Electronic Transactions on Numerical Analysis.
Volume 54, pp. 256–275, 2021.

ISSN 1068–9613.
DOI: 10.1553/etna_vol54s256

ETNA
Kent State University and
Johann Radon Institute (RICAM)

AN AUGMENTED WAVELET RECONSTRUCTOR FOR ATMOSPHERIC TOMOGRAPHY*

RONNY RAMLAU† AND BERNADETT STADLER†

Abstract. Atmospheric tomography, i.e., the reconstruction of the turbulence profile in the atmosphere, is a challenging task for adaptive optics (AO) systems for the next generation of extremely large telescopes. Within the AO community, the solver of first choice is the so-called Matrix Vector Multiplication (MVM) method, which directly applies the (regularized) generalized inverse of the system operator to the data. For small telescopes this approach is feasible, however, for larger systems such as the European Extremely Large Telescope (ELT), the atmospheric tomography problem is considerably more complex, and the computational efficiency becomes an issue. Iterative methods such as the Finite Element Wavelet Hybrid Algorithm (FEWHA) are a promising alternative. FEWHA is a wavelet-based reconstructor that uses the well-known iterative preconditioned conjugate gradient (PCG) method as a solver. The number of floating point operations and the memory usage are decreased significantly by using a matrix-free representation of the forward operator. A crucial indicator for the real-time performance are the number of PCG iterations. In this paper, we propose an augmented version of FEWHA, where the number of iterations is decreased by 50% using an augmented Krylov subspace method. We demonstrate that a parallel implementation of augmented FEWHA allows the fulfilment of the real-time requirements of the ELT.

Key words. adaptive optics, atmospheric tomography, inverse problems, augmented Krylov subspace methods

AMS subject classifications. 65R32, 65Y05, 65Y20, 65B99, 85-08, 85-10

1. Introduction.

The image quality of the new generation of earthbound Extremely Large Telescopes (ELTs) is heavily degraded by atmospheric turbulences triggered by the irregular mixing of cold and hot air. Due to these irregularities, the refractive index of air becomes inhomogeneous, and thus the wavefronts arriving at the telescope pupil are distorted. These optical distortions can be compensated by using a technique called Adaptive Optics (AO) [7, 28, 29], in which the deformations of the wavefronts of natural or laser guide stars (NGS or LGS) are measured by wavefront sensors (WFS) and, subsequently, corrected using deformable mirrors (DMs). Such a DM typically consists of a thin surface to reflect light and a set of actuators that drive the shape of the mirror. Several AO systems require the reconstruction of the turbulence profile in the atmosphere, called atmospheric tomography. We consider three of these AO systems, namely, Laser Tomography AO (LTAO), Multi Object Adaptive Object AO (MOAO), and Multi Conjugated AO (MCAO). LTAO uses several LGS and NGS with one DM to sharpen in a single direction of interest. MOAO is based on the same concept but uses different mirrors that sharpen for various directions (objects) of interest at the same time. MCAO uses several DMs, conjugated to different heights, i.e., each DM is positioned at a certain distance from the telescope along the optical axis. Within MCAO, one aims to achieve a high imaging quality in a large field of view. Detailed information about the systems can be found in [2, 5, 15, 21, 25].

Mathematically, the atmospheric tomography problem is ill-posed, i.e., the relation between measurements and the solution is unstable [4, 19]. As a consequence, regularization techniques are applied. A common way to deal with this inverse problem is the Bayesian framework because it allows to incorporate statistical information about turbulence and noise. Typically, the random variables are assumed to be Gaussian, and the maximum a posteriori estimate (MAP) is used for an optimal point estimate of the solution. The dimension of the atmospheric tomography problem depends on the number of subapertures of the WFS and on

*Received November 13, 2020. Accepted March 3, 2021. Published online on March 31, 2021. Recommended by L. Reichel.
†Industrial Mathematics Institute, Johannes Kepler University Linz, Altenbergerstraße 69, 4040 Linz, Austria (ronny.ramlau@jku.at, bernadett.stadler@indmath.uni-linz.ac.at).

ETNA
Kent State University and
Johann Radon Institute (RICAM)

the number of degrees of freedom of the DMs, which has drastically increased during the last years. Since the atmosphere is changing rapidly, the solution has to be computed in real-time, i.e., within approximately 2 ms. There are various ways to solve the atmospheric tomography problem, either directly or iteratively; see [6, 9, 11, 12, 13, 14, 22, 23, 24, 26, 30, 34, 39, 40, 43]. So far, the standard solver is the Matrix Vector Multiplication (MVM) method, i.e., the direct application of a (regularized) generalized inverse of the system operator. This direct solution method is suitable for small telescopes, however, for extremely large telescopes the computational efficiency becomes an issue. A promising alternative are iterative methods such as the Finite Element Wavelet Hybrid Algorithm. FEWHA utilizes a dual domain discretization approach in which the operators are expressed in a finite element or wavelet basis, leading to sparse representations of the underlying matrices. The dual domain discretization of the MAP estimate is then solved using a preconditioned conjugate gradient method (PCG).

As already illustrated in [36], FEWHA is not perfectly parallelizable, and for ELT-sized test configurations, the real-time requirements are difficult to fulfill with an off-the-shelf hardware system. A crucial indicator of the computational efficiency is the number of PCG iterations. Our novel method, called augmented FEWHA, speeds up the convergence of the PCG method by recycling information from previous time steps. In particular, the Krylov subspace generated when solving the previous systems is reused by utilizing orthogonal projections in subsequent systems. The augmented conjugate gradient method for solving consecutive linear systems was proposed in [8]. The concept there is based on the idea of Saad in [31] on augmented Krylov subspace methods for solving linear systems with multiple right-hand sides. However, Saad uses an Arnoldi or a Lanczos process for computing an approximate solution rather than the CG method. For a survey on subspace recycling techniques for iterative methods we refer to [35]. We want to emphasize that we are using an augmented Krylov subspace method but no recycling. What is commonly known in the literature as Krylov subspace recycling, in addition to augmentation, changes the Krylov space [35]. In our method, the Krylov subspace is not changed. By recycling we understand reusing the search directions from previous time steps. In [33] the authors propose a deflated version of the augmented CG, which further improves the convergence behavior. However, the overhead induced by their projection is large and thus not feasible for the runtime requirements of the ELT. In [1] the authors study improved seed methods for linear equations with multiple right-hand sides. We considered this method for our application as well, but it performed worse than the augmented CG. The reason is that the augmented CG applies an additional projection in every CG iteration, which further improves the approximation.

The paper is organized as follows: we start with a short overview of the atmospheric tomography problem and the Bayesian framework for regularization in Section 2. Afterwards, we recall the dual domain discretization approach of FEWHA in Section 3. In Section 4 we present the well-known PCG method and an extension based on an augmented Krylov subspace method, called augmented PCG. In Section 5 we propose the augmented FEWHA and define the ELT-sized test configuration, which is used later on in a detailed analysis. Section 6 is devoted to a study of convergence rates for FEWHA and its augmented version. The quality and robustness of the algorithm is demonstrated by numerical simulations in Section 7. We conclude with an analysis of the algorithm regarding computational performance, i.e., floating point operations, memory usage, and runtime on a CPU, in Section 8.

2. Atmospheric tomography. In atmospheric tomography, a layered system of the atmosphere is commonly assumed with the aim to reconstruct the refractive index of the turbulent layers from WFS measurements [29]. The relation between the L turbulent layers $\phi = (\phi_1, \ldots, \phi_L)$ and the wavefront sensor measurements s in a guide star direction g is

ETNA
Kent State University and
Johann Radon Institute (RICAM)

given by the tomography operator A,

$$s = (s_g)_{g=1}^{G} = A\phi. \tag{2.1}$$

We consider here the Shack-Hartmann (SH) WFS, i.e., the average slope of the wavefront over the area of the lens, called subaperture, is determined by the vertical and horizontal shift of the focal points of the subapertures. The tomography operator is composed of an SH operator Γ and a geometric propagation operator P in the direction of a certain guide star g

$$s_g = \Gamma_g P_g \phi, \qquad \text{for } g = 1, \dots, G.$$

The SH operator Γ maps wavefronts φ to SH-WFS measurements

$$s = \begin{pmatrix} s^x \\ s^y \end{pmatrix} = \begin{pmatrix} \Gamma^x \varphi \\ \Gamma^y \varphi \end{pmatrix} =: \Gamma\varphi.$$

Assuming that the wavefronts are described by piecewise continuous bilinear functions with nodal values φ_{ij}, we obtain the SH measurements in a subaperture Ω_{ij} by

$$\begin{aligned} s_{ij}^x &= \frac{(\varphi_{i,j+1} - \varphi_{i,j}) + (\varphi_{i+1,j+1} - \varphi_{i+1,j})}{2}, \\ s_{ij}^y &= \frac{(\varphi_{i+1,j} - \varphi_{i,j}) + (\varphi_{i+1,j+1} - \varphi_{i,j})}{2}. \end{aligned}$$

The wavefront aberrations φ in the direction of an NGS are defined by

$$\varphi_\theta(x) = (P_\theta^{NGS}\phi)(x) := \sum_{\ell=1}^{L} \phi_\ell(x + \theta h_\ell),$$

where $x = (x_1, x_2, 0)$ is a point on the aperture, $\theta = (\theta_1, \theta_2, 1)$ is the direction of the guide star, and h_ℓ is the layer height. For an LGS at fixed height H, the cone-effect has to be taken into account leading to

$$\varphi_\theta(x) = (P_\theta^{LGS}\phi)(x) := \sum_{\ell=1}^{L} \phi_\ell \left((1 - \frac{h_\ell}{H})x + \theta h_\ell \right).$$

For details about the definition of the geometric propagation operator, either for NGS or LGS, we refer to [9].

