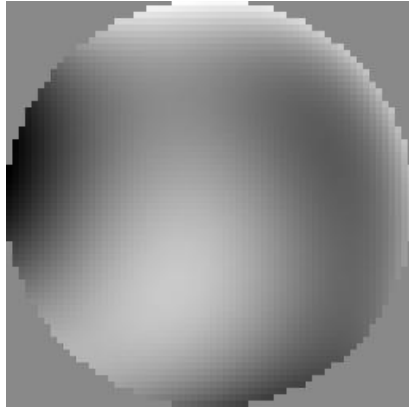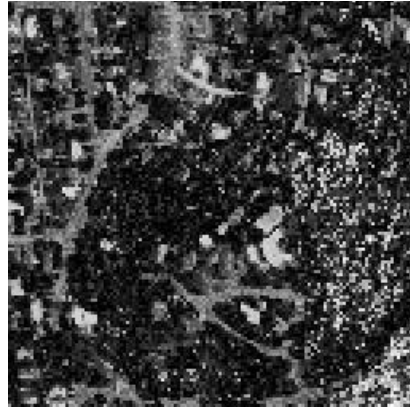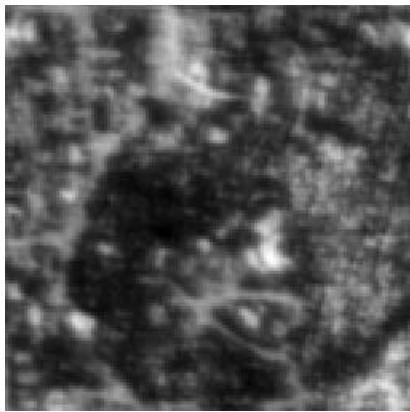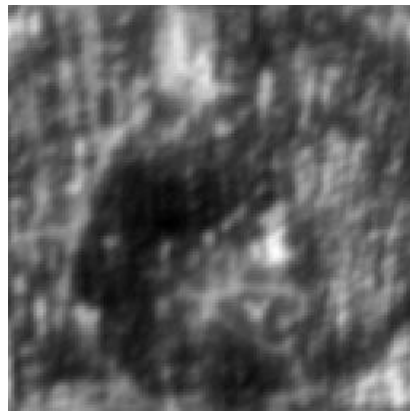
Fig. 2. (a) aberrated phase (λ/7rms), (b) true object,(c) focused image, (d) defocused image
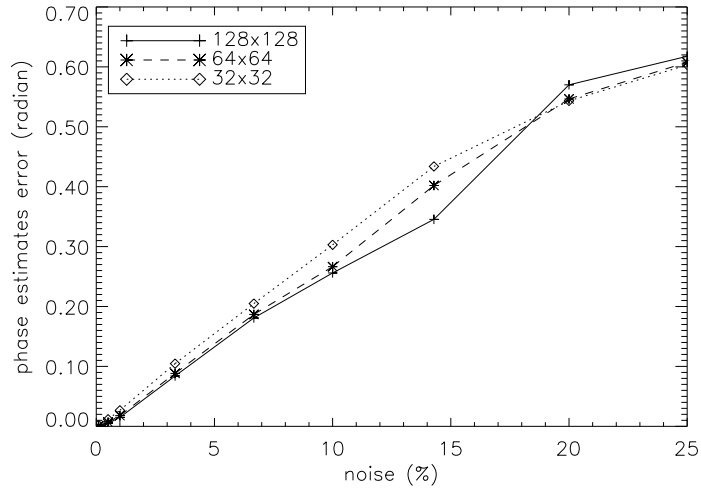
(a)
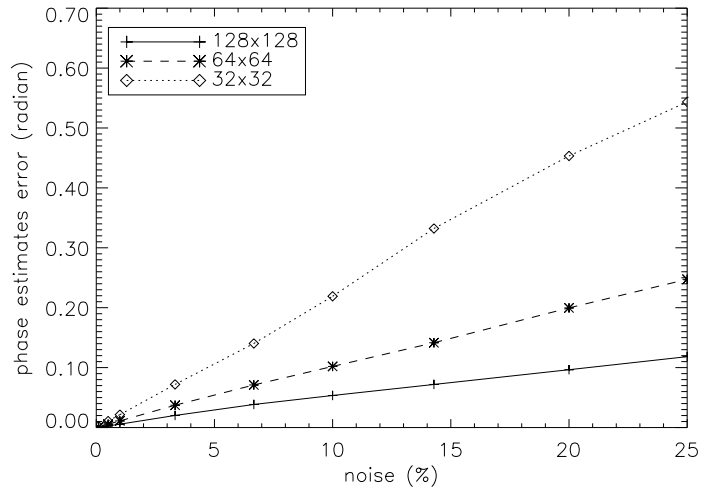
(b)

