# An Approximation of the Riccati Equation in Large-Scale Systems With Application to Adaptive Optics

Paolo Massioni, Henri-François Raynaud, Caroline Kulcsár, Jean-Marc Conan

# An Approximation of the Riccati Equation in Large-Scale Systems with Application to Adaptive Optics

Paolo Massioni, Henri-François Raynaud, Caroline Kulcsár, Jean-Marc Conan

***Abstract*—The problem of finding linear optimal controllers and estimators, like linear quadratic regulators (LQRs) or Kalman filters, is solved by means of a matrix Riccati equation. A bottleneck of such an approach is that the numerical solvers for this equation are computationally intensive for systems with a high number of states, making it difficult if not impossible to apply optimal (minimum-variance) control and/or estimation methods to large-scale systems. A specific example is adaptive optics (AO) systems for the next generation of extremely large telescopes, for which the number of states to be estimated by a Kalman filter is in the order of the tens of thousands, making the numerical solution of the Riccati equations problematic. In this article, we show that for a special class of state-space systems, the discrete-time algebraic Riccati equation can be simplified with an approximation which leads to a closed-form solution, which can be computed more quickly and used as an alternative to standard numerical solvers. The class of systems for which this approximation holds includes a class of models widely employed in AO, namely autoregressive models of order 1 or 2 (AR1, AR2). We verify a posteriori the accuracy and applicability of the proposed solution.**

***Index Terms*—Adaptive optics, autoregressive models, Kalman filtering, large-scale systems, Riccati equation.**

## I. Introduction

Recent technological developments in miniaturization and microsystems have made it possible to construct systems with a huge number of actuators and sensors. For this reason, the system and control community has started devoting significant efforts on the topic of the so-called large-scale systems [27]. These systems have a high dimensionality for which standard control techniques might be too computationally demanding, too slow or simply unfeasible. Typically, the study of large-scale system relies on the exploitation of certain properties of the system itself in order to simplify the analysis / synthesis problem: the features that can be used to this effect are either a special structure, symmetry or invariance of the system [3], [33], [32] or the sparsity of the system matrices [8], [26], [16], [15]. In this article we will focus on large-scale systems which do not have these features, but for which an abundance of low-noise measurements is available.

This assumption is justified for one of the benchmark examples of large-scale systems, namely adaptive optics (AO) systems employed on ground-based telescopes. The resolution of these telescopes is significantly limited by the presence of the atmosphere, as the wavefront of the light collected at the primary mirror is distorted due to turbulence, which causes blur in the images. Adaptive optics [35] is a technique that allows the real-time correction of the turbulence effects on the image formation, by means of a deformable mirror (DM) compensating for the wavefront distortions. In practice the DMs are often considered to have no dynamics, so the key point in the efficient control of an AO system is the minimum variance estimation of the phase distortion from the measurements coming from a wavefront sensor (WFS). For the turbulence, a dynamic model can be assumed in order to approximate at best the physics of the problem; a common choice nowadays is given by autoregressive (AR) models [30], [32]. Under the hypothesis of an underlying dynamic model, the asymptotically optimal solution of the estimation problem on the infinite horizon is given by the steady-state Kalman filter [21].

The use of the Kalman filter implies solving the associated discrete algebraic Riccati equation (DARE) [20], which can be quite time-consuming for the next generation of extremely large telescopes, even for very simple autoregressive models. For example, the envisioned European extremely large telescope (E-ELT) [18] will feature a 39.3 m primary mirror and actuators and sensors in the range of the tens of thousands. In addition to this, the changing atmospheric conditions during a long-exposure will require a regular updating of the estimator, at least once every minute, which makes it a critical requirement to have a fast method for computing it. Although methods for solving high dimensionality DAREs have been recently developed [6], [7], they are not applicable in this case as they are applicable only if *all* the matrices in the equation are sparse; this is not the case in AO applications, due to the non-sparsity of the spatial correlation of atmospheric turbulence.

In this article, we propose a new and simple approach which makes it possible to skip the numerical solution of the Riccati equation by using an approximated closed-form solution, which allows significant savings in terms of computational time. The result is obtained with a Taylor expansion, a tool of perturbation analysis, which has a rich history in

The work of the first author has been partially supported by the Netherlands Organization for Scientific Research (NWO) and the Marie Curie COFUND Action through a Rubicon grant. This work has also been supported by the French National Research Agency (ANR) through the project CHAPERSOA ANR-09-BLAN-0162-01.

P. Massioni is with Laboratoire Ampère, UMR CNRS 5005, INSA de Lyon, Université de Lyon, 69621 Villeurbanne CEDEX, France paolo.massioni@insa-lyon.fr

H.-F. Raynaud and C. Kulcsár are with Laboratoire Charles Fabry - CNRS, Institut d'Optique Graduate School, RD 128, 91127 Palaiseau CEDEX, France {henri-francois.raynaud, caroline.kulcsar}@institutoptique.fr

J.-M. Conan is with ONERA, the French Aerospace Lab, F-92322 Châtillon, France jean-marc.conan@onera.fr

optimization and control [1], [10], [23]. The linearization is made possible by a set of assumptions which are typically satisfied for AO systems, namely an abundance of low-noise measurement channels. We will apply the technique to a few test cases, and we will verify *a posteriori* the accuracy of the approximations in terms of performance. The results of this article extend the ones in [24], where we considered only a special case featuring the simplest state-space model that could be applied to AO.

The article is organized as follows. Section II describes the approximation of the Riccati equation and the assumptions under which such an approximation is valid. Section III introduces adaptive optics and its dynamic models, and Section IV contains an evaluation of performance of the proposed approximation method for representative turbulence and sensor models. Finally, Section V presents some conclusions.

## II. Approximate solution to the minimum-variance estimation problem

### A. Discrete-time dynamical systems

Let us consider a stochastic discrete-time dynamical system in standard state-space form:

$$\begin{cases} x(k+1) = Ax(k) + Bu(k) + v(k) \\ y(k) = Cx(k) + w(k), \end{cases} \tag{1}$$

where $x(k) \in \mathbb{R}^n$ is the state vector, $u(k) \in \mathbb{R}^m$ is the input vector, $y(k) \in \mathbb{R}^r$ is the output vector and $k \in \mathbb{Z}$ is the discrete-time index. The vectors $v(k) \in \mathbb{R}^n$ and $w(k) \in \mathbb{R}^r$ are respectively the process and measurement noises, and they are assumed to be uncorrelated zero-mean Gaussian white noises, with covariance matrices $\Sigma_v$ and $\Sigma_w$. The estimation DARE associated with such a system is the following:

$$P = APA^T + \Sigma_v - APC^T \left(CPC^T + \Sigma_w\right)^{-1} CPA^T, \tag{2}$$

where $P \in \mathbb{R}^{n \times n}$ is the unknown. The positive definite solution $P$ of the Riccati equation for the standard steady-state Kalman filter is the asymptotic value of the covariance matrix of the state estimation error $\hat{x}(k+1|k) - x(k+1)$, where $\hat{x}(k+1|k)$ is the estimate of the state $x(k+1)$ based on the information at time $k$. The gain of the Kalman filter is then

$$K = APC^T(CPC^T + \Sigma_w)^{-1} \tag{3}$$

which is equivalent to

$$K = A(P^{-1} + C^T\Sigma_w^{-1}C)^{-1}C^T\Sigma_w^{-1} \tag{4}$$

thanks to a well-known matrix identity [28].

The solution for $P$ can be obtained by means of standard solvers, for example employing Schur-type algorithms [2], [22]. These methods however can be very time consuming (or even unfeasible) for high dimensional systems, as their computational complexity grows with the third power of the size of the matrices involved. The fast iterative alternating directions implicit (ADI) methods recently developed [6], [7] are alas not applicable for AO systems as $\Sigma_v$ is in general a full matrix with no constraints on its rank.

