Semidefinite programming for model-based sensorless adaptive optics

Jacopo Antonello,[1,*] Michel Verhaegen,[1] Rufus Fraanje,[1] Tim van Werkhoven,[2,3]
Hans C. Gerritsen,[2] and Christoph U. Keller[3]

[1]*Delft Center for Systems and Control, Delft University of Technology, Mekelweg 2, Delft 2628 CD, The Netherlands*
[2]*Molecular Biophysics, Utrecht University, P.O. Box 80000, Utrecht 3508 TA, The Netherlands*
[3]*Leiden Observatory, Leiden University, P.O. Box 9513, Leiden 2300 RA, The Netherlands*
Corresponding author: j.antonello@tudelft.nl

Wavefront sensorless adaptive optics methodologies are widely considered in scanning fluorescence microscopy where direct wavefront sensing is challenging. In these methodologies, aberration correction is performed by sequentially changing the settings of the adaptive element until a predetermined image quality metric is optimized. An efficient aberration correction can be achieved by modeling the image quality metric with a quadratic polynomial. We propose a new method to compute the parameters of the polynomial from experimental data. This method guarantees that the quadratic form in the polynomial is semidefinite, resulting in a more robust computation of the parameters with respect to existing methods. In addition, we propose an algorithm to perform aberration correction requiring a minimum of $N + 1$ measurements, where N is the number of considered aberration modes. This algorithm is based on a closed-form expression for the exact optimization of the quadratic polynomial. Our arguments are corroborated by experimental validation in a laboratory environment.   © 2012 Optical Society of America

OCIS codes: 010.1080, 010.7350, 110.0113, 220.1000.

1. INTRODUCTION

Adaptive optics is concerned with the active suppression of disturbances in optical systems. The sources of the disturbances can be different, according to the application in question. Notable examples are atmospheric turbulence for astronomy and heterogeneity in the index of refraction within specimens for microscopy. As a consequence, phase aberrations develop in the pupil of the objective lens, severely affecting the quality of the image [1]. The principle of adaptive optics is that by measuring such phase variations with a sensor, they can be cancelled by appropriately driving an active wavefront correction element. In astronomy this practice is well established with the use of a Shack–Hartmann wavefront sensor and a deformable mirror [1].

Nonetheless, there are instances where the deployment of a wavefront sensor is challenging. This is the case for scanning fluorescence microscopy [2], due to difficulties in the rejection of out-of-focus light and in the lack of reference-point sources within specimens [3–9].

Alternatively, sensorless adaptive optics schemes have been considered, where the fluorescence emission is used as a feedback signal for the suppression of the aberrations. One approach involves the rejection of out-of-focus background [10]. More commonly, instead, aberration correction is achieved by sequentially modulating the adaptive element until a selected image quality metric is optimized. The assumption is that the global extremum of the metric is attained when the aberrations have been maximally suppressed. Examples of such metrics are, among others, sharpness measures for images [11] and the amount of fluorescence emission.

In the literature, a number of proposed solutions make use of model-free optimizations. These include hill-climbing algorithms [12,13], genetic algorithms [13–17], image-based algorithms [18,19], conjugate gradient methods [20], stochastic parallel gradient descent methods [21], and the Nelder–Mead simplex algorithm [11,22,23]. Such general methodologies require a large number of measurements of the metric [2,24,25] and may not converge to the global optimum [23,26]. Reducing the number of necessary measurements is a critical factor for the overall image acquisition time [12,13] and for inhibiting side effects, such as phototoxicity and photobleaching [2].

It has been shown [26] that physical modeling of the image quality metric allows for direct and deterministic optimization methods, requiring a reduced number of measurements with respect to model-free solutions. Initially, model-based methodologies were proposed for optical systems where the object is a point source. In [26,27], a quadratic polynomial was employed to model a Strehl-based metric. For small aberrations, it was shown that the proposed model-based approach outperforms model-free algorithms. This result was extended to encompass larger aberrations in [28], by using a metric based on the Lukosz–Zernike functions and a nonlinear detector. In [29] a generalization was provided to handle arbitrary functions other than the Lukosz–Zernike functions. The case of incoherent imaging was analyzed in [30]. Here first-principles derivations motivated employing a quadratic polynomial in order to model a metric based on the low spatial frequency content of the recorded images. Similarly, in [31], theoretical derivations supported using a quadratic polynomial to model an image quality metric, which is appropriate for

structured illumination microscopy. Experimental validation of model-based approaches was also provided for two-photon microscopy [32] and for multiharmonic microscopy [33].

One challenge of model-based approaches is found in the need to compute the parameters of the quadratic polynomial for a given real optical system. Initially, this task was performed using first principles, i.e., by computing the theoretical value of each parameter [26–30]. In this way, however, imperfections in the real optical system are not accounted for [34]. Also, experimentally computing the parameters is more suited, for example, in the case of coherent microscopies such as third-harmonic generation [34]. To address these shortcomings, experimental methods for the computation of the parameters were developed [31,34]. Such methods, nevertheless, fail to guarantee that the quadratic form in the polynomial used to model the image quality metrics is semidefinite. This latter property always follows from the theoretical analysis of the image formation processes [24,27,28,30–34]. In this paper, we present and validate a new method that guarantees that the semidefiniteness property is satisfied. We compare our procedure with the previously proposed methods [31,34] and show that a more accurate fitting of the experimental data is achieved. We remark that an inaccurate computation of the parameters of the polynomial adversely affects the performance in the correction of the aberrations as shown elsewhere [24,34].

Once the parameters of the polynomial are known, the correction of an arbitrary aberration is performed by solving an optimization problem that exploits the knowledge about the quadratic polynomial. For the imaging system considered in [26–29], an approximate solution of the optimization was proposed in [26–28], using $N + 1$ measurements. In [29] an exact solution was provided using $N + 1$ measurements. For the remaining imaging systems [24,30–34], an exact solution of the optimization was provided using a minimum of $2N + 1$ measurements. In this paper, we derive an exact solution of the optimization requiring a minimum of $N + 1$ measurements. Because our formulas are derived for a quadratic polynomial in its most general form, all the model-based approaches mentioned so far are encompassed as special cases.

This paper is organized as follows. Section 2 provides a first-principles derivation showing that a quadratic polynomial can model the image quality metric used in our experimental validation. Section 3 considers the experimental computation of the parameters of a quadratic polynomial used to model an image quality metric. Section 4 focuses on the algorithms used for aberration correction. Section 5 provides a description of the optical system used in the experimental validation. Experimental results are reported in Section 6. Finally, conclusions are found in Section 7.

2. QUADRATIC MODELING OF A WAVEFRONT SENSORLESS ADAPTIVE OPTICS SYSTEM

A. Problem Formulation
Consider the problem of correcting a static aberration in a wavefront sensorless adaptive optics system. Such a problem can be formulated as follows:

$$\max_{\mathbf{u}(k)} \tilde{y}(k), \tag{1}$$

where $\tilde{y}(k) \in \mathbb{R}$ is the value of a metric quantifying the image quality, $k \in \mathbb{Z}$ is the discrete time index, and $\mathbf{u}(k) \in \mathbb{R}^N$ is the control signal applied to an active element with N degrees of freedom. An instance of this problem is found when imaging a single focal spot in a fluorescence scanning microscope [35,36], where static specimen-induced aberrations are to be suppressed. In this case, the value of metric $\tilde{y}(k)$ depends on the amount of fluorescence emission originating from the focal spot. A phase deformation can be applied to the illumination light in the pupil of the objective lens, for instance by employing a deformable mirror that is controlled via vector $\mathbf{u}(k)$. When the deformation induced by the deformable mirror maximally suppresses the specimen-induced aberration, a solution of Eq. (1) is found.

In general we have that $\tilde{y}(k) = f(\mathbf{u}(k))$, where $f(\cdot)$ is a function with a global maximum and possibly multiple local extrema. For this reason, a general nonlinear optimization algorithm can be employed in order to solve Eq. (1) as discussed for the model-free methodologies in the introduction. Instead, model-based methodologies exploit the fact that within a suitable neighborhood of the global maximum, $f(\cdot)$ can be approximated by a quadratic polynomial. Here, metric $\tilde{y}(k) = f(\mathbf{u}(k))$ can be modeled with an approximate metric $y(k) = q(\mathbf{u}(k))$, where $q(\cdot)$ is a quadratic polynomial. The knowledge about $q(\cdot)$ allows us to efficiently solve Eq. (1). In the next section we provide a derivation for $q(\cdot)$ based on first principles for the optical system that was used in our experimental validation. This serves as an example in order to highlight the advantage of experimentally determining $q(\cdot)$ as proposed in this paper.

