Real-time adaptive optics with pyramid wavefront sensors: part II. Accurate wavefront reconstruction using iterative methods

Victoria Hutterer\textsuperscript{1}, Ronny Ramlau\textsuperscript{1,2}, and Iuliia Shatokhina\textsuperscript{2}

\textsuperscript{1} Industrial Mathematics Institute, Johannes Kepler University, Linz, Austria
\textsuperscript{2} Johann Radon Institute for Computational and Applied Mathematics, Linz, Austria

E-mail: victoria.hutterer@indmath.uni-linz.ac.at

Received 16 July 2018, revised 28 November 2018
Accepted for publication 21 February 2019
Published 21 March 2019

Abstract
In this paper, we address the inverse problem of fast, stable, and high-quality wavefront reconstruction from pyramid wavefront sensor data for adaptive optics systems on extremely large telescopes. For solving the indicated problem we apply well-known iterative mathematical algorithms, namely conjugate gradient, steepest descent, Landweber, Landweber–Kaczmarz and steepest descent-Kaczmarz iteration based on theoretical studies of the pyramid wavefront sensor. We compare the performance (in terms of correction quality and speed) of these algorithms in end-to-end numerical simulations of a closed adaptive loop. The comparison is performed in the context of a high-order SCAO system for METIS, one of the first-light instruments currently under design for the extremely large telescope. We show that, though being iterative, the analyzed algorithms, when applied in the studied context, can be implemented in a very efficient manner, which reduces the related computational effort significantly. We demonstrate that the suggested analytically developed approaches involving iterative algorithms provide comparable quality to standard matrix-vector-multiplication methods while being computationally cheaper.

Keywords: pyramid wavefront sensor, wavefront reconstruction, adaptive optics, iterative methods

(Some figures may appear in colour only in the online journal)
1. Introduction

We consider the problem of fast, stable and highly accurate wavefront correction for large-scale real-time closed loop adaptive optics (AO) systems on extremely large telescopes (ELTs). More specifically, we focus on reconstructing the wavefront \( \Phi \) from pyramid wavefront sensor data \( s = [s_x, s_y] \) with a pre-defined accuracy and within the required time slot to fit the real-time setting. The available measurements \( s \) are related to the wavefront \( \Phi \) by the non-linear integral operator \( P = [P_x, P_y] \) representing the pyramid sensor model. In part I of this paper [38], we extensively studied the physical and mathematical forward models of the pyramid wavefront sensor and the underlying operators. The major aim of this paper is to apply the theory of part I [38] in order to solve the inverse problem of reconstruction the unknown wavefront from the given sensor data. Several approximations of the full Fourier optics based, non-linear pyramid sensor model derived in [35, 38] allow for a fast numerical implementation of the wavefront sensor (WFS) operators, which suggests to apply iterative algorithms for solving the reconstruction problem.

We study and compare the performance (in terms of correction quality and speed) of well-known mathematical algorithms for solving inverse problems, namely conjugate gradient, steepest descent, Landweber, Landweber–Kaczmarz, and steepest descent–Kaczmarz iteration. The comparison is performed within the context of two specific high-order AO systems which are currently under design for the future ELT instruments. One of the systems is the SCAO (single conjugate adaptive optics) module for the METIS instrument [7] on a 39 m telescope equipped with a \( 74 \times 74 \) pyramid wavefront sensor (PWFS) sensing in the near-infrared K-band (at the wavelength \( \lambda = 2200 \) nm). The SCAO system is supposed to control \(~4000\) mirror actuators at frequencies of 500–1000 Hz. The second system of interest is the eXtreme adaptive optics (XAO) module for the EPICS instrument [46] having a \( 200 \times 200 \) PWFS as a core component. For the XAO system a huge amount of \(~30000\) mirror actuators have to be controlled at a frequency of 3 kHz, which is a challenge for the real-time control as the algorithm has to produce the reconstruction within a time of less than 0.33 ms. Both considered systems are expected to provide perfect correction quality resulting in an unprecedented image contrast required for their scientific aims.

Additionally, we compare the performance of our methods to the so far standard wavefront reconstruction algorithms used in these days. These are based on matrix-vector-multiplication (MVM) and invert the most exact Fourier optics model. However, the main drawback of any MVM method is their computational effort, which is demanding for the planned large-scale real-time AO systems. The computational complexity required for setting up the command matrix scales as \( O(N^3) \) with \( N \) being the number of controlled actuators, and the application of this command matrix on the sensor data scales as \( O(N^2) \) [19]. The indicated limitation of MVM methods makes their application on large-scale AO systems, such as the XAO system on the ELT, hardly feasible even on the hardware expected at the time of the telescope launch in around 2024. The steadily growing mirror sizes of future telescope systems imply an immense grow in the computational load of existing algorithms which means that the numerical effort has to be kept in mind when developing new algorithms. We demonstrate in this paper that all the proposed iterative algorithms provide the required high-quality and stable wavefront correction along with the heavily reduced computational complexity compared to standard matrix-vector-multiplication methods. The suggested matrix-free approaches make highly accurate real-time wavefront reconstruction easily feasible even for the XAO system.

Since the pyramid wavefront sensor was first introduced in astronomical AO by Ragazzoni [62], it has been extensively studied in optical test benches [25, 60, 63, 77] and acknowledged to possess several precious characteristics distinguishing it from other types of wavefront
sensors. Among those are the increased sensitivity, adjustable linearity range, and pupil sampling, as well as its ability to sense the segmented piston modes [25, 60], which is gaining a special importance in the era of ELTs inevitably having segmented mirrors. Due to the named advantages, the PWFS is nowadays integrated as baseline on several telescope instruments under design. Among those we can name MICADO [13], HARMONI [28, 57], METIS [7], EPICS [46], and ATLAS [29] on the ELT, SCAO and LTAO systems on the Giant Magellan Telescope (GMT) [17, 26], as well as the NFIRAOS [55, 76] and PFI [54] instruments on the Thirty Meter Telescope (TMT). Furthermore, the application area of the PWFS is not limited to adaptive optics in astronomy. Apart from astronomical observations, the pyramid sensor is also utilized in AO in ophthalmology [11, 16, 18, 43] and microscopy [41, 42]. Hence, the problem of improving the quality of wavefront reconstruction approaches for this type of sensor by using sophisticated mathematical methods has never been more interesting, challenging, and important.

In section 2 we describe algorithms existing for wavefront reconstruction from pyramid sensor data. We sketch their advantages, drawbacks and limitations. We proceed with recalling the theoretical principles of wavefront sensing using the pyramid sensor in section 3 and mention details on the discretization of the sensor in section 4. Afterwards, we describe several iterative algorithms, namely the conjugate gradient method for the normal equation and the steepest descent method, Landweber iteration as well as Kaczmarz type algorithms in section 5. This section contains details on the numerical implementation of the involved operators as well. Section 6 presents the performance of the proposed algorithms and a comparison with respect to the achieved reconstruction quality. Finally, in section 7 we evaluate and compare the computational complexities of the analyzed approaches. Both, the reconstruction quality and the speed of the algorithms are additionally compared versus those of an MVM approach.

2. Existing algorithms

For the PWFS, several types of reconstruction approaches have been considered so far [35, 40]. Frequently used are interaction-matrix-based methods which were already mentioned in section 1.

A more computationally efficient algorithm, the Fourier transform reconstructor (FTR), with a complexity of $O(N \log N)$, was suggested in [61]. The method was developed for a simplified geometrical optics based model of the roof WFS, which assumes a large amount of modulation applied to the sensor. Under these assumptions the sensor data are modeled as the derivative of the phase, similar to the Shack–Hartmann (SH) sensor. In the reported algorithms, the authors applied SH Fourier domain filters for wavefront reconstruction. A slightly worse performance of the SH-based FTR compared to MVM was demonstrated for an 8 m telescope. However, since the geometrical model is valid only for large modulations, this method requires further research on a modal optimization, which becomes especially important in the case of ELTs.

The idea of applying inverse Fourier domain filters for wavefront reconstruction from pyramid sensors was further developed in [72, 74], where the authors reported on two methods—convolution with the linearized inverse filter (CLIF) with the complexity $O(N^{3/2})$ and the pyramid Fourier transform reconstructor (PFTR) with $O(N \log N)$ complexity. Both algorithms use a precise Fourier optics forward model allowing the derivation of more exact Fourier filters connecting the incoming phase to the sensor measurements.

The (to our knowledge) fastest available reconstruction algorithm is the preprocessed cumulative reconstructor with domain decomposition (P-CuReD) [73, 75]. Based on an
analytical relation in the Fourier domain, pyramid data are transformed into SH-like data and subsequently inverted using CuReD [67, 68] which was originally developed for SH sensors and has already been tested on-sky [4, 5]. This algorithm provides in end-to-end numerical simulations the same or even better quality results than the standard interaction-matrix-based approaches and scales with a complexity of \( O(N) \).

Further algorithms are based on the fact that the wavefront reconstruction from the full pyramid sensor model can be simplified to an inversion of the finite Hilbert transform. The finite Hilbert transform reconstructor (FHTR) algorithm proposed in [71] uses the direct inversion formula of the finite Hilbert transform and a second method called singular value type reconstructor (SVTR) [35, 37] is based on an analytical singular value expansion of the finite Hilbert transform operator. Both algorithms have the complexity \( O(N^{3/2}) \).

3. Pyramid/roof wavefront sensors

In the following, we recall the theory discussed in part I of the paper [38]. Since all algorithms are based on an approximation of the full pyramid sensor, or more precisely on a linearization of the roof sensor, we will only focus on the properties of the latter and mention corresponding adjoint operators which are needed for the application of the proposed iterative methods. For a precise analysis of the full pyramid sensor model, details about the linearization procedure and proofs we refer the reader to part I of the paper [38].

The analytical models of the non- and modulated pyramid wavefront sensors are complex and therefore difficult to invert directly. However, some assumptions suggested by the physical setting of the model itself, allow to simplify the non-linear Fourier optics based model of the pyramid wavefront sensor \( P = [P_x, P_y] \) (see [38] for the definition) by substituting the pyramidal prism by two orthogonally placed two-sided roof prisms [8, 59, 77]. Due to the physical decoupling of the roof prisms and their orthogonality, the two signal sets \( s = [s_x, s_y] \) provided by the pyramid sensor as

\[
\begin{align*}
  s_x &= -\frac{1}{2}P_x \Phi \\
  s_y &= \frac{1}{2}P_y \Phi
\end{align*}
\]

become independent and contain information about the incoming phase \( \Phi \) only in \( x \)- and \( y \)-direction respectively. Because of symmetry, we only consider the roof sensor operator \( R = [R_x, R_y] \) in \( x \)-direction, i.e. \( R_x \). By interchanging \( x \) and \( y \) all assertions are derived for \( R_y \) accordingly.