Fig. 3. RMSE of phase estimates as a function of noise level given in percent (it is the ratio between the noise standard deviation and the mean flux per pixel). Figure (a) is for the joint estimator, figure (b) for the marginal estimator. The solid, dashed and dotted lines, respectively correspond to images of dimensions 128×128, 64×64 and 32×32 pixels. Such RMSE estimates have been obtained as an empirical averages on 50 independent realizations of noise.

noise 14%

(a)

noise 4%
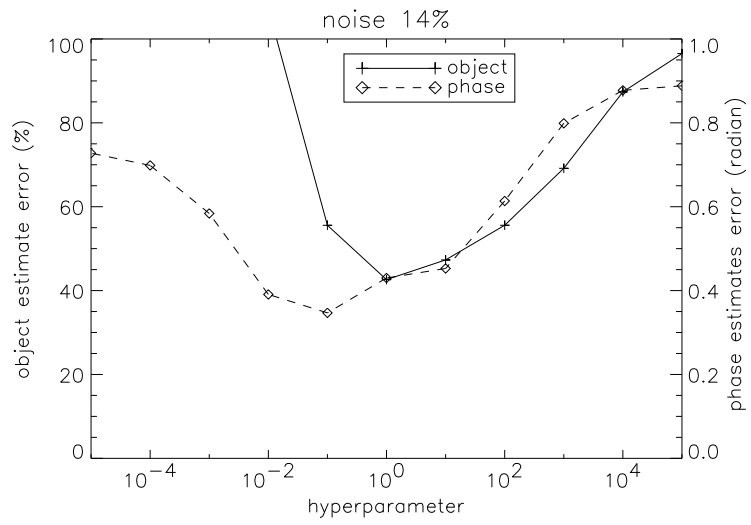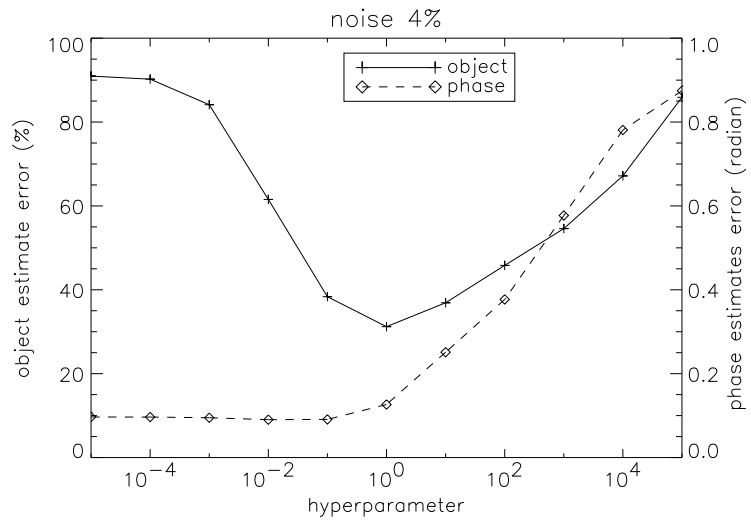
(b)

Fig. 4. Plots of RMSE for joint phase estimates (dashed line - see right vertical axis) and joint object estimate (solid line - see left vertical axis) as a function of the value of the hyperparameter μ for an image size of 32×32 pixels. figure (a) is for a noise level of 14%, figure (b) is for 4%.