B. Modeling of a Wavefront Sensorless Adaptive Optics Imaging System
Consider the optical configuration in Fig. 1. A disturbance in the entrance pupil of L_1 induces an unknown time-invariant aberration to the wavefront. The entrance pupil is reimaged by lenses L_1 and L_2 onto the membrane of the deformable mirror. An image is formed by lens L_3 onto a photodiode, which is covered by a pinhole aperture. Let $\tilde{y}$ denote the integral over the pinhole aperture of the intensity distribution in the focal plane of L_3. Such quantity will be hereafter referred to with the general term "intensity," as is commonly done in the literature [26,27,37]. The intensity is taken to be proportional to the voltage recorded at the output of the photodiode. As outlined in [26,27,37], it is assumed that $\tilde{y}$ is a valid image quality metric, i.e., maximizing $\tilde{y}$ minimizes the aberration of the wavefront.
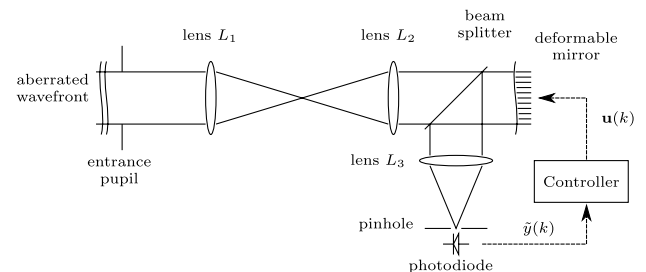

Fig. 1. Schema representing a sensorless adaptive optics system. An unknown aberration applied at the entrance pupil of the system must be corrected by a deformable mirror that is conjugated to the entrance pupil. The measurement $\tilde{y}(k)$ made with a photodiode covered by a pinhole is an indicator of the residual aberration in the wavefront. The controller changes control signal $\mathbf{u}(k)$ in order to maximize $\tilde{y}(k)$.

In Section [6] such a conjecture is experimentally verified by measuring the residual aberration with a Shack–Hartmann wavefront sensor.

The intensity $\tilde{y}(k)$ can be modeled [38] by

$$
\tilde{y}(k) = \left(\frac{A}{\lambda f}\right)^2 \int_{\Sigma_2} \\
\times \left| \int_{\Sigma_1} \exp\left[j\frac{2\pi}{\lambda} W(\xi,\eta,k) - j\frac{2\pi}{\lambda f}(\xi\alpha + \eta\beta) \right] \mathrm{d}\xi \mathrm{d}\eta \right|^2 \mathrm{d}\alpha \mathrm{d}\beta + w(k),
\tag{2}
$$

where k is the time index, A and λ are, respectively, the amplitude and wavelength of the monochromatic wave, f is the focal distance of L_3, Σ_2 is the pinhole aperture, Σ_1 is the pupil, $j = \sqrt{-1}$, $W(\xi,\eta,k)$ is the wavefront aberration, and $w(k)$ is the measurement noise.

If the effects due to the finite size of the pinhole are neglected (see [26,27]),

$$
\tilde{y}(k) \approx \left(\frac{A}{\lambda f}\right)^2 \left| \int_{\Sigma_1} \exp\left[j\frac{2\pi}{\lambda} W(\xi,\eta,k) \right] \mathrm{d}\xi \mathrm{d}\eta \right|^2 + w(k).
\tag{3}
$$

Let $\tilde{\Phi}(\xi,\eta,k) := (2\pi/\lambda)W(\xi,\eta,k)$ and $S := \int_{\Sigma_1} \mathrm{d}\xi\mathrm{d}\eta$. As done in [39,40] for k fixed, the exponential in Eq. ([3]) is expanded into a Taylor series and the terms of order higher than 2 are neglected. The approximation becomes

$$
\tilde{y}(k) \approx \left(\frac{AS}{\lambda f}\right)^2 \\
\times \left[1 - \left(\frac{1}{S}\int_{\Sigma_1} \tilde{\Phi}(\xi,\eta,k)^2 \mathrm{d}\xi\mathrm{d}\eta - \left(\frac{1}{S}\int_{\Sigma_1}\tilde{\Phi}(\xi,\eta,k)\mathrm{d}\xi\mathrm{d}\eta\right)^2 \right) \right] \\
+ w(k).
\tag{4}
$$

We assume that $\tilde{\Phi}$ can be expanded into the following series:

$$
\tilde{\Phi}(\xi,\eta,k) = \sum_{i=1}^{\infty} \Psi_i(\xi,\eta) v_i(k).
\tag{5}
$$

Possible choices for $\Psi_i(\cdot,\cdot)$ include, among others, Zernike polynomials [41], mirror modes [42], or simply the influence functions of the actuators of a deformable mirror [43]. Let $\mathcal{F}(\xi,\eta) := [\Psi_1(\xi,\eta) \ldots \Psi_N(\xi,\eta)]^T$ and $\mathbf{v}(k) := [v_1(k) \ldots v_N(k)]^T$; then the Nth order truncation of Eq. ([5]) is denoted as

$$
\tilde{\Phi}(\xi,\eta,k) \approx \mathcal{F}(\xi,\eta)^T \mathbf{v}(k).
\tag{6}
$$

Substituting Eq. ([6]) into Eq. ([4]) and performing the integrations leads to the following quadratic approximation:

$$
\tilde{y}(k) \approx \tilde{c}_0 - \mathbf{v}(k)^T \tilde{Q} \mathbf{v}(k) + w(k),
\tag{7}
$$

where $\tilde{c}_0 = (AS/\lambda f)^2$ and

$$
\tilde{Q} = \tilde{c}_0 \left(\frac{1}{S}\int_{\Sigma_1} \mathcal{F}(\xi,\eta)\mathcal{F}(\xi,\eta)^T \mathrm{d}\xi\mathrm{d}\eta \right. \\
\left. - \left(\frac{1}{S}\int_{\Sigma_1}\mathcal{F}(\xi,\eta)\mathrm{d}\xi\mathrm{d}\eta\right)\left(\frac{1}{S}\int_{\Sigma_1}\mathcal{F}(\xi,\eta)^T\mathrm{d}\xi\mathrm{d}\eta\right) \right).
\tag{8}
$$

Note that the elements of $\mathcal{F}$ need not be orthogonal over the pupil. If such elements are chosen among N Zernike polynomials, the Strehl ratio can be recognized and $\tilde{Q}$ is diagonal. It can be seen that $\tilde{Q}$ must be positive semidefinite, i.e., $\tilde{Q} \succcurlyeq 0$. Vector $\mathbf{v}(k)$ accounts for both the contribution due to the unknown aberration $\mathbf{x}$ in the entrance pupil of L_1 and the aberration $\mathbf{u}(k)$ induced by the deformable mirror. Letting $\mathbf{v}(k) = \mathbf{x} - \mathbf{u}(k)$, Eq. ([7]) is rewritten into

$$
\tilde{y}(k) \approx \tilde{c}_0 - (\mathbf{x} - \mathbf{u}(k))^T \tilde{Q}(\mathbf{x} - \mathbf{u}(k)) + w(k).
\tag{9}
$$

Quadratic polynomials, such as the right-hand side of Eq. ([9]) without the noise term $w(k)$, have been used in order to model image quality metrics in a variety of different imaging techniques [24,26–34]. This suggests that Eq. ([9]) can be employed as a generalized approximate metric. In each of these publications, a thorough analysis of the image formation process led to a quadratic polynomial where matrix $\tilde{Q}$ was positive semidefinite [respectively negative semidefinite if optimization ([1]) is formulated as a minimization problem as in [28], for example]. The semidefiniteness property of $\tilde{Q}$ stems from the fact that $\tilde{y}$ exhibits a global extremum. Nevertheless, when applying the experimental computations proposed in [31,34], such a property can be violated. This shortcoming is addressed in the procedure proposed in Section [3].

3. IDENTIFICATION OF THE PARAMETERS FOR QUADRATIC APPROXIMATE METRICS

Once metric $\tilde{y}$ has been selected, depending on the imaging system (see [24,27,28,30–34]), parameters $\tilde{c}_0$ and $\tilde{Q}$ must be computed before aberration correction can be applied. This operation is required once only. As in [31,34], the unknown aberration $\mathbf{x}$ is assumed to be zero throughout this calibration procedure.

One possibility is to compute $\tilde{c}_0$ and $\tilde{Q}$ from their first principles definitions, e.g., Eq. ([8]) for $\tilde{Q}$. Such an approach can be cumbersome. First, accurately measuring a number of quantities in a real system is both prone to errors and inconvenient. Also additional equipment may be necessary. In Eq. ([8]), amplitude A and the surface of the pupil S must be measured. Second, a different numerical integration must be solved for each different choice of the basis functions $\Psi_i(\cdot,\cdot)$ in Eq. ([5]). In addition, a first-principles computation of the parameters does not account for defects in the real optical system, such as misalignment, imperfect illumination profile, noncircular pupils, etc. (see the discussion in [34]). A final drawback is related to the fact that the overall modeling error is neither explicitly defined nor minimized. As remarked earlier, an inaccurate choice for $\tilde{Q}$ leads to a decreased performance in the correction of aberrations [24,34]. The alternative followed in this paper, instead, is to select the values of $\tilde{c}_0$ and $\tilde{Q}$ by minimizing the fitting error, e.g., $\sum_k (\tilde{c}_0 - \mathbf{u}(k)^T \tilde{Q}\mathbf{u}(k) - \tilde{y}(k))^2$, over some set of real input–output measurements. For this reason an experimental computation of $\tilde{c}_0$ and $\tilde{Q}$ is desirable.

A. Débarre's Experimental Identification Procedure

We briefly report the latest, most accurate experimental procedure to compute $\tilde{Q}$, which was proposed in [34], in order to compare it with our proposed method. The idea is to compute $\tilde{Q}$ by independently estimating each of its elements $\tilde{q}_{i,j}$. First, the diagonal elements are recovered. Afterward, the off-diagonal elements are computed by estimating each of the $N(N-1)/2$ submatrices of $\tilde{Q}$ of dimension 2×2.

