

MODEL REDUCTION IN ATMOSPHERIC TOMOGRAPHY BY OPTIMAL GROUPING OF TURBULENT LAYERS*

GÜNTER AUZINGER[†]

Abstract. In wide-field applications of adaptive optics systems, the problem of atmospheric tomography has to be solved. Commonly used methods for this purpose operate on a set of two-dimensional reconstruction layers. Due to run-time restrictions and demands on stability, in general the usable number of such reconstruction layers is less than the number of atmospheric turbulence layers. Hence, model reduction has to be applied to the profile of atmosphere layers in order to achieve a smaller number of the most relevant reconstruction layers. In continuation of earlier published and purely heuristic experiments, we concentrate on the question how the choice of the heights of these reconstruction layers influences the performance of the tomographic solver, aiming for a more rigorous analysis. We derive a function representing an approximate expected value for the best-case residual error, i.e., a limitation (in a statistical sense) for what any tomographic solver is able to reach. We provide a method for the minimization of this function, which consequently yields an algorithm for the (approximately) optimal choice of the reconstruction layer heights for a given input atmosphere model, i.e., given the turbulence strength depending on the altitude. Our implementation of the optimization algorithm has acceptable run-time, and first tests of the resulting layer profiles show that the obtained quality is significantly better than for other choices of the reconstruction layer profiles.

Key words. model reduction, adaptive optics, atmospheric tomography, layer compression, optimization

AMS subject classifications. 78M34, 78A10, 85-08

1. Introduction. The problem of finding appropriate reconstruction layer profiles for atmospheric tomography (commonly referred to as *layer compression*) is essential for wide-field applications of adaptive optics (AO) control [9]. It is known [1, 3, 7] that the resulting quality of commonly used tomographic reconstruction methods depends crucially on the choice of these reconstruction layer profiles. We present the optimal grouping method as a model reduction approach for a priori compression in the form of a method for calculating an appropriate reconstruction layer profile for a chosen number of layers from a given atmosphere profile, i.e., given heights and turbulence strengths C_n^2 per layer. The numerical tests [15] are promising, and the method seems to outperform all other competitive approaches independent of the algorithm used for the tomographic reconstruction. A simplified version of the algorithm was already published in [15]. We derive the complete and more general method in this work and provide the theoretical background.

Section 1.1 gives a short summary of the atmospheric tomography problem. In Section 1.2 we informally discuss the basic idea of the proposed compression method, which should serve as a motivation for elaborating the technical details in the subsequent discussion. In Section 2 we derive an a priori estimate of the residual error and its dependence on the reconstruction layer profile. Several approximations lead to a rather simple expression for this dependence, which is necessary for the efficiency of the resulting compression method. In Section 3 we present an optimization algorithm for finding the global minimum of the achieved error estimate function, which yields approximately optimal reconstruction layer profiles.

1.1. Atmospheric tomography: a brief problem description. This summary of the tomography problem is kept short and serves more as a clarification of the notation used in the subsequent sections rather than a real introduction. For a detailed description of the modeling of light and adaptive optics systems, the reader is referred to the introductory literature [9]. We describe light in terms of Fourier optics [8], i.e., as propagating electromagnetic waves

*Received September 5, 2017. Accepted May 29, 2018. Published online on August 28, 2018. Recommended by L. Reichel.

[†]Industrial Mathematics Institute, Johannes Kepler University Linz, Altenbergerstrasse 69, Linz, Austria (auzinger@mathconsult.co.at).

according to Maxwell's equations. The major source of disturbances of the wave-fronts is the fluctuation of the air's refractive index due to unpredictable atmospheric turbulence leading to a deformation of the originally plain wave-fronts. The essential effect on the waves are phase delays, which add up as the light propagates through the air.

In wide-field applications of AO systems, the following tomography problem arises: We obtain measurements from a number G of wave-front sensors (WFSs) by looking into the directions of G (natural or laser) guide-stars (GSs), $g = 1, \dots, G$. From these measurements, the wave-fronts (WFs) $\tilde{\varphi}_g$ are reconstructed. They represent sums over shifted optical phase delays $\Phi^{(n)}$ in the atmosphere, where $n = 1, \dots, N$. In general, the number N of atmospheric layers at the heights H_1, \dots, H_N is far too large to directly state the tomography problem (i.e., solving for the phase delays) on these layers. This is due to memory and run-time constraints as well as stability demands on the solver, as the problem of narrow angle tomography is severely ill-posed [4]. Hence, the layer structure of the atmosphere is usually approximated by a smaller number $L < N$ of reconstruction layers at the heights h_1, \dots, h_L .

We define the shift operator

$$(1.1) \quad (T^{\Delta \mathbf{x}} \Phi)(\mathbf{x}) := \Phi(\mathbf{x} - \Delta \mathbf{x}),$$

which formally denotes the horizontal shift of the phase delays Φ by a translation vector $\Delta \mathbf{x}$. The value of this translation caused by the GS with number g in a layer at height h_l is given by

$$\Delta \mathbf{x} = h_l \cdot \mathbf{d}_g,$$

where \mathbf{d}_g is the view direction of the GS number g and which is defined as the vector (d_g^1, d_g^2) in the aperture plane for the GS being located on the line $\xi(d_g^1, d_g^2, 1)$ for $\xi \geq 0$. The chosen coordinate system has its origin at the center of the telescope aperture and the z -axis is identical to the optical axis; we assume the telescope being directed to zenith here for simplicity. These view directions are bound by a GS separation angle α via

$$(1.2) \quad |\mathbf{d}_g| \leq \tan(\alpha), \quad \text{for all } g = 1, \dots, G,$$

where α is usually of the same order of magnitude as the field of view.

Using these notations, we can formulate the tomography problem as follows:

PROBLEM 1.1. *Given the measured and reconstructed wave-fronts $\tilde{\varphi}_g$, for $g = 1, \dots, G$ (number of GS). For all g , solve the system*

$$\sum_{l=1}^L T^{h_l \mathbf{d}_g} \hat{\Phi}^{(l)} = \tilde{\varphi}_g$$

for the unknown phase delays $\hat{\Phi}^{(l)}$, where $l = 1, \dots, L$ with L the number of reconstruction layers. The wave-fronts $\tilde{\varphi}_g$ are given as noise-contaminated sums of the atmospheric phase delays in the view directions

$$\tilde{\varphi}_g = \sum_{n=1}^N T^{H_n \mathbf{d}_g} \Phi^{(n)} + \mathcal{N}.$$

The noise \mathcal{N} stems from discretization (due to finitely many sub-apertures of the WFS) and measurement errors (e.g., quantum effects) of the WFSs. The reconstruction of the wave-fronts $\tilde{\varphi}$ from the sensor measurements causes additional numerical errors, which we interpret as included in \mathcal{N} here for the sake of simplicity. Problem 1.1 is in general ill-posed and a

solver can only try to find a good approximation of the least-squares solution to Problem 1.1. In the sequel we denote by \mathcal{T} such a solver that maps the input $\tilde{\varphi}_g$ to the unknowns $\hat{\Phi}^{(l)}$:

$$\mathcal{T} : (\tilde{\varphi}_1, \dots, \tilde{\varphi}_G) \mapsto (\hat{\Phi}^{(1)}, \dots, \hat{\Phi}^{(L)}).$$

Note that we have neglected the *cone effect*, which introduces an additional height-dependent scaling of the phase delays if laser GSs are in use. This is because we assume that it does not play a significant role in our context. Also the nature of \mathcal{N} is not taken into account here. The treatment of noise will be discussed in the beginning of Section 2. Further we neglect the so-called piston mode, i.e., we assume that

$$(1.3) \quad \int_{\mathbb{R}^2} \Phi(\mathbf{x}) \, d\mathbf{x} = 0$$

in all atmospheric layers and for the results $\hat{\Phi}$ of the solver \mathcal{T} on all reconstruction layers.

The phase delays Φ are assumed to be non-deterministic, and their behavior is described by a statistical model such as the widely used model according to von Karman. Such models are usually formulated using either the structure function, the autocorrelation function, or the power spectral density of the atmosphere's refractive index or the phase delay. In our context, the autocorrelation function is most practical. We assume the turbulence model to be given in form of the autocorrelation function of the atmospheric phase delays, that is,

$$\mathbf{C}_{\Phi}(\Delta\mathbf{x}) := \mathbb{E}(\langle \Phi(\mathbf{x}) - \text{meanValue}(\Phi), \Phi(\mathbf{x} - \Delta\mathbf{x}) - \text{meanValue}(\Phi) \rangle),$$

where \mathbb{E} denotes the expected value and $\langle \cdot, \cdot \rangle$ is the usual scalar product in L^2 . Under our assumption that the piston mode is zero (1.3), this takes the form

$$(1.4) \quad \mathbf{C}_{\Phi}(\Delta\mathbf{x}) := \mathbb{E}(\langle \Phi(\mathbf{x}), \Phi(\mathbf{x} - \Delta\mathbf{x}) \rangle).$$

According to [9], the autocorrelation function is a constant times the inverse Fourier transform of the power spectral density function $\mathbf{P}_n(\omega)$ of the air's refractive index n :

$$\mathbf{P}_n \propto \mathbf{P}_{\Phi} = \mathcal{F}\{\mathbf{C}_{\Phi}\}.$$

The latter is commonly specified according to the von Karman model (see, e.g., [13])

$$(1.5) \quad \mathbf{P}_n(\omega) = C \cdot C_n^2(H) \left(|\omega|^2 + \frac{1}{\tilde{L}^2} \right)^{-11/6} \cdot \exp\left(-\frac{|\omega|^2}{|\omega_m|^2}\right),$$

where $|\omega_m| = 5.92/l_{\min}$. The parameters l_{\min} and \tilde{L} denote the characteristic minimum length of an eddy and the outer scale of the spatial coherence of the turbulence, respectively. The scaling constant C does not depend on H and is not relevant in our context.

However, in simulation tools for adaptive optics systems, also a modified version of the von Karman model is used, namely

$$(1.6) \quad \mathbf{P}_n^{\text{sim}}(\omega) = C' \cdot C_n^2(H) \left(|\omega|^2 + \frac{1}{\tilde{L}^2} \right)^{-11/6}.$$

This modification changes the decay of \mathbf{P} from exponential to polynomial, from which we can expect a decrease in smoothness of the phase delays Φ as well as the autocorrelation function.

In both models (1.5), (1.6), the relative strength of the turbulence in an atmospheric layer at height H (i.e., the dependence on H) is described by a measured constant $C_n^2(H)$. For the reconstruction layers, predictors $c^2(h)$ corresponding to these constants are in general required for the tomographic solver \mathcal{T} as parameters—usually these are internally used as weights, e.g., in the penalty terms.

1.2. The idea and motivation for the proposed compression method. The focus of our interest is the question how to choose the reconstruction layer heights h_l in order to achieve the best possible performance when solving Problem 1.1 using any tomographic solver \mathcal{T} if the number of reconstruction layers L is given. Our approach is based on choosing h_l such that the minimum residual (i.e., the best residual \mathcal{T} can theoretically achieve) gets as small as possible. For this aim we need an estimate for the minimum residual that can be calculated a priori only from information about the atmosphere model.