For this reason, we propose an approximation of the Riccati equation that will yield a simplified closed-form expression for its solution. A Riccati equation has in general multiple solutions, one of which is positive definite under a certain set of conditions [20]. The solution we are interested in for the Kalman filtering is the positive definite one, so from now on we will consider this solution only when we write $P$ and matrices deriving from it.

### B. Hypotheses

In order to obtain the simplified solution, we assume the following hypotheses.

1) $(A, C)$ is detectable, $(A, \Sigma_v^{1/2})$ is controllable, $\Sigma_v \geqslant 0$ and $\Sigma_w > 0$.
2) $A$ and $C$ can be partitioned as:
$$A = \begin{bmatrix} A_1 & A_{12} \\ A_{21} & A_2 \end{bmatrix}, \; C = \begin{bmatrix} C_1 & 0 \end{bmatrix} \tag{5}$$
with $A_1 \in \mathbb{R}^{n_1 \times n_1}$, and $C_1 \in \mathbb{R}^{r \times n_1}$, where $C_1$ is full column rank (i.e. $C_1$ has rank $n_1$, which implies that $C_1$ is a "tall" matrix, with $r \geqslant n_1$).
3) $\Sigma_w = \sigma_w^2 \Sigma$, where $\sigma_w^2$ is a scalar and $\Sigma > 0$; if the measurement noise for each channel is at the same level of intensity and there is no cross-correlation between channels, then $\Sigma = I$.

The first hypothesis implies that the Riccati equation has a unique strictly positive definite solution $P$ [20]. The fact that $C_1$ is full column rank means that the first part $x_1$ of the state vector (the one corresponding to the matrix $A_1$) can be directly reconstructed with the pseudoinverse of $C_1$. The hypotheses can also account for the case of $n_1 = n$, for which $A_2$, $A_{12}$ and $A_{21}$ degenerate to zero-dimension matrices.

### C. Linearization of the DARE

If we assume that the variance of the measurement noise is tending to zero, then we can write a Taylor expansion of the inverse terms in the Riccati equation, based on the assumption that $\sigma_w^2\Sigma$ is small compared with the solution $P$. This, together with a few additional hypotheses on the system, will eventually lead to a simplified formula for the solution. Nevertheless, the Riccati equation does not have a solution for $\sigma_w^2 = 0$; in fact we will be looking for an asymptotic or limit value that the solution assumes when $\sigma_w^2 \to 0$. We then remind that $o(z)$ for $z \to c$ indicates a quantity for which $\lim_{z \to c} \frac{o(z)}{z} = 0$. With this notation, the following well-known results hold.

**Lemma 1** (Matrix inversion lemma [19])**.** *Consider four matrices $X$, $Y$, $U$ and $V$. Then the following equality holds:*

$$(X + UYV)^{-1} = X^{-1} - X^{-1}U\left(Y^{-1} + VX^{-1}U\right)^{-1}VX^{-1} \tag{6}$$

*provided that the matrix dimensions are compatible and that the indicated matrix inverses exist.*

**Lemma 2.** *Let $X$ be an invertible matrix, $Y$ a square matrix of the same dimensions and $z$ a scalar. Then for $z \to 0$ it holds that:*

$$\begin{aligned} (X + zY)^{-1} \quad = \quad & X^{-1} - zX^{-1}YX^{-1} \\ & + z^2X^{-1}YX^{-1}YX^{-1} + o(z^2) \end{aligned} \tag{7}$$

*Proof: the expressions can be derived by noticing that* $(X+zY)^{-1} = (I+zX^{-1}Y)^{-1}X^{-1}$ *and then applying the matrix identity* $(I+A)^{-1} = I - A + A^2 - A^3 + \ldots = \sum_{i=0}^{\infty}(-A)^i$ *[38], which holds true if the spectral radius of* $A$ *is smaller than* 1 *(so it does for* $z \to 0$*).* ■

We then introduce a partition for $P$ and $\Sigma_v$ as well, according to the one defined in (5):

$$P = P^T = \begin{bmatrix} P_1 & P_{12} \\ P_{12}^T & P_2 \end{bmatrix}, \ \Sigma_v = \Sigma_v^T = \begin{bmatrix} \Sigma_{v1} & \Sigma_{v12} \\ \Sigma_{v12}^T & \Sigma_{v2} \end{bmatrix}, \tag{8}$$

where $P_1$ is strictly positive definite (and so invertible) due to the fact that $P$ is strictly positive definite. Notice that with the hypotheses of the previous section, $CPC^T = C_1 P_1 C_1^T$. Then (2) can be rewritten as:

$$\begin{aligned} P = APA^T + \Sigma_v \\ -A\begin{bmatrix} P_1 \\ P_{12}^T \end{bmatrix} C_1^T \left(C_1 P_1 C_1^T + \sigma_w^2 \Sigma\right)^{-1} C_1 \begin{bmatrix} P_1 & P_{12} \end{bmatrix} A^T. \end{aligned} \tag{9}$$

We can then derive the following lemma, which concerns the term subject to matrix inversion in (9).

**Lemma 3.** *Consider (9), under the hypotheses of Section II-B, for* $\sigma_w^2 \to 0$*. Then*

$$\begin{aligned} C_1^T \left(C_1 P_1 C_1^T + \sigma_w^2 \Sigma\right)^{-1} C_1 = \\ = P_1^{-1} - \sigma_w^2 P_1^{-1}(C_1^T \Sigma^{-1} C_1)^{-1} P_1^{-1} + o(\sigma_w^2). \end{aligned} \tag{10}$$

*Proof: first apply the matrix inversion lemma (Lemma 1) to* $(C_1 P_1 C_1^T + \sigma_w^2 \Sigma)^{-1}$*. This yields*

$$\begin{aligned} C_1^T \left(C_1 P_1 C_1^T + \sigma_w^2 \Sigma\right)^{-1} C_1 = \sigma_w^{-2} \left(C_1^T \Sigma^{-1} C_1 \right. \\ \left. - C_1^T \Sigma^{-1} C_1 (P_1^{-1} \sigma_w^2 + C_1^T \Sigma^{-1} C_1)^{-1} C_1^T \Sigma^{-1} C_1\right). \end{aligned} \tag{11}$$

*Notice that* $C_1^T \Sigma^{-1} C_1$ *is invertible, due to* $\Sigma$ *being strictly positive definite and* $C_1$ *full column rank. Then use Lemma 2, for which:*

$$\begin{aligned} (P_1^{-1}\sigma_w^2 + C_1^T \Sigma^{-1} C_1)^{-1} = \\ = (C_1^T \Sigma^{-1} C_1)^{-1} - \sigma_w^2 (C_1^T \Sigma^{-1} C_1)^{-1} P_1^{-1} (C_1^T \Sigma^{-1} C_1)^{-1} \\ + \sigma_w^4 (C_1^T \Sigma^{-1} C_1)^{-1} P_1^{-1} (C_1^T \Sigma^{-1} C_1)^{-1} P_1^{-1} (C_1^T \Sigma^{-1} C_1)^{-1} \\ + o(\sigma_w^4). \end{aligned} \tag{12}$$

*Replacing (12) into (11) and simplifying yields the statement of the lemma (remember that* $o(\sigma_w^4)\sigma_w^{-2} = o(\sigma_w^2)$*).* ■

This lemma shows that it is possible to linearize a part of the Riccati equation under the hypotheses of Section II-B. Nevertheless, developing the Riccati equation (9) with a full matrix $A$ leads to a block expression of the equation which does not have a simple solution. However, we can pinpoint two special cases of practical interest for which the linearization of Lemma 3 yields a very simple expression for $P$, which can be used for approximating the solution of the equation.