As an example, consider Eq. (9) when $N = 3$,

$$y(u_1, u_2, u_3) = \tilde{c}_0 - \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}^T \begin{bmatrix} \tilde{q}_{1,1} & \tilde{q}_{2,1} & \tilde{q}_{3,1} \\ \tilde{q}_{2,1} & \tilde{q}_{2,2} & \tilde{q}_{3,2} \\ \tilde{q}_{3,1} & \tilde{q}_{3,2} & \tilde{q}_{3,3} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}. \quad (10)$$

By keeping u_2 and u_3 fixed to zero, an input–output data set is collected. Subsequently, element $\tilde{q}_{1,1}$ and $\tilde{c}_0$ can be estimated by fitting the resulting parabola $y(u_1) = \tilde{c}_0 - \tilde{q}_{1,1} u_1^2$. Repeating this step allows us to recover $\tilde{c}_0$ and all the diagonal elements of $\tilde{Q}$. This demands $p_1 N$ input–output data points altogether, where $p_1 \geq 2$.

The estimation of $\tilde{q}_{2,1}$ can be achieved by taking input–output data sets where u_3 is fixed to zero and u_2 is fixed to a constant $\bar{u}$. A parabola in u_1 results in

$$y(u_1, \bar{u}) = \tilde{c}_1 - \begin{bmatrix} u_1 \\ \bar{u} \end{bmatrix}^T \begin{bmatrix} \tilde{q}_{1,1} & \tilde{q}_{2,1} \\ \tilde{q}_{2,1} & \tilde{q}_{2,2} \end{bmatrix} \begin{bmatrix} u_1 \\ \bar{u} \end{bmatrix}. \quad (11)$$

The extremum of Eq. (11) is reached when $u_1 = -(\tilde{q}_{1,2}/\tilde{q}_{1,1})\bar{u}$. Fitting this latter linear relation allows us to compute $\tilde{q}_{1,2}$. This demands $p_3 \geq 3$ input–output data points for $p_2 \geq 1$ different fixed values of $\bar{u}$. This step needs to be repeated $N(N-1)/2$ times. Altogether, $\tilde{c}_0$ and $\tilde{Q}$ can be estimated using $p_1 N + p_2 p_3 N(N-1)/2$ input–output data points.

One shortcoming of such a methodology is that the total amount of necessary measurements can be large (see Section 6). Measurements in each input–output data set are only exploited for estimating a subset of the parameters instead of all the parameters at once. Most importantly, this procedure cannot ensure that the resulting matrix $\tilde{Q}$ be semidefinite, as predicted by the theoretical derivations [24,27,28,30–34]. Indeed, noise in the measurements of $\tilde{y}$ and numerical errors can lead to computing an indefinite $\tilde{Q}$. A more robust estimation of $\tilde{Q}$ (see Section 6) is achieved including the semidefinite constraint in the estimation. In this way indefinite matrices are excluded *a priori*.

B. Data-Driven Identification Procedure

In [44], we first proposed using semidefinite programming [45] for estimating $\tilde{Q}$ and $\tilde{c}_0$. This allows us to recast the computation of the parameters into a single mathematical optimization. The constraint $\tilde{Q} \succeq 0$ is also satisfied.

Metric (9) can be slightly generalized by including a linear term. This allows to relax the assumption that no aberration is present during the experimental computation of the parameters. Hence Eq. (9) is redefined as

$$y(k) = c_0 + \mathbf{c}_1^T (\mathbf{x} - \mathbf{u}(k)) - (\mathbf{x} - \mathbf{u}(k))^T Q(\mathbf{x} - \mathbf{u}(k)) + w(k), \quad (12)$$

where $c_0 \in \mathbb{R}$, $\mathbf{c}_1 \in \mathbb{R}^N$, $Q \in \mathbb{R}^{N \times N}$, $Q \succeq 0$. Term $w(k)$ represents the uncertainty in approximating $\tilde{y}$ with Eq. (12), and as such it cannot be measured by definition. Quantities c_0, $\mathbf{c}_1$ and Q are the new set of parameters that must be estimated.

Again, as in Section 3, we temporarily assume that $\mathbf{x} = \mathbf{0}$. A collection of input vectors $\{\mathbf{u}(k) \in \mathbb{R}^N | k = 1, \ldots, D\}$ is applied as the input to the deformable mirror and the corresponding measurement of $\tilde{y}$ is recorded. This results into the identification data set $\{(y(k), \mathbf{u}(k)) | k = 1, \ldots, D\}$. Such a collection of input vectors can be selected arbitrarily, in contrast with the methods proposed in [31,34], where specific pupil functions must be generated. In addition all the data points are used at once to estimate all the parameters.

Minimal fitting error could be attained by solving a linear least-squares problem [46]. However, such an approach does not guarantee that the constraint $Q \succeq 0$ be satisfied. For this purpose, the following constrained optimization problem is defined:

$$\min_{c_0, \mathbf{c}_1, Q} \sum_{k=1}^{D} |y(k) - (c_0 - \mathbf{c}_1^T \mathbf{u}(k) - \mathbf{u}(k)^T Q \mathbf{u}(k))|^2$$
$$\text{s.t.} \quad Q \succeq 0, \quad (13)$$

where $c_0 \in \mathbb{R}$, $\mathbf{c}_1 \in \mathbb{R}^N$, and $Q \in \mathbb{R}^{N \times N}$ are decision variables. Optimization (13) belongs to the realm of semidefinite programming [45]. A convenient tool for formulating Eq. (13) is the modeling suite YALMIP [47]. The widely used numerical solver SeDuMi [48] is employed to solve Eq. (13).

In [44], we included a regularization condition on Eq. (13). This leads to a matrix Q, which is strictly positive definite. Such a constraint, however, should be removed as some modes naturally correspond to a null space in Q. Examples of these modes are the piston mode if Zernike polynomials are used, or high-frequency aberrations that deform the Airy disk while keeping the encircled energy stationary. A better approach is to include regularization only when inverting Q, for instance truncating negligible singular values.

The computational complexity of SeDuMi is a function of the number of scalar decision variables and of the number of rows in the total linear matrix inequality constraint [49]. This latter is related to D in Eq. (13). A QR factorization can be used in order to compress the second-order cone constraint implicit in the cost function of Eq. (13). For this purpose, it is convenient to rewrite the least-squares cost function in Eq. (13) as

$$\min_{\mathbf{z}} \|\mathbf{b} - A\mathbf{z}\|_2^2, \quad (14)$$

where $\mathbf{b} = [y(1) \quad \ldots \quad y(D)]^T$,

$$A = \begin{bmatrix} 1 & -\mathbf{u}(1)^T & -\mathbf{u}(1)^T \otimes \mathbf{u}(1)^T \\ \vdots & \vdots & \vdots \\ 1 & -\mathbf{u}(D)^T & -\mathbf{u}(D)^T \otimes \mathbf{u}(D)^T \end{bmatrix} \in \mathbb{R}^{D \times (1+N+N^2)}, \quad (15)$$

and $\mathbf{z} = [c_0 \quad \mathbf{c}_1^T \quad \text{vec}(Q)^T]^T$. Here, function $\text{vec}(\cdot)$ denotes the vectorization transformation and $\otimes$ the Kronecker product [46]. Assuming the rank r of A is $1 + N + N^2$, let $A = \bar{Q}[\bar{R}^T \quad 0^T]^R$ be the QR factorization of A, where $\bar{Q} \in \mathbb{R}^{D \times D}$ is an orthogonal matrix and $\bar{R} \in \mathbb{R}^{r \times r}$ is an upper triangular matrix. Left multiplication by $\bar{Q}^T$ inside the norm in Eq. (14) lead to the following equivalent formulation for Eq. (13):

$$\min_{\mathbf{z}} \|\mathbf{b}_1 - \bar{R}\mathbf{z}\|_2^2 \text{s.t.} \qquad Q \succcurlyeq 0, \qquad (16)$$

where $\mathbf{b}_1 = [I_{r \times r} \quad 0]\bar{Q}^T\mathbf{b}$. Solving Eq. (16) is preferable to Eq. (13) as that leads to a smaller semidefinite programme (see Section 6).

4. ABBERATION CORRECTION FOR QUADRATIC APPROXIMATE METRICS

With the parameters c_0, $\mathbf{c}_1$, and Q known, we now discuss correcting for the unknown aberration $\mathbf{x}$. In the scanning microscope example, this situation corresponds to introducing the specimen and recording the fluorescence emitted from a given focal volume.

A. Independent Parabolic Optimization Algorithm

We briefly outline the correction method proposed in [30] and used in [27,28,30–33]. For simplicity, we take both $\mathbf{c}_1$ and $w(\cdot)$ to be zero. Since $Q = Q^T$, there exists an orthogonal matrix V such that $Q = V\Delta V^T$, where Δ is a diagonal matrix. Let $\mathbf{z} = V^T\mathbf{x}$ and $\mathbf{p} = V^T\mathbf{u}$ and let z_i, p_i, $\lambda_{i,j}$ denote respectively the elements of $\mathbf{z}$, $\mathbf{p}$ and Δ. Equation (12) can be rewritten as

$$y(\mathbf{p}) = c_0 - \sum_{i=1}^{N} \lambda_{i,i}(z_i - p_i)^2. \qquad (17)$$

The diagonalization of Q is referred to as "linear crosstalk removal" in [34] and has the purpose of reformulating the N-dimensional optimization of Eq. (12) into N independent one-dimensional parabola optimizations. If Q is semidefinite, a global optimum of Eq. (17) is found by composing the result of the one-dimensional optimizations. Consider the ith parabola optimization. As $\mathbf{p}$ is the independent variable, we can set $p_j = 0$ for $i \neq j$ so that optimizing Eq. (17) results in

$$\max_{p_i} \alpha_1 p_i^2 + \alpha_2 p_i + \alpha_3, \qquad (18)$$

where the coefficients α_1, α_2, α_3 are unknown as they depend on $\mathbf{z}$. If three measurements are taken—y_1 for $p_i = -b$, y_2 for $p_i = 0$, and y_3 for $p_i = b$, where $b \in \mathbb{R}_+$ is a bias—a Vandermonde system can be solved giving

$$\begin{cases} \alpha_1 = (y_1 - 2y_2 + y_3)/(2b^2) \\ \alpha_2 = (y_1 - y_3)/(2b) \\ \alpha_3 = y_2 \end{cases} . \qquad (19)$$

Consequently the extremum of the parabola is found by setting $p_i = -\alpha_2/(2\alpha_1)$, which evaluates to

$$-b(y_1 - y_3)/(2y_1 - 4y_2 + 2y_3) \qquad (20)$$

(Eq. (33) in [30]). If the measurement for $p_i = 0$ is shared among all the modes, this requires a minimum of $2N + 1$ measurements.