For the following considerations, we describe the annular telescope aperture mask by \( \Omega := B(0, D/2) \setminus B(0, r) \subset [-D/2, D/2]^2 \) for a central obstruction \( B(0, r) \) with radius \( r \) and a circular telescope primary mirror \( B(0, D/2) \) according to part I of the paper [38]. Single lines (intervals) of the annular aperture are represented by \( \Omega_x = \{(\tilde{x}, \tilde{y}) \in \Omega, \tilde{x} = x\} \) and \( \Omega_y = \{(\tilde{x}, \tilde{y}) \in \Omega, \tilde{y} = y\} \) for fixed \( x \) and \( y \) respectively and do either consist of one or two parts due to the central obstruction.

3.1. Linearized roof WFS

**Theorem 1.** Under the roof sensor assumption, the PWFS signal corresponding to no, circular and linear modulation of amplitude \( \alpha = \frac{b \lambda}{D} \) with a positive integer \( b \) is approximated by
\[ s_x^{(n,c,l) \text{lin}} = -\frac{1}{2} \left( R_x^{(n,c,l) \text{lin}} \Phi \right)(x,y) \]

with

\[
\left( R_x^{(n,c,l) \text{lin}} \Phi \right)(x,y) := \lambda_0(x,y) \frac{1}{\pi} \int_{\Omega_y} \left\{ \Phi(x',y) - \Phi(x,y) \right\} \cdot k^{(n,c,l)}(x' - x) \, dx',
\]

\[
\left( R_y^{(n,c,l) \text{lin}} \Phi \right)(x,y) := \lambda_0(x,y) \frac{1}{\pi} \int_{\Omega_x} \left\{ \Phi(x,y') - \Phi(x,y) \right\} k^{(n,c,l)}(y' - y) \, dy',
\]

indicating the linearized roof sensor operators and

\[
k^n(x) := 1 \quad \text{(no modulation)},
\]

\[
k^c(x) := J_0(\alpha x) \quad \text{(circular modulation)},
\]

\[
k^l(x) := \text{sinc}(\alpha x) \quad \text{(linear modulation)}.
\]

The modulation parameter is given by \( \alpha = \frac{2\pi}{\lambda} \) and \( J_0 \) denotes the zero-order Bessel function of the first kind.

**Proof.** See [8, 38, 77]. □

The linearized roof sensor operators \( R_x^{(n,c,l) \text{lin}} \) offer a further possibility for simplification of the model due to the splitting

\[
\left( R_x^{(n,c,l) \text{lin}} \Phi \right)(x,y) = \lambda_0(x,y) \left[ \left( L_x^{(n,c,l)} \Phi \right)(x,y) - \Phi(x,y) \cdot \left( L_x^{(n,c,l)} 1 \right)(x,y) \right]
\]

for the integral operators \( L_x^{(n,c,l)} : \mathcal{H}^{1/6}(\mathbb{R}^2) \to \mathcal{L}_2(\mathbb{R}^2) \) defined by

\[
(L_x^{(n,c,l)} \Phi)(x,y) := \frac{1}{\pi} \rho.v. \int_{\Omega_y} \Phi(x',y) k^{(n,c,l)}(x' - x) \, dx',
\]

where \( \rho.v. \) denotes the Cauchy principal value. Note that \( L_x^\alpha \) is the finite Hilbert transform operator. Dropping the second term in (3) leads to the inverse problem

\[ s_x = -\frac{1}{2} L_x^{(n,c,l)} \Phi \]

for pyramid sensor data \( s_x \).

### 3.2. Adjoint operators

All iterative methods for wavefront reconstruction from pyramid sensor data proposed below require the application of adjoint operators. As mentioned, e.g. in [38], it holds that \( \Phi \in \mathcal{H}^{11/6}, \) i.e. the underlying operators are defined as \( L_x^{(n,c,l) \text{lin}} : \mathcal{H}^{11/6} \to \mathcal{L}_2 \) and \( R_x^{(n,c,l) \text{lin}} : \mathcal{H}^{11/6} \to \mathcal{L}_2 \). Together with the embedding operator \( i_x : \mathcal{H}^{11/6} \to \mathcal{L}_2 \) we derive the corresponding adjoint operators by

\[
\left( L_x^{(n,c,l)} \right)^* = i_x^* \left( L_x^{(n,c,l)} \right)^* \]
with \( \left( \mathbf{L}^{(n,e,l)} \right)^* : \mathcal{L}_2 \rightarrow \mathcal{L}_2 \) according to [64]. Hence, it suffices to calculate \( \left( \mathbf{L}^{(n,e,l)} \right)^* \). For simplicity, we use the notation \( \left( \mathbf{L}^{(n,e,l)} \right)^* \) instead of \( \left( \mathbf{L}^{(n,e,l)} \right)^* \) in the following. The roof sensor operators are considered accordingly.

**Proposition 1.** The adjoint operators of the roof sensor and its one-term approximation in \( \mathcal{L}_2 (\mathbb{R}^2) \) are given by

\[
\left( \mathbf{L}_1^{(n,e,l)} \right)^* \Psi (x, y) = -\frac{1}{\pi} \text{p.v.}\int_{\mathcal{O}_l} \Psi(x', y) \cdot k^{(n,e,l)}(x' - x) \frac{dx'}{x' - x},
\]

\[
\left( \mathbf{R}_1^{(n,e,l), \text{fin}} \right)^* \Psi (x, y) = -\frac{1}{\pi} \text{p.v.}\int_{\mathcal{O}_l} \left[\Psi(x', y) + \Psi(x, y)\right] \cdot k^{(n,e,l)}(x' - x) \frac{dx'}{x' - x}.
\]

**Proof.** See [38].

### 4. The discrete pyramid wavefront sensor

The full continuous measurements \( s_j(x, y) \) and \( s_k(x, y) \) of the pyramid wavefront sensor are not available in practice. For the description of the discrete pyramid sensor we perform a division of the continuous two dimensional process into finitely many equispaced regions called subapertures. The data are then assumed to be averaged over every subaperture which corresponds to the finite sampling of the pyramid sensor. Note that in reality, the subaperture grid is predefined by the sensor’s physics. Following the approach in [77], we examine the sensor data as functions evaluated in the (discrete) middle points of the WFS subapertures.

In the two dimensional case we consider quadratic subapertures of size \( d \times d \) with \( d = \frac{D}{\pi} \), where \( D \) represents the telescope diameter, i.e. the primary mirror size, and \( n \) the number of subapertures in one direction.

Note that all considerations are valid for measurements both in \( x \)-direction \( s_j(x, y) \) and \( y \)-direction \( s_k(x, y) \), as well as for non-modulated, circularly, and linearly modulated data. Thus, we consider general measurements identified by \( s(x, y) \). Discretizing \( s(x, y) \) delivers \( n^2 \) data values \( s_{jk} \) with \( j, k = 1, \ldots, n \).

For the following, we use the Dirac comb \( \mathbf{III}_d \) defined as

\[
\mathbf{III}_d (x, y) := \sum_{\ell = -\infty}^{\infty} \sum_{m = -\infty}^{\infty} \delta (x - \ell d, y - m d)
\]

where \( \delta \) denotes the delta distribution.

The continuous signal is captured by the wavefront sensor as follows: first, the average of the measurements per each subaperture is calculated. This is represented as a convolution of the continuous data \( s(x, y) \) with a characteristic function \( \mathcal{X}_{[-d/2, d/2]} \) \( (x, y) \), i.e.

\[
\check{s}(x, y) = \frac{1}{d^2} \int_{x-d/2}^{x+d/2} \int_{y-d/2}^{y+d/2} s(x', y') \, dy' \, dx' = \frac{1}{d^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} s(x', y') \cdot \mathcal{X}_{[-d/2, d/2]} (x - x') \cdot \mathcal{X}_{[-d/2, d/2]} (y - y') \, dy' \, dx' = \frac{1}{d^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} s(x', y') \cdot \mathcal{X}_{[-d/2, d/2]} (x - x', y - y') \, dy' \, dx' = \frac{1}{d^2} \left( s * \mathcal{X}_{[-d/2, d/2]} \right) (x, y).
\]
The discretization is carried out as a application of the Dirac comb \( \mathbf{III}_d \) assuming that the measurements \( s \) fulfill the necessary conditions on applying the distribution \( \delta \). Herewith, we assign a discrete set of measurements \( \mathbf{\bar{s}} \) centered on the subapertures

\[
\mathbf{\bar{s}} = \langle \mathbf{III}_d, \bar{s} \rangle = \left( \sum_{\ell = -\infty}^{\infty} \sum_{m = -\infty}^{\infty} \delta (\cdot - \ell d, \cdot - md), \bar{s} \right)
\]

from floating average values \( \bar{s}(x, y) \) to the discrete set of subaperture middle points \( \{(jd, kd) : j, k \in \mathbb{Z} \} \).

Finally, we restrict the number of measurements to the size of the region captured by the sensor. For several telescope systems the pupil \( \Omega \) is annular instead of circular since a shade created by the secondary mirror prevents reliable measurements on these areas.

Hence, the last step is realized as a multiplication with a second characteristic function

\[
s_{j,k} = \langle (\mathbf{III}_d, \bar{s}) \cdot \chi(\Omega), j_k \rangle
\]

for \( j, k = 1, \ldots, n \). To be precise, we usually consider less than \( n^2 \) measurements due to the annular shape of the aperture and ignore subapertures which are too less illuminated in order to produce reliable data. However, we will not specifically mention this fact throughout the paper.

5. Iterative wavefront reconstruction methods

In this section, we adapt well-known mathematical algorithms, namely the conjugate gradient method for the normal equation, the steepest descent algorithm, and Landweber iteration as well as modifications of these methods coupled with a Kaczmarz strategy to the problem of wavefront reconstruction from pyramid sensor data. By solving the WFS equations in the operator setting instead of forming matrices we minimize the computational complexity of the proposed methods. For wavefront reconstruction we solve the two integral equations

\[
-\frac{1}{2} R_{x}^{(n,c,l)} \cdot \Phi = s_x
\]

(5)

\[
\frac{1}{2} R_{y}^{(n,c,l)} \cdot \Phi = s_y
\]

(6)

with \([R_{x}^{(n,c,l)} \cdot R_{y}^{(n,c,l)}]\) representing a linearization of the roof WFS according to (2) and \( s = [s_x, s_y] \) pyramid sensor measurements. As a further simplification, we consider the inverse problem

\[
-\frac{1}{2} L_{x}^{(n,c,l)} \cdot \Phi = s_x
\]

(7)

\[
\frac{1}{2} L_{y}^{(n,c,l)} \cdot \Phi = s_y
\]

For simplicity of notation we use \( Q := \frac{1}{2} \cdot [R_{x}^{(n,c,l)} \cdot R_{y}^{(n,c,l)}] \) in the following since the basic idea is the same for all types of modulation. Moreover, we only concentrate on solving the inverse problem (5) and (6), but mention that solutions of (1) and (7) can be calculated accordingly.