In a first step, we try to find this estimate in the simple case with only one atmospheric layer at height H , and we want to find an approximation on a single reconstruction layer at a different height h with $\Delta h := h - H \neq 0$. Let the atmospheric phase delay in layer H be $\Phi(\mathbf{x})$. The tomographic reconstructor \mathcal{T} yields an approximation $\hat{\Phi}(\mathbf{x})$ on layer h . The aim of \mathcal{T} is to find $\hat{\Phi}$ such that the difference between Φ and $\hat{\Phi}$ is small for several view directions \mathbf{d} . These different view directions cause $\hat{\Phi}$ to be shifted about $\Delta \mathbf{x} = \Delta h \cdot \mathbf{d}$ relative to Φ .

Our goal is now to find an estimate for the error $\Phi - \hat{\Phi}$ and explicate its dependence on Δh . The first rough idea, on which this work is based, is to assume that $\hat{\Phi} \approx \Phi$, and hence,

$$(1.7) \quad \Phi(\mathbf{x}) - \hat{\Phi}(\mathbf{x} - \Delta h \cdot \mathbf{d}) \approx \Phi(\mathbf{x}) - \Phi(\mathbf{x} - \Delta h \cdot \mathbf{d}) \approx |\mathbf{d}| \nabla \Phi(\mathbf{x}) \Delta h,$$

indicating that the error depends linearly on Δh . From established statistical models of the atmosphere we assume that $\nabla \Phi$ is linearly correlated to $\sqrt{C_n^2(H)}$, hence the squared minimum residual error should be proportional to $C_n^2(H) \cdot \Delta h^2$. In order to use this single-layer connection for the full profile, we make the assumption that the solution $\hat{\Phi}^{(l)}$ on every reconstruction layer h_l approximates the phase delays in the atmospheric layers H_n that are close to h_l . Thus, the problem should be solved if the values for h_l are chosen such that the sum over all single-layer estimates $C_n^2(H) \cdot \Delta h^2$ between all reconstruction layers and their neighboring atmospheric layers is small.

Of course an argumentation like (1.7) is formally not acceptable. We will show a more rigorous way of achieving such a result in the sequel: the two unknowns, the phase delay Φ and its reconstruction $\hat{\Phi}$, are treated in different ways. In a first step we assume $\hat{\Phi}$ to be the best possible solution that the tomographic solver \mathcal{T} can theoretically find, i.e., the least-squares solution to Problem 1.1 for a given Φ —this is motivated by our aim to choose the reconstruction profile such that the *possibilities* for \mathcal{T} are optimal. Subsequently, the expected value for this minimal error is estimated under the assumption that Φ is in accordance with an appropriate turbulence model of the atmosphere. From the rough idea sketched in (1.7), it can be expected that we need some kind of smoothness assumptions on Φ (e.g., Φ has to be differentiable or at least Lipschitz continuous in order to allow the last step in (1.7))—this will turn out to appear in the form of a smoothness condition on the autocorrelation function in the vicinity of its maximum. This is more consistent with the statistical nature of Φ .

The idea to simply add up the single-layer estimates in order to achieve an estimate for the complete profile turns out to be hard to justify theoretically. We have to base it on a conjecture, the validity of which we can only show heuristically; see Section 2.2.

2. A profile-dependent function for estimating the minimum residual error. As mentioned in Section 1.1, the tomographic solver \mathcal{T} has to operate on a number L of reconstruction layers that is in general less than N , the number of turbulence layers in the atmosphere. Hence, the essential question is how to choose the heights h_l and the weights c_l^2 for the reconstruction layer profile for a given L . It turned out during earlier experiments [1] that the variation of the heights h_l has significantly more impact than the variation of c_l^2 . Hence this article concentrates primarily on the heights; for the weights c_l^2 , see Section 2.2.3.

We want to achieve optimal results for the tomographic solver \mathcal{T} without any knowledge about \mathcal{T} . This can be formulated in two ways:

- Amongst all error sources that restrict the performance of the solver \mathcal{T} (such as instability due to the ill-posedness of the tomography problem, various kinds of noise from WFS measurements and actuation of deformable mirrors, disturbing effects from laser guide stars, etc.), identify those stemming directly from the fact that not every atmospheric layer coincides with a reconstruction layer. Subsequently, choose the reconstruction layer profile h_1, \dots, h_L such that this profile-caused error restriction is minimal.
- Neglect all physical and numerical error sources and assume that the solver \mathcal{T} yields the best possible result, i.e., a solution with minimum residual error in the reconstruction layer profile. Subsequently, choose h_1, \dots, h_L such that this residual error gets minimal.

We will work with the second formulation, which is formally easier to handle. In Section 2.1 the simplified case of having only one atmospheric layer and one reconstruction layer is analyzed, and the dependence of the residual error on the height difference of these layers can be shown. In Section 2.2 we expand this result to the general case of N atmospheric layers and L reconstruction layers.

2.1. Error analysis for a single layer. In this section we derive a formula similar to (1.7) in a more rigorous way. However, we have to use several approximations in order to achieve a formula that is sufficiently simple. This is motivated by the need for a practical method for layer compression with acceptable run-time.

2.1.1. The definition of the problem geometry. Let $\mathcal{D} \subset \mathbb{R}^2$ be the telescope aperture and V the number of view directions \mathbf{d}_v with $v = 1, \dots, V$. In a standard tomography setup, these could be GS directions and $V = G$. However, we understand the term view direction more general since for controlling MCAO (Multi Conjugate Adaptive Optics, see [9]) systems, the view directions can be completely detached from the GSs. H is the height of the single atmospheric layer carrying a phase delay Φ , h is the height of the reconstruction layer on which the solver \mathcal{T} yields a reconstructed phase delay $\hat{\Phi}$, and $\Delta h := h - H$. We define Ω_H as the intersection of the atmospheric layer at height H with all the telescope's cylinders of view:

$$\Omega_H := \{ \mathbf{x} \in \mathbb{R}^2 \mid \forall v : \mathbf{x} - H \cdot \mathbf{d}_v \in \mathcal{D} \} = \bigcap_{v=1}^V (\mathcal{D} + H \cdot \mathbf{d}_v).$$

This is the subregion of the layer at height H , where the fields of view corresponding to all view directions are overlapping. The region that is actually relevant for the tomography problem is in general larger (see Section 2.1.4), but we concentrate on Ω_H first.

Each view direction causes $\hat{\Phi}$ to be shifted about $\Delta \mathbf{x}$ relative to Φ due to $\Delta h \neq 0$. For a certain view direction \mathbf{d}_v with $|\mathbf{d}_v| \leq \tan(\alpha)$, this shift is given by $\Delta \mathbf{x} = \Delta h \cdot \mathbf{d}_v$, and consequently, the L^2 -error has the form

$$(2.1) \quad E_{\mathbf{d}_v}^2 = \int_{\Omega_H} \left(\Phi(\mathbf{x}) - \hat{\Phi}(\mathbf{x} - \Delta h \cdot \mathbf{d}_v) \right)^2 d\mathbf{x},$$

with $|\Delta \mathbf{x}| \leq |\Delta h| \cdot \tan(\alpha)$; see Figure 2.1.

Note that under our restriction to Ω_H , the atmospheric phase delay Φ is only known (and relevant) in Ω_H , whereas the reconstructed $\hat{\Phi}$ can in general be any function in $L^2(\mathbb{R}^2)$. We need a reformulation of this expression to a form with the shift applied to Φ rather than $\hat{\Phi}$. For this purpose, we expand the domain of integration to the set of all values \mathbf{x} at which the evaluation of $\hat{\Phi}$ can contribute to (2.1):

$$\Omega_H^{\Delta h} := \{ \mathbf{x} \in \mathbb{R}^2 \mid \exists v : \mathbf{x} - \Delta h \cdot \mathbf{d}_v \in \Omega_H \} = \bigcup_{v=1}^V (\Omega_H + \Delta h \cdot \mathbf{d}_v).$$

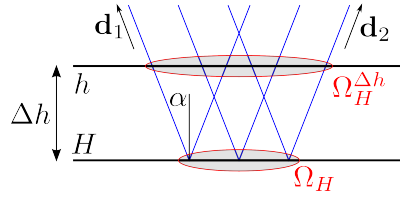


FIG. 2.1. Single atmospheric layer at height H (below) and reconstruction layer at height h (above) intersected by light rays (blue) in 2 view directions $\mathbf{d}_{1,2}$ at a maximum angle α . Ω_H is the intersection of the atmospheric layer with the cylinders of view corresponding to the guide stars; $\Omega_H^{\Delta h}$ is the extension to the reconstruction layer.

Further, let Φ^0 be the continuation of Φ from Ω_H to $\Omega_H^{\Delta h}$ with constant 0 outside Ω_H :

$$(2.2) \quad \Phi^0(\mathbf{x}) := \begin{cases} \Phi(\mathbf{x}) & \text{for } \mathbf{x} \in \Omega_H, \\ 0 & \text{else.} \end{cases}$$

Now $E_{\mathbf{d}_v}$ can be written as

$$E_{\mathbf{d}_v}^2 = \int_{\Omega_H^{\Delta h}} \left(\Phi^0(\mathbf{x} + \Delta h \cdot \mathbf{d}_v) - \hat{\Phi}(\mathbf{x}) \right)^2 d\mathbf{x}.$$

2.1.2. The central least-squares solution. According to our concept of optimizing the solver \mathcal{T} , we now assume that $\hat{\Phi}$ minimizes the overall (with respect to the view directions) residual error, which we define in the l^2 -sense: let

$$\mathbf{E}^2 := \sum_{v=1}^V E_{\mathbf{d}_v}^2.$$

Then our minimization problem is of the following form: for a given phase delay Φ on the atmospheric layer H and for the view directions $\mathbf{d}_1, \dots, \mathbf{d}_V$, find a reconstruction $\hat{\Phi}$ on the reconstruction layer h such that

$$(2.3) \quad \mathbf{E}^2 = \sum_{v=1}^V \int_{\Omega_H^{\Delta h}} \left(\Phi^0(\mathbf{x} + \Delta h \cdot \mathbf{d}_v) - \hat{\Phi}(\mathbf{x}) \right)^2 d\mathbf{x} \rightarrow \min.$$

If we denote by $\|\cdot\|_{H, \Delta h}$ and $\langle \cdot, \cdot \rangle_{H, \Delta h}$ the $L^2(\Omega_H^{\Delta h})$ -norm and the associated scalar product, respectively, and by

$$(2.4) \quad \Phi_v^0(\mathbf{x}) := \Phi^0(\mathbf{x} + \Delta h \cdot \mathbf{d}_v)$$

the residual minimization, then (2.3) can be written in the compact form

$$(2.5) \quad \mathbf{E}^2 = \sum_{v=1}^V \left\| \Phi_v^0 - \hat{\Phi} \right\|_{H, \Delta h}^2 \rightarrow \min.$$