We note that in this way not all information derived from $Q = V\Delta V^T$ has been exploited. In fact, α_1 is known *a priori* as the opposite of the ith eigenvalue of Q, namely, $-\lambda_{i,i}$. Henceforth, only two coefficients α_2 and α_3 are unknown for each mode. Sharing one measurement among all the

modes, one sees that $N + 1$ measurements are sufficient to exactly optimize Eq. (17).

We illustrate this fact by examining Eq. (17) for $N = 1$. Taking 0 and $\bar{p} \neq 0$ for the independent variable p, we have

$$\begin{cases} y(0) & = c_0 - \lambda z^2 \\ y(\bar{p}) & = c_0 - \lambda z^2 + 2\lambda \bar{p}z - \lambda \bar{p}^2 \end{cases} . \qquad (21)$$

Considering the difference $\delta := y(\bar{p}) - y(0)$ between two measurements of $\tilde{y}$, we have $p^* = (\delta + \lambda \bar{p}^2)/(2\lambda \bar{p})$. A generalization to N dimensions is reported in the next section. Note that this result is achieved via a closed-form expression, i.e., without resorting to an approximate solution as was proposed in [26–28].

B. Linear Least-Squares Optimization

In this section we provide formulas for the exact optimization of Eq. (12) in a minimum of $N + 1$ measurements of $\tilde{y}$. In our proposed solution, there is no need to diagonalize Q and the optimization is solved simultaneously for all the modes in a linear least-squares sense.

The system is excited with M input vectors $\{\mathbf{u}(k) \in \mathbb{R}^N | k = 1, ..., M\}$, where $M \in \mathbb{N}_+$, $M \geq 2$, and the rank of $[\mathbf{u}(1) \ ... \ \mathbf{u}(M)]$ is $\min(N, M)$. The corresponding output values are collected $\{y(k) \in \mathbb{R} | k = 1, ..., M\}$.

Define $\delta y(k, l) := y(k) - y(l)$. Then, we obtain

$$\begin{aligned} \delta y(k, l) = &-\mathbf{c}_1^T(\mathbf{u}(k) - \mathbf{u}(l)) + 2(\mathbf{u}(k) - \mathbf{u}(l))^T Q\mathbf{x} \\ &+ \mathbf{u}(l)^T Q\mathbf{u}(l) - \mathbf{u}(k)^T Q\mathbf{u}(k) + w(k) - w(l). \end{aligned} \qquad (22)$$

It can be seen that Eq. (22) is linear in the unknown $\mathbf{x}$. By stacking a number of such expressions together, a linear set of equations in $\mathbf{x}$ is established.

From $\{y(k) \in \mathbb{R} | k = 1, ..., M\}$, $M - 1$ values of $\delta(\cdot, \cdot)$ can be computed, i.e., $\delta(M, 1), ..., \delta(M, M - 1)$. The linear set of equations in $\mathbf{x}$ is arranged as

$$F_M\mathbf{x} + L_M\mathbf{e}_M = \mathbf{d}_M, \qquad (23)$$

where

$$F_M = \begin{bmatrix} 2(\mathbf{u}(M) - \mathbf{u}(1))^T Q \\ \vdots \\ 2(\mathbf{u}(M) - \mathbf{u}(M - 1))^T Q \end{bmatrix} \in \mathbb{R}^{(M-1) \times N}, \qquad (24)$$

$L_M = [-I_{(M-1) \times (M-1)} \quad \mathbf{1}_{M-1}]$ ($\mathbf{1}_{M-1}$ denoting a vector of ones), $\mathbf{e}_M^T = [w(1) \quad ... \quad w(M)] \in \mathbb{R}^M$, and

$$d_M = \begin{bmatrix} \delta y(M,1) + \mathbf{c}_1^T(\mathbf{u}(M) - \mathbf{u}(1)) - \mathbf{u}(1)^T Q\mathbf{u}(1) + \mathbf{u}(M)^T Q\mathbf{u}(M) \\ \vdots \\ \delta y(M,M-1) + \mathbf{c}_1^T(\mathbf{u}(M) - \mathbf{u}(M-1)) \\ -\mathbf{u}(M-1)^T Q\mathbf{u}(M-1) + \mathbf{u}(M)^T Q\mathbf{u}(M) \end{bmatrix}$$
$$\in \mathbb{R}^{M-1}. \qquad (25)$$

The solution $\hat{\mathbf{x}}_M$ of the weighted least-squares problem

$$\min_{\mathbf{x}} \mathbf{e}_M^T \mathbf{e}_M$$
$$\text{s.t.} \qquad F_M \mathbf{x} + L_M \mathbf{e}_M = \mathbf{d}_M \qquad (26)$$

satisfies the normal equation

$$(F_M^T W_M F_M)\hat{\mathbf{x}}_M = F_M^T W_M \mathbf{d}_M, \qquad (27)$$

where $W_M = (L_M L_M^T)^{-1}$.

We consider the case where the set of equations (23) is not underdetermined, i.e., $M \geq N + 1$. Between time $k = 1$ and $k = N + 1$ inclusive, the system is excited with an $N + 1$ input sequence. In Section 6, the column vectors of $b[\mathbf{0} \quad V]$ were used for such an input sequence, where b is a bias and V a base of eigenvectors of Q as in Section 4. At time $k = N + 2$, output $y(N + 1)$ has been acquired and $\hat{\mathbf{x}}_{N+1}$ is computed. Hence, an estimate of the input $\mathbf{u}(N + 2)$ maximizing the intensity $y(N + 2)$ is

$$\mathbf{u}(N + 2) = \hat{\mathbf{x}}_{N+1} - \frac{1}{2}Q^{-1}\mathbf{c}_1. \qquad (28)$$

Similarly, in the following time instants, a refined estimation of $\mathbf{x}$ is obtained by solving the overdetermined system, where $M > N + 1$. The control law for sample time $k > N + 1$ is

$$\mathbf{u}(k) = \hat{\mathbf{x}}_{k-1} - \frac{1}{2}Q^{-1}\mathbf{c}_1. \qquad (29)$$

Note that if the additive noise $w(k) \in \mathbb{R}$ is assumed to be white noise with zero mean and covariance $E[w(k)w(j)] = \delta(k - j)$, where $\delta(\cdot)$ is the unit pulse, then $\hat{\mathbf{x}}_{N+1}$ is the minimum variance unbiased linear estimator of $\mathbf{x}$ (see [46] for further details). The choice of fixing index k in Eq. (22) to M is arbitrary, and different arrangements are possible as long as at least $M - 1$ nonzero values of $\delta y(\cdot, \cdot)$ can be computed.

Equation (12) models the image quality metric for small aberrations, e.g., when the approximations made in Eq. (4) are valid (see Section 6). For larger aberrations, the error in modeling $\tilde{y}$ with a quadratic polynomial is not negligible and therefore applying Eq. (29) fails to maximally suppress the aberration. Nevertheless, experimental evidence suggests that applying multiple iterations of Eq. (29) enables us to gradually suppress large amounts of aberration. This experimental observation is also reported in [24]. One feasible approach is to implement a window-based aberration correction. Let i denote the ith iteration. Each iteration consists of a data acquisition part followed by a correction part. In the first part, input–output data is collected by exciting the system with input vectors taken from a neighborhood centred at the estimate of $\mathbf{x}$ in the previous iteration. In the second part, aberration correction takes place by solving Eq. (26) over the input–output data acquired in the first part. This process is illustrated in Fig. 2.