To specify the representation of the incoming phase \( \Phi \) and the measurements \( s \) we consider an \( n \times n \) pyramid wavefront sensor and denote the number of subapertures in one row/column.
by \( n \). There are various possible representations for the phase and the measurements, e.g. Zernike polynomials or bilinear spline functions. We choose a representation that guarantees maximum computational efficiency, and thus assume that the incoming phase and the measurements are piecewise constant on the subapertures, i.e.

\[
\Phi(x, y) = \sum_{i=1}^{n} \phi_i \mathcal{X}_{\Omega_i}(x), \quad s_i(x, y) = \sum_{i=1}^{n} s_{ix} \mathcal{X}_{\Omega_i}(x),
\]

where \((\phi_i)_{1 \leq i \leq n}, (s_{ix})_{1 \leq i \leq n}\) denote basis coefficients and \( \Omega_i = [y_{i-1}, y_i] \) the \( i \)th subaperture of a row for fixed \( x \). As the wavefront sensor provides two measurements (one in \( x \)- and one in \( y \)-direction), for every single subaperture, the suggestion of representing the measurements via piecewise constant functions describing the subaperture grid is reasonable. We calculate the involved operators as, e.g.

\[
(Q, \Phi)(x, y) = -\mathcal{X}_0(x, y) \frac{1}{2\pi} \int_{\Omega_0} \frac{[\Phi(x', y) - \Phi(x, y)] k^{(x, y)}(x' - x)}{x' - x} \, dx' \\

= -\mathcal{X}_0(x, y) \frac{1}{2\pi} \int_{\Omega_0} \sum_{l=1}^{n} \phi_l \mathcal{X}_{\Omega_l}(x') \sum_{l=1}^{n} \phi_l \mathcal{X}_{\Omega_l}(x) k^{(x, y)}(x' - x) \, dx' \\

= -\mathcal{X}_0(x, y) \frac{1}{2\pi} \sum_{l=1}^{n} \phi_l \mathcal{X}_{\Omega_l}(x) \sum_{l=1}^{n} \phi_l \mathcal{X}_{\Omega_l}(x') k^{(x, y)}(x' - x) \, dx' + \mathcal{X}_0(x, y) \sum_{l=1}^{n} \mathcal{X}_{\Omega_l}(x) k^{(x, y)}(x' - x) \, dx' \\

= -\mathcal{X}_0(x, y) \frac{1}{2\pi} \sum_{l=1}^{n} \phi_l \alpha^{(x, y)}_l(x, y), \quad \alpha^{(x, y)}_l(x, y) = \frac{1}{2\pi} \int_{\Omega_0} \mathcal{X}_{\Omega_l}(x') k^{(x, y)}(x' - x) \, dx'.
\]

The functions \( \alpha^{(x, y)}_l(x, y) \) are computed offline and do not influence the computational speed of the proposed methods. The implementation of all involved operators is performed analogously when choosing the basis representation (8).

Now, we focus on concrete wavefront reconstruction algorithms for pyramid sensor data using iterative methods. In particular, we consider the conjugate gradient algorithm for the normal equation (CGNE), the steepest descent (SD), and Landweber iteration. Since the pyramid sensor provides two measurements \( s_x \) and \( s_y \), the above named approaches deliver two solutions \( \Phi = [\Phi_x, \Phi_y] \), one in \( x \)- and one in \( y \)-direction. The final reconstruction \( \Phi_{rec} \) is then computed as the average of the two temporary solutions. An alternative combination of the two measurements \( s_x \) and \( s_y \) using Kaczmaz loops is investigated in section 5.4. Note that the theory of the presented algorithms is mainly based on [21, 53]. The considered norms are the \( L_2 \)-norms.

5.1. CGNE approach

The conjugate gradient (CG) method is one of the most powerful algorithms for solving self-adjoint, positive (semi-)definite linear equations [6, 21, 31, 32, 33, 45]. For solving the wavefront reconstruction problem we apply the conjugate gradient method to the normal equation

\[
Q^* \Phi = Q^* s.
\]

Let \( Q^\dagger \) denote the Moore–Penrose generalized inverse. The CG-iterates \( \Phi_i \) converge to \( Q^\dagger s \) for all \( s \in \mathcal{D}(Q^\dagger) \) [21] by requiring the fewest iterations among all semiiterative methods.
The CGNE method (algorithm 1) applied to the inverse problem of wavefront reconstruction from pyramid wavefront sensor data is described by:

**Algorithm 1.** CGNE for pyramid sensors.

choose $\Phi_0$, initialize $d_0 = s - Q\Phi_0$, $p_1 = s_0 = Q^*d_0$

for $i = 1, \ldots, K$ do

$q_i = Qp_i$

$\alpha_i = ||s_{i-1}||^2/||q_i||^2$

$\Phi_i = \Phi_{i-1} + \alpha_i p_i$

$d_i = d_{i-1} - \alpha_i q_i$

$s_i = Q^*d_i$

$\beta_i = ||s_i||^2/||s_{i-1}||^2$

$p_{i+1} = s_i + \beta_i p_i$

endfor

$\Phi_{rec} = (\Phi_{x,K} + \Phi_{y,K})/2$

---

5.2. Steepest descent approach

For solving the system (5) and (6) we are additionally interested in the method of steepest descent where we consider different choices of the step sizes in the iterative process. For pyramid sensors, we use the SD method (algorithm 2) applied to the least-squares functional

$$J(\Phi) = ||Q\Phi - s||^2_{L2} \to \min.$$  \hspace{1cm} (10)

The method of steepest descent was originally introduced by Cauchy [9] as one of the most basic procedures to minimize a differentiable functional. A popular step size is determined by an exact line search in the direction of the negative gradient. Alternative choices of the step size have already been considered, e.g. in [69, 70] and will be discussed below for the problem of wavefront reconstruction from pyramid data using the SD method. The gradient of the classical least-squares functional is given by

$$J'(\Phi) = Q^*(Q\Phi - s)$$  \hspace{1cm} (11)

and the resulting algorithm reads as:

**Algorithm 2.** Steepest descent method for pyramid sensors.

choose $\Phi_0$

for $i = 1, \ldots, K$ do

$d_{i-1} = -J'(\Phi_{i-1})$

$\tau_{i-1} = \min_{t \in [0,\infty)} J(\Phi_{i-1} + td_{i-1})$

$\Phi_i = \Phi_{i-1} + \tau_{i-1} d_{i-1}$

endfor

$\Phi_{rec} = (\Phi_{x,K} + \Phi_{y,K})/2$

---

5.2.1. Step size choices and convergence. The speed of convergence of the gradient iteration

$$\Phi_i = \Phi_{i-1} + \tau_{i-1} d_{i-1}$$

$$d_{i-1} = -J'(\Phi_{i-1})$$
depends highly on the choice of the step size $\tau_i$. We consider the classical steepest descent (line search) step size that is defined by

$$
\tau_{i}^{SD} = \min_{t \in [0, \infty)} J(\Phi_i + td_i).
$$

This means that an exact line search is performed in the direction of steepest descent which corresponds to the direction of the negative gradient. For the least-squares functional (10) and corresponding derivative (11) the steepest descent step size with $d_i = -Q^* (Q\Phi - s)$ reads as

$$
\tau_{i}^{SD} = \frac{||d_i||^2}{||Qd_i||^2},
$$

and results in the so-called Cauchy method.

For finite dimensions, the convergence of the Cauchy method was proven for $Q^*Q$ being a (symmetric and positive definite) matrix [1, 27]. More precisely, we have R-linear convergence with convergence factor $\frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max} + \lambda_{\min}}$, where $\lambda_{\max}$ and $\lambda_{\min}$ indicate the largest and smallest eigenvalues of $Q^*Q$.

If we minimize the gradient norm along the search direction, we obtain another line search method for finite dimensions, namely the method of minimal gradient (MG) [15] given by

$$
\tau_{i}^{MG} = \frac{||Qd_i||^2}{||Q^*Qd_i||^2},
$$

From Cauchy–Schwarz inequality it follows $\tau_{i}^{MG} \leq \tau_{i}^{SD}$.

Because of zigzagging between consecutive steps the SD method suffers from slow convergence in some cases. To overcome these effects, a fast and efficient alternative step size choice was introduced by Barzilai and Borwein (BB) in [3]. The BB technique is motivated by quasi-Newton methods and derived from a two-point approximation to the secant equation. There exist two versions of the BB method which are defined by

$$
\tau_{i}^{BB1} = \frac{\langle \Delta \Phi_i, \Delta d_i \rangle}{\langle \Delta d_i, \Delta d_i \rangle} \quad \text{and} \quad \tau_{i}^{BB2} = \frac{\langle \Delta \Phi_i, \Delta \Phi_i \rangle}{\langle \Delta \Phi_i, \Delta d_i \rangle}
$$

with $\Delta \Phi_i = \Phi_i - \Phi_{i-1}$ and $\Delta d_i = d_i - d_{i-1}$. Plugging in the calculations corresponding to algorithm 2 we obtain

$$
\Delta \Phi_i = \Phi_i - \Phi_{i-1} = \tau_{i-1} d_{i-1},
\Delta d_i = d_i - d_{i-1} = -Q^*Q (\Phi_i - \Phi_{i-1}) = -\tau_{i-1} Q^*Q d_{i-1},
\quad d_i = (I - \tau_{i-1} Q^*Q) d_{i-1}
$$

since the involved operators are linear. Therefore, the BB step sizes are rewritten as

$$
\tau_{i}^{BB1} = \frac{||d_{i-1}||^2}{||Q^*Qd_{i-1}||^2} \quad \text{and} \quad \tau_{i}^{BB2} = \frac{||d_{i-1}||^2}{||Qd_{i-1}||^2},
$$

i.e.

$$
\tau_{i}^{BB1} = \tau_{i-1}^{MG} \quad \text{and} \quad \tau_{i}^{BB2} = \tau_{i-1}^{SD}.
$$

The idea is to use additional information of the previous iteration to compute the step size for the current iteration. Once again with the Cauchy–Schwarz inequality we obtain $\tau_{i}^{BB1} \leq \tau_{i}^{BB2}$. The convergence of the BB method in finite dimensions is R-superlinear for quadratic functionals in two dimensions [3]. Generally, while the SD and MG method...
decrease monotonically, the BB step size choices are non-monotone as the error behaves non-monotonously, i.e. \(|\Phi - \Phi_{i+1}| \leq |\Phi - \Phi_i|\) for the true solution \(\Phi\) is not fulfilled for every iteration \(i\). Nevertheless, the BB method converges to a solution of (10) as found in [65].