31
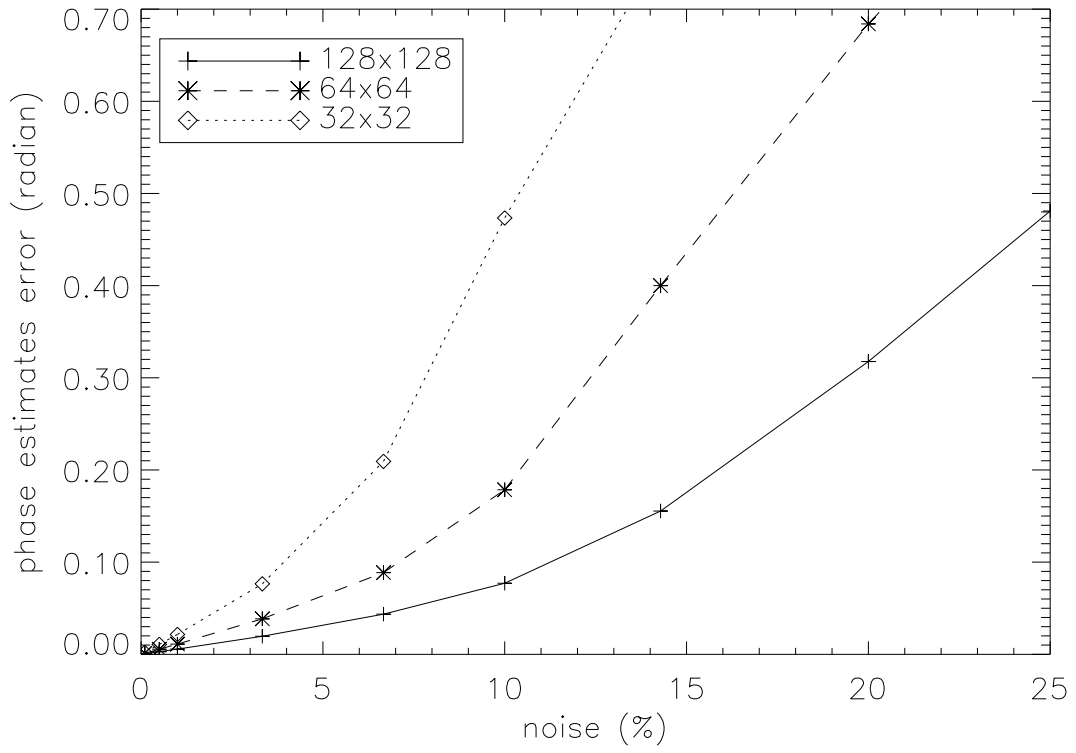
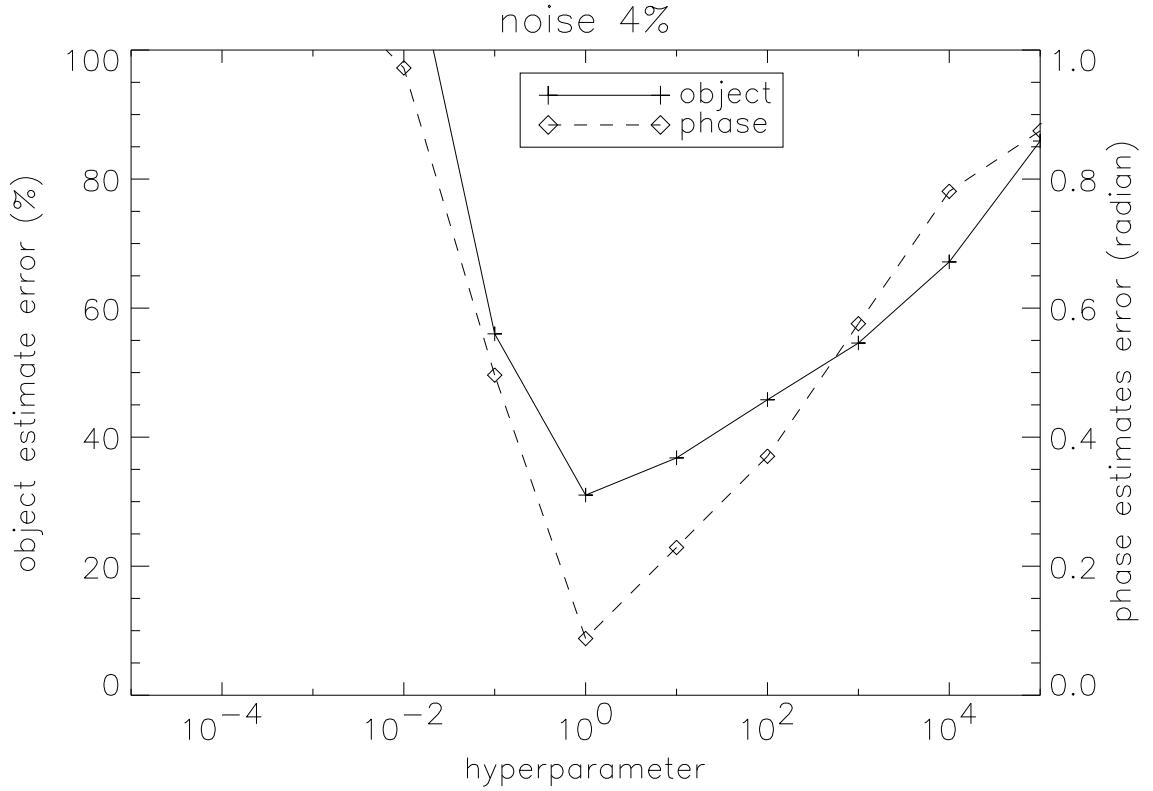Fig. 5. RMSE of joint phase estimates as a function of noise level for a near null regularization for 3 image sizes.