Since the functional $\|\Phi_v^0 - \cdot\|^2$ as well as a sum over such expressions are strictly convex, there exists a unique minimum. This minimum can be found by setting the Fréchet-derivative to zero, which leads to

$$(2.6) \quad \hat{\Phi}_{\min} = \frac{1}{V} \sum_{w=1}^V \Phi_w^0.$$

Inserting $\hat{\Phi} = \hat{\Phi}_{\min}$ into the expression for \mathbf{E} in (2.5), we get a Φ -dependent minimal error that is theoretically achievable by \mathcal{T} :

$$(2.7) \quad \mathbf{E}_{\min}^2 = \mathbf{E}_{\min}^2(\Phi, \Delta h) := \sum_{v=1}^V \left\| \Phi_v^0 - \hat{\Phi}_{\min} \right\|_{H, \Delta h}^2.$$

For our further argumentation we require that \mathbf{E}_{\min} is expressed in terms of differences between the shifted phase delays Φ_v^0 . First we note that for f, g elements of any real Hilbert space it trivially holds that $\|f - g\|^2 = \|f\|^2 + \|g\|^2 - 2\langle f, g \rangle$. We define the differences $\Psi_v := \Phi_v^0 - \hat{\Phi}_{\min}$ and see that for all $v, w \in \{1, \dots, V\}$,

$$\left\| \Phi_v^0 - \Phi_w^0 \right\|_{H, \Delta h}^2 = \|\Psi_v\|_{H, \Delta h}^2 + \|\Psi_w\|_{H, \Delta h}^2 - 2\langle \Psi_v, \Psi_w \rangle_{H, \Delta h}.$$

Summing up this equation over all v and w leads to a representation without $\hat{\Phi}_{\min}$:

$$\begin{aligned} \sum_{v,w} \left\| \Phi_v^0 - \Phi_w^0 \right\|_{H, \Delta h}^2 &= 2V \sum_{k=1}^V \|\Psi_k\|_{H, \Delta h}^2 - 2 \sum_v \sum_w \langle \Psi_v, \Psi_w \rangle_{H, \Delta h} \\ &\stackrel{(2.7)}{=} 2V \cdot \mathbf{E}_{\min}^2 - 2 \left\langle \sum_v \Psi_v, \sum_w \Psi_w \right\rangle_{H, \Delta h} \\ &= 2V \cdot \mathbf{E}_{\min}^2 - 2 \left\langle -V \hat{\Phi}_{\min} + \sum_v \Phi_v^0, -V \hat{\Phi}_{\min} + \sum_w \Phi_w^0 \right\rangle_{H, \Delta h}. \end{aligned}$$

Due to (2.6), the factors in the scalar product vanish, and we get

$$(2.8) \quad \mathbf{E}_{\min}^2 = \frac{1}{2V} \sum_{v \neq w} \left\| \Phi_v^0 - \Phi_w^0 \right\|_{H, \Delta h}^2.$$

In this form the minimum residual error is only dependent on shifted copies of Φ , and we can use the atmosphere model—especially the autocorrelation function—as a distribution for the random quantity Φ in order to find an expected value.

2.1.3. The approximation of the expected value for the minimum residual error.

Since for all $\hat{\Phi}$ it holds that $\mathbf{E} \geq \mathbf{E}_{\min}$, a lower bound for \mathbf{E}_{\min} seems to be of interest. However, a usable lower bound cannot be found because \mathbf{E}_{\min} could theoretically be zero for special choices of Φ . For example, if the view directions \mathbf{d}_v have the property that there exists a $\mathbf{p} \in \mathbb{R}^2$ such that $(\mathbf{d}_v - \mathbf{d}_w) \cdot \mathbf{p} \in \mathbb{N}$ for all v, w (e.g., view directions on a regular square or hexagonal grid fulfill this condition), then a phase delay $\Phi(\mathbf{x}) = \sin(2\pi \mathbf{x} \cdot \mathbf{p} \Delta h)$ would lead to $\mathbf{E}_{\min} = 0$ except for the slim region $\Omega_H^{\Delta h} \setminus \Omega_H$. However, the probability that Φ is of such a special form is negligible in reality. What we really need is not an estimate for \mathbf{E}_{\min} in the sense of a best or worst case but an expression for the average case under the assumption that Φ is realistic. This means that we search for the expected value $\mathbb{E}(\mathbf{E}_{\min}^2)$ using the atmosphere models (1.5) or (1.6) as a distribution of the random 'variable' Φ .

In order to achieve an efficient layer compression method, we aim for an approximation of $\mathbb{E}(\mathbf{E}_{\min}^2)$ that can be evaluated fast, i.e., a formula similar to (1.7). Hence, we have to use several approximations and heuristic arguments in this section. The main justification lies in the fact that we are talking about expected values of completely unknown functions, the exact values of which are probably not of more practical use than estimates. Further, the satisfying performance of the resulting method has already been shown [15].

For functions Φ resulting from statistical models like (1.5) or (1.6), norms and scalar products over the domain \mathbb{R}^2 are in general infinite. However, in our subsequent derivations the need for a well-defined L^2 -norm of the phase delays will arise. Hence, we define a circular region $\Lambda \subset \mathbb{R}^2$ with diameter $d_\Lambda \gg \text{diam}(\mathcal{D})$ around a midpoint $m \in \mathcal{D}$. Further, we state the statistical model (especially the autocorrelation) of the phase delays to be given on Λ instead of \mathbb{R}^2 and assume that the error due to this simplification is negligible if Λ is large enough. In this sense we begin our analysis with the scalar product $\langle \cdot, \cdot \rangle_\Lambda$ and the norm $\|\cdot\|_\Lambda$ and for the shifted phase delays

$$\Phi_v(\mathbf{x}) := \Phi(\mathbf{x} + \Delta h \cdot \mathbf{d}_v)$$

without the truncation to Ω_H as in (2.2), (2.4).

Due to the representation (2.8) of \mathbf{E}_{\min} , it suffices to find expected values for the discrepancies between the shifted phase delays:

$$(2.9) \quad \mathbb{E} \left(\|\Phi_v - \Phi_w\|_\Lambda^2 \right) = \mathbb{E} \left(\|\Phi_v\|_\Lambda^2 + \|\Phi_w\|_\Lambda^2 - 2 \langle \Phi_v, \Phi_w \rangle_\Lambda \right).$$

Since Λ is very large compared to the shift $\Delta h \cdot \mathbf{d}_v$, it can be shown that the expected value of the difference between the norms $\|\Phi_v\|_\Lambda^2$ and $\|\Phi_w\|_\Lambda^2$ is negligible:

Let $\Lambda_v := \Lambda - \Delta h \cdot \mathbf{d}_v$, $\Lambda_v^+ := \Lambda_v \setminus \Lambda$, and $\Lambda_v^- := \Lambda \setminus \Lambda_v$. Then

$$\mathbb{E} \left(\|\Phi_v\|_\Lambda^2 \right) = \mathbb{E} \left(\|\Phi\|_{\Lambda_v}^2 \right) = \mathbb{E} \left(\|\Phi\|_\Lambda^2 + \|\Phi\|_{\Lambda_v^+}^2 - \|\Phi\|_{\Lambda_v^-}^2 \right).$$

Since $|\Lambda| = O(d_\Lambda^2)$ and $|\Lambda_v^\pm| = O(d_\Lambda)$, the relative error between $\mathbb{E} \left(\|\Phi_v\|_\Lambda^2 \right)$ and $\mathbb{E} \left(\|\Phi\|_\Lambda^2 \right)$ can be made arbitrarily small if d_Λ is chosen large enough. Hence, $\Phi_{v,w}$ can be replaced by Φ in the Λ -norms.

According to (1.4), the expected value of the scalar product $\langle \Phi_v, \Phi_w \rangle$ is nothing else than the autocorrelation function evaluated at $\Delta \mathbf{x} = \Delta h(\mathbf{d}_v - \mathbf{d}_w)$. Thus, (2.9) can be simplified to

$$(2.10) \quad \mathbb{E} \left(\|\Phi_v - \Phi_w\|_\Lambda^2 \right) \approx 2\mathbb{E} \left(\|\Phi\|_\Lambda^2 \right) - 2\mathbf{C}_\Phi(\Delta h(\mathbf{d}_v - \mathbf{d}_w)).$$

The autocorrelation function, when restricted to the domain Λ , takes its global maximum at $\Delta \mathbf{x} = 0$, and this maximum is identical to $\mathbb{E} \left(\|\Phi\|_\Lambda^2 \right)$. An additional property of \mathbf{C}_Φ is its rotational symmetry (isotropy of atmospheric turbulence), i.e., it depends only on $|\Delta \mathbf{x}|$. Equation (2.10) could be used in this form for all subsequent derivations, however, since we would like to obtain a method that solves the layer compression problem in acceptable run-time, we try to simplify (2.10) further. For that, we use the following approximation of \mathbf{C}_Φ by a power function in the vicinity of the maximum under the assumption that $|\Delta \mathbf{x}|$ is small:

$$(2.11) \quad \mathbf{C}_\Phi(\Delta \mathbf{x}) \approx \mathbb{E} \left(\|\Phi\|_\Lambda^2 \right) \cdot (1 - \gamma |\Delta \mathbf{x}|^\kappa),$$

with some constants $\gamma, \kappa > 1$. The essential property of this term is the fact that the difference $(\mathbf{C}_\Phi(0) - \mathbf{C}_\Phi(\cdot))$ is approximated by a multiplicative function, which will enable us to essentially simplify the resulting target function. The order of magnitude for $|\Delta \mathbf{x}|$ can be estimated in the following heuristic way: For future extremely large telescopes like the ELT of ESO, the field of view has a diameter of maximal $2\alpha \leq 7$ arc-minutes. According to the C_n^2 -profiles of commonly used atmosphere models, 95% of the turbulence occurs in the altitude range

0–20 km, and usually $L \geq 10$ reconstruction layers will be used by the tomographic solvers. Hence, assuming an equidistant distribution of the reconstruction layers, $\Delta h \leq 1000\text{m}$, and due to (1.2), we have for all g that $|\mathbf{d}_g| \leq 10^{-3}$. Thus, $|\Delta \mathbf{x}| = |\Delta h(\mathbf{d}_v - \mathbf{d}_w)| \leq 2\text{m}$. (In general, the optimal reconstruction layers are not equidistant, but we will see that distances $\Delta h \geq 20\text{km}/(2L)$ occur only in regions of weak turbulence). Using this estimate, we would like to assess the quality of the approximation (2.11). Algebraic representations of the inverse Fourier transforms of the von Karman power spectral density can be found in the literature, however, these formulas are rather complicated, and we found no easy way to apply one formalism to both models (1.5) and (1.6). Hence we prefer a numerical calculation by the two-dimensional IFFT for typical parameter values ($l_{\min} = 1\text{m}$ and $\tilde{L} = 20\text{m}$), noting that also the estimate for $|\Delta \mathbf{x}|$ is only approximate.