The Cauchy–Barzilai–Borwein (CBB) step size is based on the idea to use the SD and BB step size alternating. The method, which was introduced in 2003 and is also called alternate step size (AS) gradient method, aims at reducing the zigzag-effect of the Cauchy method [14], and therefore leads to a faster convergence. The promising alternative to the BB method reads as

\[
\tau_i^{\text{CBB}} = \begin{cases} 
\tau_i^{\text{AS}}, & \text{for } i \text{ odd}, \\
\tau_i^{\text{BB2}}, & \text{for } i \text{ even}.
\end{cases}
\]

Due to \(\tau_i^{\text{BB2}} = \tau_i^{\text{SD}} - 1\), we use the same step size twice in two consecutive iterations. For strictly convex quadratic functionals, the AS method is proven to be two-step-Q-superlinearly convergent. Furthermore, an alternate version of the MG method called alternate minimization (AM) gradient method was proposed in [15] having an SD iteration for every second step. Generally, the SD method becomes faster when one non-monotone (e.g. BB) step is made even after several SD steps [78]. In addition, a variety of step size choices have been introduced using combinations or shortened step size versions of the above mentioned options.

To reduce the computational effort of the algorithms, one can use a fixed step size. Step sizes which are tuned heuristically depend mainly on the size of the telescope and the resolution of the wavefront sensor (discretization). For a fixed step size \(\tau_i = \beta\) the steepest descent algorithm for the least-squares functional (10) reduces to the standard Landweber iteration.

5.3. Landweber approach

For the Landweber iteration [48], the normal equation (9) is transformed into the equivalent fixed point equation

\[
\Phi = \Phi + Q^* (s - Q\Phi).
\]

In order to ensure convergence by \(|Q| \leq 1\) we introduce a relaxation parameter \(0 < \beta \leq |Q|^{-2}\) and iterate by

\[
\Phi_i = \Phi_{i-1} + \beta Q^* (s - Q\Phi_{i-1}), \quad i \in \mathbb{N}.
\]

Then, \((\Phi_i)\) converges to a least-squares solution of (5) and (6) for \(s \in \mathcal{D}(Q^*)\) [21].

The Landweber iteration modified for wavefront reconstruction based on pyramid sensor measurements (algorithm 3) reads as:

**Algorithm 3.** Landweber iteration for pyramid sensors.

choose \(\Phi_0\), set relaxation parameter \(\beta\)

for \(i = 1, \ldots, K\) do

\[
\Phi_i = \Phi_{i-1} + \beta Q^* (s - Q\Phi_{i-1})
\]

endfor

\(\Phi_{\text{rec}} = (\Phi_{x,K} + \Phi_{y,K}) / 2\)

Besides the above discussed methods for pyramid sensors, there already exist several algorithms providing two reconstructions, one from data \(s_x\) and one from data \(s_y\) [37, 71, 74].
Since in the reconstructions obtained by averaging the two solutions we experienced distinct horizontal and vertical artifacts, the aim is to combine the reconstructions already during the iteration steps. For this reason we investigate Kaczmarz methods in which the two data sets are used alternating.

5.4. Kaczmarz methods for wavefront reconstruction from pyramid sensor data

For the reconstruction of the incoming wavefront, the pyramid sensor provides two data sets \( s_x \) and \( s_y \). Although wavefront reconstruction could in principle be performed by utilizing only measurements in one direction, there are several facts that support the usage of both data sets. On the one hand, we expect better reconstruction quality in case we utilize more information. This argument is additionally strengthened by the presence of noise in the sensor measuring process. On the other hand, deeper investigations of the underlying operators in \( x \)- and \( y \)-direction show that they have different null spaces, i.e. depending on the underlying model there exist modes that cannot be sensed/reconstructed. For instance, pyramid and roof wavefront sensors are not able to detect a constant added to the incoming phase \( \Phi \). This undetectable constant, called piston mode (mode of order 0), has no influence on the measurements \( s \). In order to characterize effects that are invisible in sensor data we discuss selected wavefront modes (of order 0 and 1) which are elements of the null space of the roof wavefront sensor operators, i.e. phase elements that deliver measurements equal to zero. For the following investigations, we will consider the mathematical forward model of the linearized roof sensor \( R_{\text{lin}} \) and analyze the null spaces of the corresponding operators described by

\[ \mathcal{N}(R_{\text{lin}}) := \{ \Phi \in \mathcal{H}^{11/6}([\mathbb{R}^2]) | R_{\text{lin}}\Phi = 0 \}. \]

We study the response of the linearized roof sensor to a global piston mode shown in figure 1 left. Hence, we define

\[ \Phi_{\text{piston}}(x, y) = c \cdot \mathcal{X}_\Omega(x, y), \]

where \( c \in \mathbb{R} \) is a constant. Furthermore, we analyze how the sensor responses to modes of order 1 called tip and tilt modes (see figure 1 middle and right) represented by

\[ \Phi_{\text{tip}}(x, y) = ax \cdot \mathcal{X}_\Omega(x, y) \quad \text{and} \quad \Phi_{\text{tilt}}(x, y) = by \cdot \mathcal{X}_\Omega(x, y) \]

for \( a, b \in \mathbb{R} \).

**Proposition 2.** Constant functions \( c \cdot \mathcal{X}_\Omega \) with \( c \in \mathbb{R} \) are elements of the null space of the linearized roof sensor operators \( R_{\text{lin}}^{(n,c,l)} \) with \( R_{\text{lin}}^{(n,c,l)} = \begin{bmatrix} R_{l,\text{lin}}^{(n,c,l)} \end{bmatrix} \) defined in (2). Moreover, tip signals \( cx \cdot \mathcal{X}_\Omega(x, y) \) are in the null space of \( R_{\text{lin}}^{(n,c,l)} \) and tilt signals \( cy \cdot \mathcal{X}_\Omega(x, y) \) are in the null space of \( R_{\text{lin}}^{(n,c,l)} \).

**Proof.** Because of the definition of the operators

\[ \left(R_{\text{lin}}^{(n,c,l)} \Phi \right)(x, y) = \mathcal{X}_\Omega(x, y) \frac{1}{\pi} \int_{\Omega} \frac{[\Phi(x', y) - \Phi(x, y)] k^{(n,c,l)}(x' - x)}{x' - x} \, dx' \]

it suffices to consider if the numerator \([\Phi(x', y) - \Phi(x, y)]\) equals zero.
As \[ \Phi_{\text{piston}}(x', y) - \Phi_{\text{piston}}(x, y) = 0 \] and \[ \Phi_{\text{tilt}}(x', y) - \Phi_{\text{tilt}}(x, y) = 0 \] but \[ \Phi_{\text{tip}}(x', y) - \Phi_{\text{tip}}(x, y) \neq 0 \] we conclude that global piston and tilt mode are in the null space of \( R^{\{n, c, l\}}_{\text{lin}} \), and by symmetry global piston and tip mode are in the null space of \( R^{\{n, c, l\}}_{y, \text{lin}} \).

If we further simplify the roof sensor model by excluding the second term in (2) and consider the operators \( L^{\{n, c, l\}} \) defined in (4), we note that these operators are injective, i.e. \( \mathcal{N}(L^{\{n, c, l\}}) = \{0\} \). This assertion follows from the injectivity of the finite Hilbert transform shown in [20].

Note that the results of proposition 2 are directly transferred to the full non-linear roof sensor operators \( R^{\{n, c, l\}} \) discussed in part I of the paper [38]. As soon as we consider functions including various powers of \( x \) or \( y \) in \( \Phi_{\text{tip}} \) or \( \Phi_{\text{tilt}} \), the corresponding measurements in \( x \)- or \( y \)-direction are equal to zero as well. The fact that \( R^{\{n, c, l\}}_{\text{lin}} \) and \( R^{\{n, c, l\}}_{y, \text{lin}} \) have different null spaces intensifies the requirement of an appropriate combination of the two data sets \( s_x \) and \( s_y \) for reconstruction methods which are based on the roof sensor model.

One idea to appropriately combine both data sets is to reconstruct independently in both directions and average the reconstructions at the end as already considered for the CGNE, SD, and Landweber iteration above. However, it is not guaranteed that the final (averaged) solution \( \Phi_{\text{rec}} \) fulfills both equations (5) and (6). Another possibility is to consider \( [Q_x, Q_y] \) as one single operator and a third one is to use a Kaczmarz strategy [44, 56] which is computationally cheaper and for which it is guaranteed that the equations (5) and (6) are fulfilled for the final solution. Kaczmarz methods, in general, have been developed for solving linear systems of equations. We have decided to implement Kaczmarz strategies for the pyramid sensor in combination with several of the above discussed algorithms.

5.4.1 Landweber–Kaczmarz approach. In practice, the Landweber algorithm is used because it is simple and each iteration is cheap. Though, the process usually requires a high number of iterations. Anyway, we do not experience slow convergence for reconstruction from pyramid data due to a close similarity between adjoint and inverse operators as investigated in [37] for the non-modulated sensor, i.e. for the finite Hilbert transform operator. When using proper basis functions for the representation of the incoming wavefront \( \Phi \) and the measurements \( s = [s_x, s_y] \) as derived in (8), the involved operators can be precomputed offline. These facts make the Landweber iteration coupled with a Kaczmarz strategy interesting for wavefront reconstruction from pyramid sensor data. A general convergence analysis of the linear Landweber–Kaczmarz method can be found in [47].
In the linear setting the Landweber–Kaczmarz method for wavefront reconstruction from pyramid wavefront sensor measurements (4) reads as:

**Algorithm 4.** Landweber–Kaczmarz iteration for pyramid sensors.

choose $\Phi_0$, set relaxation parameters $\beta_1, \beta_2$

for $i = 1, \ldots, K$ do

$\Phi_{i,0} = \Phi_{i-1}$  
$\Phi_{i,1} = \Phi_{i,0} + \beta_1 Q_x^* (s_x - Q_x \Phi_{i,0})$  
$\Phi_{i,2} = \Phi_{i,1} + \beta_2 Q_y^* (s_y - Q_y \Phi_{i,1})$  
$\Phi_i = \Phi_{i,2}$

endfor

$\Phi_{rec} = \Phi_K$

**5.4.2. Steepest descent-Kaczmarz approach.** The idea of modified steepest descent algorithms coupled with a Kaczmarz strategy is comparable to the method described in [10] for non-linear problems. As in the previous method, we cyclically consider each measurement equation (5) and (6).