1) $A_{12} = 0$; this implies that the second block of the state is non observable, making (5) coincide precisely with the canonical observability decomposition.
2) $A_2 = 0$, $\Sigma_{v2} = 0$ and $\Sigma_{v12} = 0$; this implies that the second block of the state is just a delayed/scaled version of the first block, yielding a vector-valued second-order system.

The hypotheses of these special cases lead to the following theorems.

**Theorem 4.** *Consider (9) for* $\sigma_w^2 \to 0$*, under the hypotheses of Section II-B. If* $A_{12} = 0$ *and* $\Sigma_{v1}$ *is invertible, then*

$$P_1 = \Sigma_{v1} + \sigma_w^2 A_1 (C_1^T \Sigma^{-1} C_1)^{-1} A_1^T + o(\sigma_w^2) \tag{13}$$

$$P_{12} = \Sigma_{v12} + \sigma_w^2 A_1 (C_1^T \Sigma^{-1} C_1)^{-1} (\Sigma_{v1}^{-1} \Sigma_{v12} A_2^T + A_{21}^T) + o(\sigma_w^2). \tag{14}$$

**Theorem 5.** *Consider (9) for* $\sigma_w^2 \to 0$*, under the hypotheses of Section II-B. If* $A_2 = 0$*,* $\Sigma_{v2} = 0$ *and* $\Sigma_{v12} = 0$*, then*

$$\begin{aligned} P_1 \ = \ & \Sigma_{v1} + \sigma_w^2 A_1 (C_1^T \Sigma^{-1} C_1)^{-1} A_1^T \\ & + \sigma_w^2 A_{12} A_{21} (C_1^T \Sigma^{-1} C_1)^{-1} A_{21}^T A_{12}^T + o(\sigma_w^2) \end{aligned} \tag{15}$$

$$P_{12} = \sigma_w^2 A_1 (C_1^T \Sigma^{-1} C_1)^{-1} A_{21}^T + o(\sigma_w^2). \tag{16}$$

*Proof: both theorems can be proven by applying Lemma 3 to (13).* ■

These theorems can be used to derive a very simple approximation of $P_1$ and $P_{12}$ by taking the first-order terms in $\sigma_w^2$; both cases are of practical interest for our application, as it will be shown in Section III. Notice that the main result of [24] is a special case of Theorem 4, in the case where $A$ is a scalar times the identity matrix ($A = a_{\mathrm{R}} I$); Theorem 5 is not present in [24].

### *D. Kalman filter gain with first-order approximation*

If we approximate the matrix $P$ with its first-order expansion in $\sigma_w^2$, then in both cases (Theorem 4 and Theorem 5) the approximate Kalman gain is given by:

$$K_1 = \begin{bmatrix} A_1 P_{1,1} + A_{12} P_{12,1}^T \\ A_{21} P_{1,1} + A_2 P_{12,1}^T \end{bmatrix} C_1^T \left(C_1 P_{1,1} C_1^T + \sigma_w^2 \Sigma\right)^{-1}, \tag{17}$$

where $P_{1,1}$ and $P_{12,1}$ are the first-order approximations of $P_1$ and $P_{12}$, defined in (13) and (14) (or (15) and (16) according to the case). Notice that $K_1$ does not depend on $P_2$, which therefore does not need to be computed.

Using (4), $K_1$ can be equivalently rewritten as

$$K_1 = A \left(\sigma_w^2 P^{-1} + \begin{bmatrix} C_1^T \Sigma^{-1} C_1 & 0 \\ 0 & 0 \end{bmatrix}\right)^{-1} \begin{bmatrix} C_1^T \Sigma^{-1} \\ 0 \end{bmatrix}. \tag{18}$$

This formulation involves the inversion of a matrix of size $n \times n$, whereas (17) requires the inversion of $\left(C_1 P_{1,1} C_1^T + \sigma_w^2 \Sigma\right)$, which is $r \times r$. So this second formulation is in general computationally cheaper considering our hypothesis of $r \geqslant n_1$. Moreover, notice that $P_2$ will be simplified from the computations even if we use (18), as we know from (17) that the result does not depend on it.

This first-order Kalman filter gain $K_1$ can be used to implement an observer on the system that approximates the optimal Kalman filter. There is no guarantee that the estimator will be stable, but the stability can be easily verified *a posteriori*. In practice, the approximation is applicable when $\sigma_w^2$ is small compared with the singular values of $\Sigma_v$, which represent the intensity of the process noise.

## III. Adaptive Optics

AO systems compensate for the phase distortion of the light at the telescope pupil using a deformable mirror (Figure 1). The distortion is caused by the passage of light through the atmosphere, which turns the once flat incoming optical wavefront into a non-flat smooth surface.

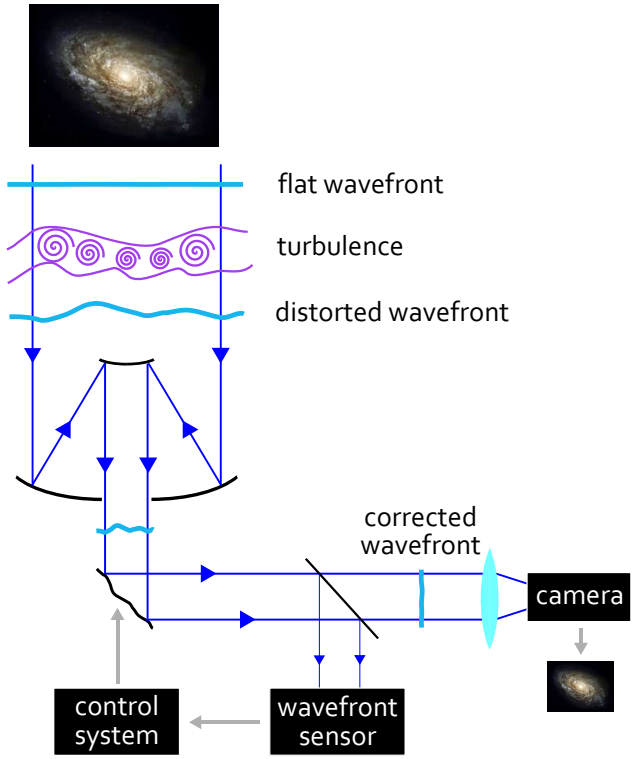



Fig. 1. Scheme of an AO system.

### A. Wavefront sensing

In classical AO systems, the phase distortion is assumed to be caused by a single equivalent turbulent layer located at the ground level, and such a distortion is measured with a single WFS monitoring the light from a single guide star (either natural or laser). We consider a Shack-Hartmann WFS [31], consisting of an array of small lenses (or lenslets) which partitions the light beam according to a regular grid. Each lenslet focalises its beam on a photon sensor; the position of the local image of the guide star depends on the local gradient of the wavefront. For this reason, the WFSs cannot measure the wavefront directly, but only its gradient in a number of points. We employ here a "zonal basis" for the turbulent phase $\phi(k)$, which means that we consider it as sampled on a regular square grid. For the Shack-Hartmann sensors, the output signal is the gradient of the phase sampled at the center of the squares delimited by four sample points of the phase (Figure 2). If for example we call the phase at the four corners $\phi_{\rm a}$, $\phi_{\rm b}$, $\phi_{\rm c}$ and $\phi_{\rm d}$, then the gradient at that point can be expressed as

$$s = \frac{1}{2}\begin{bmatrix} (\phi_{\rm b}+\phi_{\rm d})-(\phi_{\rm a}+\phi_{\rm c}) \\ (\phi_{\rm a}+\phi_{\rm b})-(\phi_{\rm c}+\phi_{\rm d}) \end{bmatrix}. \tag{19}$$

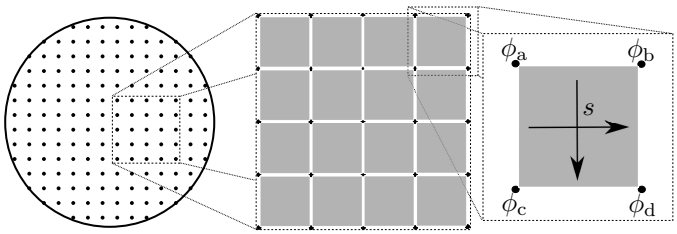



Fig. 2. Sensor geometry. The black dots represent the phase sampling points, and the gray areas represent the lenslets or subapertures of the sensor.