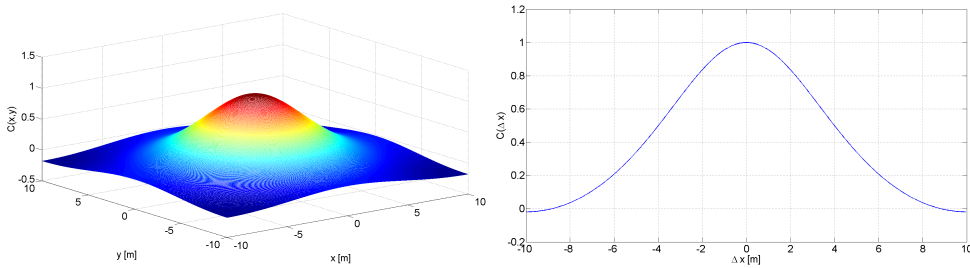


FIG. 2.2. The autocorrelation function $\mathbf{C}_{\Phi}(\Delta \mathbf{x})$ for the turbulence model according to von Karman (1.5) for the parameters $l_{\min} = 1\text{m}$ and $\tilde{L} = 20\text{m}$: 3D plot (left) and a section through the maximum (right). The plots are scaled such that the value at the central maximum is 1.

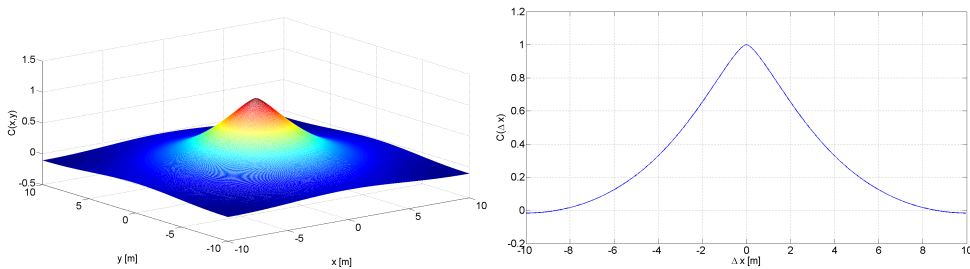


FIG. 2.3. The autocorrelation function $\mathbf{C}_{\Phi}(\Delta \mathbf{x})$ for the turbulence model (1.6) used for simulation with $\tilde{L} = 20\text{m}$: 3D plot (left) and a section through the maximum (right). The plots are scaled such that the value at the central maximum is 1.

In Figure 2.2 we display a three-dimensional plot (left) of \mathbf{C}_{Φ} , calculated by the IFFT of \mathbf{P}_n according to (1.5) as well as a cross-section (right) through the maximum at $\Delta \mathbf{x} = 0$. Within the range of interest $|\Delta \mathbf{x}| \leq 2\text{m}$, we can see that $\mathbf{C}_{\Phi}(\Delta \mathbf{x})$ can be approximated by a parabolic function $\mathbf{C}_{\Phi}(0) \cdot (1 - \gamma|\Delta \mathbf{x}|^2)$, i.e., (2.11) is applicable with $\kappa = 2$. For the model (1.6) used for simulations, the resulting autocorrelation function is displayed in Figure 2.3. In this case, a cone function $\mathbf{C}_{\Phi}(0) \cdot (1 - \gamma|\Delta \mathbf{x}|)$ appears to be a reasonable approximation within the range $|\Delta \mathbf{x}| \leq 2\text{m}$, i.e., (2.11) is applicable with $\kappa = 1$.

Of course, these arguments are quite heuristic. We note that the reason for using (2.10) and (2.11) is mainly a significant increase of efficiency of the resulting method (see Section 3) since the optimization algorithm has to evaluate $\mathbb{E} \left(\|\Phi_v - \Phi_w\|_{\Lambda}^2 \right)$ frequently. In spite of all

these simplifications, the results are still convincing [15].

Using $\Delta \mathbf{x} = \Delta h \cdot (\mathbf{d}_v - \mathbf{d}_w)$, (2.10), and (2.11), we can express the error as

$$(2.12) \quad \mathbb{E} \left(\|\Phi_v - \Phi_w\|_\Lambda^2 \right) \approx 2\gamma |\mathbf{d}_v - \mathbf{d}_w|^\kappa \cdot \mathbb{E} \left(\|\Phi\|_\Lambda^2 \right) |\Delta h|^\kappa.$$

In both statistical models (1.5) and (1.6), the power spectral density $\mathbf{P}_n(\omega)$ is linearly dependent on C_n^2 . Hence, there is a constant $\theta > 0$ such that

$$(2.13) \quad \mathbb{E} \left(\|\Phi\|_\Lambda^2 \right) = \theta \cdot C_n^2(H).$$

The last piece we need for finishing this section is a connection between the expected values of integrals over Λ and subsets of Λ . First we note that for any integrable function $f : \Lambda \rightarrow \mathbb{R}_0^+$ and any subset $T \subset \Lambda$ it holds that $\mathbb{E} \left(\int_T f \, d\mathbf{x} \right) = \frac{|T|}{|\Lambda|} \mathbb{E} \left(\int_\Lambda f \, d\mathbf{x} \right)$ if f obeys a statistical model independent of \mathbf{x} . Hence, using the notation $\Omega_{H|vw} := (\Omega_H - \Delta h \mathbf{d}_v) \cap (\Omega_H - \Delta h \mathbf{d}_w)$ and $\eta := \frac{1}{|\Lambda|}$,

$$\mathbb{E} \left(\langle \Phi_v^0, \Phi_w^0 \rangle_\Lambda \right) = \mathbb{E} \left(\langle \Phi_v, \Phi_w \rangle_{\Omega_{H|vw}} \right) = \eta \cdot |\Omega_{H|vw}| \cdot \mathbb{E} \left(\langle \Phi_v, \Phi_w \rangle_\Lambda \right).$$

Now we use the argument that $|\Delta \mathbf{x}| = |\Delta h(\mathbf{d}_v - \mathbf{d}_w)|$ is small ($\leq 2m$, see above) compared to the aperture diameter ($\approx 40\text{m}$ for the ELT). Further, the strongest turbulence occurs at lower altitudes, where $\text{diam}(\Omega_H)$ is still significantly larger than $|\Delta \mathbf{x}|$. Hence, we assume that $|\Omega_{H|vw}| \approx |\Omega_H|$ and approximate the expected difference of the truncated phase delays by the expected difference of the phase delays on Λ scaled by $\eta|\Omega_H|$:

$$(2.14) \quad \begin{aligned} \mathbb{E} \left(\|\Phi_v^0 - \Phi_w^0\|_{H,\Delta h}^2 \right) &= \mathbb{E} \left(\|\Phi_v^0\|_\Lambda^2 + \|\Phi_w^0\|_\Lambda^2 - 2 \langle \Phi_v^0, \Phi_w^0 \rangle_\Lambda \right) \\ &= \mathbb{E} \left(\eta|\Omega_H| \left(\|\Phi_v\|_\Lambda^2 + \|\Phi_w\|_\Lambda^2 \right) - 2\eta|\Omega_{H|vw}| \langle \Phi_v, \Phi_w \rangle_\Lambda \right) \\ &\approx \eta|\Omega_H| \cdot \mathbb{E} \left(\|\Phi_v - \Phi_w\|_\Lambda^2 \right). \end{aligned}$$

Now the expected residual error can finally be approximated by a rather simple expression that shows the dependence on Δh similar to the original idea (1.7):

$$(2.15) \quad \begin{aligned} \mathbb{E} \left(\mathbf{E}_{\min}^2 \right) &\stackrel{(2.8)}{=} \frac{1}{2V} \sum_{v \neq w} \mathbb{E} \left(\|\Phi_v^0 - \Phi_w^0\|_{H,\Delta h}^2 \right) \\ &\stackrel{(2.14)}{\approx} \frac{\eta|\Omega_H|}{2V} \sum_{v \neq w} \mathbb{E} \left(\|\Phi_v - \Phi_w\|_\Lambda^2 \right) \\ &\stackrel{(2.12)}{\approx} \frac{\eta|\Omega_H|}{2V} \left(\sum_{v \neq w} 2\gamma |\mathbf{d}_v - \mathbf{d}_w|^\kappa \right) \mathbb{E} \left(\|\Phi\|_\Lambda^2 \right) |\Delta h|^\kappa \\ &\stackrel{(2.13)}{\approx} \eta\theta\gamma \frac{|\Omega_H|}{V} \left(\sum_{v \neq w} |\mathbf{d}_v - \mathbf{d}_w|^\kappa \right) \cdot C_n^2(H) \cdot |\Delta h|^\kappa. \end{aligned}$$

The constant κ stems from the approximation (2.11) and depends on the shape of the auto-correlation function in the vicinity of its maximum and hence on the choice of the turbulence model. We have seen that for the original physical model (1.5), we can choose $\kappa = 2$, whereas for (1.6) as used in the simulation, the setting $\kappa = 1$ seems more suitable. Concerning the

other constants η, θ, γ , we note that their actual values are not important (as long as they are positive) since they do not effect the location of the minimum of the target function that we derive in Section 2.2.2. The next necessary step is to expand this result to the complete layer region of interest.

2.1.4. The regions of incomplete overlap. Until now we have analyzed the expected minimum residual error for the region Ω_H where all cylinders of view intersect. The complete region that is relevant for the tomographic solver is of course larger as it includes all regions where some—but not all—cylinders intersect, which creates a mosaic of sections with subsets of the view directions involved. Figure 2.4 displays a schematic example for $V = 4$ under the simplifying assumption that the aperture \mathcal{D} is a circle.

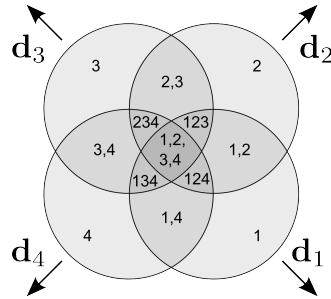


FIG. 2.4. A simple example for regions of incomplete overlap for the case $V = 4$ with a circular aperture \mathcal{D} . The indices of involved view directions are listed in each region.

For each one of these regions, our estimate (2.15) is in principle applicable with two modifications: First, the sum over the view directions has to be restricted to the ones relevant in each region, and V has to be adapted accordingly. Second, $|\Omega_H|$ in (2.15) has to be replaced by the area of the region.