Hence, for

$$J_x(\Phi) := \|Q_x \Phi - s\|_{L_2}^2, \quad J_y(\Phi) := \|Q_y \Phi - s\|_{L_2}^2,$$

(14)

the steepest descent-Kaczmarz (SD-K) method for wavefront reconstruction using pyramid sensors (algorithm 5a) is described by:

**Algorithm 5a.** Steepest descent-Kaczmarz method for pyramid sensors.

choose $\Phi_0$

for $i = 1, \ldots, K$ do

$\Phi_{i-1,0} = \Phi_{i-1}$  
$d_{i-1,1} = -J'_x(\Phi_{i-1,0})$  
$\tau_{i-1,1} = \min_{t \in [0,\infty)} J_x(\Phi_{i-1,0} + td_{i-1,1})$  
$\Phi_{i-1,1} = \Phi_{i-1,0} + \tau_{i-1,1}d_{i-1,1}$  
$d_{i-1,2} = -J'_y(\Phi_{i-1,1})$  
$\tau_{i-1,2} = \min_{t \in [0,\infty)} J_y(\Phi_{i-1,1} + td_{i-1,2})$  
$\Phi_{i-1,2} = \Phi_{i-1,1} + \tau_{i-1,2}d_{i-1,2}$  
$\Phi_i = \Phi_{i-1,2}$

endfor

$\Phi_{rec} = \Phi_K$

During an observation, the reconstructions have to be repeated up to 0.3 milliseconds. Assuming that the incoming wavefront do not change much from one time steps to the next and, in particular, tip and tilt do not change significantly, another idea (implemented in algorithm 5b) would be to reconstruct in $x$-direction for even time steps $t$ and in $y$-direction for the proximate odd time $t + 1$ steps. The big advantage of algorithm 5b consists in the reduction of the computational demand by more than 50% compared to the normal SD approach.
Algorithm 5b. Modified steepest descent-Kaczmarz method for pyramid sensors.

if \((t \mod 2 = 0)\)
  apply algorithm 2 in \(x\)-direction only
else if
  apply algorithm 2 in \(y\)-direction only
endif

The post-loop step of algorithms 1–3, i.e. the averaging of the two reconstructions is not necessary for algorithms 4, 5a, and 5b since we only obtain one reconstruction \(\Phi_K\). Please note that for the Kaczmarz-type methods it is merely necessary to choose one initial guess \(\Phi_0\) instead of two as required for algorithms 1–3.

6. Numerical results

We test the quality of the reconstruction approaches by continuously correcting the incoming wavefront in closed loop AO. In this setting, the wavefront sensor measures the incoming phase after passing the deformable mirror, i.e. the sensor sees the difference between the incoming wavefront and the correction induced by the mirror. For numerical simulations, we use the end-to-end simulation tool Octopus developed by ESO [50, 51]. As already mentioned in the introduction, we test the performance of the proposed methods for an ELT-sized telescope system. In particular, we consider the METIS instrument on the 39 m sized ELT for non-modulated and modulated pyramid wavefront sensors having a \(74 \times 74\) spatial sampling. Although the observing facility has a primary mirror diameter of 39 m, for METIS only the inner 37 m are used. The incoming wavefronts are simulated by a realization of the von Karman atmospheric model having 35 layers. The system runs at a frequency of 1 kHz for the non-modulated sensor and at a frequency of 500 Hz for the modulated sensor. The mirror geometry in the simulations corresponds to the M4 geometry planned for the ELT which was just recently incorporated in Octopus. For the temporal control of the algorithms we use a simple integrator and optimize the gains with a resolution of 0.1.

As a quality measure we use the long-exposure (LE) Strehl ratio, which is computed as the average on-axis Strehl ratio for all performed time steps. The Strehl ratio is defined as the ratio of the peak aberated image intensity from a point source compared to the maximum attainable intensity using an ideal optical system limited only by diffraction over the system’s aperture. The maximum achievable value is 1.

The quality results of the algorithms are expressed in terms of long-exposure Strehl ratios at an observing wavelength of 2.2 \(\mu m\) (K-band). Note that according to the specifications of the METIS instrument, K-band is not included in the science range. Instead, observations are performed in L-band (at \(\lambda_1 = 3.0 \mu m, \lambda_2 = 3.7 \mu m\)), in M-band (at \(\lambda = 4.7 \mu m\)) and in N-band (at \(\lambda = 10.0 \mu m\)). For analysis purposes, however, we find it useful to have the output at a shorter wavelength as well. As such, we use \(\lambda = 2.2 \mu m\) in the K-band where the imaging is performed.

In our numerical tests we evaluate the reconstruction quality in a range of photon flux levels between 50 and 10000 photons per subaperture per frame for median atmospheric conditions. The simulation parameters are summarized in table 1. In order to speed up convergence to the closed loop, in the first 13 time steps we apply the CuReD reconstructor [67, 68], which corrects mainly for the low frequencies in the wavefront.
Before we compare the reconstruction quality of all proposed methods, we investigate the optimal step size choice for the steepest descent algorithm applied to WF reconstruction. For that analysis we consider the METIS instrument on the ELT having a pyramid sensor without modulation incorporated. The simulation parameters are identical to those listed in table 1. As photon flux, we use 10000 photons per subaperture per frame. The reconstruction quality is evaluated after 500 time steps using 5 SD-iterations for each reconstruction in order to find the optimal choice of the step size. As listed in table 2, best results are obtained for the SD iteration combined with the classical steepest descent step size. The reason for the small number of performed iterations is (amongst others) related to the roof sensor approximation for modeling a pyramid sensor and discussed below in more detail.

6.2. Simulated closed loop performance

Let us analyze the closed loop performance of the developed algorithms and compare their reconstruction quality. Our reconstruction methods are all based on a simplification of the full pyramid sensor model. As a consequence, after some iteration steps, the reconstructions suffer from an approximation error and depart from the true solution of the full pyramid sensor model although the residuals

| Table 1. Test case setting. |
|-----------------------------|
| **Simulation parameters**   |
| Telescope diameter          | 37 m                               |
| Central obstruction         | 30%                                |
| Science target              | on-axis (SCAO)                     |
| WFS                         | PWFS                               |
| Sensing band                | K (2.2 µm)                         |
| Evaluation bands            | L (3.0, 3.7 µm)                    |
|                             | M (4.7 µm)                         |
|                             | N (10.0 µm)                        |
| Modulation                  | [0, 4] λ/D                         |
| Controller                  | integrator                         |
| Atmospheric model           | von Karman                          |
| Number of simulated layers  | 35                                 |
| Outer scale L₀              | 25 m                               |
| Atmosphere                  | median                             |
| Fried radius r₀ at λ = 500 nm | 0.157 m                           |
| Number of subapertures      | 74 × 74                            |
| Number of active subapertures | 3912 out of 5476                   |
| Frame rate                  | [1000, 500] Hz                     |
| DM delay                    | 1                                  |
| Detector read-out noise     | 1 electron/pixel                   |
| Background flux             | 0.000 321 photons/pixel/frame      |
| Photon flux                 | [50, 100, 1000, 10 000]            |
| Iterations per simulation   | 500                                |
with respect to the simplified model continue to scale down during the iterations. Due to the fact that the full non-linear pyramid sensor model \( P \) consists of two terms \( P = P_{\text{lin}} + P_{\text{rest}} \), where the first term again contains two terms \( P_{\text{lin}} = R_{\text{lin}} + S_{\text{lin}} \) (see part I of the paper [38] for more details), a reduction of the roof sensor residual (15) can imply an error increase of

\[
S_{\text{lin}} \Phi_{\text{rest}}
\]

in the residual corresponding to the full pyramid sensor model

\[
\left\| s - \frac{1}{2} R_{\text{lin}} \Phi \right\| = \left\| s - \frac{1}{2} R_{\text{lin}} \Phi_{\text{rest}} \Phi \right\|.
\]

Besides the approximation error, another error source, the data error, is present in the reconstruction process. It is inevitable to search for an adequate stopping criterion taking into account both the difference between the real pyramid sensor operator \( P \) providing the measurements \( s \) and the approximate operator \( R_{\text{lin}} \), which builds the foundation of the model-based reconstruction algorithms, as well as data errors. For choosing the regularization parameter in the generally non-linear problem of wavefront reconstruction from pyramid data, we discuss the usage of Morozov’s discrepancy principle. Assume that the pyramid sensor provides noisy data \( s^\delta \) fulfilling \( \|s - s^\delta\| < \delta \) for some noise level \( \delta > 0 \). The iteration is terminated with stopping index \( k_* \) when for the first time the residual is below \( \tau \delta \) for some \( \tau > 1 \), i.e.

\[
\left\| s^\delta - \frac{1}{2} R_{\text{lin}} \Phi_i \right\| > \tau \delta \quad 0 \leq i < k_* \quad \text{and} \quad \left\| s^\delta - \frac{1}{2} R_{\text{lin}} \Phi_{k_*} \right\| \leq \tau \delta.
\]

The discrepancy principle combined with a criterion for controlling the approximation error can be transferred to the application of only a few CGNE- or SD-iterations resulting in a very low value for \( k_* \) as confirmed by a huge number of numerical simulations performed within this study. In particular, one iteration suffices to provide high reconstruction quality when using a warm restart of the system. That is, in the first time step, the initial guess is chosen as zero, i.e. \( \Phi_{0,0} := 0 \), and at time step \( t > 0 \) the initial value is set to the reconstructed phase of the previous step, i.e. \( \Phi_{t,0} := \Phi_{t-1,0} \). By employing the reconstruction of the previous step as initial guess \( \Phi_0 \), we significantly decrease the computational complexity since algorithms 1, 2, 5a and 5b are scaled down to non-iterative gradient based methods by applying only one corresponding iteration step. The warm restart technique improves the convergence speed of the iterative solvers and additionally slightly increases the quality performance. The suitable number of iterations is also depending on the number of incident photons, since a high photon flux results in reduced data noise and vice versa.