As a consequence, the output equation for the whole system can be written as

$$y(k) = G\phi(k) + w(k), \tag{20}$$

where $y(k)$ is the measurement output, $G$ is the matrix that expresses the gradient formula in (19) for all the points in the pupil and $w(k)$ is a white measurement noise; $G$ can be derived directly from (19). Each sensor channel is affected by a noise that can be assumed to have the same statistical properties and to be uncorrelated from the other channels, so that the covariance of $w(k)$ is $\sigma_w^2 I$. The output matrix is a gradient operator, so there are two modes of the phase that yield a zero output all the times, being thus unobservable. These are the piston mode, with constant phase on all the pupil (for which $\phi_{\rm a} = \phi_{\rm b} = \phi_{\rm c} = \phi_{\rm d}$) and the waffle mode (for which $\phi_{\rm a} = -\phi_{\rm b} = -\phi_{\rm c} = \phi_{\rm d}$ in each subaperture). The matrix $G$ is therefore not full column rank; it is in general "tall", because we have approximately two gradient measures per each phase sampling point ($r \approx 2n$), but it is not full rank due to the presence of the invisible modes (piston and waffle).

### B. Turbulence models

The complete AO system model requires also a model of atmospheric turbulence. A well-established option is Von Kármán's model [35], from which it is possible to compute the correlation of the phase distortion in two points of the layer as a function of their distance and a couple of parameters depending on the weather conditions, namely the Fried parameter $r_0$ and the outer scale $L_0$ [11]. The Fried parameter is a length that is roughly inversely proportional to the intensity of the turbulence; it is the maximum size of a telescope primary for which the turbulence effects are negligible, so that the images are diffraction-limited. The outer scale is the maximum size of the turbulent structures. Defining as $\phi(k)$ the column vector containing the values of the turbulent phase on a layer at a time instant $k$, it is then possible to construct the covariance of the turbulent phase vector $\Sigma_\phi = E[\phi(k)\phi(k)^T]$.

While the determination of the spatial correlations of the turbulence is a well-established field, the choice of a dynamic model that describes the temporal evolution of the phase is still object of research (see for example the recent contribution in [5] and the survey on AO control-oriented models in [21]). The use of Kalman filtering for the estimation of the turbulent wavefront requires assuming an underlying linear model for its dynamics, whose accuracy in portraying the turbulence is then a key element for the efficiency of AO control. A very

popular model choice, both for simplicity and effectiveness, is given by discrete-time autoregressive models, either of order 1 (AR1) [30], [32] or order 2 (AR2) [13].

*1) AR1 turbulence model:* The AR1 turbulence model is given by the expression

$$\phi(k+1) = A_{\mathrm{R}}\phi(k) + v(k), \tag{21}$$

where $v(k)$ is a white process noise. The matrix $A_{\mathrm{R}}$ is chosen as diagonal. If we assume a single turbulence layer, $A_{\mathrm{R}}$ has the same autoregression coefficient $a_{\mathrm{R}}$ (with $|a_{\mathrm{R}}| < 1$ ) in all its diagonal entries, so $A_{\mathrm{R}} = a_{\mathrm{R}}I$: the coefficient has to be chosen equal for all the phase points as the turbulence is a spatially homogeneous phenomenon. If we consider multiple turbulence layers located at different altitudes, then it is possible to assign a different coefficient for each layer, which can be used to take into account different dynamics (e.g. different wind speeds) at different altitudes. The covariance matrix $\Sigma_v$ of the process noise must be consistent with the chosen $\Sigma_\phi$. Computing the covariance of both sides of (21), we simply obtain

$$\Sigma_v = \Sigma_\phi - A_{\mathrm{R}}\Sigma_\phi A_{\mathrm{R}}^T. \tag{22}$$

We can discard the invisible modes from the zonal basis by employing an invertible state transformation $S$ such as $[\overline{\phi}\ \phi_{pw}]^T = S\phi$, where $\phi_{pw}$ is the vector containing the piston and waffle modes, and $\overline{\phi}$ is a vector of modes that are orthogonal to the invisible ones; then the system for $A_{\mathrm{R}} = a_{\mathrm{R}}I$ can be turned into the form of Theorem 4, as $GS^{-1} = [C\ 0]$ and $SAS^{-1} = a_{\mathrm{R}}I$. This allows the use of Theorem 4 to approximate the solution of the Riccati equation by taking the first-order truncation.

An additional simplification can be employed in the AR1 case; according to [14], [12], the inverse of $\Sigma_\phi$ can be approximated by the fourth-order derivative operator, i.e. the square of the Laplacian operator: $\Sigma_\phi^{-1} \approx \alpha\nabla^4$, (see [14] for indications on how to compute the constant $\alpha$). This implies, through (22), that $\Sigma_v = \alpha/(1-a_{\mathrm{R}}^2)\nabla^4 = \gamma\nabla^4$. It is important to remark that $\nabla^4$ is sparse. By using Lemma 2, we can then write the first-order approximation of the inverse of $P_1$ (according to Theorem 4) as

$$P_1^{-1} \approx \gamma\nabla^4 - \sigma_w^2\gamma^2\nabla^4 A_1(C_1^T\Sigma^{-1}C_1)^{-1}A_1^T\nabla^4 + o(\sigma_w^2). \tag{23}$$

We then neglect the piston and waffle terms, which does not degrade significantly the result, as the piston does not influence the quality of the images and the waffle is of very low power according to Von Kármán's model. Neglecting pistons and waffle is the same as taking $C_1 = C$, which is sparse, and $P_1 = P$. If we truncate the expression to the zero-order, we can arrive at the following for the inverse of $P$:

$$P_0^{-1} \approx \gamma\nabla^4 + \delta I \tag{24}$$

in which the term $\delta I$, with $\delta > 0$ an arbitrary small number, is a regularisation term that needs to be added in order to have the right hand side of (24) invertible. Then from (18) the zero-order Kalman gain is given by

$$K_0 = a_{\mathrm{R}}\left(C^T\Sigma^{-1}C + \sigma_w^2 P_0^{-1}\right)^{-1}C^T\Sigma^{-1}. \tag{25}$$

Notice that both (24) and (25) involve only sparse matrices, which can considerably boost the speed of the computations.