For notational purposes, we define the set of all view direction indices as $\mathcal{V} := \{1, 2, \dots, V\}$ and the intersection of the cylinder of view in direction \mathbf{d}_v with the atmospheric layer at height H as

$$\Omega_{H,v} := \{\mathbf{x} \in \mathbb{R}^2 \mid \mathbf{x} - H \cdot \mathbf{d}_v \in \mathcal{D}\} = \mathcal{D} + H \cdot \mathbf{d}_v$$

for $v \in \mathcal{V}$. The complete relevant region is now given by $\bar{\Omega}_H := \bigcup_{v=1}^V \Omega_{H,v}$. Next, for any given index subset $\sigma \subset \mathcal{V}$ corresponding to a subset of view directions, we define the regions where only the view directions σ are 'active' and all others are not:

$$\Omega_{H,\sigma} := \left[\bigcap_{v \in \sigma} \Omega_{H,v} \right] \setminus \left[\bigcup_{v \notin \sigma} \Omega_{H,v} \right].$$

For example, in Figure 2.4 the region marked with '3,4' is $\Omega_{H,\sigma}$ for $\sigma = \{3, 4\}$. The complete region $\bar{\Omega}_H$ is the disjoint union of all these sets:

$$\bar{\Omega}_H = \bigcup_{\sigma \subset \mathcal{V}} \Omega_{H,\sigma} \quad \text{and} \quad \forall \sigma_1 \neq \sigma_2 : |\Omega_{H,\sigma_1} \cap \Omega_{H,\sigma_2}| = 0.$$

Now we add up the errors on the disjoint subregions to obtain the error on the complete set $\bar{\Omega}_H$. The approximate expected minimum residual error on the layer at height H takes the

form

$$\mathbb{E}(\mathbf{E}_{\min}^2) \approx \eta\theta\gamma \cdot \sum_{\emptyset \neq \sigma \subset \mathcal{V}} \frac{|\Omega_{H,\sigma}|}{|\sigma|} \left(\sum_{v,w \in \sigma} |\mathbf{d}_v - \mathbf{d}_w|^\kappa \right) \cdot C_n^2(H) \cdot |\Delta h|^\kappa,$$

where $|\sigma|$ denotes the number of elements contained in σ . In order to split this expression into parts with a comfortable visibility of dependencies, we define the layer-independent constant

$$(2.16) \quad B := \eta\theta\gamma \cdot \frac{|\mathcal{D}|}{V} \cdot \sum_{v,w \in \mathcal{V}} |\mathbf{d}_v - \mathbf{d}_w|^\kappa$$

and the layer-dependent weights

$$(2.17) \quad \omega(H) := \left(\frac{|\mathcal{D}|}{V} \cdot \sum_{v,w \in \mathcal{V}} |\mathbf{d}_v - \mathbf{d}_w|^\kappa \right)^{-1} \cdot \sum_{\emptyset \neq \sigma \subset \mathcal{V}} \frac{|\Omega_{H,\sigma}|}{|\sigma|} \sum_{v,w \in \sigma} |\mathbf{d}_v - \mathbf{d}_w|^\kappa.$$

Because for $H = 0$ it holds that $\Omega_H = \mathcal{D} = \bar{\Omega}_H$ and $|\Omega_{H,\sigma}| = 0$ for all $\sigma \neq \mathcal{V}$, we can see that $\omega(0) = 1$. Thus, the weights $\omega(H)$ have no geometry dependency and can be conveniently used as H -dependent weights. The expected minimum residual error can now be written as

$$\mathbb{E}(\mathbf{E}_{\min}^2) \approx B \cdot \omega(H) \cdot C_n^2(H) \cdot |\Delta h|^\kappa.$$

We did not find an appropriate (i.e., sufficiently simple) way to give analytical expressions for the areas $|\Omega_{H,\sigma}|$ of the subregions in the general case of an arbitrary aperture shape \mathcal{D} . Hence, a method for the numerical calculation of the weights $\omega(H)$ was implemented in MATLAB using the possibility of approximately representing subsets of \mathbb{R}^2 pixelwise by large Boolean matrices, which can be easily shifted and added or subtracted; the areas can then be approximated by simple pixel counting.

Some example plots for values of $\omega(H)$ as functions of H are displayed in Figure 2.5. For the sake of simplicity we chose a simple ring-shaped aperture, where the relative diameter of the central obstruction is varied (0, 0.2, 0.4, and 0.6) but other geometrical complications are neglected. It can be seen that for H larger than ≈ 40 km, the weights are zero. This is the limiting height H_{lim} , where the last overlap disappears; there is only at most one view direction active, and the sums over differences of view directions consequently yield zero. This corresponds to the fact that above this limiting height H_{lim} , tomography is not possible anymore. Consequently our method will ignore atmospheric layers at higher altitudes. Typically, realistic atmosphere models do not contain such layers anyway; this is achieved by choosing the GS separation appropriately. However, if any tomographic operation is wanted above H_{lim} , then the approximation option (2.19) should be used. A general analytic expression for the function $\omega(H)$ does not seem to be achievable—not even in this simple case of a purely circular geometry.

2.1.5. Optional approximations of the region weights. Using (2.17), the weights $\omega(H)$ can be calculated numerically for arbitrary aperture shapes and view directions. However, there are practical situations where one would like to run several simulations for varying telescope geometries but with the same reconstruction profiles. Moreover, our first experiments of AO simulations with generated profiles showed that in the end, the resulting reconstruction quality is quite insensitive to the values of $\omega(H)$. We suspect that this phenomenon is a consequence of the fact that in practice the function $\omega(H)$ is significantly smoother than the function $C_n^2(H)$. We suggest as a first approximation (see also Figure 2.5)

$$(2.18) \quad \omega(H) \approx \omega_1(H) := \begin{cases} 1 - H/H_{\text{lim}} & \text{for } H \leq H_{\text{lim}}, \\ 0 & \text{else.} \end{cases}$$

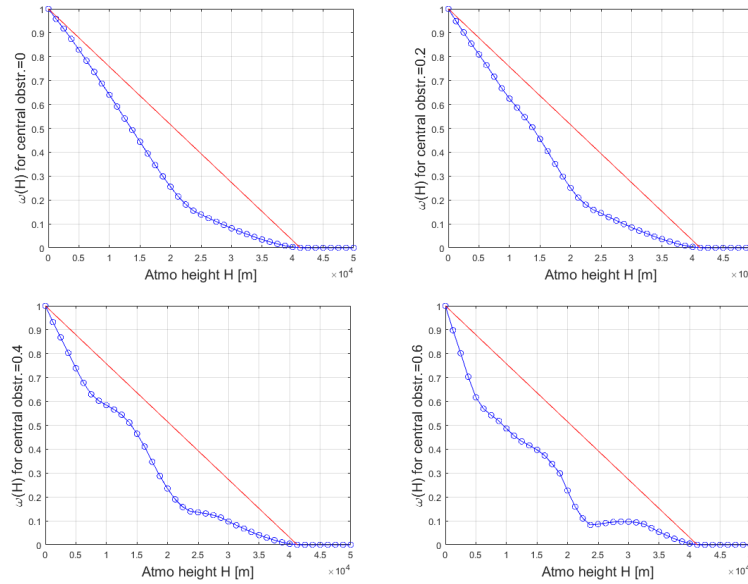


FIG. 2.5. The weights $\omega(H)$ (blue) for a typical E-ELT telescope geometry: 6 guide stars in hexagonal arrangement, guide star separation 3.5 arcmins, and aperture diameter 42m. The relative central obstruction diameter was chosen 0 (up left), 0.2 (up right), 0.4 (low left), and 0.6 (low right). The approximation $\omega_1(H)$ from (2.18) is shown in red.

H_{lim} is the limiting height for tomography, i.e., the height above which no more intersection of any cylinders of view can happen. The simple formula (2.18) is much easier to evaluate than the original definition (2.17). Further, the results of AO simulation seem to be extremely insensitive to the replacement of ω by ω_1 .

As a second possibility we mention a crude approximation:

$$(2.19) \quad \omega(H) \approx \omega_2(H) := 1.$$

Although this does not really look acceptable, the loss of quality in the results is remarkably small, and this trivial approximation has a clear advantage: it can be used in the case when no knowledge about the telescope geometry is available or if in a certain state of research it is required that a fixed profile is used for tests on, e.g., varying GS separations. This was the case for the investigations in [15], hence, ω_2 was used there. In addition, this choice should be used if tomographic operations above H_{lim} are to be performed.

2.1.6. A summary of the error estimate for one layer. In the preceding sections we investigated the best-case possibilities for a tomography solver \mathcal{T} for the simplified one-layer case: a single phase delay Φ in an atmospheric layer at height H has to be approximated by a reconstructed phase delay $\hat{\Phi}$ on a reconstruction layer at height h with $\Delta h = h - H$. Our approach was to assume that $\hat{\Phi}$ is the best approximating solution $\hat{\Phi}_{\text{min}}$. The resulting minimum residual error \mathbf{E}_{min} is dependent on Φ , which underlies a statistical turbulence model given in the form of the autocorrelation function. Using this model, we have derived an approximation for the expected value of \mathbf{E}_{min} ,

$$(2.20) \quad \mathbb{E}(\mathbf{E}_{\text{min}}^2) \approx \mathbf{E}_{\text{exp}}^2 := B \cdot \omega(H) \cdot C_n^2(H) \cdot |\Delta h|^\kappa.$$

The constant κ describes the shape of the autocorrelation function in the vicinity of its central maximum (2.11). It takes the value 2 for the original von Karman model (1.5) and $\kappa = 1$

for the simulation model (1.6). The constant B is defined by (2.16) based on the constants η , θ , and γ stemming from approximations discussed in Section 2.1.3. As we will see in Section 2.2.2, the actual value of B is not important—we only require B to be positive and independent of H and $C_n^2(H)$.

2.1.7. A corresponding solution with approximately minimum residual. Until now we have only derived an error estimate for the complete region, but a generalization of the solution itself is still missing. For this aim we apply the considerations from Section 2.1.2 leading to (2.6) to any of the subregions $\Omega_{H,\sigma}$ for $\sigma \in \mathcal{V}$: if we assume that the phase delay Φ in the atmospheric layer H has non-zero values only in $\Omega_{H,\sigma}$, then the solution $\hat{\Phi}$ in the reconstruction layer h with minimum residual is given by

$$\hat{\Phi}_{\min,\sigma} = \frac{1}{|\sigma|} \sum_{v \in \sigma} \Phi_v.$$

Because the subregions are disjoint, the complete phase delay Φ can be uniquely represented as a sum of functions supported by single subregions. Since the solution depends linearly on the phase delay, we suggest to set

$$(2.21) \quad \hat{\Phi}_{\min}^{H,h} := \sum_{\emptyset \neq \sigma \subset \mathcal{V}} \frac{1}{|\sigma|} \sum_{v \in \sigma} T^{\Delta h \cdot \mathbf{d}_v} (\Phi \cdot \chi_{\Omega_{H,\sigma}})$$

as an approximation of a solution with minimum residual. Here, χ_{Ω} denotes the characteristic function of Ω and the translation operator T is defined by (1.1). We do not need this solution explicitly for the single-layer analysis, but it will serve as a basis for our construction of a hypothetical solver in Section 2.2.1.

In the sequel we generalize our result (2.20) to the situation of N atmospheric layers and a tomographic reconstruction algorithm \mathcal{T} with the purpose of finding a solution to Problem 1.1 on L reconstruction layers with $L < N$.

2.2. The error estimate function for a complete reconstruction layer profile. We now consider an atmosphere model given by a finite sequence of increasing heights $0 < H_1 < H_2 < \dots < H_N$, each of them carrying a strength of turbulence given as $C_n^2(H)$. Further a number $L < N$ of reconstruction layers is given, usually chosen in order to meet real-time and stability demands for the tomographic reconstructor \mathcal{T} . The crucial question is now how to choose the heights $h_1 < h_2 < \dots < h_L$ of the reconstruction layer profile and the weights c_1^2, \dots, c_L^2 corresponding to the C_n^2 -values of the atmosphere, which usually play the role of weight parameters for \mathcal{T} . For that purpose, we aim at deriving an estimate for the minimum residual error dependent on this complete profile, which can subsequently be minimized.