5. EXPERIMENTAL SETUP

Experimental validation was performed with the system depicted in Fig. 3. As was done in [24,28,30], a deformable mirror was simultaneously used as the source of the aberration and as the correcting element. Aberrations were generated
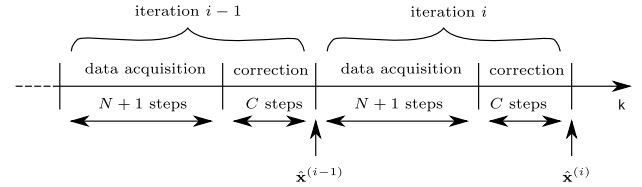

Fig. 2. Timeline of the iterative aberration correction algorithm. Each iteration consists of a data acquisition part where $N + 1$ data points are acquired and a correction part where correction is performed using Eq. (29) for C time instants. The vectors in the ith data acquisition are taken from a neighborhood of the estimate of $\mathbf{x}$ at iteration $i - 1$, i.e., $\hat{\mathbf{x}}^{(i-1)}$.

by adding an offset $\mathbf{x}$ to the control signal of the deformable mirror. Assuming $\mathbf{x}$ to be unknown, an aberration correction algorithm was subsequently applied. The performance of the aberration correction was quantitatively assessed by measuring the residual aberration with a Shack–Hartmann wavefront sensor. Obviously, the measurements from the photodiode were the only information that was supplied to the aberration correction algorithms. With reference to Fig. 3, light from a He–Ne laser source (632.8 nm wavelength) is spatially filtered and collimated using lens L_1 (11 mm), pinhole P_1 (30 μm), lens L_2 (500 mm), and I_1 in order to fill two thirds (10 mm) of the membrane of the deformable mirror DM (MMDM37, OKOTech, The Netherlands). The pupil is demagnified and reimaged by L_3 (200 mm) and L_4 (100 mm) onto the hexagonal microlens array MLA (127 microlenses, 18 mm focal distance, 300 μm pitch, OKOTech, The Netherlands). The image of the microlens array is recorded with the camera C_1 (svs340, 648×492 pixels, 7.4 μm pixel size, SYS-VISTEK, Germany). The beam splitter BS_2 divides light between the Shack–Hartmann wavefront sensor and the pinhole-photodiode
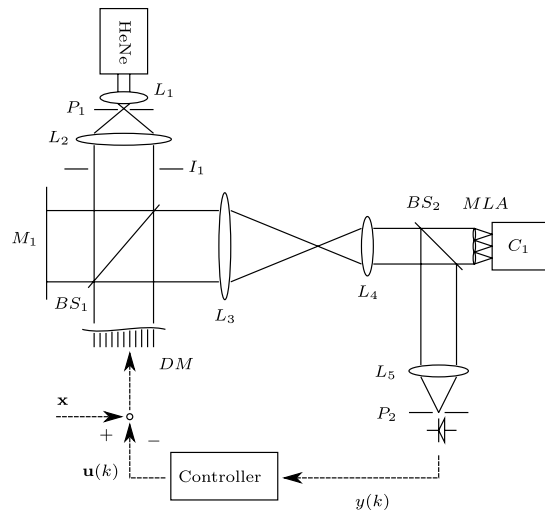

Fig. 3. The spherical wavefront is generated by spatially filtering a laser beam with lens L_1 and pinhole P_1. The beam is collimated by lens L_2 and clipped by iris I_1 to fill 10 mm of the aperture of the deformable mirror DM. The membrane of DM is reimaged by lenses L_3 and L_4 onto a microlens array MLA. C_1 and MLA implement a Shack–Hartmann wavefront sensor. Lens L_5 focuses the beam onto a photodiode that is covered by a pinhole P_2. Flat mirror M_1 is used to calibrate the Shack–Hartmann wavefront sensor. An aberration is introduced as an unknown offset $\mathbf{x}$ to the control signal of DM. An aberration correction experiment consists of suppressing $\mathbf{x}$ when only the measurements of the photodiode are available. Afterward, a measurement of the residual aberration is obtained with the wavefront sensor to assess the performance of the correction.

sensor. Lens L_5 (200 mm) focuses the beam onto the photo-diode (TSL250R-LF, TAOS, Korea), which is covered by the 50 μm pinhole P_2. Voltage to the electrodes of the deformable mirror is supplied by a high-voltage amplifier (OKOTech, The Netherlands) with 40 channels. An external power supply provides 150 V to the high-voltage amplifier. The system is operated using a desktop PC running Linux. The high voltage amplifier is controlled with a 16 bit analog output card (PD2-AO-96/16A, United Electronic Industries, United States). Voltage from the photodiode is acquired with a 16 bit analog input card (PCI-6220, National Instruments, United States). A framegrabber card (Leonardo CL Full, Arvoo, The Netherlands) is used to acquire images from camera C_1. Customized software written in C and MATLAB (Version R2011a, The MathWorks, United States) is used to perform the experiments.

A modal wavefront reconstruction method was implemented [50] using the first 15 Zernike polynomials defined and enumerated as in [41]. We estimated with a least-squares fit a linear relationship between the square root of the voltage applied to the electrodes of the deformable mirror and the Zernike coefficients [51]. As done in [24,31], 11 Zernike coefficients (Z_5 to Z_{15}, see [41]) were controlled, so that $N = 11$ in the previous formulas. As suggested by OKOtech, bidirectional operation of the deformable mirror was achieved by slightly misaligning L_3 to compensate for the defocus that is introduced when the mirror is biased [51].

6. EXPERIMENTAL RESULTS

Experimental validation has been performed using the system described in the previous section.

A. Comparison of the Identification Procedures for the Approximate Metric

A comparison was made between Débarre's experimental computation method (see Section 3) and our proposed procedure (see Section 3). First the system was initialized by flattening the deformable mirror with the Shack–Hartmann wavefront sensor. Subsequently, the Nelder–Mead simplex method [23] was briefly applied to correct for relative misalignment between the Shack–Hartmann wavefront sensor and the pinhole-photodiode detector. In this way, any residual aberration was removed from the system so that both **x** and c_1 are zero in Eq. (12). From this initial condition, Débarre's method and Eq. (16) were repeatedly applied to compute c_0 and Q.

The result of each identification experiment was assessed by evaluating the variance accounted for (VAF). This is defined as VAF($\mathbf{y}, \hat{\mathbf{y}}$) = max (1 − (var($\mathbf{y} - \hat{\mathbf{y}}$)/var($\mathbf{y}$)), 0) × 100%, where the elements of vector **y** are taken from real output measurements of the system, whereas the elements of vector $\hat{\mathbf{y}}$ are the corresponding output predictions computed using the right-hand side of Eq. (12). If the VAF is 100% for one identification experiment, then that indicates that metric $\tilde{y}$ has been perfectly modeled. Forty different identification experiments were executed. In each, an input–output data set was collected and both Débarre's procedure and Eq. (16) were applied in order to compute c_0 and Q. The VAF was computed using this identification data. Afterward, a second input–output data set was collected in order to perform a cross validation. In this second set, the input was randomly selected. The VAF was then computed using this cross-validation data.

In Fig. 4(a), the mean value, maximum, minimum, and standard deviation of the VAFs computed with the identification data sets are reported. These show that using Eq. (16) guarantees a higher mean value for the VAF with respect to Débarre's procedure. The maximum values for the VAFs show comparable performance between the two procedures. Instead the minimum values and standard deviations for the VAF show that a robust performance is achieved by including the semidefinite constraint in the identification procedure. In fact an indefinite matrix Q was recovered using Débarre's procedure in 32 out of the 40 trials. In Fig. 4(b), the mean value, maximum, minimum, and standard deviation of the VAFs computed with the cross-validation data sets are reported. These numbers also support that our identification procedure produces an accurate result. Figure 4(c) shows the mean value of Q for both identification procedures. As expected, both matrices have large elements in the diagonal. Nonetheless, some cross-talk elements are also present due to the finite size of the detection pinhole (see [52]). Whereas the diagonal elements are quite similar for both identification procedures, differences are found in the off-diagonal elements. The consequences of employing an inaccurate matrix Q in the correction of aberrations have already been discussed elsewhere [24,34].

B. Empirical Analysis of the Quadratic Approximation

Optimization (16) can also be used to empirically study the region of validity within which metric $\tilde{y}$ can be approximated with Eq. (12). For this purpose, a large input–output data set with 50,000 tuples was recorded where the maximum input aberration was 2 rad rms. Afterward, optimization (16) was solved by choosing 10 different subsets of the identification data. In each subset, the maximum rms for the input aberrations, denoted by $\rho_{\max}$, was increased and a quarter of the tuples were reserved for validation. The resulting identification and validation VAFs are reported in Fig. 5. It can be seen that between 0.4 and 0.6 rad rms, the effect of the modeling error becomes appreciable and $\tilde{y}$ begins to deviate from its value as predicted by a quadratic polynomial.

Experimentally, we found that employing an iterative aberration correction scheme (e.g., the one depicted in Fig. 2) provided a better performance instead of solving Eq. (26) for $M \gg N + 1$. This empirical observation is in agreement with Fig. 5. Assuming that after each iteration the residual aberration is decreased, then from Fig. 5 one can see that the VAF computed with the input–output tuples within each iteration increases. Henceforth, the input–output tuples of the previous iterations should not be reused to solve Eq. (26) as they are fitted less accurately by Eq. (12).

An additional observation regards the selection of the bias b in the initial $N + 1$ excitation sequence (see Section 4). Using a small bias (e.g., <0.5 rad rms) is preferable as this ensures that the data points collected during the $N + 1$ data acquisition (see Fig. 2) have a similar VAF.