*2) AR2 turbulence model:* The AR2 turbulence model takes into account 2 past values of the state:

$$\phi(k+1) = A_{\mathrm{R1}}\phi(k) + A_{\mathrm{R2}}\phi(k-1) + v_1(k). \tag{26}$$

The matrices $A_{\mathrm{R1}}$ and $A_{\mathrm{R2}}$ are again chosen as diagonal, with the same diagonal entry for each layer. This can be translated into the following state-space model:

$$\begin{aligned}\begin{bmatrix}\phi(k+1)\\ \phi(k)\end{bmatrix} &= \begin{bmatrix}A_{\mathrm{R1}} & A_{\mathrm{R2}}\\ I & 0\end{bmatrix}\begin{bmatrix}\phi(k)\\ \phi(k-1)\end{bmatrix} + \begin{bmatrix}I\\ 0\end{bmatrix}v_1(k)\\ y(k) &= \begin{bmatrix}G & 0\end{bmatrix}\begin{bmatrix}\phi(k)\\ \phi(k-1)\end{bmatrix} + w(k).\end{aligned} \tag{27}$$

An empirical formula for obtaining the coefficients has been proposed in [29]. The covariance matrix $\Sigma_v$ consistent with $\Sigma_\phi$ is slightly more complicated to compute compared with the AR1 case. It can be obtained from the Lyapunov equation:

$$\begin{aligned}\begin{bmatrix}\Sigma_\phi & \Sigma_{\phi'}\\ \Sigma_{\phi'}^T & \Sigma_\phi\end{bmatrix} &= \begin{bmatrix}A_{\mathrm{R1}} & A_{\mathrm{R2}}\\ I & 0\end{bmatrix}\begin{bmatrix}\Sigma_\phi & \Sigma_{\phi'}\\ \Sigma_{\phi'}^T & \Sigma_\phi\end{bmatrix}\begin{bmatrix}A_{\mathrm{R1}}^T & I\\ A_{\mathrm{R2}}^T & 0\end{bmatrix}\\ &\quad + \begin{bmatrix}\Sigma_{v1} & 0\\ 0 & 0\end{bmatrix},\end{aligned} \tag{28}$$

where $\Sigma_\phi = E[\phi(k)\phi(k)^T] = E[\phi(k-1)\phi(k-1)^T]$ and $\Sigma_{\phi'} = E[\phi(k)\phi(k-1)^T]$. In the special case of $A_{R1}$ and $A_{R2}$ diagonal, with entries $a_{R1,i}$, $a_{R2,i}$ respectively in the diagonal, we have that the entries in position $(i,j)$ of $\Sigma_{\phi'}$ are simply:

$$(\Sigma_{\phi'})_{ij} = \frac{a_{R1,i} + a_{R2,i}a_{R1,j}}{1 - a_{R2,i}a_{R2,j}}(\Sigma_\phi)_{ij}, \tag{29}$$

where $(\Sigma_\phi)_{ij}$ is the entry in position $(i,j)$ of $\Sigma_\phi$. In the even simpler case of $A_{R1} = a_{R1}I$ and $A_{R2} = a_{R2}I$, we have $\Sigma_{\phi'} = a_{R1}(1-a_{R2})^{-1}\Sigma_\phi$.

The system in (27) fits into the hypotheses of Theorem 5 if the invisible modes (piston and waffle) are extracted (by the same state transformation $S$) and discarded (this leads from the non-full rank $G$ to a full-column rank $C_1$). As we have already stated, the piston does not influence the images and the waffle is of very low power, so this approximation is acceptable. This allows the use of the formulas of Theorem 5 for getting a first-order approximation of the Riccati equation.

## IV. Results

### A. Performance evaluation method

In this section, we will evaluate the effectiveness of the proposed method on two test cases by estimating the theoretical loss of performance due to the use of the proposed observer instead of the standard Kalman filter. As the piston component of the phase does not influence the image quality, we have to remove the piston contribution from the covariance matrix in order to have a meaningful evaluation of performance. We define $\Pi_{\mathrm{p}}$ as the projection matrix that removes the piston contribution, and $C_\phi$ as the output matrix which returns the phase $\phi$ from the state $x$ of the system, so that $\phi(k) = C_\phi x(k)$ ($C_\phi = I$ for AR1 models and $C_\phi = [I\ 0]$ for AR2).

When using the exact Kalman filter, $P$ is the variance of the state estimation error. So the mean variance of the phase

estimation error (without piston) is simply the average of the entries in the diagonal of $\Pi_{\rm p} C_\phi P C_\phi^T \Pi_{\rm p}^T$:

$$\sigma^2_{{\rm err},K} = \mathrm{mean}(\mathrm{diag}(\Pi_{\rm p} C_\phi P C_\phi^T \Pi_{\rm p}^T)). \tag{30}$$

This quantity determines the quality of the image. If a different observer is used, like the one featuring the gain $K_1$ that can be obtained with the proposed method, then the covariance of the state estimation error $\hat{x}(k+1|k) - x(k+1)$ will assume a value that we call $P_{K_1}$; under the hypotheses of using the turbulence temporal models of Section III, we have

$$\begin{array}{c}(A - K_1 C) P_{K_1} (A - K_1 C)^T - P_{K_1} \\ +\sigma_w^2 K_1 K_1^T + (1 - a_{\rm R}^2)\Sigma_\phi = 0.\end{array} \tag{31}$$

The expected covariance of the phase estimation error is then simply:

$$\sigma^2_{{\rm err},K_1} = \mathrm{mean}(\mathrm{diag}(\Pi_{\rm p} C_\phi P_{K_1} C_\phi^T \Pi_{\rm p}^T)). \tag{32}$$

In an AO system, the residual phase $\phi_{\rm r}$ (the one determining the image quality) is given by the phase $\phi$ of the light entering the pupil plus a correction term $\phi_{\rm c}$ generated by the deformable mirror:

$$\phi_{\rm r}(k) = \phi(k) + \phi_{\rm c}(k). \tag{33}$$

Under the hypothesis of being able to fully compensate for the estimated values of the phase with the deformable mirror ($\phi_{\rm c}(k) = -\hat{\phi}(k|k-1)$), the variance of the residual phase is the variance of the estimation error. The performance of the tested algorithm is then evaluated in terms of additional residual rms (root mean square) error with respect to the exact Kalman filter, which is the theoretical optimal.

### *B. First test: AR1 turbulence model*

We first consider a classical AO configuration with AR1 models stemming from standard classical AO situations with typical atmospheric profiles. We have considered telescope diameters from 2 to 42 m (the size of the first E-ELT proposed design), all of them with a phase sampling distance of 0.5 m. We have chosen a Fried parameter $r_0 = 0.53$ m (at the imaging wavelength), an outer scale $L_0 = 25$ m and an autoregression parameter $a_{\rm R} = 0.99$; the WFS has a measurement noise standard deviation of $\sigma_w = 45$ nm rms. For the chosen atmospheric condition, Von Kármán's model predicts that the standard deviation of the turbulence intensity is roughly 1900 nm rms, whereas the variance of the measurement signals at each WFS lenslet is around 460 nm rms. This confirms the assumption of a relatively low measurement noise. The controller runs at 250 Hz, and there is a one-step delay between the acquisition of the measurement and their use for computations.

The standard deviation of the measurement noise, as well as $r_0$ and $L_0$, have been taken from the specifications of the CANARY project [9], [37], [36], [34]; this project consists in a small-scale demonstrator for the E-ELT, so the intensity of the measurement noise that we have assumed can be considered as representative of current technology.

In the tests, we compare performance and computational times of the following methods.