2.2.1. A hypothetical solver with strictly valid locality properties. The result of Section 2.1.6 can only be of use if we can assume that a solution $\hat{\Phi}^{(l)}$ on the reconstruction layer l contributes to the sum in Problem 1.1 mainly as a compensation of $\Phi^{(n)}$ on those atmospheric layers at the heights H_n that are near to h_l . This motivates the notation of corresponding subsets of atmospheric layers:

For given H_n and h_l , we define the Voronoi intervals

$$I_1 := \left[0, \frac{h_1 + h_2}{2} \right), \quad I_L := \left[\frac{h_{L-1} + h_L}{2}, \infty \right), \quad \text{and} \quad I_l := \left[\frac{h_{l-1} + h_l}{2}, \frac{h_l + h_{l+1}}{2} \right),$$

for $l = 2, \dots, L-1$,

and

$$\tilde{G}_l := \{1 \leq n \leq N : H_n \in I_l\} \quad \text{for } l = 1, \dots, L.$$

Now \tilde{G}_l contains the indices n of those atmosphere layers with height H_n that are closer to h_l than to any other h_k with $k \neq l$. Using this notation, we require that the following locality properties to hold for all l :

- The result $\hat{\Phi}^{(l)}$ on the reconstruction layer at height h_l is mainly determined by the data stemming from atmospheric layers at height H_n with $n \in \tilde{G}_l$.
- All other atmospheric layers do not significantly contribute to $\hat{\Phi}^{(l)}$.

Of course this is not guaranteed in general since we have no information about the solver \mathcal{T} . In order to circumvent this problem, we construct a solver \mathcal{L} that has these properties and whose residual error can be estimated and is expected to be approximately minimal, at least in a statistical sense.

Let $\Phi = (\Phi^{(1)}, \dots, \Phi^{(N)})$ be the given phase delays in the atmospheric layers H_1, \dots, H_N , and let $\hat{\Phi} = (\hat{\Phi}^{(1)}, \dots, \hat{\Phi}^{(L)})$ denote the reconstructed delays in the reconstruction layers h_1, \dots, h_L . Then we define \mathcal{L} as the operator that assigns to each h_l the sum of the single-layer solutions $\hat{\Phi}_{\min}^{H, h}$ from (2.21) for $H = H_n$ over $n \in \tilde{G}_l$:

$$(2.22) \quad \hat{\Phi} = \mathcal{L}\Phi$$

with

$$(2.23) \quad \hat{\Phi}^{(l)} := \sum_{n \in \tilde{G}_l} \hat{\Phi}_{\min}^{H_n, h_l}.$$

We note that this solver is only of theoretical interest and cannot be used in practice since it uses information that is not available: in a running AO system, a tomographic solver has only information about sensor measurements of accumulated wave-fronts but not about specific phase delays in separated atmospheric layers. Further, \mathcal{L} will in general not yield the exact least-square solution; especially the optical energy $\|\hat{\Phi}\|^2$ will in general be significantly larger than the minimum. The advantage of \mathcal{L} is that its residual error can be estimated, and we can assume that this residual error is sufficiently close to the actual minimum residual error. We were not able to prove this. We therefore formulate it as a conjecture:

CONJECTURE 2.1. *The expected value of the residual error of \mathcal{L} defined by (2.22), (2.23) is not significantly larger than the expected value of the minimum residual error of the complete tomography problem Problem 1.1.*

At least a heuristic argumentation for the validity of this conjecture could be found in one of the following ways: First, tomography on finitely many two-dimensional layers is in a certain sense a discretization of a fully three-dimensional tomography problem. Hence a 'good' solver should reconstruct the three-dimensional distribution of the atmospheric phase delay correctly and consequently local structures of this function should not be reconstructed at a completely wrong height. Second, it might be possible to construct an atmospheric phase delay in a layer H_n that can be reconstructed with a small residual in a completely different reconstruction layer h_l if the special structure of the view directions is used for periodicity properties—similarly to the argumentation in the beginning of Section 2.1.3. But again, we can argue that such specially constructed phase delays will not occur in nature and can thus be ignored in our statistical approach: according to (2.20), the expected error is increasing if the distance between an atmospheric layer at height H_n carrying $\Phi^{(n)}$ and the reconstruction layer

on which $\hat{\Phi}^{(l)}$ is reconstructed gets larger. Hence, a smaller residual is likely to be reached if another reconstruction layer—one that is closer to H_n —tries to approximate $\Phi^{(n)}$. This means that the smallest residual can be expected if $n \in G_l$.

However, under the assumption that Conjecture 2.1 holds, the expected minimum residual error of the complete tomography problem can be approximately minimized by choosing the profile (h_1, \dots, h_L) if such a choice minimizes the expected minimum residual error of our hypothetical solver \mathcal{L} .

From the definition of \mathcal{L} it follows that each reconstruction layer at height h_l can in the sense of the locality properties be called ‘responsible’ for a certain subset of the atmospheric layers. Several methods for layer compression somehow use this principle, e.g., [1] by a similar introduction of Voronoi intervals. Usually, these subsets of responsibility are chosen a priori, and the values h_l and c_l are calculated by a simple and fast method. The idea of our approach is now to give up the advantage of having something fast and simple but instead invest calculation time to optimize the configuration of these subsets. We will refer to this configuration as *grouping* in the sequel.

2.2.2. The grouping of atmospheric layers and the overall sum of errors. We define ‘groups’ G_1, \dots, G_L as non-empty subsets of $\{1, 2, \dots, N\}$ with the properties

$$(2.24) \quad \bigcup_{l=1}^L G_l = \{1, 2, \dots, N\}, \quad \min(G_1) = 1; \quad \max(G_L) = N;$$

$$\forall 1 \leq l < L : \max(G_l) + 1 = \min(G_{l+1})$$

and call $\Gamma := \{G_1, \dots, G_L\}$ a ‘grouping’. Every group G_l contains the indices of those atmospheric layers for which reconstruction layer number l is responsible in the sense described above. The setup is displayed in Figure 2.6.

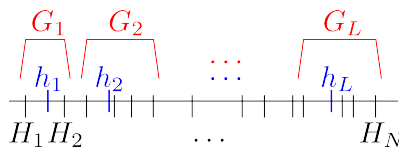


FIG. 2.6. Example for a grouping of atmospheric layers: $G_1 = \{1, 2\}$, $G_2 = \{3, 4, 5, 6\}$, ... Reprinted from [15].

REMARK 2.2. Note that G_l plays the same role as \tilde{G}_l in Section 2.2.1. We use a different notation here because G_l is defined independently of h_l . According to the definition of the optimization problem (2.25) and (3.1), it can be shown that G_l and \tilde{G}_l are actually the same if (Γ, \mathbf{h}) is a solution; see Remark 3.1.

We now would like to give an estimate for the minimal error that \mathcal{L} can achieve. According to the locality properties defined in Section 2.2.1, we assume that for the reconstruction layer h_l , only the data of the atmospheric layers with the indices in G_l are relevant. Hence, the contribution of the reconstruction layer l to the overall residual error is the sum of the single-layer errors within the group G_l . We have stated in (2.20) that the expected minimum residual error stemming from the discrepancy $\Delta h_{k,l} = h_l - H_k$ between an atmospheric layer at height H_k and a reconstruction layer at height h_l can be approximated by

$$\mathbf{E}_{\text{exp}}^2 = B \cdot \omega(H_k) \cdot C_n^2(H_k) \cdot |h_l - H_k|^\kappa.$$

Now the question arises, how to sum these error estimates in each group $k \in G_l$ and, further, how to sum over all groups. Since $\mathbf{E}_{\text{exp}}^2$ is an expected value of a squared residual norm, the

square root, \mathbf{E}_{exp} , has the character of a quantity that is linearly correlated to the error. These quantities form a vector (with indices k, l), the q -norm of which we want to minimize, taking into account the grouping of the layers, i.e.,

$$B \cdot \left(\sum_{l=1}^L \sum_{k \in G_l} (\omega(H_k) \cdot C_n^2(H_k))^q \cdot |h_l - H_k|^{q\kappa} \right)^{1/q} \rightarrow \min .$$

Since we are interested in the values h_l for which the minimum is attained, the value of B is irrelevant and we can ignore the exponent $1/q$. With the definitions

$$\rho_k := \omega(H_k) \cdot C_n^2(H_k), \quad p := q\kappa, \quad \text{and} \quad \beta := \frac{p}{\kappa},$$

we can define the *scaled error estimator*

$$(2.25) \quad E_{\mathbf{A}}^{p,\beta}(\Gamma, \mathbf{h}) := \sum_{l=1}^L \sum_{k \in G_l} \rho_k^\beta \cdot |H_k - h_l|^p,$$

where $\mathbf{A} := (H_1, \dots, H_N; C_n^2(H_1), \dots, C_n^2(H_N))$ denotes the information about the atmosphere profile, Γ is the grouping, and $\mathbf{h} = (h_1, \dots, h_L)$. The reason for introducing the parameters p and β is a practical one: On the one hand, it is not clear in which norm the vector of errors should be minimized, i.e., how to choose q . On the other hand, the resulting optimization method (see Section 3.1) can be implemented significantly faster if $p = 1$ or $p = 2$. Hence, it is more convenient to choose $p \in \{1, 2\}$ and let q be defined by $q = p/\kappa$. If the turbulence is modeled by (1.5), then $\kappa = 2$, and we have to choose $p = 2$ in order to guarantee that $q \geq 1$. The parameter β can in principle be chosen arbitrarily in (2.25) and is left open in the implementation for experimentation. It turned out during these tests (using the atmosphere model (1.6), i.e., $\kappa = 1$), that for both $p \in \{1, 2\}$, the setting $\beta = p/1$ is indeed the best choice with respect to the resulting quality in AO simulations; see also Section 3.5.

Since our principle aim is to derive a reconstruction layer profile \mathbf{h} that provides the best possible chance for \mathcal{L} to solve the tomography problem in Problem 1.1, we have to minimize the scaled error estimator (2.25) for appropriately chosen parameters p and β . A method for solving this minimization problem will be presented in Section 3. First we have to finish the theoretical part by illustrating how the weights c_1^2, \dots, c_L^2 are selected.

2.2.3. The choice of the predicting weights c_l^2 . According to our definition of \mathcal{L} , the reconstruction layer at height h_l carries a solution $\hat{\Phi}^{(l)}$ that has to compensate all phase delays $\Phi^{(k)}$ for $k \in G_l$; we have called this *responsibility*. Under our assumption that the piston mode is zero (1.3), the turbulence strength C_n^2 has the character of a variance. Since variances are additive, we can expect the turbulence strength of $\hat{\Phi}^{(l)}$ to be the sum of the turbulence strengths on the atmospheric layers in the corresponding group. This motivates the setting

$$(2.26) \quad c_l^2 := \sum_{k \in G_l} C_n^2(H_k), \quad \text{for all } l = 1, \dots, L,$$

where (G_l) is already minimizing (2.25). Experience [1] shows that tomographic solvers \mathcal{T} are usually rather insensitive to the weights c_l^2 , and it will probably not pay off to use a more sophisticated method than (2.26) for the calculation of the c_l^2 .