The underlying mathematical problem is ill-posed [4], i.e., there is an unstable relation between the measurements and the solution. In the AO-literature it is common to use the Bayesian framework for regularization, which allows the incorporation of statistical information regarding the turbulence model and noise. We define S and Φ as random variables that correspond to the vectors of SH-WFS measurements and turbulent layers. Further, we assume the presence of noise modeled via a random variable η. This leads to the following reformulation of equation (2.1) in the Bayesian framework

$$S = A\Phi + \eta. \tag{2.2}$$

We assume the random variables to be Gaussian and use the maximum a posteriori (MAP) estimate to compute an optimal point estimate for the solution [9]. The solution of equation (2.2) is then given by

$$x_{MAP} = \underset{\phi \in \mathbb{R}^n}{\operatorname{argmin}} \left\{ \|\phi\|^2_{C_\Phi^{-1}} + \|s - A\phi\|^2_{C_\eta^{-1}} \right\}, \tag{2.3}$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

where C_Φ^{-1} and C_η^{-1} are the inverse covariance matrices of layers Φ and noise η, respectively. For details about the definition of C_Φ and C_η we refer to [42]. The norms in equation (2.3) induced by the covariance matrices, which are symmetric and positive definite, are defined by

$$\|x\|_C^2 := (Cx, x).$$

The solution to the minimization problem (2.3) is given by the solution of the linear system of equations

$$(A^* C_\eta^{-1} A + C_\phi^{-1})\phi = A^* C_\eta^{-1} s, \tag{2.4}$$

where A^* denotes the adjoint operator of A.

3. Dual domain discretization. In order to numerically compute a solution of equation (2.4), discretization is required. There are certain fundamental advantages of using wavelets, as already extensively studied for FEWHA in [45]. The main idea of the previously developed FEWHA is to use compactly supported orthonormal wavelets for representing the turbulent layers. The properties in the frequency domain allow a diagonal approximation of C_ϕ. The atmospheric tomography operator A has a more efficient representation in a finite element domain, where continuous piecewise bilinear functions are utilized to represent wavefronts and layers.

We use the same dual domain discretization approach for our method. We utilize a square grid with equidistant spacing on Ω, the subaperture domain at the telescope pupil, to define the piecewise bilinear wavefront functions. The piecewise bilinear layer functions are defined using a square mesh with equidistant spacing on Ω_l, the domain on which the turbulent layers are defined. This mesh consists of 2^{2J_l} points, where J_l denotes the number of wavelet scales. See Figure 3.1 for a graphical representation of Ω and Ω_l. We obtain the following discretized MAP equation

$$(\boldsymbol{W}^{-T} \hat{A}^T C_\eta^{-1} \hat{A} \boldsymbol{W}^{-1} + \alpha D)c = \boldsymbol{W}^{-T} \hat{A}^T C_\eta^{-1} s, \tag{3.1}$$

where $\hat{A}$ is the atmospheric tomography operator in the finite element domain and $\hat{A}^T$ is the transposed matrix. We denote the linear mapping between the finite element and the wavelet domain by $\boldsymbol{W} = \operatorname{diag}(\delta_1 W, \ldots, \delta_L W)$, where W is the discrete wavelet transform and δ_l is the scaling constant at layer l. The operator C_η^{-1} denotes the inverse covariance matrix of the noise, and D is a diagonal approximation of C_ϕ^{-1} in the frequency domain. We introduce a scalar factor α for tuning the balance between the fitting and the regularizing terms. If α is increased, then it stabilizes the modeling error, however, a too large α reduces the quality of the reconstruction. An automatic adjustment of α with, e.g., the discrepancy principle, is not possible because of the runtime requirements of the application. In practice, α is tuned by hand via numerical simulations for a specific test configuration. For a detailed study we refer to [16]. The vector c is a concatenation of all wavelet coefficients of all turbulence layers, and the vector s is the concatenation of all SH sensor measurements from all guide star directions. In all our simulations we use periodic Daubechies-3 wavelets, which are an orthogonal wavelet family with compact support.

For the sake of simplicity, we define the left-hand side operator of equation (3.1) by

$$M := (\boldsymbol{W}^{-T} \hat{A}^T \hat{C}_\eta^{-1} \hat{A} \boldsymbol{W}^{-1} + \alpha D) \tag{3.2}$$

and the right-hand side as

$$b := \boldsymbol{W}^{-T} \hat{A}^T C_\eta^{-1} s.$$

Note, that the matrix $M \in \mathbb{R}^{2^{2J_\ell} L \times 2^{2J_\ell} L}$ is symmetric and positive definite.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

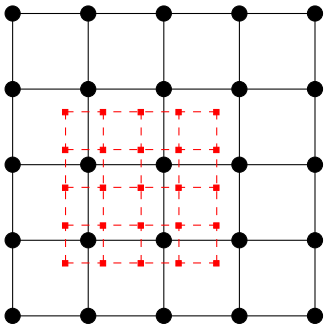

FIG. 3.1. *Square grid of layers Ω_l in black with $2^{2J_l} = 2^4$ points and equidistant spacing. Projected grid of subapertures Ω in red with $n_s^2 = 16$ subapertures.*

4. Preconditioned conjugate gradient method for atmospheric tomography. We are considering the case where the matrix M in (3.2) is very large, in particular, too large for direct solution methods, such as the MVM algorithm, to be attractive or feasible. A direct solver requires the factorization of the left-hand side operator, which is computationally very expensive for large matrices. We focus on the most prominent iterative solver for M being symmetric and positive definite, the conjugate gradient (CG) algorithm. An iterative method applies the forward operator M repeatedly to obtain the solution. Since all the operators involved in the definition of M in (3.2) have a sparse structure and can be represented in a matrix-free way (see [17] for details), iterative methods are especially efficient here. The CG algorithm was already used for the previously developed FEWHA [45]. The most time consuming part, which is the application of M, is well-parallelizable. For more details on the parallelization possibilities for FEWHA we refer to [36].

The idea behind the CG method is that for a symmetric and positive definite matrix M, solving the linear system $Mc = b$ is equivalent to minimizing $\frac{1}{2}(Mc, c) - (b, c)$. This minimization is performed iteratively over a subspace spanned by the search directions p. Theoretically, the CG algorithm needs at most $2^{2J_\ell} L$ iterations, corresponding to the size of M, until the exact solution is obtained. However, this would be far too much to meet the real-time requirements of the ELT. In practice, the method provides already a good approximation after a few iterations. Therefore, the number of CG iterations is often fixed for real-time applications, i.e., the algorithm terminates after a predetermined number of $\mathrm{maxIter}$ iterations. Note that under these circumstances, the method is not guaranteed to reduce the error below a given threshold. To improve the approximation we use the residual vector $r_0 = b - Mc_0$ with an initial guess c_0 as a starting value of the CG method. For the initial guess the solution of the previous time step is used. This is also referred to as warm restart. The convergence behavior of the CG method is affected by the eigenvalue distribution of the operator, which can be modified by preconditioning. We prefer a mixed preconditioner $J^{-1} = (J^{-1/2})^T J^{-1/2}$ here to preserve the symmetry of the matrix and thus be still able to apply the CG method. We obtain the following preconditioned version of equation (3.1)

$$(J^{-1/2})^T M J^{-1/2} y = (J^{-1/2})^T b. \tag{4.1}$$

The solution c to the original problem is given by

$$c = J^{-1/2} y.$$

Within FEWHA, a modified Jacobi preconditioner is utilized with a different weighting of the low and high frequency regimes; see [44]. In general, the Jacobi preconditioner is

ETNA
Kent State University and
Johann Radon Institute (RICAM)

a diagonal matrix with $J = \operatorname{diag}(M)$, thus, easy to invert and very efficient to apply. For FEWHA, the definition is slightly modified and given by

$$J = \operatorname{diag}((\boldsymbol{W}^{-T}\hat{A}^T\hat{C}_\eta^{-1}\hat{A}\boldsymbol{W}^{-1}) + \alpha \max(D, \tau I)),$$

with a non-negative constant τ. The benefit of such an approach is the reduction of CG iterations and an increased robustness and stability of the whole method. For computing J, the diagonal entries of the left-hand side matrix M are required. Computing the explicit form of M is computationally very demanding since it involves several matrix-matrix multiplications. However, in our application the update time for M is fixed to 6 minutes. Hence, J is precomputed and reused in the forthcoming time steps, in which only the right-hand side b changes. Inside the PCG method, the dense matrix M is never used explicitly. Instead, the sparse matrices involved in the definition of M are applied in a matrix-free way. Algorithm 1 displays the PCG method with the split Jacobi preconditioner $J^{-1} = J^{-1/2}J^{-1/2}$.