- **Standard Kalman filter**, computed using Matlab's `dare` function. The cost of `dare` is approximately $75n^3$ floating point operations (FLOPs) [22], so we can estimate that the total cost of computing the Kalman gain, including (3), is $81n^3$ FLOPS. If we consider that the block $C^T\Sigma^{-1}C$ can be pre-computed only once (it depends only on the geometry and technology of the telescope and not on the atmospheric conditions), the number of total FLOPS reduces to $79n^3$.
- **Minimum mean square estimator (MMSE)**, the minimum-variance static estimator [4], [39], according to which the estimate of the phase is

  $$\hat{\phi}_{\rm MMSE}(k|k-1) = (C^T\Sigma^{-1}C + \sigma_w^2\Sigma_\phi^{-1})^{-1} C^T \Sigma^{-1} y(k-1). \tag{34}$$

  The cost of computing the estimator can be evaluated to be $6n^3$ FLOPS, or only $4n^3$ if $C^T\Sigma^{-1}C$ is pre-computed once and for all.
- **First-order Kalman filter**, computed with the formulas of Theorem 4 and (18). The total computational cost for obtaining the Kalman gain $K_1$ is $8.33n^3$ FLOPS. If $C_1^T\Sigma^{-1}C_1$ and $(C_1^T\Sigma^{-1}C_1)^{-1}$ are pre-computed, the total cost is only $4n^3$, roughly the same as the MMSE.
- **Zero-order Kalman filter**, computed with (24) and (25), which exploits the sparse approximation of the inverse of the turbulence variance matrix. In principle, the computational cost of this estimator is of the order of $n^2$, as all the matrices involved in the computations are sparse.
- **Distributed Kalman filter**, another approximate method based on a distributed control technique [25], [17], which approximates the telescope aperture with an infinite-sized one. The Kalman filter takes thus the form of a spatial convolution:

  $$\begin{cases} \hat{\phi}_{x_1,x_2}(k+1|k) = a_1 \hat{\phi}_{x_1,x_2}(k|k-1) \\ \quad + \displaystyle\sum_{x_1'=-z}^{+z} \sum_{x_2'=-z}^{+z} k_{x_1-x_1',x_2-x_2'} \big(y_{x_1',x_2'}(k) - \hat{y}_{x_1',x_2'}(k|k-1)\big) \\ \hat{y}_{x_1,x_2}(k|k-1) = F_{x_1,x_2}(\hat{\phi}(k|k-1)) \end{cases} \tag{35}$$

  where $x_1$, $x_2$ are the two ortogonal spatial coordinates of the telescope aperture's grid, $A = a_1 I$ and $F_{x_1,x_2}$ is basically the gradient function which generates the WFS output. This method has the advantage of being of order $n^0$: the cost of computing the convolution kernel $k_{x_1-x_1',x_2-x_2'}$ is the same for any size, being based on the infinite approximation. The only size-dependent cost is the one of reassembling the kernel in order to build the Kalman gain matrix.

The computational times given above are approximated to the higher order, and they have to be considered as a "worst case", as some special matrix structures could help reduce them. Parallelization of the solution algorithms might also be used to speed up the computations. As a last remark, we point out that the reduction in the computational complexity does not ease the problem of storing full, high dimension matrices, which might be another bottleneck in dealing with large-scale systems.

The first set of results, in Figure 3, shows the computational times that are required to compute the first- and zero-order approximation of the Kalman filter (with Theorem 4) compared with other methods, on a quad-core 3.6 GHz computer (running Matlab). The experimental computational times are the ones that are necessary to compute the Kalman gain, and they exploit, when possible, the possibility of pre-computing some parts.

It is possible to verify that the exact Kalman filter is the slowest to compute. The first-order approximation has a computational cost that is close to the one of the MMSE. The distributed method has an approximately constant computational complexity which becomes competitive for extremely large diameters. Remember that the size of the state vector is proportional to the square of the diameter (for $D = 42$ m, the state has $5540$ entries). We can also remark that the time needed for computing the first-order approximation is more than two orders of magnitude smaller with respect to the exact Kalman filter, whereas from the FLOPs computations above we could have expected a difference of one order of magnitude; this is probably due to Matlab's `dare` being slower than expected.

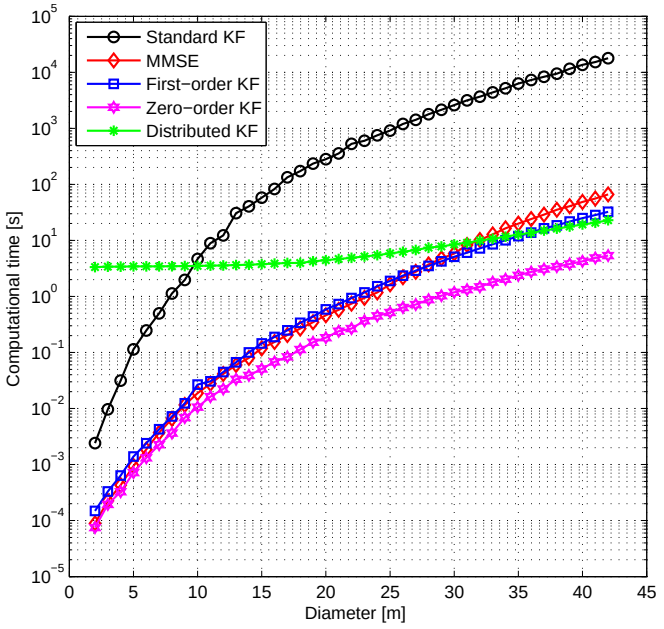



Fig. 3. Computational times for the different methods, on a quad-core 3.6 GHz computer running Matlab (test #1).

For what concerns performance, Figure 4 shows the additional residual error for the different approximate methods, which achieve stability in all cases. In order to have an idea of the global error budget, we can say that the total residual error due to the turbulence goes from $94$ nm rms (for $D = 2$ m) to $235$ nm rms (for $D = 42$ m) for the exact Kalman filter case. We can see that, after the exact KF, the best performance is given by the first-order approximation proposed in this paper; in particular, it performs better than the very fast distributed technique. This also justifies the choice of taking a first-order development without going further into higher orders: performance is already quite good for order one, and going into higher orders, which would feature a growing number of matrix terms, would significantly reduce the computational time advantages.

The zero-order method, as expected, is slightly faster to compute than the first-order one, at the cost of a little extra additional error, but still it outperforms the MMSE.

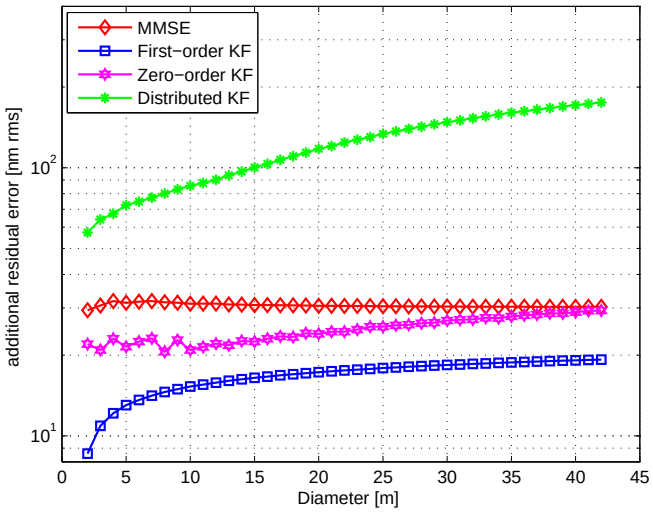



Fig. 4. Additional residual error (with respect to the exact Kalman filter), for different approximation methods, for $\sigma_w = 45$ nm rms, classical AO with an AR1 model (test #1).

We consider also the case of a higher measurement noise level, namely $\sigma_w = 90$ nm rms which can considered as a worst-case scenario for an ELT ([34]). From the results, (Figure 5) we can see that the first- and zero-order filters shows a slight performance degradation, as the assumption of a low level of noise becomes less applicable.