C. Aberration Correction using the Quadratic Approximate Metric

We report a number of aberration correction experiments where we compare our proposed method with the model-based 3N and 5N algorithms (see [24] and Section 4) and the model-free Nelder–Mead simplex method [23]. For the model-based algorithms, i.e., the 3N, 5N, and our proposed
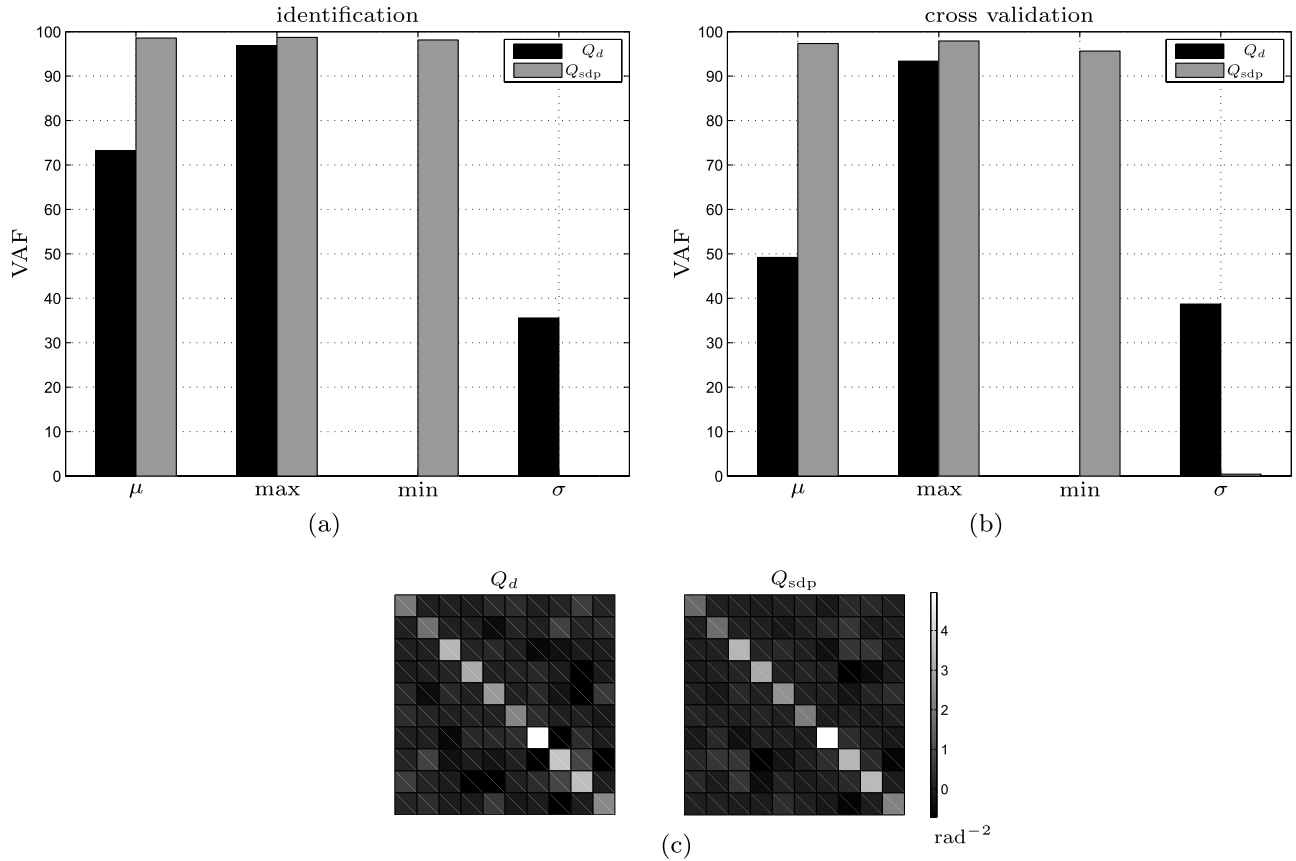
(a) (b) (c)

Fig. 4.   Comparison of the experimental computation of matrix Q with Débarre's method (see Section 3) and our proposed procedure (see Section 3). First a data set of 15246 input–output tuples is acquired ($p_1 = 21$, $p_2 = 13$, and $p_3 = 21$) and Q is computed with Débarre's method, resulting in Q_d. The same input–output data set is used to compute Q with Eq. (16), resulting in Q_{sdp}. The VAFs for Q_d and Q_{sdp} are computed over the identification data set. Subsequently, a new input–output data set with 15,000 tuples is acquired for cross validation. In this second set the input aberrations are chosen randomly. The VAFs for Q_d and Q_{sdp} are computed using this latter validation set. Such steps are repeated 40 times. (a) Mean value μ, maximum max, minimum min, and standard deviation σ of the identification VAFs for Q_d and Q_{sdp}. (b) Mean value μ, maximum max, minimum min, and standard deviation σ of the cross-validation VAFs for Q_d and Q_{sdp}. (c) Mean value of Q_d and Q_{sdp} over the 40 realizations. The color map is scaled to the maximum and minimum of the elements of Q_{sdp} in order to preserve contrast between the two matrices. Matrix Q_d resulted indefinite 32 times out of the 40 trials.
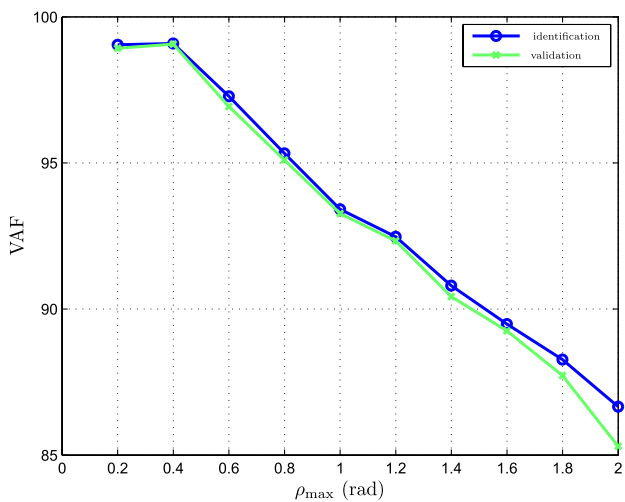

Fig. 5.   (Color online) Optimization (16) was solved for 10 different input–output data sets where the maximum rms of the input aberration (ρ_{max}) is linearly increasing up to 2 rad rms. The VAF is reported for both identification (3750 data points) and validation (1250 data points). Between 0.4 and 0.6 rad rms, the difference between $\tilde{y}$ and Eq. (12) becomes noticeable.

solution, the same parameters c_0, $\mathbf{c}_1$, and Q were used. Such parameters were identified by collecting an input–output data set with 6000 tuples. The input aberrations were randomly chosen with a maximum rms of 0.5, as motivated in Section 6. The data set was split into two sets, reserving 4500 tuples for identification and 1500 tuples for validation. Identification was completed in a couple of seconds by solving Eq. (16) using YALMIP [47] and SeDuMi [48]. Note that SeDuMi took less time to solve Eq. (16) instead of Eq. (13) (the ratio between the two computation times was 0.17). A VAF of 98.27% and 98.16% was found respectively for identification and validation.

Figures 6(a)–6(d) show the results of the correction of random aberrations with magnitudes of 0.3, 0.4, 0.6, and 0.8 rad rms. Such magnitudes are of interest in microscopy applications, where moderate amounts of aberrations are corrected at intermediate depth levels as one focuses deep within a sample (see Section 5 of [24]). For each figure, 50 random aberrations were generated, where the Zernike coefficients of each aberration were randomly selected so that the resulting rms phase profile had a determined value. In the upper plot, the initial aberration and the final residual aberrations are plotted. The circles, vertical bars, and horizontal bars
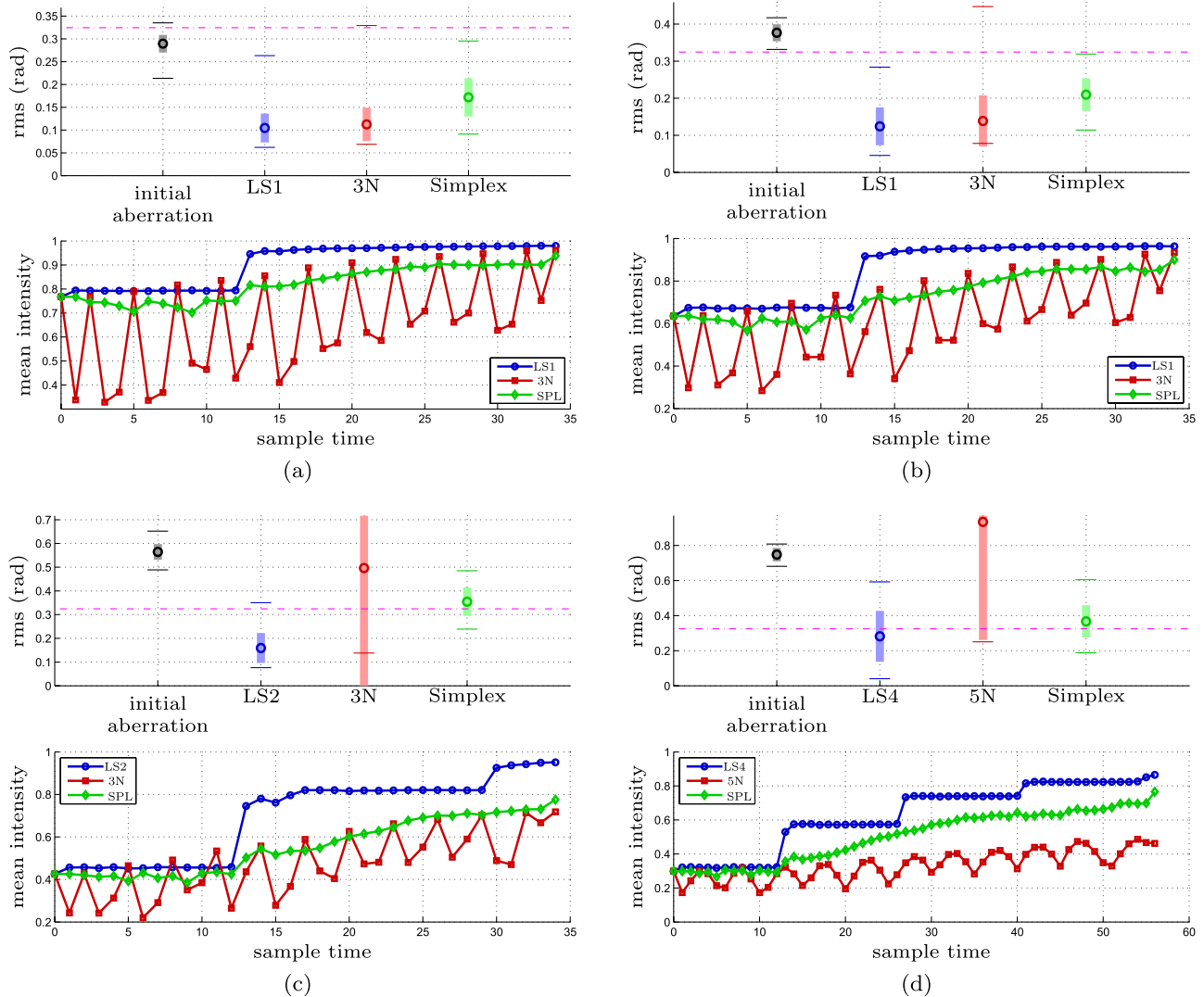
Fig. 6. (Color online) Each figure reports a summary of the correction of a set of 50 random aberrations. In the upper plot, the mean value, standard deviation, minimum, and maximum of the residual aberrations after the correction are reported in radians. These are denoted, respectively, by a circle, a thick vertical bar, and thin horizontal lines. The same indicators are also reported for the random initial aberrations before correction. (LS1), (LS2), and (LS4) denote respectively 1, 2, and 4 iterations of Eq. (29) as depicted in Fig. 2. (3N) and (5N) are described in Section 4 and [24]. (Simplex) and (SPL) denote the Nelder–Mead simplex method [23]. The horizontal dashed–dotted magenta line denotes a Strehl ratio of 0.9. The lower plots report the mean value of the normalized intensity against sample time for the 50 aberration correction experiments.