Fig. 6. Plots of RMSE for marginal phase estimates (dashed line - see right vertical axis) and marginal object estimate (solide line - see left vertical axis) as a function of the value of the hyperparameter μ for an image size of 32×32 pixels for a noise level of 4%.
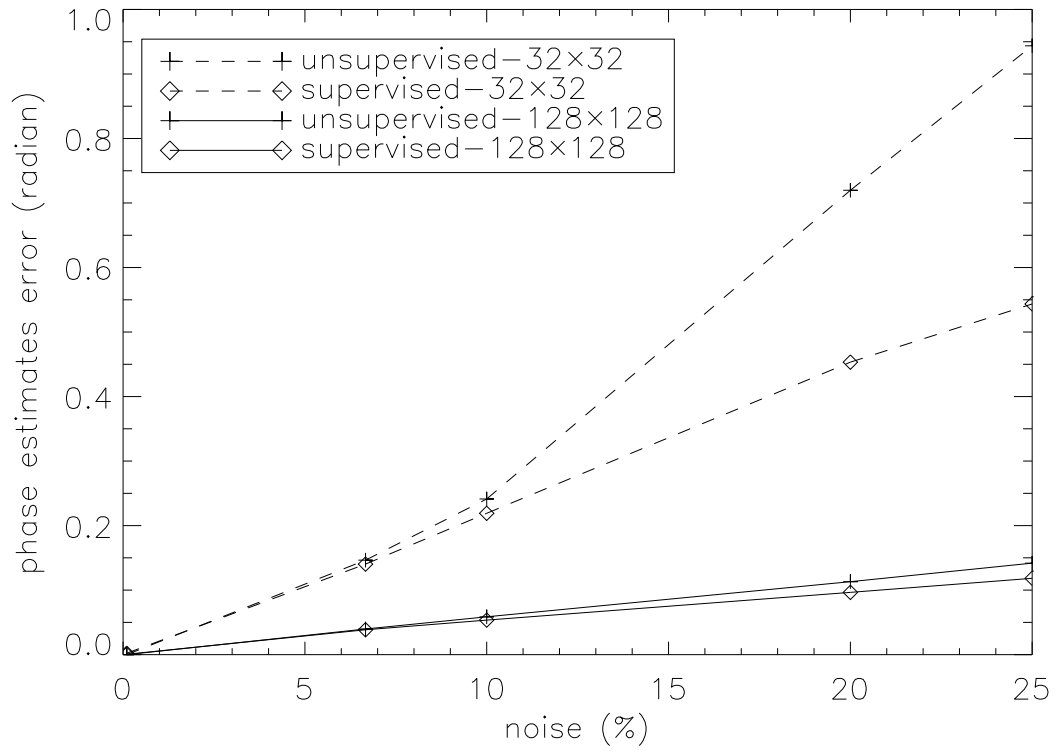
Fig. 7. Performance of marginal estimation with the true hyperparamaters (pluses) and for unsupervised estimation (diamonds) as measured by RMSE of phase estimates vs. noise level.