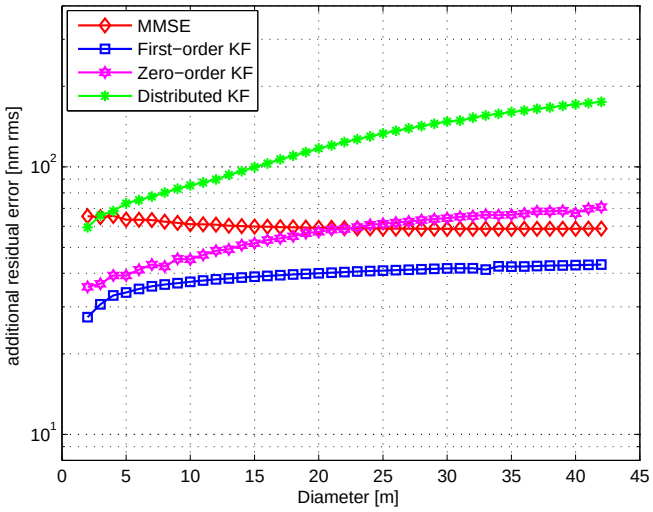



Fig. 5. Additional residual error (with respect to the exact Kalman filter), for different approximation methods, for $\sigma_w = 90$ nm rms, classical AO with an AR1 model (test #1).

### C. Second test: AR2 turbulence model

We have also evaluated the loss of performance in the case of an AR2 model, with $A_{\mathrm{R1}} = 1.98I$ and $A_{\mathrm{R2}} = -0.99I$, and the same atmospheric parameters $r_0$ and $L_0$ as in the previous case, and for $\sigma_w = 45$ nm rms. In Figure 6, the loss of performance of the MMSE is compared with the one of the first-order approximation (with Theorem 5). We can see that in the AR2 case, the first-order approximation is verified to be stable again and with acceptable performance, whereas the MMSE performs very poorly as it neglects the more complex dynamics. We have to stress that the MMSE is heavily penalized here due to the significant difference between the implicit absence of dynamics that it assumes and the AR2 model. In general the MMSE will have acceptable performance in real on-sky applications [9]; nevertheless, it has been noted through simulation that AR2 models can have better performance even when the turbulence is assumed to be the superimposition of translating frozen layers (the so-called Taylor's hypothesis), which models the reality better [13]. The formulas of Theorem 5 (for $n_1 = n/2$), taking advantage of precomputed blocks, require approximately $10n_1^3 \approx 1.2n^3$ FLOPs, compared with the $75n^3$ FLOPs required by the Riccati solver. The distributed filter and the zero-order approximation have not been evaluated because they are not directly applicable to the AR2 case, and we have not reached the maximum diameter because the computations of the standard Kalman gain are now too intensive, compared with the first case, as the order reaches $11080$ for $42$ m.

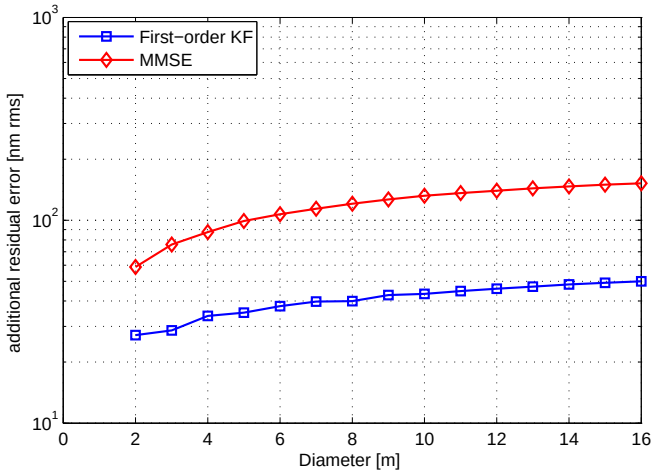



Fig. 6. Additional residual error (with respect to the exact Kalman filter), classical AO with an AR2 model, for $\sigma_w = 45$ nm rms (test #2).

## V. Conclusions

In this paper we have shown that for discrete-time state-space systems with abundance of low-noise measurements, it is possible to obtain, through linearization, a simple approximate solution of the Riccati equation associated with the minimum-variance estimation problem. The technique is applicable to dynamical systems that can be modeled as autoregressive stochastic processes; in this paper, we have applied it to the adaptive optics of astronomical imaging systems, where a Kalman filter can be used for reconstructing and predicting wavefront aberrations. We can foresee application in other real-time imaging systems which are governed by similar models and similar output equations (e.g. microscopy or medical imaging). For the AO application, we have verified in a realistic test case that the proposed first-order approximated solution is the most accurate compared with the distributed approximation method, allowing substantial savings in computational times with respect to a standard solver. The ability of quickly recomputing the Kalman filter gain is of critical importance in AO systems, as it allows the update of the control system in real-time in order to take into account changes in the atmospheric conditions (eg. change of wind velocity, temperature, etc.). The proposed method can then be considered as a good candidate for obtaining a fast and accurate solution of large-scale AO problems, allowing more trade-off choices between performance and speed of computations in AO controller design.

Further research might look into the dual optimal controller synthesis problem, for systems with a high number of input channels and a low penalty on the control effort. Incidentally, this case might still be useful in AO systems: in fact the active mirrors are systems with a number of input channels of the same order as the number of output channels of a WFS; moreover the control effort is not present in the AO cost function, which is the variance of the residual phase.

## References

[1] P. Apkarian and D. Noll. Controller design via nonsmooth multidirectional search. *SIAM J. Control and Opt.*, 44(6):1923–1949, 2006.

[2] W.F. III Arnold and A.J. Laub. Generalized eigenproblem algorithms and software for algebraic Riccati equations. *Proc. of the IEEE*, 72(12):1746–1754, 1984.

[3] B. Bamieh, F. Paganini, and M.A. Dahleh. Distributed control of spatially invariant systems. *IEEE Trans. Aut. Contr.*, 47(7):1091–1107, 2002.

[4] C. Béchet, M. Tallon, and É. Thiébaut. Closed-loop AO performance with FrIM. In *Adaptive Optics: Analysis and Methods/Computational Optical Sensing and Imaging/Information Photonics/Signal Recovery and Synthesis Topical Meetings*, Vancouver, Canada, 2007. OSA.

[5] A. Beghi, A. Cenedese, and A. Masiero. Stochastic realization approach to the efficient simulation of phase screens. *J. Opt. Soc. Am. A*, 25(2):515–525, 2008.

[6] P. Benner. Solving large-scale control problems. *IEEE Control Systems Magazine*, 24(1):44–59, February 2004.

[7] P. Benner and H. Fassbender. On the numerical solution of large-scale sparse discrete-time Riccati equations. *Advances in Computational Mathematics*, 35(2-4):119–147, November 2011.

[8] F. Borrelli and T. Keviczky. Distributed LQR design for identical dynamically decoupled systems. *IEEE Trans. Aut. Contr.*, 53(8):1901–1912, 2008.

[9] M. Brangier, F. Vidal, T. Morris, E. Gendron, Z. Hubert, A. Basden, G. Rousset, R. Myers, F. Chemla, A. Longmore, et al. CANARY MOAO demonstrator: on-sky first results. In *Adaptive Optics: Methods, Analysis and Applications (AO)*, Toronto, Canada, 2011. OSA.

[10] C. Büskens and H. Maurer. SQP-methods for solving optimal control problems with control and state constraints: adjoint variables, sensitivity analysis and real-time control. *J. Comput. Appl. Math.*, 120(1):85–108, 2000.

[11] R. Conan. Mean-square residual error of a wavefront after propagation through atmospheric turbulence and after correction with Zernike polynomials. *J. Opt. Soc. Am. A*, 25(2):526–536, 2008.

[12] C. Correia, J.-M. Conan, C. Kulcsár, H.-F. Raynaud, and C. Petit. Adapting optimal LQG methods to ELT-sized AO systems. In *Proc. of the 1st Conference on Adaptive Optics for Extremely Large Telescopes*, Paris, France, 2010.

[13] C. Correia, J.-P. Véran, and L. Poyneer. Gemini planet imager minimum-variance tip-tilt controllers. In *Adaptive Optics: Methods, Analysis and Applications (AO)*, Toronto, Canada, 2011. OSA.