Algorithm 1 PCG Algorithm.

Input: c_0 (previous wavelet coefficients)
r_0 (previous residual vector)
$\operatorname{maxIter}$ (number of PCG iterations)
J^{-1} (preconditioner)
Output: $c^{(i+1)}$ (new wavelet coefficients)
$r^{(i+1)}$ (new residual)
$p_0 = r_0$
$z_0 = J^{-1}r_0$
for $k = 0, \ldots, \operatorname{maxIter}$ **do**
  $q_k = Mp_k$
  $\alpha = (r_k, z_k)/(p_k, q_k)$
  $c_{k+1} = c_k + \alpha p_k$, $r_{k+1} = r_k - \alpha q_k$
  $z_{k+1} = J^{-1}r_{k+1}$
  $\beta = (r_{k+1}, z_{k+1})/(r_k, z_k)$
  $p_{k+1} = z_{k+1} + \beta p_k$
end for
$c^{(i+1)} = c_{k+1}$, $r^{(i+1)} = r_{k+1}$

In the following, we list some important properties of the CG method (see, e.g., [8]), which we will use in subsequent sections. The residuals are orthogonal to each other

$$(r_i, r_j) = 0, \qquad \text{for } i, j \geq 0, i \neq j.$$

The vectors p_i are search directions and M-orthogonal to each other, i.e.,

$$(J^{-1/2}MJ^{-1/2}p_i, p_j) = 0, \qquad \text{for } i, j \geq 1, i \neq j.$$

Let us assume that we have performed m CG-iterations. We define $R_m := (r_0, \ldots, r_m)$ as the matrix of residuals and $P_m := (p_0, \ldots, p_m)$ as the matrix of conjugate search directions. The following relations hold:

$$P_m^T M P_m = D_m \quad \text{and} \quad \operatorname{Span}(R_m) = \operatorname{Span}(P_m) = \mathcal{K}_m(J^{-1/2}MJ^{-1/2}, r_0), \tag{4.2}$$

where D_m is a diagonal matrix of size $m \times m$ and $\mathcal{K}_m(J^{-1/2}MJ^{-1/2}, r_0)$ is the Krylov subspace of size $m + 1$ generated by the initial residual r_0.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

4.1. Augmented Krylov subspace method. Within AO, we are dealing with several right-hand sides b of equation (3.1), available consecutively in each time step. Since the CG method is an iterative solver we need to reapply it in every single time step. This is costly in terms of computational speed compared to direct solvers, where the factorization can be reused independently of the right-hand side as long as the left-hand side matrix does not change. However, within atmospheric tomography we can also exploit the fact that the right-hand side only changes slightly in every time step. The basic idea is to speed up the convergence of the current time step by using the search directions in P_m obtained from the previous system.

We define equation (4.1) for several right-hand sides, which correspond to the WFS measurements from time steps $i = 1, 2, \ldots$, by

$$J^{-1/2} M J^{-1/2} c^{(i)} = J^{-1/2} b^{(i)}, \tag{4.3}$$

where $b^{(i)} := \boldsymbol{W}^{-T} \hat{A}^T C_\eta^{-1} s^{(i)}$. The main idea is to use the information obtained when solving equation (4.3) for a certain time step i with m iterations of the PCG method for the upcoming time steps and reuse the Krylov subspace $\mathcal{K}_m(J^{-1/2} M J^{-1/2}, r_0^{(i)})$ generated in the previous system.

A first idea in this direction is to improve the convergence of the PCG method by choosing a more optimized initial guess c_0 as a starting vector. In fact, $c_0^{(i+1)}$ can be chosen such that the initial residual $r_0^{(i+1)}$ is orthogonal to the Krylov subspace $\mathcal{K}_m(J^{-1/2} M J^{-1/2}, r_0^{(i)})$ generated in the previous time step. This Galerkin projection technique was first proposed by Saad in [31] for the Lanczos process and adapted in [8] for the CG method. We enforce this orthogonality condition by

$$P_m^T r_0^{(i+1)} = 0.$$

To satisfy this condition we choose

$$c_0^{(i+1)} = c_0^{(i)} + P_m D_m^{-1} P_m^T r_0^{(i+1)},$$

where P_m and D_m are defined by equation (4.2). For the proof we refer to [8].

Another way to improve the convergence behavior is to keep the orthogonality of the residual vectors with respect to the Krylov subspace throughout the iterations of the PCG method. As before, this method was first proposed for the Lanczos process in [31] (called modified Lanczos process) and adapted for the CG method in [8] (referred to as augmented CG). Within this approach we use two subspaces: the Krylov subspace generated for the first system $\mathcal{K}_m(J^{-1/2} M J^{-1/2}, r_0^{(i)})$ and the subspace $\mathrm{Span}(r_0^{(i+1)}, \ldots, r_k^{(i+1)})$. The residual $r_{k+1}^{(i+1)}$ must be orthogonal to both subspaces, and the directions p_{k+1} must be M-orthogonal to both subspaces. Hence, this method is a projection method onto the space

$$\mathcal{K}_{m,k}(J^{-1/2} M J^{-1/2}, r_0^{(i)}, r_0^{(i+1)}) = \mathcal{K}_m(J^{-1/2} M J^{-1/2}, r_0^{(i)}) + \mathrm{Span}(r_0^{(i+1)}, \ldots, r_k^{(i+1)}),$$

which is not a Krylov-subspace. The projection is defined by the following three conditions (see [8]):

1. $p_0^{(i+1)} = (I - P_m D^{-1} (M P_m)^T) r_0^{(i+1)}$,
2. $c_{k+1} - c_k \in \mathcal{K}_{m,k}(J^{-1/2} M J^{-1/2}, r_0^{(i)}, r_0^{(i+1)})$,
3. $(r_{k+1}^{(i+1)}, z) = 0$ for all $z \in \mathcal{K}_{m,k}(J^{-1/2} M J^{-1/2}, r_0^{(i)}, r_0^{(i+1)})$.

The last condition is called a Petrov-Galerkin condition and is satisfied if the current residual $r_{k+1}^{(i+1)}$ is orthogonal to $r_k^{(i+1)}$ and the current direction $p_{k+1}^{(i+1)}$ is M-conjugate to $p_k^{(i+1)}$ and $p_m^{(i)}$.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

The augmented PCG method is given in Algorithm 2. The method starts by computing an initial guess $c_0^{(i+1)}$ (line 5) such that the initial residual $r_0^{(i+1)}$ is orthogonal to the Krylov subspace $\mathcal{K}_m(J^{-1/2}MJ^{-1/2}, r_0^{(i)})$. Further, the initial descent direction $p_0^{(i+1)}$ is calculated in lines 7–11 such that it is conjugate to the descent directions in P_m. In each PCG iteration the new search direction $p_k^{(i+1)}$ is orthogonalized against the last search direction of the previous system $p_m^{(i)}$ in line 20.

Algorithm 2 Augmented PCG Algorithm [8].

1: **Input:** c_0, r_0 (previous wavelet coefficients and residual)
maxIter (number of PCG iterations)
J^{-1} (preconditioner)
$p^{(i)}, q^{(i)}$ (previous descent directions)
2: **Output:** $c^{(i+1)}, r^{(i+1)}$ (new wavelet coefficients and residual)
$p^{(i+1)}, q^{(i+1)}$ (new descent directions)
3: **for** $j = 0, \ldots, m$ **do**
4: $\sigma_j = (r, p_j^{(i)})/(p_j^{(i)}, q_j^{(i)})$
5: $c_0 = c_0 + \sigma_j p_j^{(i)}$, $r_0 = r_0 - \sigma_j q_j^{(i)}$
6: **end for**
7: $z_0 = J^{-1} r_0$
8: **for** $j = 0, \ldots, m$ **do**
9: $z_0 = z_0 - (z, q_j^{(i)})/(p_j^{(i)}, q_j^{(i)})$
10: **end for**
11: $p_0^{(i+1)} = z_0$
12: **for** $k = 0, \ldots, \mathrm{maxIter}$ **do**
13: $q_k^{(i+1)} = M p_k^{(i+1)}$
14: $\alpha = (r_k, z_k)/(p_k^{(i+1)}, q_k^{(i+1)})$
15: $c_{k+1} = c_k + \alpha p_k^{(i+1)}$, $r_{k+1} = r_k - \alpha q_k^{(i+1)}$
16: $z_{k+1} = J^{-1} r_{k+1}^{(i+1)}$
17: $\mu = (z_{k+1}, q_m^{(i)})/(p_m^{(i)}, p_m^{(i)})$
18: $z_{k+1} = z_{k+1} - \mu p_m^{(i)}$
19: $\beta = (r_{k+1}, z_{k+1})/(r_k, z_k)$
20: $p_{k+1}^{(i+1)} = z_{k+1} + \beta p_k^{(i+1)}$
21: **end for**
22: $c^{(i+1)} = c_{k+1}$, $r^{(i+1)} = r_{k+1}$