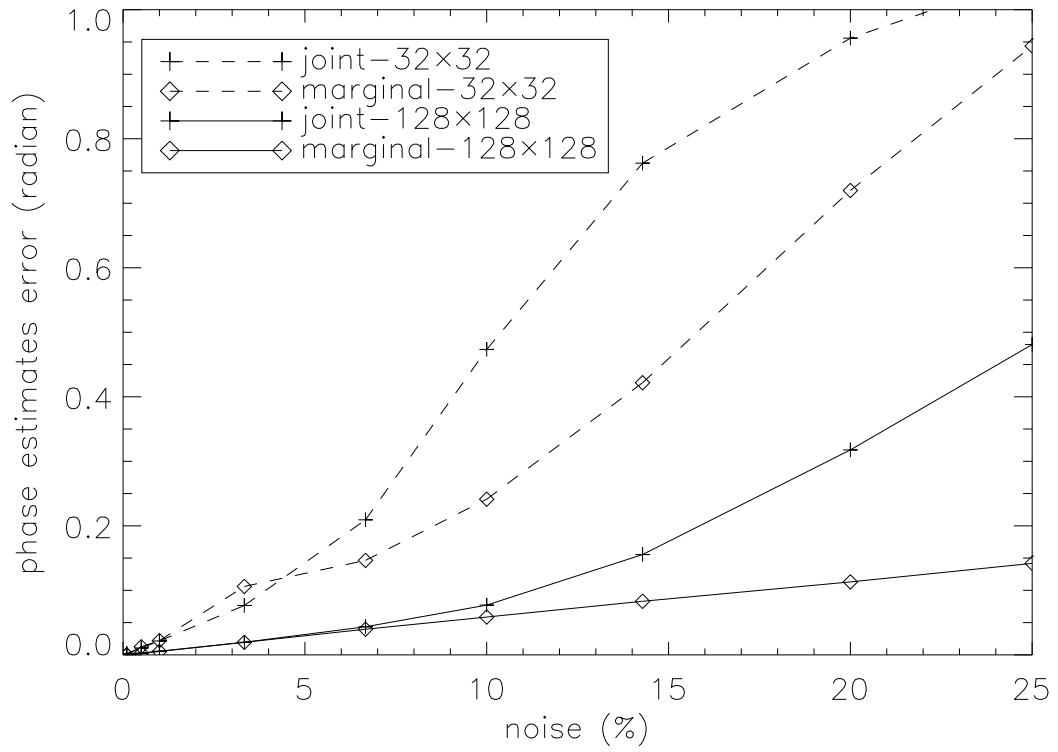
Fig. 8. RMSE of unsupervised marginal estimator (diamonds) and the joint one with near null regularization (pluses) aberrations estimates as a function of noise level.
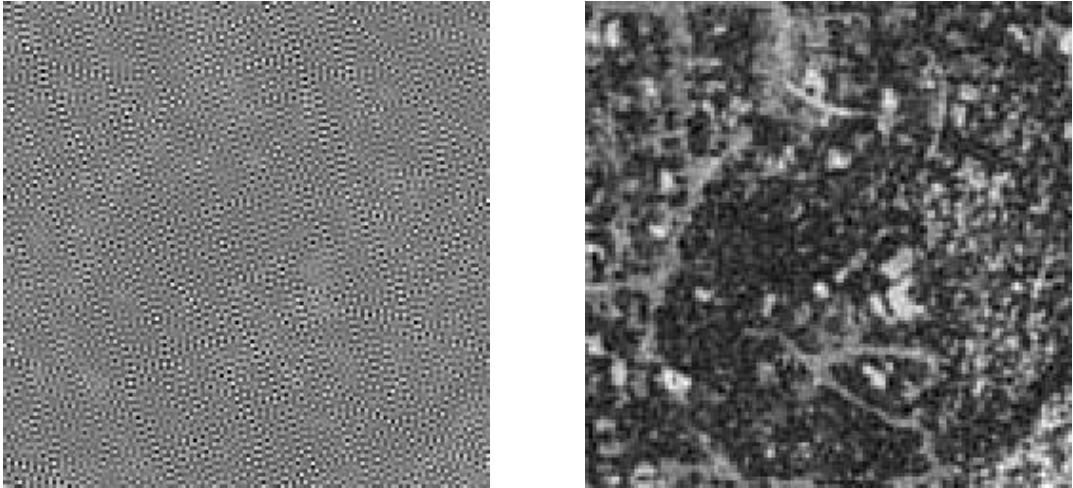
Fig. 9. Restorated object with the joint method (near null regularization used, left) and with the unsupervised marginal method (right). The noise level is equal to 4%.