3. The minimization of the error estimation function. In Section 2 we have given the theoretical background for our paradigm that an approximately optimal reconstruction profile (h_1, \dots, h_L) can be found by minimizing the scaled error estimator (2.25). We will now propose a method for carrying out this minimization in practice.

Given an atmosphere profile $\mathbf{A} = (H_1, \dots, H_N; C_n^2(H_1), \dots, C_n^2(H_N))$, parameters $p \geq 1, \beta > 0$, and the number L of desired reconstruction layers, we want to find a grouping Γ and heights $\mathbf{h} = (h_1, \dots, h_L)$ such that

$$(3.1) \quad E_{\mathbf{A}}^{p,\beta}(\Gamma, \mathbf{h}) \rightarrow \min .$$

We will now reduce this target function to a purely discrete one, suggest a practical data structure for handling the groupings, and present an iterative algorithm for solving (3.1) in acceptable run-time.

3.1. The reduction to a discrete optimization problem. In our target function (3.1) we have to deal with continuous variables \mathbf{h} as well as discrete ones Γ . In a first step we simplify the problem to a purely discrete optimization problem. This can be done easily if the parameter p is chosen as 1 or 2. To see that, we assume for the moment that the grouping Γ is fixed. The minimization problem for \mathbf{h} has in this case a very simple solution $\mathbf{h}_{\min}(\Gamma)$:

- Let $p = 1$: Obviously, for each group G_l the internal sum $\sum_{k \in G_l} \rho_k^\beta \cdot |H_k - h_l|$ as a function of h_l is a convex piecewise linear function, and hence it takes its minimum at $h_l^* = H_{k^*}$ for some $k^* \in G_l$. This k^* is found by a simple evaluation for all $k \in G_l$ and taking k^* as the index for which the minimum is attained. We can do this independently for all groups, which yields the minimizing height profile $\mathbf{h}_{\min}(\Gamma) = (h_1^*, \dots, h_L^*)$ for any given grouping Γ .
- Let $p = 2$: In this case h_l minimizes $\sum_{k \in G_l} \rho_k^\beta (H_k - h_l)^2$ within each group G_l , which yields the unique solution

$$h_l^* = \frac{\sum_{k \in G_l} \rho_k^\beta H_k}{\sum_{k \in G_l} \rho_k^\beta},$$

defining the desired $\mathbf{h}_{\min}(\Gamma) = (h_1^*, \dots, h_L^*)$.

- For other values of p , the minimizer $\mathbf{h}_{\min}(\Gamma)$ can be calculated as well, but in general by far not as fast as for $p = 1, 2$. Since this calculation has to be performed very often within the process of finding the optimal grouping Γ , our experiments done so far were restricted to $p = 1$ or $p = 2$.

Having a fast implementation for the computation of $\mathbf{h}_{\min}(\Gamma)$ at hand, (3.1) reduces to the purely discrete problem of finding Γ such that

$$(3.2) \quad f(\Gamma) := E_{\mathbf{A}}^{p,\beta}(\Gamma, \mathbf{h}_{\min}(\Gamma)) \rightarrow \min .$$

REMARK 3.1. As an alternative, we could as well reduce (3.1) to a purely continuous optimization problem by calculating $\Gamma_{\min}(\mathbf{h})$ for given \mathbf{h} and by defining the target function $g(\mathbf{h}) := E_{\mathbf{A}}^{p,\beta}(\Gamma_{\min}(\mathbf{h}), \mathbf{h})$. It can be shown that $\Gamma_{\min}(\mathbf{h})$ is determined by the index subsets \tilde{G}_l of Section 2.2.1. However, it turned out that computing the global minimum of $f(\Gamma)$ can be implemented easier and more efficiently than minimizing $g(\mathbf{h})$. Thus, we choose the former approach.

sense that the iteration tended to get stuck at local minima and required a large number of iterations. The improved vicinity presented above seems to avoid this problem by performing larger jumps through the space of valid groupings though still containing the antecedent one.

Using the representation of Γ as a binary number b from Section 3.2, the vicinity can be expressed in the following way: If b corresponds to Γ and b' to Γ' , then the relation between Γ and Γ' defined above has a simple equivalent: b' is constructed from b by 'moving' one of the ones to another position, i.e., changing one digit from one to zero and another one from zero to one. In terms of the Hamming distance

$$D(b, b') := |\{0 < j < N : b_j \neq b'_j\}| = W(b \text{ XOR } b')$$

this can be written in the compact form $D(b, b') = 2$.

It can easily be seen that for any given b , there are $(L - 1) \cdot (N - L)$ possibilities for the choice of b' : simply set one of the $(L - 1)$ bits carrying one to zero, and set one of the bits that have been zero from the beginning to one. Consequently, we have

$$|V(\Gamma)| = (L - 1)(N - L).$$

Hence it is feasible to evaluate $f(\Gamma')$ from (3.2) for all $\Gamma' \in V(\Gamma)$ for any given Γ —this is exploited by our minimization Algorithm 1 in each iteration step.

Algorithm 1 Minimization of the error estimator (published earlier in [15]).

Input: L , Atmosphere profile \mathbf{A} , parameters p, β , number of starting values R .
 Choose arbitrary grouping Γ_{best} . ▷ Will be overwritten.
for $r = 1, \dots, R$ **do**
 Choose initial grouping Γ_r^0 randomly or according to Section 3.4.
 Set $\Gamma_{new} := \Gamma_r^0$.
 repeat
 Set $\Gamma_{old} := \Gamma_{new}$.
 Create the vicinity $U := V(\Gamma_{old}) \cup \{\Gamma_{old}\}$.
 Find $\Gamma_{new} = \operatorname{argmin}_{\Gamma' \in U} (f(\Gamma'))$. ▷ NOTE: f (3.2) depends on p and β .
 until $f(\Gamma_{new}) = f(\Gamma_{old})$
 if $f(\Gamma_{new}) < f(\Gamma_{best})$ **then**
 Set $\Gamma_{best} := \Gamma_{new}$.
 end if
end for
for $l = 1, \dots, L$ **do**
 $c_l^2 := \sum_{k \in G_l} C_n^2(H_k)$ according to (2.26), where G_l are the groups of Γ_{best} .
end for
Output: Reconstruction layers $\mathbf{h} := \mathbf{h}_{\min}(\Gamma_{best})$ and the weights c_l^2 .

An iterative algorithm of this form is of course not guaranteed to find the global minimum. The question remains, how to choose the starting values Γ^0 .

3.4. Recommended starting values. The problem of finding the global minimum of (3.2) can simply be solved by applying Algorithm 1 for a sufficiently large number R of different randomly chosen starting values Γ^0 . However, this leads to the questions how many starting values should be tried, which probability for finding the global minimum is acceptable, and if the dependence on the random generator is problematic. Hence we propose a small

number of recommended deterministic starting values by which an almost optimal result can be achieved in a fast way even without an additional use of random starting values.

A first approach is a grouping $\Gamma^{0,eq}$ that uses the same number of indices in each group and distributes the remaining indices over the lower groups if N is not an integral multiple of L . Let $q \in \mathbb{N}$ and $0 \leq r < L$ be such that $N = qL + r$. Then

$$|G_l^{0,eq}| = \begin{cases} q + 1 & \text{for } l \leq r, \\ q & \text{else,} \end{cases} \quad \text{for all } l = 1, \dots, L.$$

This choice of $\Gamma^{0,eq}$ is motivated by the following consideration: keeping the number of indices approximately the same in all groups will produce a reconstruction layer height profile \mathbf{h} that approximately inherits the layer density of the atmosphere in the sense of 'layers per unit length'. We call $\Gamma^{0,eq}$ the *standard starting value* (SSV). It turned out that this starting value seems to find the global minimum in many but not all cases. In order to decrease this risk of failure, an additional sequence of starting values was constructed: for every $m = 1, \dots, L$ we define $\Gamma^{0,\delta;m}$ such that

$$|G_l^{0,\delta;m}| = \begin{cases} N - L + 1 & \text{for } l = m, \\ 1 & \text{else,} \end{cases}, \quad \text{for all } l = 1, \dots, L.$$

$\Gamma^{0,\delta;m}$ is called the *mth delta starting value* (DSV). There is no real physical motivation for these values; they were found by trial-and-error.

In addition, we propose two more deterministic starting values, which are physically more meaningful than the SSV. For that aim, we first define two operations on groupings containing a number L' of groups with L' not necessarily equal to L :

- Greedy Merge: For a given grouping containing L' groups, find the best merge of two neighboring groups (i.e., leading to the minimal increase of f) giving a new grouping containing $L' - 1$ groups.
- Greedy Split: For a given grouping containing L' groups, find the best split of one group into two groups (i.e., leading to the maximal decrease of f) giving a new grouping containing $L' + 1$ groups.

Using these fast and simple operations, two new starting values ('Greedy Starting Values', GSVs) can be defined: The first GSV, $\Gamma^{0,g\uparrow}$, is generated from the trivial grouping with $L' = 1$ (containing only one group $G_1 = \{1, \dots, N\}$) by an $(L - 1)$ -fold application of the Greedy Split routine. The second GSV, $\Gamma^{0,g\downarrow}$, is generated from the opposite trivial grouping with $L' = N$ (containing N one-element groups) by an $(N - L)$ -fold application of the Greedy Merge routine. In terms of the representation of groupings by binary numbers, Greedy Merge simply sets one digit from one to zero, and Greedy Split sets one digit from zero to one.

Our current implementation begins by testing the SSV, the DSVs (there are L of them), and the two GSVs, which gives $L + 3$ deterministic starting values. In addition, a chosen number of random starting values (RSVs) is used. All local minima that are attained are stored together with the information at which starting values they were found. It turned out that the overall number of local minima is quite small, and we do not have to use a very large number of RSVs. Even without RSVs, the deterministic starting values find the global minimum with high probability, and if not, then they find at least a local minimum with the function value f very close to the global minimum.

3.5. The choice of parameters. The method suggested in Sections 3.3 and 3.4 turned out [15] to be suitable for the minimization of our scaled error estimator (2.25) when the parameters are simply chosen as $p = \beta = 1$ and $\omega(H)$ is approximated by the constant 1. For more investigations on the variation of these parameters and the behavior of the minimization

algorithm, Algorithm 1, and the starting values Γ^0 , a follow-up publication is planned. We only summarize our first findings here.

It turned out that the method is rather insensitive to the variation of the parameters except for a strong numerical evidence to choose $\beta = p$. This coincides with the theoretical recommendation to set $\beta := p/\kappa$ (see Section 2.2.2) and the usage of (1.6) in the simulations, i.e., $\kappa = 1$ in (2.11). Switching between $p = \beta = 1$ and $p = \beta = 2$ (note that we made the restriction $p \in \{1, 2\}$ in Section 3.1) seems to have remarkably low impact on the resulting performance of a full AO simulation. Since the algorithm is slightly faster for $p = 2$, we recommend this choice—and consequently $\beta = 2$ —as a default value.