5. An augmented finite element wavelet hybrid algorithm. Our augmented FEWHA differs from the algorithm proposed in [17] only in that we use the augmented PCG algorithm described in Algorithm 2 instead of the classical PCG method. Algorithm 3 is the augmented wavelet reconstructor for time step $(i+1)$. The main input is the measurement vector $s^{(i+1)}$ and the output is the new shape of the mirrors $a^{(i+1)}$. The AO system can either operate in closed or open loop fashion. If open loop control is applied, then the measurements are

ETNA
Kent State University and
Johann Radon Institute (RICAM)

directly obtained from the wavefronts, whereas in closed loop control, the pseudo-open loop measurements are calculated as a first step of the algorithm in line 3. Due to the two-step delay (see [20] for details), the DM shape from the previous step is used. The right-hand side $b^{(i+1)}$ is computed in line 6 with the new measurement vector $s^{(i+1)}$, and, subsequently, the initial residual $r_0^{(i+1)}$ is updated in line 7. The atmospheric reconstruction takes place in line 8, where $p^{(i)}$ and $q^{(i)}$ are used within the PCG method to improve the solution by projection. Afterwards, the layers are fitted to actuator commands in line 9, and the closed or open loop control is applied in lines 10–14. The new DM shapes are a linear combination of the current and the reconstructed DM shapes, weighted by a scalar value called gain with values between zero and one. This gain control improves the stability of the reconstruction. For closed loop control, the artificially added DM shapes $a^{(i-1)}$ are subtracted from the computed mirror shapes $\tilde{a}$ such that the difference is simply the reconstruction from the closed loop measurements.

Algorithm 3 Augmented Wavelet Reconstructor.

1: **Input:** $s^{(i+1)} = (s_g)_{g=1}^{G}$ (measurement vector)
$gain$ (scalar weight)
$c^{(i)}$ (previous wavelet coefficients)
$b^{(i)}$ (previous right-hand side)
$r^{(i)}$ (previous residual vector)
$a^{(i-1)}, a^{(i)}$ (previous two DM shape)
maxIter (number of PCG iterations)
$J^{-1/2}$ (preconditioner)
$p^{(i)}$,$q^{(i)}$ (previous descent directions)
2: **Output:** $a^{(i+1)}$ (next DM shape)
3: **if** loop = closed **then**
4: $s^{(i+1)} = s^{(i+1)} + \Gamma a^{(i-1)}$
5: **end if**
6: $b^{(i+1)} = \mathbf{W}^{-T} A^T C_\eta^{-1} s^{(i+1)}$
7: $r_0 = b^{(i+1)} - Mc^{(i)} = (b^{(i+1)} - b^{(i)}) + r^{(i)}$
8: $[c^{(i+1)}, r^{(i+1)}, p^{(i+1)}, q^{(i+1)}] = augPCG(c^{(i)}, r_0, J^{-1/2}, p^{(i)}, q^{(i)}, \text{maxIter})$
9: $\tilde{a} = FW^{-1}c^{(i+1)}$
10: **if** loop = closed **then**
11: $a^{(i+1)} = a^{(i)} + \text{gain} \cdot (\tilde{a} - a^{(i-1)})$
12: **else if** loop = open **then**
13: $a^{(i+1)} = (1 - \text{gain}) \cdot a^{(i)} + \text{gain} \cdot \tilde{a}$
14: **end if**

5.1. Test configuration. To verify the quality and computational performance of the proposed algorithm we use the design of the ELT with an assumed telescope diameter of 39 m, where about 28% are obstructed. Our numerical simulations are executed via the software package MOST (see [3]), which was developed in-house as an alternative to OCTOPUS [18]. We model the vertical density profile of the laser beam scatter in the sodium layer by a Gaussian random variable with mean H =90 km and a full width at half maximum (FWHM) of 11.4 km.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

The table on the left side of Figure 5.1 summarizes the general parameter settings for our numerical simulations. We examine the performance of FEWHA versus its augmented version for the AO systems LTAO, MOAO, and MCAO. All these systems use atmospheric tomography to correct the optical distortions. For the LTAO and MOAO configuration we assume to have 3 WFS with a number of 80×80 subapertures assigned to 3 NGS positioned in a circle of 10 arcmin diameter. Further, we assume to have 6 WFS with a number of 80×80 subapertures assigned to 6 LGS positioned in a circle of 7.5 arcmin diameter. For the MCAO test case we have two WFS with 1×1 subapertures and one with 2×2 assigned to 3 NGS positioned in a circle of 2 arcmin diameter. The 6 WFS with 80×80 subapertures are assigned to 6 LGS positioned in a circle of 8/3 arcmin diameter. A graphical representation of the MCAO star asterism is shown on the right-hand side of Figure 5.1. In the LTAO case we simulate a 9-layer atmosphere and use a single DM with 81×81 actuators to correct for atmospheric distortions. The MOAO simulations have a similar setup, only the NGS and LGS fluxes differ from the LTAO setting. For the MCAO tests we simulate a 3- and a 9-layer atmosphere. In contrast to LTAO and MOAO, the MCAO system uses 3 DMs with 81×81, 47×47, and 53×53 actuators at different altitudes and actuator spacings. In the 3-layer approach we reconstruct the atmosphere directly at the deformable mirrors, whereas in the 9-layer case we solve the fitting equation using an unpreconditioned CG algorithm with 4 iterations. The same setup was already used in [17] for FEWHA. Table 5.1 summarizes the system specific configurations.

Parameter	**Value**
Telescope diameter D	39 m
Central obstruction	$\sim 28\ \%$
Fried parameter r_0	0.129 m
Na-layer thickness $FWHM$	11.4 km
Na-layer height H	90 km
Evaluation criterion	LE Strehl
Evaluation wavelength	K band
Gain	0.4
Number of wavelet scales J_ℓ	6
Regularization parameter α	2
Number of loop iterations	500

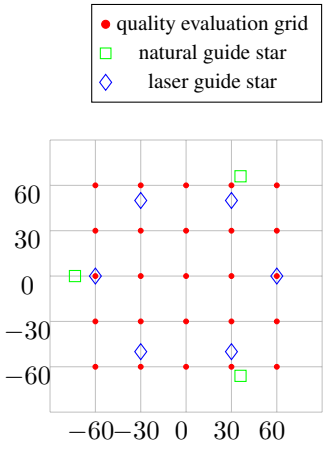

FIG. 5.1. *General parameter setting for ELT test case and the star asterism of the MCAO NGS and LGS setting with a* 5×5 *quality evaluation grid.*

6. Analysis of convergence rates. It is shown in [8] that the augmented CG computes the minimum over the subspace $\mathcal{K}_{m,k}(M, r_0^{(i)}, r_0^{(i+1)})$ and that this minimum is smaller than or equal to the minimum over the subspace $\mathcal{K}_k(H^T M H, r_0^{(i+1)})$. The matrix of M-orthogonal projections onto $\mathcal{K}_m(J^{-1/2} M J^{-1/2}, r_0)^{\perp_M}$ is defined by

$$H := I - PD^{-1}(MP)^T. \tag{6.1}$$

Hence, the augmented CG is a classical CG applied to the matrix $H^T M H$. We utilize this result to calculate an asymptotic error bound for the classical CG, the classical PCG, and the augmented PCG at a certain iteration k by applying the well-known theory of convergence rates from the classical method. The augmented PCG can be seen as CG preconditioned by

ETNA
Kent State University and
Johann Radon Institute (RICAM)

TABLE 5.1
LTAO, MOAO, and MCAO system specific parameters.

Parameter	**LTAO**	**MOAO**	**MCAO**
Number of Layer L	9	9	3 or 9
Number of NGS G_{NGS}	3	3	3
Number of WFS subap. n_{lay}	84×84	84×84	one 2×2, two 1×1
Direction, circle of	10	10	$8/3$ arcmin diameter
Wavelength	589 nm	589 nm	589 nm
Flux	500	5–500	300 photons/subap./frame
Number of LGS G_{LGS}	6	6	6
Number of WFS subap. n_{lay}	84×84	84×84	84×84
Direction, circle of	7.5	7.5	2 arcmin diameter
Wavelength	500 nm	500 nm	1650 nm
Flux	50–500	5–500	50-500 photons/subap./frame
Number of DMs	1	1	3
Number of actuators	85×85	85×85	85×85, 47×47, 53×53
DM altitude	0 m	0 m	0 m, 4 km, 12.7 m
Actuator spacing	0.5 m	0.5 m	0.5 m, 1 m, 1 m
PCG iterations	8	10	4

HH^T and J^{-1}. Let κ be the condition number of $H^T M H$, then we can use the theory of the classical CG (see, e.g., [41]) to obtain the error bound for the augmented CG at iteration k by,

$$\|c_k - c\|_M \leq 2 \|c_0 - c\|_M \left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1}\right)^k.$$

The condition number of $H^T M H$ is less or equal to the condition number of M [8]. Thus, the asymptotic convergence rate of the augmented PCG is less or equal to the asymptotic convergence rate of the classical PCG. We want to verify this statement for the MCAO test case defined in Table 5.1. We restrict ourself to this special test case because it is related to MAORY, a first-light instrument of the ELT. The atmospheric tomography problem for MAORY provides a suitable real-world example. The condition numbers of M, $J^{-1/2} M J^{-1/2}$, and $J^{-1/2} H^T M H J^{-1/2}$ are given in Table 6.1. Note, that the condition number of M is influenced by the regularization parameter α, which is 2 for our test setting. We observe that $\kappa_0 > \kappa_1 > \kappa_2$. However, this upper bound for the error is loose and can be improved.