---

Table 2. SD-reconstruction (algorithm 2) results for the non-modulated sensor in the K-band after 500 time steps using different step sizes. The used photon flux is 10000 photons per subaperture per frame.

| Step size choice       | LE Strehl ratio |
|------------------------|-----------------|
| Classical SD           | 0.8322          |
| Minimal gradient       | 0.8310          |
| Barzilai–Borwein 1     | 0.8311          |
| Barzilai–Borwein 2     | 0.8316          |
| Cauchy–Barzilai–Borwein 1 | 0.8317       |

---
In our applications a total number of \( K = 1 \) iterations turned out to be optimal with respect to the reconstruction quality and the computational complexity of the method. Except for the Landweber type approaches (algorithms 3 and 4), we use more than one iteration, but already \( K = 5 \) Landweber steps combined with an adapted choice of the relaxation parameter and the warm restart technique are enough to obtain satisfying reconstruction quality.

In case of one CGNE- or SD-iteration the two algorithms coincide when using the classical steepest descent step size (12). Additionally, the step sizes in the SD method discussed in the previous section do not differ for one SD-iterate except for the classical SD step size and the MG step size. Hence, for the numerical simulations with results provided in tables 3 and 4 we used the minimal gradient step size (13) in order to have an additional comparison of step size choices as well. Since algorithm 5b has a reduced computational complexity compared to algorithm 5a, we only consider the modified SD-Kaczmarz algorithm in our numerical tests. As above, one SD-Kaczmarz iteration suffices as well.

It is important to mention here that in our numerical tests we have seen that the reconstruction quality of the presented iterative algorithms strongly depends on the accuracy of integration when computing the involved integral operators. In particular, a significant difference was observed for a large-scale (XAO) and a moderate-scale (METIS) AO systems. From the reconstruction quality point of view, it was crucial for the METIS system to assume a finer data resolution using an interpolation to a finer grid than given by the subaperture spacing. Usage of the subaperture spacing only resulted in a decreased reconstruction quality. On the contrary, for the XAO case subaperture resolution was enough for a good quality and further interpolation did not bring a significant improvement. This may be related to the difference in subaperture sizes of both systems (21 cm in XAO versus 50 cm in METIS).

Corresponding results having a cold start \( (\Phi_0 = 0) \) for every time step \( t \) can be found in table 2 for algorithm 2 utilizing a pyramid sensor without modulation while results using the warm restart technique for all presented algorithms are summarized in table 3 for the non-modulated pyramid sensor and in table 4 for the modulated pyramid sensor. Hence, the warm

| Table 3. Long-exposure Strehl ratios in the K-band obtained with the presented algorithms after 500 closed loop simulation steps for a pyramid sensor without modulation. Best results are obtained for the CGNE approach (algorithm 1), the SD (algorithm 2), and Landweber–Kaczmarz iteration (algorithm 4). |
|---------------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Photon flux | Algorithm 1 | Algorithm 2 | Algorithm 3 | Algorithm 4 | Algorithm 5b |
| 50 | 0.8374 | 0.8376 | 0.8332 | 0.8371 | 0.8331 |
| 100 | 0.8407 | 0.8409 | 0.8384 | 0.8415 | 0.8393 |
| 1000 | 0.8414 | 0.8413 | 0.8395 | 0.8420 | 0.8412 |
| 10000 | 0.8415 | 0.8412 | 0.8396 | 0.8419 | 0.8413 |

| Table 4. Long-exposure Strehl ratios in the K-band obtained with the presented algorithms after 500 closed loop simulation steps for a pyramid sensor with modulation. Here, the CGNE approach (algorithm 1) provides the highest reconstruction quality in most of the cases. |
|---------------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Photon flux | Algorithm 1 | Algorithm 2 | Algorithm 3 | Algorithm 4 | Algorithm 5b |
| 50 | 0.8432 | 0.8434 | 0.8427 | 0.8439 | 0.8340 |
| 100 | 0.8524 | 0.8520 | 0.8517 | 0.8510 | 0.8454 |
| 1000 | 0.8597 | 0.8579 | 0.8590 | 0.8562 | 0.8570 |
| 10000 | 0.8604 | 0.8581 | 0.8595 | 0.8577 | 0.8580 |
restart technique improves the reconstruction quality of the SD approach from an LE Strehl ratio of 0.8322 having a cold start to 0.8412 with the warm restart. In case of zero modulation, best reconstruction quality is obtained for the Landweber–Kaczmarz approach using the two measurement sets alternating and for the gradient based approaches (algorithms 1 and 2) calculating two reconstructions and averaging at the end. For the sensor having modulation $4 \frac{\lambda}{D}$, surprisingly, the CGNE approach even outperforms the Landweber–Kaczmarz algorithm except for the simulations with 50 photons per subapertures per frame. Anyhow, the differences in the results are very small and we conclude that using the Strehl ratio as a quality criterion, all methods provide equivalent reconstruction quality. In addition to the K-band results shown in table 3, we provide the long-exposure Strehl ratios in other science bands as defined by the instrument specifications. Table 5 shows the quality in L-, M-, and N-bands obtained with the Landweber–Kaczmarz algorithm in the high flux case (10 000 ph/subaperture/frame) for the non-modulated sensor.

The simulations for a modulated sensor whose results are presented in table 4 were performed with a frame rate of 500 Hz. In order to have a direct comparison of the non-modulated and modulated sensor, we additionally run a simulation at a frame rate of 1 kHz (instead of 500 Hz) using a pyramid sensor with modulation $4 \frac{\lambda}{D}$. For the modulated sensor with the CGNE method we obtain the LE Strehl ratio of 0.8782 in the K-band in the high flux case after 500 time steps and the LE Strehl ratio of 0.8415 for the non-modulated sensor. This result fits well our previous experiences with other model-based reconstruction algorithms according to which the modulated sensor provides a higher quality compared to the non-modulated one.

All in all, the developed reconstruction algorithms deliver comparable quality and allow for robust and accurate wavefront reconstruction with low computational costs.

### 6.3. Comparison to interaction-matrix-based approaches

In the literature, there are many variants of interaction-matrix-based approaches: statistical estimators or least-squares methods; zonal or modal [49] control approaches (i.e. the degrees of freedom are actuators/subapertures or modes). The least-squares approach applied to the pyramid wavefront sensor reaches high correction accuracy without regularization, at least if the number of degrees of freedom is small as demonstrated in [22–24]. However, for large-scale AO systems the least-squares attempt turned out to be less accurate and the minimum variance estimator allowing statistical regularization is preferred [2, 19] (see also MAP [12, 30, 49, 52], MMSE [2]). One should note that regularization typically requires optimization (fine tuning) of the regularization parameters. Moreover, the regularized control matrix has to be recomputed each time the seeing conditions or the photon flux change, which is a rather time-consuming task. (As already mentioned, the computational complexity required

| Sensing wavelength ($\mu$m) | LE Strehl       |
|-----------------------------|-----------------|
| 2.2                         | 0.8419          |
| 3.0                         | 0.9107          |
| 3.7                         | 0.9401          |
| 4.7                         | 0.9623          |
| 10.0                        | 0.9915          |
for setting up the command matrix scales as \(O(N^3)\) and the application of this command matrix on the sensor data as \(O(N^2)\).

Often, in practice and also in simulations, the non-modulated sensor being operated with an MVM approach, is reported to be unstable, see, e.g. [30, 52]. One can for instance apply some tricks, like using a ‘wrong’ command matrix derived for the modulated sensor, or heavily fine-tune the regularization parameters to filter out the unstable modes in the correct interaction matrix (measured or computed for the sensor with modulation 0), which has to be performed on the fly and is a very time-consuming task.

Recently, there was a modal MAP result published in [52] for the non-modulated sensor running in Octopus with a modal MVM at 1kHz frame rate. The achieved quality in the K-band was reported to be 0.62 for the high flux case (10 000 photons/subaperture/frame). For comparison, the pyramid sensor with modulation \(4\lambda/D\) was reported to provide in the same environment the LE Strehl ratio of 0.80.

As reported in [34], another variant of MVM, the zonal minimum variance reconstructor in the YAO simulation tool [66], which is a zonal regularized approach, achieves LE Strehl of 0.89 for the modulated pyramid sensor (with modulation 4) and high photon flux.

Comparing the performances of the described algorithms, we can draw the following conclusions. For the pyramid sensor without modulation our reconstruction algorithms, which use the forward model of the sensor, allow not only to close the loop easily, but also to achieve a stable correction over time with a quality significantly higher compared to the interaction-matrix-based reconstructor utilized in [52]. In case of the pyramid sensor with the optimal amount of modulation, our algorithms achieve a reconstruction quality which is slightly (only 0.012 points of LE Strehl for the simulations at a frame rate of 1kHz) below the best (known) result obtained with the zonal MMSE variant of MVM.

7. Computational complexity

We define the computational complexity of the algorithm as a number of required floating point operations (flops), where one flop is assumed to be either one complex multiplication or one summation. Let \(n\) denote the number of subapertures in one direction, then \(N = n^2\) indicates approximately the number of unknowns to be found.

7.1. Complexity of Landweber iteration and Landweber–Kaczmarz iteration for pyramid sensors

We only consider the complexity of the operations that have to be performed online and exclude the pre-calculations needed in the application of the operators \(Q\) and \(Q^*\) from our considerations. The number of floating point operations for every step in the Landweber iteration approach for wavefront reconstruction using pyramid sensors (algorithms 3 and 4) is provided in table 6. Computationally the most expensive step is calculating \(Q\Phi\), i.e. the application of the pyramid sensor operator (and of the corresponding adjoint). As this step composes of the multiplication of two \(n \times n\)-matrices, we need \(n^3\) multiplications and \(n^2(n-1)\) summations, thus, altogether \(2n^3 - n^2\) flops.

Since we perform the mentioned operations twice (in \(x\)- and in \(y\)-direction), for \(K\) iterations we obtain the complexity

\[C_4(n; K) = (8n^3 + 2n^2) \cdot K\]

for the Landweber–Kaczmarz approach (algorithm 4) and
$C_3(n; K) = (8n^3 + 2n^2) \cdot K + 2n^2$

flops for the application of the Landweber iteration (algorithm 3) having the additional step of averaging.