denote respectively the mean value, standard deviation, and minimum and maximum for the correction of the 50 aberrations. The horizontal dashed–dotted magenta line denotes the rms corresponding to a Strehl ratio of 0.9. The lower plot reports the mean value of the intensity against sample time for the 50 correction experiments. The vertical axis is normalized to the maximum intensity recorded when no aberration is applied.

Figure 6(a) reports a summary of the correction of 50 random aberrations of 0.3 rad rms with one iteration of our proposed method (LS1) [see Eq. (29)], the 3N (3N) algorithm (see [24] and Section 4) and the model-free Nelder–Mead simplex method [23] (Simplex). The residual aberration is comparable for the three algorithms. Nevertheless, a value of 0.95 for the normalized intensity is already achieved at sample time 13 for (LS1), whereas the other two algorithms reach about 0.94 at sample time 34. A bias of 0.5 rad is used for the 3N algorithm, as suggested in [24] for small aberrations.

Instead, a smaller bias of 0.02 rad was used for our proposed solution, as motivated in Section 6. Such a small bias can be advantageous in scanning image acquisition processes, as it leads to less excitation of the dynamics of the deformable mirror and to a smoother variation of the image as aberration correction is being applied.

Figure 6(b) shows a summary of the correction of 50 random aberrations of 0.4 rad rms with one iteration of our proposed method (LS1), the 3N algorithm (3N), and the simplex method (Simplex). Also in this case the two model-based approaches outperform the model-free one. Lower mean values of the residual aberrations are reached for both (LS1) and (3N). In addition, the average normalized intensity is also higher than in (Simplex). This time, at sample time 13, (LS1) reaches a normalized mean intensity of 0.91, which is exceeded by (3N) only at sample time 32. (Simplex) instead stops at 0.9 at sample time 34. Note that both in Fig. 6(a) and in Fig. 6(b), the mean intensity of (LS1) does not improve

significantly after the first correction is applied at sample time $N + 2$, as was discussed in Section 6.

Figure 6(c) reports a summary of the correction of 50 random aberrations of 0.6 rms with two iterations of our proposed method (LS2) (see Fig. 2), the 3N algorithm (3N), and the simplex method (Simplex). In this case, the 3N algorithm is underperforming. A similar behavior was also reported in [24] (see Fig. 4 therein). Nevertheless, a robust performance is shown by (LS2), as can be seen by the reduced standard deviation of the residual aberration and the mean value of the normalized intensity.

In Fig. 6(d), 50 random aberrations with a magnitude of 0.8 rms were corrected. Here, four iterations of our proposed method (LS4), the 5N (5N) algorithm (see [24]), and the simplex method [23] (Simplex) are applied. The performance of (5N) is worse than what is reported in [24] (see, for instance, Fig. 5 in [24]). Such a discrepancy can be explained by three facts. Firstly, in [24], both the imaging system and the selected metric were different from our case. Secondly, a maximum bias of 0.5 rad was used, instead of the suggested maximum of 2 rad as in [24]. We found that the deformable mirror could not reproduce such a large deformation without saturating the actuators or producing an inaccurate phase profile, especially for coma, spherical aberration, and second-order astigmatism. Finally, in [24], the square root of a Lorentzian curve was used instead of Eq. (12) (see also the discussion in the next section). Nevertheless, on average, (LS4) reaches a Strehl ratio higher than 0.9 (see the horizontal dashed–dotted line in the upper plot) and outperforms (Simplex).

D. Aberration Correction using Nonquadratic Approximate Metrics

As seen in Section 6, approximating $\tilde{y}$ with a quadratic polynomial is accurate within a restricted neighborhood of the global maximum of $\tilde{y}$. Nonetheless, empirically, it was found that a broader range of aberrations could be encompassed when using Gaussian or Lorentzian functions to approximate $\tilde{y}$ (see [31,33] and [24,30,32], respectively). Neglecting issues related to numerical computations and the measurement noise, the quadratic polynomial is recovered by assuming that $t(\tilde{y}(k)) \approx c_0 + \mathbf{c}_1^T(\mathbf{x} - \mathbf{u}(k)) - (\mathbf{x} - \mathbf{u}(k))^T Q(\mathbf{x} - \mathbf{u}(k))$, where $t(\cdot)$ is the logarithm when Gaussian functions are used. The use of such an output transformations to recover the quadratic polynomial follows what has been done in [24,30–33].

In Fig. 7 we report a summary of the corrections of 50 random aberrations with a magnitude of 1.0 rad rms. Here a Gaussian function was used to model $\tilde{y}$ and a new identification was performed by solving Eq. (16). A comparison is made between three iterations of our proposed solution (LS3E) and the simplex method (Simplex). It can be seen that (LS3E) outperforms (Simplex), since a lower mean value is achieved for the residual aberration and the mean intensity is consistently higher than the one of (Simplex) after sample time 11. A study of the amount of aberration that can be corrected with model-based approaches is found in [24] and is not worth repeating here. Similar results should be expected if our aberration correction algorithm is employed.

7. CONCLUSION

In this paper, a new experimental procedure to compute the parameters of quadratic approximate metrics in wavefront sensorless adaptive optics has been presented. Such metrics are applicable to a broad spectrum of different imaging techniques [24,26–34]. Our proposed procedure has been shown to produce a more robust computation of the parameters with respect to existing procedures [31,34]. Arbitrary input–output data can be used without the need to generate specific pupil functions as in [31,34]. An additional benefit is found in the possibility to empirically study the region of applicability of the quadratic approximate metric.

A second contribution is found in the algorithm used for aberration correction. Formulas have been provided where aberration correction is achieved by exactly optimizing the quadratic approximate metric using a closed-form expression in a minimum of $N + 1$ measurements. Since these expressions are given for a quadratic polynomial in its most general form, they are widely applicable [24,26–34] and they represent an improvement to previously employed algorithms that required a minimum of $2N + 1$ measurements [24,30–34]. Finally, our arguments have been corroborated by experimental validation in a laboratory environment.
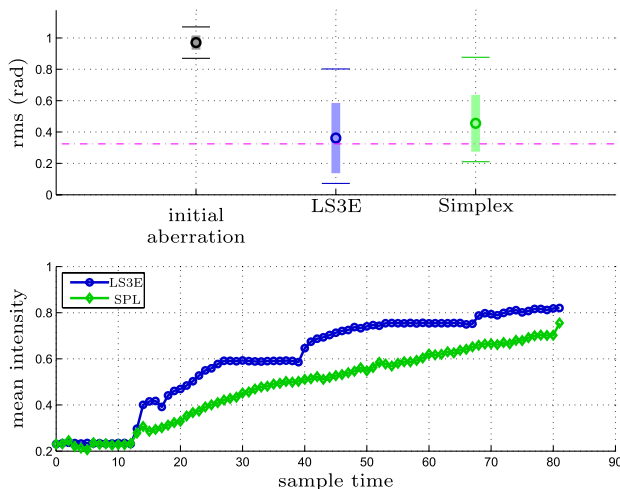
ACKNOWLEDGMENTS

Fig. 7. (Color online) Summary of 50 random aberration correction experiments. The same conventions as in Fig. 6 are employed to report the results. In this case, a Gaussian function instead of Eq. (12) was used to model $\tilde{y}$, as outlined in Section 6. (LS3E) denotes three iterations of Eq. (29) as depicted in Fig. 2.