Table 1. Values of the coefficients used for simulations.

Coefficient	Value (radian)
a_4	-0.2
a_5	0.3
a_6	-0.45
a_7	0.4
a_8	0.3
a_9	-0.25
a_{10}	0.35
a_{11}	0.2
a_{12}	0.1
a_{13}	0.05
a_{14}	-0.05
a_{15}	0.05
a_{16}	0.02
a_{17}	0.01
a_{18}	-0.01
a_{19}	-0.02
a_{20}	0.01
a_{21}	0.01

diaphragm

plate

CCD
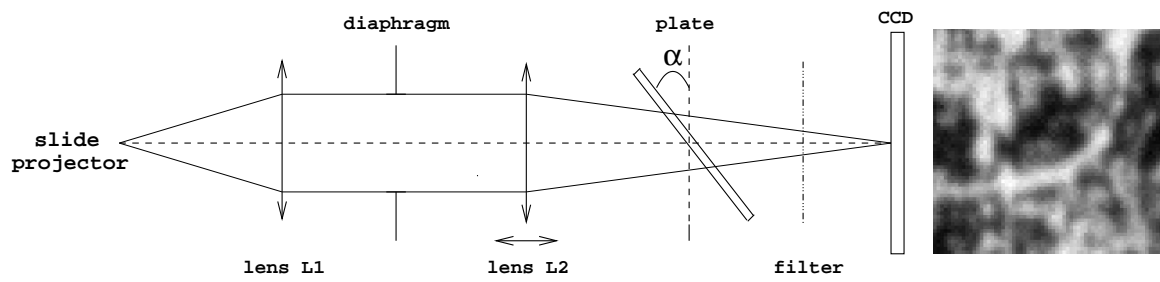
α

slide
projector

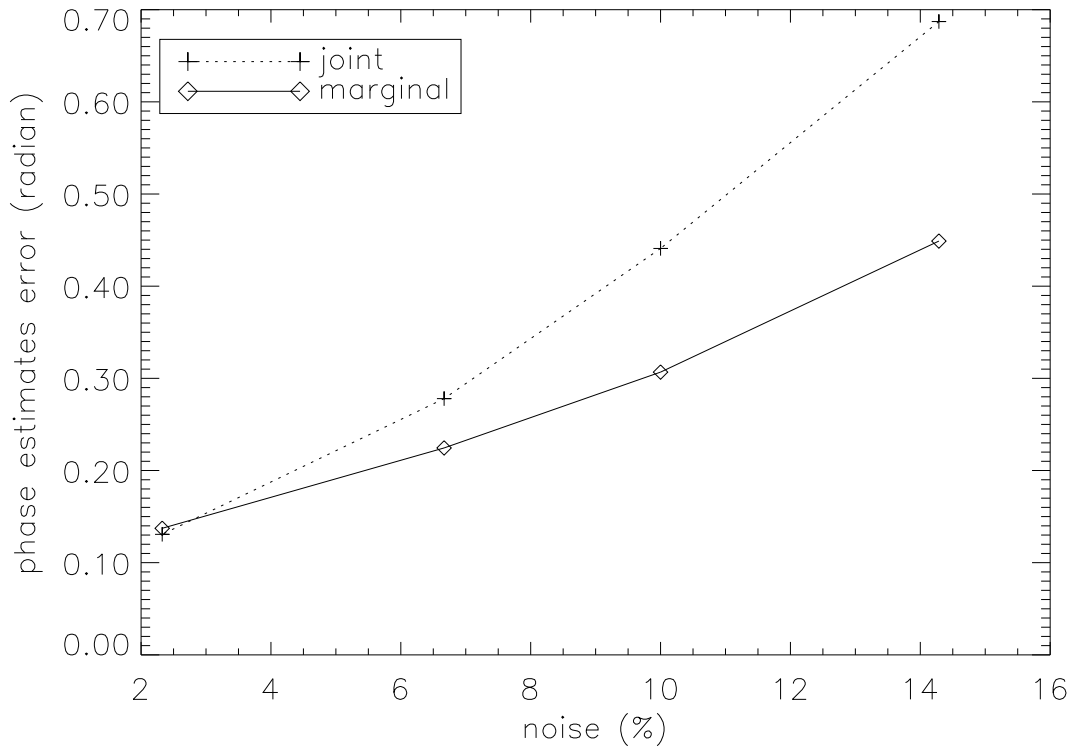lens L1

lens L2

filter

Fig. 10. The optical setup

Fig. 11. RMSE of unsupervised marginal estimator (diamonds) and the joint one with null regularization (pluses) aberrations estimates as a function of noise level.