[14] B.L. Ellerbroek. Efficient computation of minimum-variance wavefront reconstructors with sparse matrix techniques. *J. Opt. Soc. Am. A*, 19(9):1803–1816, 2002.

[15] W.R. Ferng, W.-W. Lin, and C.-S. Wang. The shift-inverted J-Lanczos algorithm for the numerical solutions of large sparse algebraic Riccati equations. *Computers & Mathematics with Applications*, 33(10):23–40, 1997.

[16] R. Fraanje, N. Doelman, and M. Verhaegen. Sparse prediction of turbulence in time and space. In *Proc. of the 1st Conference on Adaptive Optics for Extremely Large Telescopes*, Paris, France, 2010.

[17] L. Gilles, P. Massioni, C. Kulcsár, H.-F. Raynaud, and B. Ellerbroek. Distributed Kalman filtering compared to Fourier domain preconditioned conjugate gradient for laser guide star tomography on extremely large telescopes. *J. Opt. Soc. Am. A*, 30(5):898–909, June 2013.

[18] R. Gilmozzi and J Spyromilio. The European extremely large telescope (E-ELT). *The Messenger, ESO*, 127:11–19, 2007.

[19] G.H. Golub and C.F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, 1996.

[20] T. Kailath, A.H. Sayed, and B. Hassibi. *Linear Estimation*. Prentice Hall, 2000.

[21] C. Kulcsár, H.-F. Raynaud, C. Petit, and J.M. Conan. Minimum variance prediction and control for adaptive optics. *Automatica*, 48(9):1939–1954, 2012.

[22] A.J. Laub. A Schur method for solving algebraic Riccati equations. *IEEE Trans. Aut. Contr.*, 24(6):913–921, September 1979.

[23] F. Lin, M. Fardad, and M.R. Jovanovic. Optimal control of vehicular formations with nearest neighbor interactions. *IEEE Trans. Aut. Control*, 57(9):2203–2218, 2012.

[24] P. Massioni, C. Kulcsár, H.-F. Raynaud, and J.-M. Conan. Approximating the Riccati equation solution for optimal estimation in large-scale adaptive optics systems. In *Proc. of the 2012 American Control Conference*, Montréal, Canada.

[25] P. Massioni, C. Kulcsár, H.-F. Raynaud, and J.-M. Conan. Fast computation of an optimal controller for large-scale adaptive optics. *J. Opt. Soc. Am. A*, 28(11):2298–2309, June 2011.

[26] P. Massioni and M. Verhaegen. Distributed control for identical dynamically coupled systems: a decomposition approach. *IEEE Trans. Aut. Contr.*, 54(1):124–135, January 2009.

[27] K.M. Mohammadpour, J.; Grigoriadis, editor. *Efficient Modeling and Control of Large-Scale Systems*. Springer, 2010.

[28] K.B. Petersen and M.S. Pedersen. The matrix cookbook. *Technical University of Denmark*, 2008.

[29] C. Petit. Definition of AR2 models for LQG control. Technical Report RF 1/17838 DOTA, ONERA, 2010.

[30] C. Petit, J.-M. Conan, C. Kulcsár, and H.-F. Raynaud. Linear quadratic gaussian control for adaptive optics and multiconjugate adaptive optics: experimental and numerical results. *J. Opt. Soc. Am. A*, 26(6):1307–1325, June 2009.

[31] B.C Platt and R. Shack. History and principles of Shack-Hartmann wavefront sensing. *Journal of Refractive Surgery*, 17(5):S573–S577, 2001.

[32] L.A. Poyneer and J.-P. Véran. Optimal modal Fourier-transform wavefront control. *J. Opt. Soc. Am. A*, 22(8):1515–1526, August 2005.

[33] J.K. Rice and M. Verhaegen. Distributed control: A sequentially semi-separable approach for heterogeneous linear systems. *IEEE Trans. Aut. Contr.*, 54(6):1270–1283, 2009.

[34] 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(11):A201–A215, 2010.

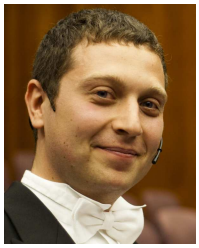
[35] F. Roddier, editor. *Adaptive optics in astronomy*. Cambridge University Press, 1999.

[36] G. Sivo, C. Kulcsár, J.-M. Conan, H.-F. Raynaud, E. Gendron, A. Basden, F. Vidal, T. Morris, S. Meimon, C. Petit, et al. Full LQG control with vibration mitigation: From theory to first on-sky validation on the CANARY MOAO demonstrator. In *Adaptive Optics: Methods, Analysis and Applications*, pages OTu2A–2. Optical Society of America, 2013.

[37] G. Sivo, C. Kulcsár, H.-F. Raynaud, J.-M. Conan, E. Gendron, and F. Vidal. MOAO real-time LQG implementation on CANARY. *AO4ELT,(Victoria, Canada)*, 2011.

[38] G. Stewart. *Matrix Algorithms: Basic decompositions*. SIAM, 1998.

[39] C.R. Vogel and Q. Yang. Fast optimal wavefront reconstruction for multi-conjugate adaptive optics using the Fourier domain preconditioned conjugate gradient algorithm. *Opt. Express*, 14(17):7487–7498, 2006.

**Paolo Massioni** got his MSc in Aerospace Engineering from Politecnico di Milano, Italy, in 2005, and his PhD in Control Engineering from Delft University of Technology, the Netherlands, in 2010. From 2012 he is assistant professor at Laboratoire Ampère at the French National Institute for Applied Science (INSA) of Lyon, France. Dr. Massioni has been visiting scientist in the European Space Research and Technology Center (ESTEC), Noordwijk, the Netherlands, and he is a past recipient of a Rubicon grant from the Netherlands Organisation for Scientific Research (NWO) and the Marie Curie COFUND Action. His main interests are distributed control of large-scale systems and adaptive optics.

**Henri-François Raynaud** is a researcher at Laboratoire Charles Fabry (Institut d'Optique Graduate School) and associate professor at Université Paris 13. He obtained a master's degree in statistics and stochastic modeling and a PhD in applied mathematics and control theory at Université Paris 11 - Orsay in 1990. He has been involved since the early 2000s in the design of high-performance controllers for adaptive optics systems in astronomy and retinal imaging, and he is now a member of the adaptive optics research team at Institut d'Optique. His research interests include adaptive control, stochastic modeling, linear systems theory, identification for control, fractional filters, $\mathcal{H}_\infty$ control and non-linear feedforward control, with applications to aerospace, industrial processes and telecommunication systems.

**Caroline Kulcsár** is professor at Institut d'Optique Graduate School, Université Paris 11. After graduating from the École Supérieure d'Informatique-Électronique-Automatique (Paris) and after a master's degree in control theory and signal processing at Université Paris 11 - Orsay, she received her PhD in control theory in 1995 at the Laboratoire des Signaux et Systèmes. She has been head of the adaptive optics (AO) team since 2001, initially with L2TI - Université Paris 13 until 2012, and with Laboratoire Charles Fabry - CNRS since then. Her research activities include identification and control of dynamical systems with applications to adaptive optics systems for astronomy and retinal imaging and performance modelling of super-resolution imaging systems.

**Jean-Marc Conan** is a senior research scientist at ONERA (Châtillon, France) with more than 20 years' experience in adaptive optics (AO). He graduated from Institut d'Optique Graduate School and received his PhD in 1994 (Université Paris 11 - ONERA) on AO theory. He is currently leading research activities in AO (wavefront sensing, control) for various fields of application: astronomy, ophthalmology, free space optics and defense. He was recently co-leader of a design study for the wide field AO module of the European Extremely Large Telescope.