REFERENCES

1. J. W. Hardy, *Adaptive Optics for Astronomical Telescopes* (Oxford University, 1998).
2. M. J. Booth, "Adaptive optics in microscopy," Philos. Trans. Ser. A **365**, 2829–2843 (2007).
3. O. Azucena, J. Crest, J. Cao, W. Sullivan, P. Kner, D. Gavel, D. Dillon, S. Olivier, and J. Kubby, "Wavefront aberration

measurements and corrections through thick tissue using fluorescent microsphere reference beacons," Opt. Express **18**, 17521–17532 (2010).

4. J. W. Cha, J. Ballesta, and P. T. C. So, "Shack–Hartmann wavefront-sensor-based adaptive optics system for multiphoton microscopy," J. Biomed. Opt. **15**, 046022 (2010).

5. M. Feierabend, M. Rückel, and W. Denk, "Coherence-gated wave-front sensing in strongly scattering samples," Opt. Lett. **29**, 2255–2257 (2004).

6. S. A. Rahman and M. J. Booth, "Adaptive optics for high-resolution microscopy: wave front sensing using back scattered light," Proc. SPIE **8253**, 82530I (2012).

7. T. van Werkhoven, H. Truong, J. Antonello, R. Fraanje, H. Gerritsen, M. Verhaegen, and C. Keller, "Coherence-gated wave-front sensing for microscopy using fringe analysis," Proc. SPIE **8253**, 82530E (2012).

8. R. Aviles-Espinosa, J. Andilla, R. Porcar-Guezenec, O. E. Olarte, M. Nieto, X. Levecq, D. Artigas, and P. Loza-Alvarez, "Measurement and correction of in vivo sample aberrations employing a nonlinear guide-star in two-photon excited fluorescence microscopy," Biomed. Opt. Express **2**, 3135–3149 (2011).

9. X. Tao, B. Fernandez, O. Azucena, M. Fu, D. Garcia, Y. Zuo, D. C. Chen, and J. Kubby, "Adaptive optics confocal microscopy using direct wavefront sensing," Opt. Lett. **36**, 1062–1064 (2011).

10. A. Leray and J. Mertz, "Rejection of two-photon fluorescence background in thick tissue by differential aberration imaging," Opt. Express **14**, 10565–10573 (2006).

11. L. Murray, J. C. Dainty, and E. Daly, "Wavefront correction through image sharpness maximisation," Proc. SPIE **5823**, 40–47 (2005).

12. P. Marsh, D. Burns, and J. Girkin, "Practical implementation of adaptive optics in multiphoton microscopy," Opt. Express **11**, 1123–1130 (2003).

13. A. J. Wright, D. Burns, B. A. Patterson, S. P. Poland, G. J. Valentine, and J. M. Girkin, "Exploration of the optimisation algorithms used in the implementation of adaptive optics in confocal and multiphoton microscopy," Microsc. Res. Tech. **67**, 36–44 (2005).

14. O. Albert, L. Sherman, G. Mourou, T. B. Norris, and G. Vdovin, "Smart microscope: an adaptive optics learning system for aberration correction in multiphoton confocal microscopy," Opt. Lett. **25**, 52–54 (2000).

15. W. Lubeigt, S. P. Poland, G. J. Valentine, A. J. Wright, J. M. Girkin, and D. Burns, "Search-based active optic systems for aberration correction in time-independent applications," Appl. Opt. **49**, 307–314 (2010).

16. L. Sherman, J. Y. Ye, O. Albert, and T. B. Norris, "Adaptive correction of depth-induced aberrations in multiphoton scanning microscopy using a deformable mirror," J. Microsc. **206**, 65–71 (2002).

17. S. P. Poland, A. J. Wright, and J. M. Girkin, "Evaluation of fitness parameters used in an iterative approach to aberration correction in optical sectioning microscopy," Appl. Opt. **47**, 731–736 (2008).

18. L. Murray, "Smart optics: wavefront sensor-less adaptive optics—image correction through sharpness maximisation," Ph.D. thesis (National University of Ireland, 2007).

19. N. Ji, D. E. Milkie, and E. Betzig, "Adaptive optics via pupil segmentation for high-resolution imaging in biological tissues," Nat. Methods **7**, 141–147 (2009).

20. J. R. Fienup and J. J. Miller, "Aberration correction by maximizing generalized sharpness metrics," J. Opt. Soc. Am. A **20**, 609–620 (2003).

21. M. A. Vorontsov, "Decoupled stochastic parallel gradient descent optimization for adaptive optics: integrated approach for wave-front sensor information fusion," J. Opt. Soc. Am. A **19**, 356–368 (2002).

22. G. Vdovin, "Optimization-based operation of micromachined deformable mirrors," Proc. SPIE **3353**, 902–909 (1998).

23. J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright, "Convergence properties of the Nelder–Mead simplex method in low dimensions," SIAM J. Optim. **9**, 112–147 (1998).

24. A. Facomprez, E. Beaurepaire, and D. Débarre, "Accuracy of correction in modal sensorless adaptive optics," Opt. Express **20**, 2598–2612 (2012).

25. A. J. Wright, S. P. Poland, J. M. Girkin, C. W. Freudiger, C. L. Evans, and X. S. Xie, "Adaptive optics for enhanced signal in CARS microscopy," Opt. Express **15**, 18209–18219 (2007).

26. M. J. Booth, "Wave front sensor-less adaptive optics: a model-based approach using sphere packings," Opt. Express **14**, 1339–1352 (2006).

27. M. J. Booth, "Wavefront sensorless adaptive optics, modal wavefront sensing, and sphere packings," Proc. SPIE **5553**, 150–158 (2004).

28. M. J. Booth, "Wavefront sensorless adaptive optics for large aberrations," Opt. Lett. **32**, 5–7 (2007).

29. H. Linhai and C. Rao, "Wavefront sensorless adaptive optics: a general model-based approach," Opt. Express **19**, 371–379 (2011).

30. D. Débarre, M. J. Booth, and T. Wilson, "Image based adaptive optics through optimisation of low spatial frequencies," Opt. Express **15**, 8176–8190 (2007).

31. D. Débarre, E. J. Botcherby, M. J. Booth, and T. Wilson, "Adaptive optics for structured illumination microscopy," Opt. Express **16**, 9290–9305 (2007).

32. D. Débarre, E. J. Botcherby, T. Watanabe, S. Srinivas, M. J. Booth, and T. Wilson, "Image-based adaptive optics for two-photon microscopy," Opt. Lett. **34**, 2495–2497 (2009).

33. N. Olivier, D. Débarre, and E. Beaurepaire, "Dynamic aberration correction for multiharmonic microscopy," Opt. Lett. **34**, 3145–3147 (2009).

34. D. Débarre, A. Facomprez, and E. Beaurepaire, "Assessing correction accuracy in image-based adaptive optics," Proc. SPIE **8253**, 82530F (2012).

35. J. B. Pawley, ed., *Handbook of Biological Confocal Microscopy* (Plenum, 1995).

36. W. Denk, J. Strickler, and W. Webb, "Two-photon laser scanning fluorescence microscopy," Science **248**, 73–76 (1990).

37. H. Song, R. Fraanje, G. Schitter, H. Kroese, G. Vdovin, and M. Verhaegen, "Model-based aberration correction in a closed-loop wavefront-sensor-less adaptive optics system," Opt. Express **18**, 24070–24084 (2010).

38. J. W. Goodman, *Introduction to Fourier Optics*, 3rd ed. (Roberts & Company, 2004).

39. M. Born and E. Wolf, *Principles of Optics*, 7th ed. (Cambridge University, 1999).

40. T. S. Ross, "Limitations and applicability of the Maréchal approximation," Appl. Opt. **48**, 1812–1818 (2009).

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

42. B. Wang and M. J. Booth, "Optimum deformable mirror modes for sensorless adaptive optics," Opt. Commun. **282**, 4467–4474 (2009).

43. G. Vdovin and M. Loktev, "Deformable mirror with thermal actuators," Opt. Lett. **27**, 677–679 (2002).

44. J. Antonello, R. Fraanje, H. Song, M. Verhaegen, H. Gerritsen, C. U. Keller, and T. van Werkhoven, "Data driven identification and aberration correction for model based sensorless adaptive optics," Proc. SPIE **8436**, 84360S (2012).

45. L. Vandenberghe and S. Boyd, "Semidefinite programming," SIAM Rev. **38**, 49–95 (1996).

46. M. Verhaegen and V. Verdult, *Filtering and System Identification: A Least Squares Approach*, 1st ed. (Cambridge University, 2007).

47. J. Löfberg, "YALMIP : a toolbox for modeling and optimization in MATLAB," in *Proceedings of the 2004 IEEE International Symposium on Computer Aided Control System Design* (IEEE, 2004), pp. 284–289.

48. J. F. Sturm, "Using SEDUMI 1.02, a MATLAB toolbox for optimization over symmetric cones," Optim. Methods Softw. **11**, 625–653 (1999).

49. Y. Labit, D. Peaucelle, and D. Henrion, "SEDUMI INTERFACE 1.02: a tool for solving LMI problems with SEDUMI," in *Proceedings of the 2002 IEEE International Symposium on Computer Aided Control System Design* (IEEE, 2002), pp. 272–277.

50. G.-M. Dai, "Modified Hartmann-Shack wavefront sensing and iterative wavefront reconstruction," Proc. SPIE **2201**, 562–573 (1994).

51. M. Loktev, O. Soloviev, and G. Vdovin, *Adaptive Optics Guide*, 3rd ed. (OKO Technologies, 2008).

52. M. A. A. Neil, M. J. Booth, and T. Wilson, "New modal wave-front sensor: a theoretical analysis," J. Opt. Soc. Am. A **17**, 1098–1107 (2000).