If phase delays in a real atmosphere are to be reconstructed and these physical phase delays are more in accordance with the original von Karman model (1.5), then our theory suggests the choice $\kappa = 2$ and consequently $p = 2$ and $\beta = p/2 = 1$. However, until now we have no experience with ‘on-sky’ reconstruction, i.e., the profiles produced by Algorithm 1 have not been tested on real data from a telescope so far.

Another setup that has to be specified is how the subregion weights $\omega(H)$, discussed in Section 2.1.4, are to be calculated or approximated. It turned out that the results are remarkably insensitive to this choice. We recommend to use the linear approximation (2.18) if at least the limiting height for overlap, H_{lim} , is known from the telescope geometry (using the exact values seems to bring no significant benefit and their calculation takes time) or the constant approximation (2.19) if no information about the telescope is available or independence of the telescope geometry is required for some reason. It has to be noted that if the linear approximation or the exact values are used, then the atmospheric layers above H_{lim} are completely ignored. If there is a need to take these layers into account, then it is also recommended to use the constant approximation.

3.6. Implementation. Algorithm 1 was implemented[‡] in MATLAB and used for the results presented in [15]. The number of RSVs (random starting values) can be chosen as well as the parameters p , β , and the approximation mode for ω ; the choice of p is restricted to the values 1 or 2. The atmosphere profile has to be provided in the form of two vectors of dimension N containing the heights H_i and the C_n^2 -values, respectively, together with the desired number of reconstruction layers L . The resulting reconstruction layer profile is also provided as two corresponding vectors of dimension L . In addition, a list containing all found local minima of $f(\Gamma)$ is produced as output, each one with the information by how many RSVs and for which deterministic starting values it was found.

The code is not optimized for speed yet; for 50 starting values (this was sufficient in all cases tested so far) it takes a few seconds of run-time on a usual desktop PC dependent on the atmosphere profile and L . This is acceptable as a pre-processing step in the context of AO simulation, where tests usually run for several hours.

4. Conclusion and outlook. The basic idea of the optimal grouping method is to optimize the possibilities for a tomography solver to reach a small residual by choosing the reconstruction layer profile such that the theoretic limit for this residual is minimized. For that purpose we constructed a hypothetical solver \mathcal{L} that yields a solution to the tomography problem Problem 1.1 with a residual error that is supposedly close to the best possible one. Subsequently, the expected value of the residual of \mathcal{L} is minimized. It depends on the solver \mathcal{T} that is used in practice if this decreased limit can be exploited. Hence, there is no general guarantee that every solver will benefit from the reconstruction layer profiles generated by the optimal grouping method.

[‡]The implementation was entitled FRROG (Forecasting Residual Reduction by Optimal Grouping).

However, the results in [15] obtained with ESO's end-to-end simulation tool OCTOPUS [10] show that three investigated reconstruction solvers (FEWHA [20], FrIM [17, 18], and the gradient based method [16]) can indeed profit significantly from the profiles produced by Algorithm 1, especially in the case of a small number L of reconstruction layers and for large guide star separations. It is shown that using reconstruction layer profiles generated by Algorithm 1, all solvers perform superior in comparison to various other methods for layer compression under a variation of the guide star separation and the atmosphere model. All these tests were carried out with the simple parameter setting $p = \beta = 1$ and a constant ω -approximation (2.19). In addition, the effects of layer compression on the global profile properties like turbulence moments and generalized isoplanatic angles are investigated in [15].

The run-time of Algorithm 1 depends moderately on the atmosphere model and the number of requested reconstruction layers L but most significantly on the number of starting values used for the optimization routine proposed in Section 3. It turned out that in practice, few deterministic starting values (as suggested in Section 3.4) are sufficient and there is no need for a large number of random starting values. This yields a run-time of 3-10 seconds for our current implementation on a standard PC, which is acceptable for a pre-processing step in AO simulation.

Algorithm 1 has three parameters (p , β , and the approximation mode for ω), but it turned out that the results of the tomographic reconstruction are rather insensitive to their variation, and we can give clear recommendations on how to choose them; see Section 3.5. However, since the concept seems to be quite new, there are still several issues left for future research.

4.1. Open questions. An interesting special case is multi-conjugate adaptive optics (MCAO): in the case that a two-step method [6, 17, 21] or a three-step method [12, 14] is applied for the reconstruction, two subsequent tomography problems have to be solved, the second one resulting in commands for controlling the deformable mirrors (DMs). In this context, Algorithm 1 could be used for choosing optimal conjugation heights for the DMs. Another interesting question is, of course, how Algorithm 1 could be used for on-sky control, i.e., when applying it to data from a real telescope. In that case run-time requirements will arise that are not so important as long as Algorithm 1 is used as a pre-processing step preceding an AO simulation loop. In reality there might be updates for the atmosphere model available from measurements that are performed at a certain frequency, e.g., from SCIDAR [2, 5, 11, 19]. In this case the run-time of the optimization process must be significantly less than the time interval between two such updates. We first mention that there is a trivial possibility for parallelization since the outermost loop of the optimization Algorithm 1 is simply testing different starting values. Second, since the turbulence characteristics will probably not change fundamentally between two such measurements, the profile and grouping resulting from the old measurements will probably serve as good starting values for the new data as input.

Acknowledgments. When the first experiments with varying layer profiles [1] (it was a purely heuristic trial and error method without any perspective towards practical applicability) were presented in a project meeting with the ESO staff back in 2013, the importance and difficulty of choosing good layer profiles got clear to me when Enrico Fedrigo (ESO) said: 'You opened the box'.

Concerning the recent research presented in this article and in [15], I have to thank my colleagues in Linz (especially A. Obereder, S. Raffetseder, D. Saxenhuber, and M. Yudytskiy) and Miska Le Louarn (ESO) for running lots of AO simulations with the first profiles from Algorithm 1 as well as for helping with information and experience in several discussions. Further, I owe thanks to Ronny Ramlau, who encouraged me and gave me enough time for these investigations in spite of necessary research work in several other fields.

The research presented in this article was conducted in the context of the project "Astrophysics in the ELT era" funded by the Austrian Federal Ministry of Science, Research, and Economy.

REFERENCES

- [1] G. AUZINGER, *On choosing layer profiles in atmospheric tomography*, J. Phys. Conf. Ser., 595 (2015), Art. 012001 (6 pages).
- [2] R. AVILA, J. VERNIN, AND E. MASCIADRI, *Whole atmospheric-turbulence profiling with generalized scidar*, Appl. Opt., 36 (1997), pp. 7898–7905.
- [3] A. COSTILLE AND T. FUSCO, *Impact of Cn2 profile on tomographic reconstruction performance: application to E-ELT wide field AO systems*, in Adaptive Optics Systems III, B. L. Ellerbroek, E. Marchetti, J. P. Véran, eds., Proceedings of SPIE 8447, SPIE, Bellingham, 2012, Art. 844757 (9 pages).
- [4] H. ENGL, M. HANKE, AND A. NEUBAUER, *Regularization of Inverse Problems*, Kluwer, Dordrecht, 1996.
- [5] A. FUCHS, *Contribution à l'Étude de l'Apparition de la Turbulence Optique dans les Couches Minces. Concept du SCIDAR Généralisé*, PhD. Thesis, Université de Nice-Sophia Antipolis, Nice, 1995.
- [6] 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. Amer. A, 18 (2001), pp. 2527–2538.
- [7] T. FUSCO AND A. COSTILLE, *Impact of Cn2 profile structure on wide-field AO performance*, in Adaptive Optics System II, B. L. Ellerbroek, M. Hart, N. Hubin, and P. L. Wizinowich, eds., Proceedings of SPIE 7736, SPIE, Bellingham, 2010, Art. 77360J (10 pages).
- [8] J. GOODMAN, *Introduction to Fourier Optics*, 3rd ed., Roberts & Company, Greenwood Village, 2004.
- [9] J. W. HARDY, *Adaptive Optics for Astronomical Telescopes*, Oxford University Press, Oxford, 1998.
- [10] M. LE LOUARN, C. VÉRINAUD, V. KORAKOSKI, N. HUBIN, AND E. MARCHETTI, *Adaptive optics simulations for the European Extremely Large Telescope*, in Advances in Adaptive Optics II, B. L. Ellerbroek, D. B. Calia, eds., Proceedings of SPIE 6272, SPIE, Bellingham, 2006, Art. 627234 (9 pages).
- [11] G. LOMBARDI AND M. SARAZIN, *Using MASS for AO simulations: a note on the comparison between MASS and generalized SCIDAR techniques*, Monthly Notices Roy. Astronom. Soc., 455 (2016), pp. 2377–2386.
- [12] 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).
- [13] M. C. ROGGEMANN AND B. WELSH, *Imaging Through Turbulence*, CRC Press, New York, 1996.
- [14] M. ROSENSTEINER AND R. RAMLAU, *The Kaczmarz algorithm for multi-conjugate adaptive optics with laser guide stars*, J. Opt. Soc. Amer. A, 30 (2013), pp. 1680–1686.
- [15] D. SAXENHUBER, G. AUZINGER, M. L. LOUARN, AND T. HELIN, *Comparison of methods for the reduction of reconstructed layers in atmospheric tomography*, Appl. Opt., 56 (2017), pp. 2621–2629.
- [16] D. SAXENHUBER AND R. RAMLAU, *A gradient-based method for atmospheric tomography*, Inverse Probl. Imaging, 10 (2016), pp. 781–805.
- [17] M. TALLON, C. BÉCHET, I. TALLON-BOSC, M. LE LOUARN, E. THIÉBAUT, R. CLARE, AND E. MARCHETTI, *Performances of MCAO on the E-ELT using the fractal iterative method for fast atmospheric tomography*, in Second International Conference on Adaptive Optics for Extremely Large Telescope, On-line Proceedings Contributed Abstracts: AO Real-Time Control, 2011. <http://ao4elt2.lesia.obspm.fr/spip.php?article590>.
- [18] E. THIÉBAUT AND M. TALLON, *Fast minimum variance wavefront reconstruction for extremely large telescopes*, J. Opt. Soc. Amer. A, 27 (2010), pp. 1046–1059.
- [19] J. VERNIN, M. CROCHET, M. AZOUIT, AND O. GHEBREBRHAN, *SCIDAR/radar simultaneous measurements of atmospheric turbulence*, Radio Sci., 25 (1990), pp. 953–959.
- [20] M. YUDYTSKIY, *Wavelet Methods in Adaptive Optics*, PhD. Thesis, Industrial Mathematics Institute, Johannes Kepler University Linz, Linz, 2014.
- [21] M. YUDYTSKIY, T. HELIN, AND R. RAMLAU, *Finite element-wavelet hybrid algorithm for atmospheric tomography*, J. Opt. Soc. Amer. A, 31 (2014), pp. 550–560.