TABLE 6.1
Condition numbers for the left-hand side matrix of FEWHA and augmented FEWHA.

$\kappa_0 :=$	$cond(M)$	$\approx 3 \cdot 10^8$
$\kappa_1 :=$	$cond(J^{-1/2} M J^{-1/2})$	$\approx 2.76 \cdot 10^7$
$\kappa_2 :=$	$cond(J^{-1/2} H^T M H J^{-1/2})$	$\approx 1 \cdot 10^7$

The approximate solution at a certain iteration c_k has a non-linear relation to the initial vector c_0; see [38]. Let P_k denote the space of polynomials of degree at most k. Then there exists a polynomial $q \in P_k$ with $q(0) = 1$ that fulfills

$$c_k - c = q(M)(c_0 - c) \qquad \text{and} \qquad r_k = q(M) r_0.$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

It is shown, e.g., in [38] that

$$\|c_k - c\|_M^2 = \min_{q \in P_k, q(0)=0} \sum_{j=0}^{n-1} \frac{(u_j, r_0)^2}{\lambda_j} q(\lambda_j)^2, \tag{6.2}$$

where u_j is the j-th eigenvector of M and λ_j the corresponding eigenvalue. We assume $\lambda_j \geq \lambda_i$ for $j < i$. As a consequence, the rate of convergence of the CG is influenced by the eigenvalue distribution of the left-hand side matrix and the decomposition of the residual with respect to the eigenvectors.

We conclude that the clustering of the eigenvalues is important. If the value of q in (6.2) is small for a certain λ_j, then by the continuity of the polynomial we know that the value is small at all eigenvalues clustered around λ_j. Figure 6.1 illustrates the eigenvalue distribution of M, $J^{-1/2}MJ^{-1/2}$, and $J^{-1/2}H^TMHJ^{-1/2}$. Jacobi preconditioning improves the structure of the matrix in the sense that all diagonal elements become 1 (see [32]), which implies that all Gerschgorin discs are centered around 1. By Gerschgorin's circle theorem [10] we know that all the eigenvalues are contained in at least one of these Gerschgorin discs. However, we do not have any information about the radius of the discs. In Figure 6.1 we observe that the eigenvalues of the Jacobi preconditioned matrix (red) are clustered around 1. The projection operator H has only a marginal influence on the eigenvalue distribution, which is not visible in the graph. In fact, the eigenvalue distribution of $J^{-1/2}MJ^{-1/2}$ (red) almost completely overlaps with that of $J^{-1/2}H^TMHJ^{-1/2}$ (dashed green). We observe the decay of the eigenvalues towards 0, a characteristic property of an ill-posed problem.

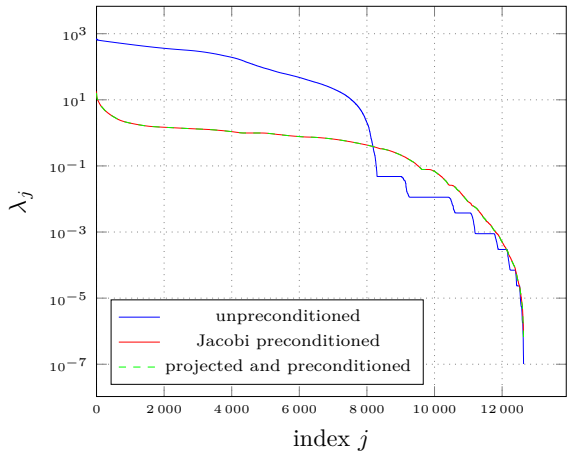

FIG. 6.1. *Logarithmic plot of the eigenvalues of the left-hand side matrix M (blue), the preconditioned matrix (red), and the projected preconditioned matrix (dashed green) as a function of the index number.*

In Figure 6.2 we illustrate the influence of the projection, used within augmented FEWHA, in the decomposition of the initial residual with respect to the eigenvectors. The initial residual, unprojected or projected, is taken after 100 time steps with 2 PCG iterations. Note, that the structure of the initial residual for the original FEWHA is already improved by the warm restart technique, i.e., by utilizing the solution from the previous time step as initial vector. The upper left graph shows the energy of the initial residual with respect to the j-th eigenvector u_j of the preconditioned matrix $J^{-1/2}MJ^{-1/2}$, i.e., for the classical FEWHA. The upper right graph illustrates the influence of the Galerkin projection of the initial residual r_0, again

ETNA
Kent State University and
Johann Radon Institute (RICAM)

with respect to the preconditioned matrix. In the last graph, the projection by the matrix H is taken into account by considering the j-th eigenvector u_j of the projected and preconditioned matrix $J^{-1/2}H^TMHJ^{-1/2}$. We observe how augmentation changes the value of (r_0, u_j). The structure of the initial residual for the classical FEWHA, i.e., with 4 PCG iterations, is similar to the one in the upper right graph. The influence of the matrix H, as shown in the last graph, in general, decreases the value of (r_0, u_j).

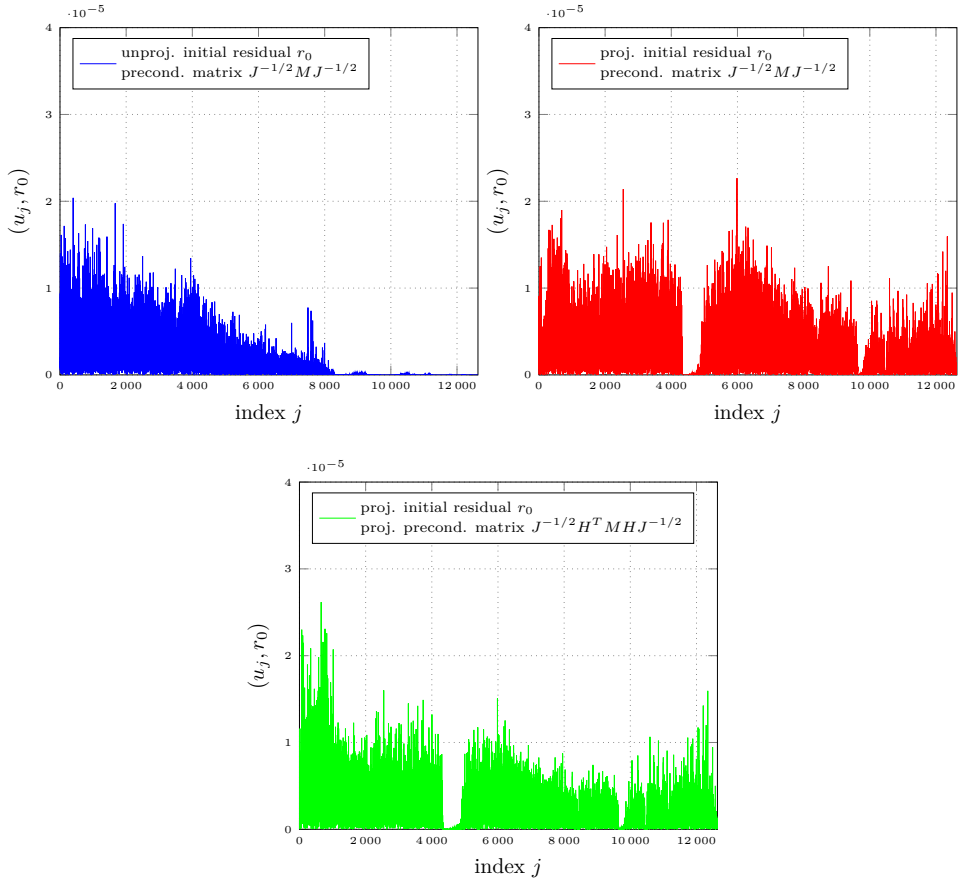

FIG. 6.2. *Plot of the energy of the unprojected and projected initial residual with respect to the eigenvectors of the preconditioned matrix $J^{-1/2}MJ^{-1/2}$ and the projected preconditioned matrix $J^{-1/2}H^TMHJ^{-1/2}$ as a function of the index j.*

Summarizing, we hypothesize that the augmented FEWHA requires a lower number of iterations than the original algorithm to obtain a similar quality of the reconstruction. We verify this hypothesis by numerical simulations in the upcoming section.