7.2. Complexity of SD and SD-Kaczmarz algorithm for pyramid sensors

We again only consider the operations performed online. The complexity of one steepest descent iteration (algorithms 2, 5a and 5b) is indicated in table 7. In case of the classical steepest descent iteration (algorithm 2) a subsequent averaging $\Phi = 1/2(\Phi_x + \Phi_y)$ has to be performed, which costs additionally $2n^2$ flops. Therefore, the number of flops for the steepest descent-Kaczmarz approach applied to pyramid sensors (algorithm 5a) is given by

$$C_{5a}(n; K) = (12n^3 + 8n^2 + 2) \cdot K,$$

for the modified algorithm 5b by

$$C_{5b}(n; K) = (6n^3 + 4n^2 + 1) \cdot K,$$

and for the steepest descent approach (algorithm 2) by

$$C_2(n; K) = (12n^3 + 8n^2 + 2) \cdot K + 2n^2,$$

where $K$ indicates the number of steepest descent steps.

7.3. Complexity of CGNE for pyramid sensors

The CGNE method consists of three steps.

1. A pre-computation and initialization step which have to be done for both $x$- and $y$-direction once.
2. The CG-loop for $K$ iterations performed twice in $x$- and $y$-direction.
3. A post loop step in which we average the two obtained reconstructions.

The number of flops in the CGNE algorithm for pyramid sensors for operations which are not pre-computed offline is indicated in table 8. Summing up the specified operations for both $s_x$ and $s_y$ data, we see that the initialization step consists of $8n^3$ flops, the loop of $(8n^3 + 20n^2 + 4) \cdot K$ and the post loop step of $2n^2$ operations. Hence, the CGNE complexity for pyramid sensors sums up as
However, since the CG method is known to require the fewest number of iterations, \( K \) usually is smaller compared to, e.g. the Landweber iteration.

### 7.4. Comparison to MVM

In our notations, the complexity of standard MVM methods scales as \( \mathcal{O}(N^2) = \mathcal{O}(n^4) \). For these studies, we calculate the reconstructions (mirror actuator commands) at the corners of the subapertures, and thus need to consider approximately \( n' = n + 1 \) phase values. For e.g. \( n = 200 \) subapertures, the complexity of MVM is roughly given by

\[
C_{\text{MVM}}(200) = 201^4 \approx 16 \cdot 10^8 = 1600 \cdot 10^6.
\]

The complexities of the developed methods are estimated in table 9. Here, we assume that the mirror actuators are equidistantly spaced on a squared shape although this is not employed in practice since not all actuators are actively controlled. However, for a theoretical comparison of complexities such assumptions are still relevant. Note that in principle \( C_1(n; 1) = C_2(n; 1) \) because the CGNE algorithm can already be terminated after the calculation of \( \Phi_1 \). However, we consider one full CGNE step in the table. As a remark, we mention that in our comparison we have omitted the additional computational effort required in the presented model-based algorithms for computation of deformable mirror commands from the reconstructed wavefront shape. This step can be represented as a bilinear interpolation from the \( n \times n \) grid of

| Table 7. The number of flops in the steepest descent and steepest descent-Kaczmarz method for pyramid sensors. |
|---------------------------------------------------------------|
| Operation                                      | # of flops     |
| Loop                                          |                |
| \( J'(\Phi_x) \)                                 | \( 4n^3 - n^2 \) |
| Evaluation of \( \tau_{SD} \)                  | \( 2n^3 + 3n^2 + 1 \) |
| \( \Phi_x - \tau_{SD}J'(\Phi_x) \)            | \( 2n^2 \)     |
| Post loop step                                 |                |
| \( \Phi = \Phi_x + \Phi_y \)                    | \( 2n^2 \)     |

| Table 8. The number of flops in the CGNE algorithm for pyramid sensors considering operations which are not precomputed offline. |
|---------------------------------------------------------------|
| Operation                                      | # of flops     |
| Init                                          |                |
| Computation of \( d_{i,0} \)                   | \( 2n^3 \)     |
| Computation of \( s_{i,0} \)                   | \( 2n^3 - n^2 \) |
| Initialization of \( p_{i,1} \)                | \( n^2 \)      |
| Loop                                          |                |
| Computation of \( q_{i} \)                    | \( 2n^3 - n^2 \) |
| Computation of \( \alpha \)                   | \( 4n^2 + 1 \)  |
| Computation of \( \Phi_x \)                   | \( 2n^2 \)     |
| Computation of \( d_{i} \)                    | \( 2n^2 \)     |
| Computation of \( s_{i} \)                    | \( 2n^3 - n^2 \) |
| Computation of \( \beta \)                    | \( 2n^2 + 1 \)  |
| Computation of \( p_{i} \)                    | \( 2n^2 \)     |
| Post loop step                                 |                |
| \( \Phi = \frac{1}{2} (\Phi_x + \Phi_y) \)    | \( 2n^2 \)     |

\[
C_1(n; K) = (8n^3 + 20n^2 + 4) \cdot K + 8n^3 + 2n^2.
\]
subapertures to the \((n + 1) \times (n + 1)\) grid of DM actuators, which requires \(4(n + 1)^2\) flops to be performed. Also, we would like to mention that the time-saving features of MVM approaches like parallelizability and pipelineability are valid in our algorithms as well.

The developed algorithms allow to significantly reduce the numerical effort of the wavefront reconstruction step in an AO loop compared to the computational load related to the solvers based on matrix-vector multiplication. This is illustrated especially well for the XAO system having a huge number of active actuators. The computational effort of MVM-based wavefront estimators is extremely demanding in this case. In contrast, the usage of analytically developed wavefront reconstructors allows one to heavily reduce the numerical effort of the AO loop. For instance as shown in Table 9, the modified steepest descent algorithm reduces the computational load of the wavefront reconstruction step in the XAO loop to approximately 3% of the MVM effort while still providing high reconstruction quality.

Please note that the gain in computational load of iterative solvers compared to MVM-based reconstructors is highly depending on the used number of iterations \(K\). This number, in turn, is depending on atmospheric conditions (noise) or telescope/sensor parameters. However, in closed loop AO the usage of a warm restart of the system allows to keep the value \(K\) small.

For the METIS simulations studied in the context of this paper, a maximum number of 10 iterations was appropriate for all introduced algorithms with the Landweber type methods always needing the highest amount of iterations and the CGNE approach the smallest as confirmed from general theoretical investigations of the iterative algorithms. Also, if the number of iterations is chosen too large, the semi-convergence of the iterative methods will affect the reconstruction quality adversely. Alternatively to choosing a predefined fixed number of iterations, the usage of the discrepancy principle as discussed in section 6.2 or the development of stopping strategies taking sensor or atmospheric parameters into account are possible as well.

8. Conclusion and outlook

In this paper we have studied the application of well-known iterative algorithms for solving the inverse problem of wavefront reconstruction from pyramid wavefront sensor data in the field of astronomical adaptive optics. From the performed end-to-end numerical simulations we can draw the conclusion that all studied algorithms deliver very similar reconstruction

| Approach | Complexity | Flops | METIS \((n = 74)\) | XAO \((n = 200)\) | XAO in (%) |
|----------|------------|-------|-------------------|-----------------|-----------|
| MVM      | \(O(N^2)\) | \(C_{\text{MVM}}(n)\) | 30 \(\cdot\) \(10^6\) | 1600 \(\cdot\) \(10^6\) | 100       |
| CGNE (algorithm 1) | \(O(\sqrt{N}/2)\) | \(C_{\text{CGNE}}(n;1)\) | 6.9 \(\cdot\) \(10^6\) | 130.9 \(\cdot\) \(10^6\) | 8         |
| SD (algorithm 2) | \(O(\sqrt{N}/2)\) | \(C_{\text{SD}}(n;1)\) | 5.1 \(\cdot\) \(10^6\) | 97.9 \(\cdot\) \(10^6\) | 6         |
| Landweber iteration (algorithm 3) | \(O(\sqrt{N}/2)\) | \(C_{\text{Landweber}}(n;5)\) | 16.9 \(\cdot\) \(10^6\) | 325.3 \(\cdot\) \(10^6\) | 20        |
| Landweber–Kaczmarz it. (algorithm 4) | \(O(\sqrt{N}/2)\) | \(C_{\text{Landweber-Kaczmarz}}(n;5)\) | 16.3 \(\cdot\) \(10^6\) | 325.2 \(\cdot\) \(10^6\) | 20        |
| SD-Kaczmarz (algorithm 5a) | \(O(\sqrt{N}/2)\) | \(C_{\text{SD-Kaczmarz}}(n;5)\) | 5.1 \(\cdot\) \(10^6\) | 97.8 \(\cdot\) \(10^6\) | 6         |
| modified SD-Kaczmarz (algorithm 5b) | \(O(\sqrt{N}/2)\) | \(C_{\text{modified SD-Kaczmarz}}(n;1)\) | 2.6 \(\cdot\) \(10^6\) | 48.9 \(\cdot\) \(10^6\) | 3         |
quality. However, it is preferable to apply the Kaczmarz versions of the algorithms or the CGNE approach, since they provide a slightly better reconstruction quality, though, the difference in the achieved quality between all the methods is minor. The best quality is obtained with the CGNE approach (algorithm 1) and with the Landweber–Kaczmarz iteration (algorithm 4), which at the same time is part of the slowest among the algorithms under comparison. If one decides to go for speed at the cost of a negligible quality loss, one should choose the modified steepest descent-Kaczmarz version combined with the classical step size choice (algorithm 5b).

As shown by numerical results presented in this study, the proposed algorithms, which are partially iterative methods, allow to keep the numerical effort of the wavefront reconstruction step in an AO loop low compared to the computational load of solvers based on matrix-vector-multiplication. This has an especially big impact for the considered XAO system having a huge number of active actuators. For instance, the modified steepest descent algorithm reduces the computational load of the wavefront reconstruction step in the XAO loop to approximately 3% of the MVM effort while still providing high reconstruction quality.

Even when using simplifications of the pyramid sensor model, all proposed algorithms provide stable high-quality reconstruction and (almost) reach the quality of interaction-matrix-based approaches in which the full pyramid model is assumed. Especially for the non-modulated sensor, the linear iterative algorithms give stable and very accurate wavefront reconstructions. If we compare the methods presented in this paper with the P-CuReD, all of them are outmatched by the P-CuReD with respect to both speed and quality [35, 40]. Nevertheless, in the proposed iterative methods, there is a possibility to investigate the full pyramid sensor model for future developments. Remarkable quality improvements are hoped for those adaptions. A big advantage of the iterative methods over the P-CuReD is that the full pyramid sensor model or real life features such as telescope spiders or the low wind effect can be incorporated. For the P-CuReD it may even be impossible to adapt the algorithm to a more sophisticated pyramid sensor model.

Finally, we would like to mention that investigations of the behavior of iterative algorithms in the presence of the so-called ‘spiders’ (support structures of the secondary mirror segmenting the telescope pupil into disjoint parts), the application of non-linear iterative methods for wavefront reconstruction as well as further quality evaluations to meet specifications of the METIS instrument will be part of our further research [35, 36, 39, 58].