7. Quality evaluation. We evaluate the quality of augmented FEWHA using the Strehl ratio, which is a frequently-used quality evaluation criterion within AO. The short exposure (SE) Strehl is defined as the ratio between the real energy distribution of the incoming light in the image plane $I(x, y)$ over the hypothetical distribution $I_D(x, y)$, which stems from the

ETNA
Kent State University and
Johann Radon Institute (RICAM)

assumption of diffraction-limited imaging,

$$S_{\tilde{\varphi}} := \frac{\max_{(x,y)} I(x,y)}{\max_{(x,y)} I_D(x,y)} \in [0,1],$$

for the phase of the light $\tilde{\varphi}$. The higher the Strehl ratio, the better the quality of the AO system. The maximum of 1 is reached only in the diffraction-limited case. A tip/tilt-mode, which leads to a horizontal shift of $I(x,y)$, does not influence S. In order to detect a tip/tilt mode, the long exposure (LE) Strehl ratio is used. The LE Strehl represents the quality of the image from the start of the loop until a certain time step. We use the LE Strehl in the K band, i.e., for a wavelength of 2200 nm. The short and long exposure Strehl ratio is often indicated in $\%$. For more details about the Strehl ratio, we refer to [28].

In our simulation environment the SE Strehl ratio is estimated by the so called Marèchal approximation (see [27, 28]), which is used for directions close to the zenith to approximate the SE Strehl ratio by

$$S_{\tilde{\varphi}} \approx \exp(-\frac{1}{|\Omega|} \|\tilde{\varphi} - \bar{\varphi}\|^2_{L_2(\Omega)}),$$

where

$$\bar{\varphi} := \frac{1}{|\Omega|} \int_\Omega \tilde{\varphi}(x) dx$$

denotes the average phase at the aperture Ω with area $|\Omega|$. Hence, this approximation relates the Strehl ratio at the aperture with the L_2-error of the phase. The phase $\tilde{\varphi}$ is related to the wavefront φ by

$$\tilde{\varphi}(x) = \frac{2\pi}{\lambda} \varphi(x),$$

with λ being the wavelength and $x \in \mathbb{R}^2$ a point on the telescope aperture. In the framework of this work, we are interested in the wavefront aberrations φ.

Figure 7.1 displays the LE Strehl for the LTAO simulations. For FEWHA we vary the number of PCG iterations n_{iter} between 4 and 8, whereas for augmented FEWHA, the number of iterations $n_{augIter}$ varies between 3 and 5. We observe that for a higher flux, the classical FEWHA requires almost double the iterations compared to the augmented version to obtain a similar LE Strehl. For the low flux tests, the situation is slightly different, however, also here we can see a reduction in the number of iterations. This leads to a non-negligible reduction of FLOPs for the augmented FEWHA (see Section 8 for a detailed analysis), hence, a considerable speed-up of the whole algorithm. Figure 7.2 presents the center LE Strehl for the MOAO simulations. We increase the number of FEWHA PCG iterations up to $n_{iter} = 10$. This makes the augmented FEWHA even more efficient here because saving approximately half the iterations saves more FLOPs than for the LTAO test case. Figure 7.3 displays the LE center Strehl for the MCAO 3-layer setting on the left and the 9-layer configuration on the right. Here, $n_{iter} = 4$ is enough to meet the quality requirements. For augmented FEWHA, we vary the number of iterations $n_{augIter}$ between 1 and 2. The augmented version with half the iterations provides the same quality as FEWHA also for the low flux cases.

Hence, we are able to verify the hypothesis stated in the previous section that the augmented FEWHA requires a lower number of iterations when providing the same quality in terms of the Strehl ratio. In fact, our numerical simulations show that the number of iterations is reduced to half.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

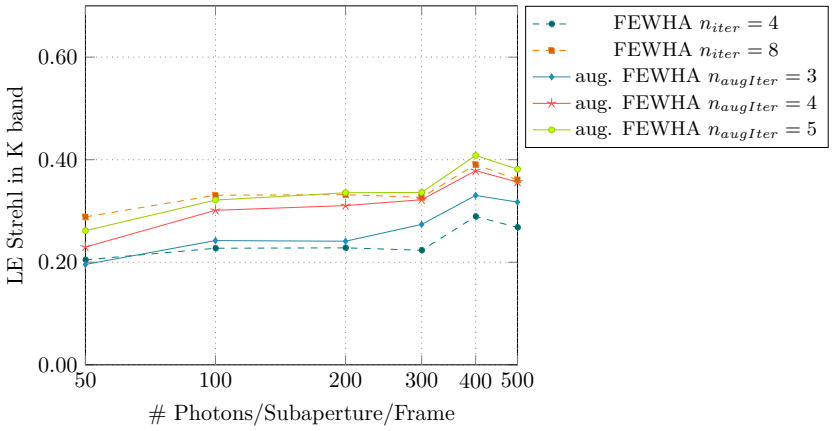

FIG. 7.1. *LTAO LE Strehl over the course of 500 iterations with a different number of n_{iter} and $n_{augIter}$ for FEWHA and its augmented version.*

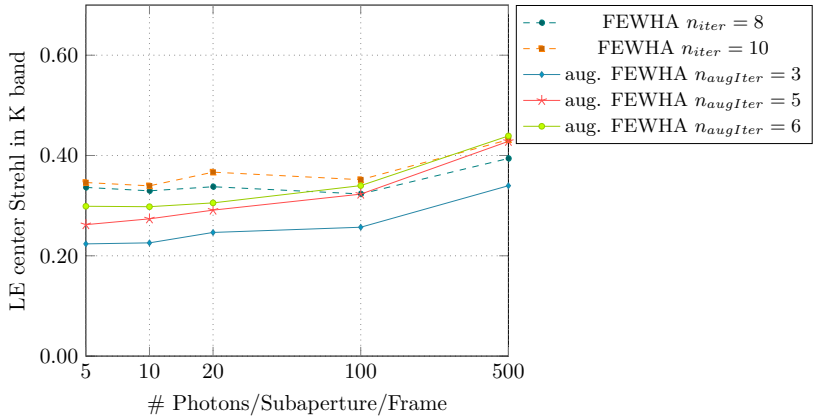

FIG. 7.2. *MOAO center LE Strehl over the course of 500 iterations with a different number of n_{iter} and $n_{augIter}$ for FEWHA and its augmented version.*

8. Computational performance. We analyze FEWHA and its augmented version regarding floating point operations, memory usage, and runtime on CPU. We use the LTAO, MOAO, and MCAO 3-layer setting as defined in the previous section. We omit the 9-layer MCAO configuration here because if we reconstruct more layers than DMs, then we have to solve an additional minimization problem, which is costly in terms of speed. Further, we fix n_{iter} and $n_{augIter}$ as shown in Table 8.1, based on our observations from the numerical simulations in Section 7, to obtain a similar quality. For details about FLOPs and memory usage for the MVM method and FEWHA, we refer to [37].

8.1. Floating point operations. The FLOPs for FEWHA and its augmented version differ only in the PCG step. There are two additional parts for the augmented PCG: the projection of the initial guess c_0 and the initial residual r_0 onto the Krylov subspace of the previous system and the projection of the descent directions p_k onto the last vector of the descent directions of the previous system p_m. In every iteration one dot-product and one vector update is added. We want to stress that the augmented FEWHA does not introduce

ETNA
Kent State University and
Johann Radon Institute (RICAM)

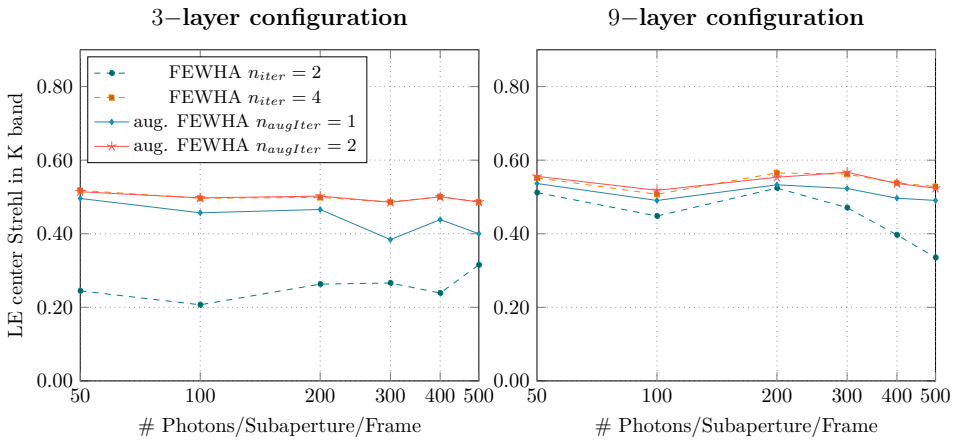

FIG. 7.3. *MCAO center LE Strehl over the course of 500 iterations for the 3-layer configuration (on the left) and the 9-layer setting (on the right) with different number of n_{iter} and $n_{augIter}$ for FEWHA and its augmented version.*

TABLE 8.1
Number of PCG iterations for FEWHA and augmented FEWHA.