Acknowledgments

This work has been partly supported by the Austrian Federal Ministry of Science and Research (HRSM) and the Austrian Science Fund (F68-N36, project 5).

ORCID iDs

Victoria Hutterer 🐸 https://orcid.org/0000-0002-8952-2731
Ronny Ramlau 🐘 https://orcid.org/0000-0002-0277-9615

References

[1] Akaike H 1998 On a Successive Transformation of Probability Distribution and its Application to the Analysis of the Optimum Gradient Method (New York: Springer) pp 37–52
De Cezaro A, Haltmeier M, Leitao A and Scherzer O 2008 On steepest-descent-kaczmarz methods

Chamot S R, Dainty C and Esposito S 2006 Adaptive optics for ophthalmic applications using a

Clare R and Le Louarn M 2011 Numerical simulations of an extreme AO system for an ELT

Dai Y-H and Yuan Y-X 2003 Alternate minimization gradient method

Dai Y-H 2003 Alternate step gradient method

Daly E M and Dainty C 2010 Ophthalmic wavefront measurements using a versatile pyramid

Ellerbroek B L 2002 Efficient computation of minimum-variance wave-front reconstructors with

Esposito S 2005 Pyramid sensor for segmented mirror

Esposito S, Pinna E, Puglisi A, Tozzi A and Stefanini P 2005 Pyramid sensor for segmented mirror

Esposito S

Forsythe G E and Motzkin T S 1951 Asymptotic properties of the optimum gradient method Bull. Am. Math. Soc. 57 183

Fusco T, Thatte N, Meimon S, Tecca M, Clarke F and Swinbank M 2010 Adaptive optics systems for HARMONE: a visible and near-infrared integral field spectrograph for the E-ELT Proc. SPIE 7736 773633

Fusco T et al 2010 ATLAS: the E-ELT laser tomographic adaptive optics system Proc. SPIE 7736 77360D

Garcia-Rissmann A and Le Louarn M 2010 Scao simulation results with a pyramid sensor on an elt-like telescope 1st AO4ELT Conf.—Adaptive Optics for Extremely Large Telescopes Proc.

Gilyazov S F 1977 Iterative solution methods for inconsistent operator equations Moscow Univ. Comput. Math. Cybern. 3 78–84
[32] Hanke M 1995 *Conjugate Gradient Type Methods for Ill-Posed Problems* (Harlow: Longman Scientific and Technical)
[33] Hestenes M R and Stiefel E 1952 Methods of conjugate gradients for solving linear systems *J. Res. Natl Bur. Stand.* **49** 409–36
[34] Hippler S, Feldt M, Bertram T, Brandner W, Cantalloube F, Carломagno B, Absil O, Obereder A, Shatokhina I and Stuik R 2018 Single conjugate adaptive optics for the ELT instrument METIS (submitted)
[35] Hutterer V 2018 Model-based wavefront reconstruction approaches for pyramid wavefront sensors in adaptive optics PhD Thesis Johannes Kepler University Linz
[36] Hutterer V and Ramlau R 2018 Nonlinear wavefront reconstruction methods for pyramid sensors using Landweber and Landweber–Kaczmarz iteration *Appl. Opt.* **57** 8790–804
[37] Hutterer V and Ramlau R 2018 Wavefront reconstruction from non-modulated pyramid wavefront sensor data using a singular value type expansion *Inverse Problems* **34** 035002
[38] Hutterer V, Ramlau R and Shatokhina I 2018 Real-time adaptive optics with pyramid wavefront sensors: part I. A theoretical analysis of the pyramid sensor *Inverse Problems* **35** 045007
[39] Hutterer V, Shatokhina I, Obereder A and Ramlau R 2018 Advanced wavefront reconstruction methods for segmented extremely large telescope pupils using pyramid sensors *J. Astron. Telesc. Instrum. Syst.* **4** 049005
[40] Hutterer V, Shatokhina I, Obereder A and Ramlau R 2018 Wavefront reconstruction for ELT-sized telescopes with pyramid wavefront sensors *Proc. SPIE* **10703** 1070344
[41] Iglesias I 2011 Pyramid phase microscopy *Opt. Lett.* **36** 3636–8
[42] Iglesias I and Vargas-Martin F 2013 Quantitative phase microscopy of transparent samples using a liquid crystal display *J. Biomed. Opt.* **18** 026015
[43] Iglesias I, Ragazzoni R, Julien Y and Artal P 2002 Extended source pyramid wave-front sensor for the human eye *Opt. Express* **10** 419–28
[44] Kaczmarz S 1937 Angenäherte Auflösung von systemen linearer Gleichungen *Bull. Int. Acad. Pol. Sci. Lett.* **35** 355–7
[45] Kammerer W J and Nashed M Z 1972 On the convergence of the conjugate gradient method for singular linear operator equations *SIAM J. Numer. Anal.* **9** 165–81
[46] Korkiakoski V and Vérinaud C 2010 Extreme adaptive optics simulations for EPICS *Proc. 1st AO4ELT Conf.* p.03007
[47] Kowar R and Scherzer O 2002 Convergence analysis of a Landweber–Kaczmarz method for solving nonlinear ill-posed problems *J. Ill-Posed Inverse Problems* **23** 253–70
[48] Landweber L 1951 An iteration formula for fredholm integral equations of the first kind *Am. J. Math.* **73** 615–24
[49] Law N F and Lane R G 1996 Wavefront estimation at low light levels *Opt. Commun.* **126** 19–24
[50] Le Louarn M, Vérinaud C, Korkiakoski V, Hubin N and Marchetti E 2006 Adaptive optics simulations for the European extremely large telescope *Proc. SPIE* **6272** 627234
[51] Le Louarn M, Vérinaud C, Korkiakoski V and Fedrigo E 2004 Parallel simulation tools for AO on ELTs *Proc. SPIE* **5490** 705–12
[52] Louarn M L, Madec P Y, Oberti S, Paufique J, Sarazin M, Stroebele S and Esselborn M 2017 Latest AO simulation results for the E-ELT *Poster Presented at AO4ELTS Conf.*
[53] Louis A 1989 *Inverse and Schlecht Gestelle Probleme* (Stuttgart: B.G Teubner)
[54] Macintosh B et al 2006 Extreme adaptive optics for the thirty meter telescope *Proc. SPIE* **6272** 62720N
[55] Mieda E, Rosensteiner M, van Kooten M, Shatokhina I and Obereder A 2016 The adaptive optics modes for HARMONI: from classical to laser assisted tomographic AO *Proc. SPIE* **9909** 990909
[56] Natterer F 1986 *The Mathematics of Computerized Tomography* (New York: Wiley)
[57] Neichel B et al 2016 The adaptive optics modes for HARMONI: from classical to laser assisted tomographic AO *Proc. SPIE* **9909** 990909
[58] Obereder A, Raffetseder R, Shatokhina I and Hutterer V 2018 Dealing with spiders on ELTs: using a pyramid wfs to overcome residual piston effects *Proc. SPIE* **10703** 107031D
[59] Phillion D W and Baker K 2006 Two-sided pyramid wavefront sensor in the direct phase mode *Proc. SPIE* **6272** 627228
[60] Pinna E, Quiroso-Pacheco F, Esposito S, Puglisi A and Stefanini P 2007 Signal spatial filtering for co-phasing in seeing-limited conditions *Opt. Lett.* **32** 3465–7
[61] Quirós-Pacheco F, Correia C and Esposito S 2010 Fourier transform wavefront reconstruction for the pyramid wavefront sensor 1st AO4ELT Conf.—Adaptive Optics for Extremely Large Telescopes Proc. vol 07005 (EDP Sciences) pp 1–6
[62] Ragazzoni R 1996 Pupil plane wavefront sensing with an oscillating prism J. Mod. Opt. 43 289–93
[63] Ragazzoni R, Diolaiti E and Vernet E 2002 A pyramid wavefront sensor with no dynamic modulation Opt. Commun. 208 51–60
[64] Ramla R and Teschke G 2004 Regularization of sobolev embedding operators and applications part I: Fourier and wavelet based methods Sampling Theory Signal Image Process. 3 175–96
[65] Rayden M 1993 On the Barzilai and Borwein choice of steplength for the gradient method IMA J. Numer. Anal. 13 321–6
[66] Rigault F and Van Dam M 2013 Simulating astronomical adaptive optics systems using yao Proc. AO4ELT3 p 18
[67] Rosensteiner M 2011 Cumulative reconstructor: fast wavefront reconstruction algorithm for extremely large telescopes J. Opt. Soc. Am. A 28 2132–8
[68] Rosensteiner M 2012 Wavefront reconstruction for extremely large telescopes via CuRe with domain decomposition J. Opt. Soc. Am. A 29 2328–36
[69] Saxenhuber D 2016 Gradient-based reconstruction algorithms for atmospheric tomography in adaptive optics systems for extremely large telescopes PhD Thesis Johannes Kepler University Linz
[70] Saxenhuber D and Ramla R 2016 A gradient-based method for atmospheric tomography Inverse Problems Imaging 10 781–805
[71] Shatokhina I 2014 Fast wavefront reconstruction algorithms for extreme adaptive optics PhD Thesis Johannes Kepler University Linz
[72] Shatokhina I, Hutterer V and Ramla R 2017 Two novel algorithms for wavefront reconstruction from pyramid sensor data: convolution with linearized inverse filter and pyramid Fourier transform reconstructor Proc. AO4ELT5
[73] Shatokhina I, Obereder A and Ramla R 2014 Fast algorithm for wavefront reconstruction in XAO/SCAO with pyramid wavefront sensor Proc. SPIE 9148 91480P
[74] Shatokhina I and Ramla R 2017 Convolution- and Fourier-transform-based reconstructors for pyramid wavefront sensor Appl. Opt. 56 6381–90
[75] Shatokhina I, Obereder A, Rosensteiner M and Ramla R 2013 Preprocessed cumulative reconstructor with domain decomposition: a fast wavefront reconstruction method for pyramid wavefront sensor Appl. Opt. 52 2640–52
[76] Véran J P, Esposito S, Spanò P, Herriot G and Andersen D 2015 Pyramid versus Shack–Hartmann: trade study results for the NFIRAOS NGS WFS Proc. 4th AO4ELT Conf.
[77] Véronaud C 2004 On the nature of the measurements provided by a pyramid wave-front sensor Opt. Commun. 233 27–38
[78] Zhou B, Gao L and Dai Y H 2006 Gradient methods with adaptive step-sizes Comput. Optim. Appl. 35 69–86