Test setting	n_{iter}	$n_{augIter}$
LTAO	8	4
MOAO	10	5
MCAO 3 layer	4	2

new matrix-vector products, which are the most time consuming steps of the algorithm. In Section 6 we show that the asymptotic convergence rate of the augmented PCG is not higher than for the classical PCG. In fact, our numerical simulations in Section 7 reveal that we can save approximately half the PCG iterations $n_{augIter} \approx n_{iter}/2$ with using the augmented Krylov subspace method. Hence, the overhead of the augmented FEWHA should be easily compensated by the reduction in the number of iterations. Table 8.2 provides the theoretical FLOPs for augmented FEWHA. The first row corresponds to the classical PCG method within FEWHA, whereas the second and third line refer to the additional computations that have to be done within augmented PCG. In Table 8.3 the results for the ELT configurations are presented. We can see immediately that we save approximately half the FLOPs when using the augmented PCG instead of the classical PCG. This is what we expect when choosing $n_{augIter} = n_{iter}/2$.

TABLE 8.2
Theoretical FLOPs for FEWHA and augmented FEWHA.

Computation step	**Theoretical FLOPs**
PCG iterations	$\{[14G_{LGS}+2G_{NGS}+12G]n_s^2+4Gn_s$ $+[(L-1)(15.6G_{LGS}+18G_{NGS})+L(G-1)](n_s+1)^2$ $+L\frac{176}{3}(4n_{lay}^2-1)+13n_{lay}^2L\}n_{iter}$
Initialization	$(9n_{lay}^2L)n_{augIter}$
Projection in every iteration	$(4n_{lay}^2L)n_{augIter}$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

8.2. Memory usage. For augmented FEWHA we have to save the descent directions p_k and $q_k = Mp_k$, for $k = 1, \ldots, n_{augIter}$. Additionally, to decrease the number of FLOPs and avoid unnecessary re-computations, we save the inner products (p_k, q_k). Both vectors are of size $n_{lay}^2 L$, hence, in total we need $(2n_{lay}^2 L + 1)n_{augIter}$ units of additional storage. The last column of Table 8.3 shows the additional memory usage of augmented FEWHA over the classical version, assuming single precision (32 bit) floating point numbers. Compared to the overall memory usage of FEWHA and especially compared to the memory intensive MVM method, the additional memory usage for augmented FEWHA is almost negligible.

TABLE 8.3
Performance evaluation of augmented FEWHA versus classical FEWHA.

Test setting	FLOPs PCG FEWHA	FLOPs PCG aug. FEWHA	Add. memory usage aug. FEWHA
LTAO	376 MFLOPs	196 MFLOPs	4.7 MB
MOAO	470 MFLOPs	244 MFLOPs	5.9 MB
MCAO	63.5 MFLOPs	33.5 MFLOPs	0.78 MB

8.3. Performance on a multi-core CPU. For ELT-sized test settings FEWHA performs better on CPUs than on GPUs. This is mainly because of its efficient matrix-free implementation. The bottleneck is not the computational throughput but the memory latency. For a detailed study we refer to [36]. That is why we focus here on the implementation of augmented FEWHA on a single CPU with multiple compute cores. In particular, we run the parallel implementation of our method on one compute node of the high performance cluster of the Radon Institute for Computational and Applied Mathematics in Linz, called Radon1, that has two 8-core Intel Haswell processors (Xeon 403 E5-2630v3, 2.4Ghz) and 128 GB of memory. For our numerical simulations we utilize 12 cores. We reuse the parallel implementation of FEWHA from [36] but change the PCG method to augmented PCG.

The last two columns of Table 8.4 refer to the runtime of the various test configurations for FEWHA and its augmented version. We simulate 1000 time steps and take the average runtime. We gain a speed-up of augmented FEWHA compared to the classical method between 1.63 and 1.8. Of course the speed-up of the MCAO 3-layer setting is less compared to the other two settings because we use a smaller number of iterations, and thus the computations outside the PCG algorithm have a bigger influence on the overall performance. This test case is particularly important as it is related to MAORY, which is an adaptive optics module for the ELT operating in MCAO. The real-time requirement of MAORY is 2 ms for each reconstruction. To the best of our knowledge the augmented FEWHA is the first method that is able to solve the 3-layer MCAO test setting within this real-time requirement.

TABLE 8.4
Runtime of augmented FEWHA versus classical FEWHA on Radon1.

Test setting	Runtime FEWHA	Runtime aug. FEWHA	Speed-up
LTAO	6.0 ms	3.5 ms	1.7
MOAO	8.0 ms	4.4 ms	1.8
MCAO	3.1 ms	1.9 ms	1.6

ETNA
Kent State University and
Johann Radon Institute (RICAM)

9. Conclusion. In this paper, we continued the work of [16, 17, 36, 44, 45] on the iterative solver FEWHA by considering an augmented Krylov subspace method. This method enables us to use information from previous time steps to speed up the convergence of the iterative PCG method. The inner products and matrix-vector products within FEWHA are well parallelizable, however, the number of PCG iterations are a crucial indicator for the real-time performance. An analysis of convergence rates for FEWHA and its augmented version suggests that the augmentation technique improves the convergence behavior for ELT-sized problems. We validate this hypothesis by numerical simulations. In fact, we are able to decrease the number of iterations to 50% with the proposed augmented FEWHA. Based on this outcome, we give a detailed study on the computational performance, i.e., FLOPs, memory usage, and runtime on a multi-core CPU. The only drawback of the augmented FEWHA is the additional used memory to store information from previous time steps. However, this is almost negligible compared to the overall memory usage. Most importantly, we showed that with augmented FEWHA, we are able to meet the real-time requirements of MAORY, an adaptive optics module for the ELT operating in MCAO. To the best of our knowledge, augmented FEWHA is the first method that fulfills this requirement.

Acknowledgments. The project has received funding by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement No. 765374 and the Austrian Science Fund (FWF) F6805-N36 (Tomography in Astronomy).

REFERENCES

[1] A. M. ABDEL-REHIM, R. B. MORGAN, AND W. WILCOX, *Improved seed methods for symmetric positive definite linear equations with multiple right-hand sides*, Numer. Linear Algebra Appl., 21 (2014), pp. 453–471.

[2] D. R. ANDERSEN, S. S. EIKENBERRY, M. FLETCHER, B. L. WILLIAM GARDHUOSE, J.-P. VERAN, D. GAVEL, R. CLARE, R. G. L. JOLISSAINT, R. JULIAN, AND W. RAMBOLD, *The MOAO system of the IRMOS near-infrared Multi-Object Spectrograph for TMT*, in Proceedings of SPIE 6269, Ground-based and Airborne Instrumentation for Astronomy, I. S. McLean and M. Iye, eds., Art. 62694K, 2006, SPIE, Bellingham, 2006.

[3] G. AUZINGER, *New Reconstruction Approaches in Adaptive Optics for Extremely Large Telescopes*, PhD. Thesis, Industrial Mathematics Institute, Johannes Kepler University Linz, 2017.

[4] M. E. DAVISON, *The ill-conditioned nature of the limited angle tomography problem*, SIAM J. Appl. Math., 43 (1983), pp. 428–448.

[5] E. DIOLAITI, A. BARUFFOLO, M. BELLAZZINI, V. BILIOTTI, G. BREGOLI, C. BUTLER, P. CILIEGI, J.-M. CONAN, G. COSENTINO, S. D'ODORICO, B. DELABRE, H. FOPPIANI, T. FUSCO, N. HUBIN, M. LOMBINI, E. MARCHETTI, S. MEIMON, C. PETIT, C. ROBERT, P. ROSSETTINI, L. SCHREIBER, AND R. TOMELLERI, *MAORY: A Multi-conjugate Adaptive Optics RelaY for the E-ELT*, The ESO Messenger, 140, (2010), pp. 28–29.

[6] B. ELLERBROEK, L. GILLES, AND C. VOGEL, *A computationally efficient wavefront reconstructor for simulation or multi-conjugate adaptive optics on giant telescopes*, in Proceedings of SPIE 4839, Adaptive Optical System Technologies II, P. L. Wizinowich and D. Bonaccini, eds., SPIE, Bellingham, 2002.

[7] B. L. ELLERBROEK AND C. R. VOGEL, *Inverse problems in astronomical adaptive optics*, Inverse Problems, 25 (2009), Art. 063001, 37 pages.

[8] J. ERHEL AND F. GUYOMARC'H, *An augmented conjugate gradient method for solving consecutive symmetric positive definite linear systems*, SIAM J. Matrix Anal. Appl., 21 (2000), pp. 1279–1299.

[9] T. FUSCO, J.-M. CONAN, G. ROUSSET, L. MUGNIER, AND V. MICHAU, *Optimal wave-front reconstruction strategies for multi conjugate adaptive optics*, J. Opt. Soc. Am. A, 18 (2001), pp. 2527–2538.

[10] S. A. GERSHGORIN, *Über die Abgrenzung der Eigenwerte einer Matrix*, Bull. Acad. Sci. URSS, 1931 (1931), pp. 749–754.

[11] L. GILLES AND B. ELLERBROEK, *Split atmospheric tomography using laser and natural guide stars*, J. Opt. Soc. Am. A, 25 (2008), pp. 2427–2435.

[12] L. GILLES, B. ELLERBROEK, AND C. VOGEL, *Layer-oriented multigrid wavefront reconstruction algorithms for multi-conjugate adaptive optics*, in Proceedings of SPIE 4839, Adaptive Optical System Technologies II, P. L. Wizinowich and D. Bonaccini, eds., SPIE, Bellingham, 2002.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

[13] ———, *Preconditioned conjugate gradient wave-front reconstructors for multiconjugate adaptive optics*, Appl. Opt., 42 (2003), pp. 5233–5250.
[14] ———, *A comparison of multigrid V-cycle versus Fourier domain preconditioning for laser guide star atmospheric tomography*, in Adaptive Optics: Analysis and Methods/Computational Optical Sensing and Imaging/Information Photonics/Signal Recovery and Synthesis Topical Meetings on CD-ROM, Art. JTuA1, Optical Society of America, Washington, 2007.
[15] F. HAMMER, F. SAYÈDE, E. GENDRON, T. FUSCO, D. BURGARELLA, V. CAYATTE, J.-M. CONAN, F. COURBIN, H. FLORES, I. GUINOUARD, L. JOCOU, A. LANÇON, G. MONNET, M. MOUHCINE, F. RIGAUD, D. ROUAN, G. ROUSSET, V. BUAT, F. ZAMKOTSIAN, *The FALCON concept: multi-object spectroscopy combined with MCAO in near-IR*, in Scientific Drivers for ESO Future VLT/VLTI Instrumentation, J. Bergeron and G. Monnet, eds., ESO Astrophysics Symposia, Springer, Berlin, 2002, pp. 139–148.
[16] T. HELIN AND M. YUDYTSKIY, *Wavelet methods in multi-conjugate adaptive optics*, Inverse Problems, 29 (2013), Art. 085003, 18 pages.
[17] ———, *Wavelet methods in multi-conjugate adaptive optics*, Inverse Problems, 29 (2013), Art. 085003, 18 pages.
[18] M. LE LOUARN, C. VERINAUD, V. KORKIAKOSKI, N. HUBIN, AND E. MARCHETTI, *Adaptive optics simulations for the European Extremely Large Telescope*, in Proceedings of SPIE 6272, Advances in Adaptive Optics II, N. Hubin, C. E. Max, P. L. Wizinowich, eds., SPIE, Bellingham, 2006, pp. U1048–U1056.
[19] F. NATTERER, *The Mathematics of Computerized Tomography*, B. G. Teubner, Stuttgart, 1986.
[20] M. PÖTTINGER, R. RAMLAU, AND G. AUZINGER, *A new temporal control approach for SCAO systems*, Inverse Problems, 36 (2020), Art. 015002, 29 pages.
[21] M. PUECH, H. FLORES, M. LEHNERT, B. NEICHEL, T. FUSCO, P. ROSATI, J.-G. CUBY, AND G. ROUSSET, *Coupling MOAO with integral field spectroscopy: specifications for the VLT and the E-ELT*, Mon. Not. R. Astron. Soc., 390 (2008), pp. 1089–1104.
[22] S. RAFFETSEDER, R. RAMLAU, AND M. YUDYTSKIY, *Optimal mirror deformation for multi conjugate adaptive optics systems*, Inverse Problems, 32 (2016), Art 025009, 21 pages.
[23] R. RAMLAU, A. OBEREDER, M. ROSENSTEINER, AND D. SAXENHUBER, *Efficient iterative tip/tilt reconstruction for atmospheric tomography*, Inverse Probl. Sci. Eng., 22 (2014), pp. 1345–1366.
[24] R. RAMLAU AND M. ROSENSTEINER, *An efficient solution to the atmospheric turbulence tomography problem using Kaczmarz iteration*, Inverse Problems, 28 (2012), Art. 095004, 23 pages.
[25] F. RIGAUT, B. ELLERBROEK, AND R. FLICKER, *Principles, limitations and performance of multiconjugate adaptive optics*, in Proceedings of SPIE 4007, Adaptive Optical Systems Technology, P. L. Wizinowich, ed., SPIE Bellingham, 2000, pp. 1022–1031.
[26] C. ROBERT, J.-M. CONAN, D. GRATADOUR, L. SCHREIBER, AND T. FUSCO, *Tomographic wavefront error using multi-LGS constellation sensed with Shack-Hartmann wavefront sensors*, J. Opt. Soc. Am. A, 27 (2010), pp. A201–A215.
[27] L. C. ROBERTS JR., M. D. PERRIN, F. MARCHIS, A. SIVARAMAKRISHNAN, R. B. MAKIDON, J. C. CHRISTOU, B. A. MACINTOSH, L. A. POYNEER, M. A. VAN DAM, AND M. TROY, *Is that really your Strehl ratio?*, in Proceedings of SPIE Vol. 5490, Advancements in Adaptive Optics, D. B. Calia, B. L. Ellerbroek, and R. Ragazzoni, eds., SPIE Bellingham, 2004, pp. 504–515.
[28] F. RODDIER, *Adaptive Optics in Astronomy*, Cambridge University Press, Cambridge, 1999.
[29] M. C. ROGGEMANN AND B. WELSH, *Imaging Through Turbulence*, CRC Press, Boca Raton, 1996.
[30] M. ROSENSTEINER AND R. RAMLAU, *The Kaczmarz algorithm for multi-conjugate adaptive optics with laser guide stars*, J. Opt. Soc. Am. A, 30 (2013), pp. 1680–1686.
[31] Y. SAAD, *On the Lánczos method for solving symmetric linear systems with several right-hand sides*, Math. Comp., 48 (1987), pp. 651–662.
[32] ———, *Iterative Methods for Sparse Linear Systems*, 2nd ed., SIAM, Philadelphia, 2003.
[33] Y. SAAD, M. YEUNG, J. ERHEL, AND F. GUYOMARC'H, *A deflated version of the conjugate gradient algorithm*, SIAM J. Sci. Comput., 21, 2000, pp. 1909–1926.
[34] D. SAXENHUBER AND R. RAMLAU, *A gradient-based method for atmospheric tomography*, Inverse Probl. Imaging, 10 (2016), pp. 781–805.
[35] K. M. SOODHALTER, E. DE STURLER, AND M. E. KILMER, *A survey of subspace recycling iterative methods*, GAMM-Mitt., 43 (2020), Art. e202000016, 29 pages.
[36] B. STADLER, R. BIASI, M. MANETTI, AND R. RAMLAU, *Real-time implementation of an iterative solver for atmospheric tomography*, Preprint on arXiv, 2020. https://arxiv.org/abs/2009.00946
[37] B. STADLER, R. BIASI, AND R. RAMLAU, *Feasibility of standard and novel solvers in atmospheric tomography for the ELT*, in Proceedings of Adaptive Optics for Extremely Large Telescopes, AO4ELT6, Quebec, 2019.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

[38] Z. STRAKOŠ AND P. TICHÝ, *On error estimation in the conjugate gradient method and why it works in finite precision computations*, Electron. Trans. Numer. Anal., 13 (2002), pp. 56–80. http://etna.ricam.oeaw.ac.at/vol.13.2002/pp56-80.dir/pp56-80.pdf

[39] M. TALLON, I. TALLON-BOSC, C. BÉCHET, F. MOMEY, M. FRADIN, AND E. THIÉBAUT, *Fractal iterative method for fast atmospheric tomography on extremely large telescopes*, in Proceedings of SPIE 7736, Adaptive Optics Systems II, B. L. Ellerbroek, M. Hart, N. Hubin, and P. L. Wizinowich, eds., SPIE Bellingham, 2010, pp. 77360X–77360X–10.

[40] E. THIÉBAUT AND M. TALLON, *Fast minimum variance wavefront reconstruction for extremely large telescopes*, J. Opt. Soc. Am. A, 27 (2010), pp. 1046–1059.

[41] A. VAN DER SLUIS AND H. A. VAN DER VORST, *The rate of convergence of conjugate gradients*, Numer. Math., 48 (1986), pp. 543–560.

[42] T. VON KARMAN, *Mechanische Ähnlichkeit und Turbulenz*, Proceedings of the Third International Congress of Applied Mechanics, C. W. Oseen and W. Weibull, eds., Sveriges Litografiska Tryckerie, Stockholm, 1930.

[43] Q. YANG, C. VOGEL, AND B. ELLERBROEK, *Fourier domain preconditioned conjugate gradient algorithm for atmospheric tomography*, Appl. Opt., 45 (2006), pp. 5281–5293.

[44] M. YUDYTSKIY, T. HELIN, AND R. RAMLAU, *A frequency dependent preconditioned wavelet method for atmospheric tomography*, in Proceedings of the Third AO4ELT Conference, S. Esposito and L. Fini, eds., 2013. http://ao4elt3.arcetri.astro.it/proceedings/

[45] ———, *Finite element-wavelet hybrid algorithm for atmospheric tomography*, J. Opt. Soc. Am. A, 31 (2014), pp. 550–560.