Numerische Mathematik manuscript No.
(will be inserted by the editor)

The ADI method for bounded real and positive real Lur'e equations

Arash Massoudi · Mark R. Opmeer · Timo Reis

Received: 08 October 2014 / Accepted: date

Abstract We propose an algorithm for the numerical solution of the Lur'e equations in the bounded real and positive real lemma for stable systems. The algorithm provides approximate solutions in low-rank factored form. We prove that the sequence of approximate solutions is monotonically increasing with respect to definiteness. If the shift parameters are chosen appropriately, the sequence is proven to be convergent to the minimal solution of the Lur'e equations. The algorithm is based on the ideas of the recently developed ADI iteration for algebraic Riccati equations [10]. In particular, the matrices obtained in our iteration express the optimal cost in a certain projected optimal control problem.

Keywords Lur'e equation · ADI iteration · numerical method in control theory · linear-quadratic optimal control · bounded real lemma · positive real lemma

Mathematics Subject Classification (2000) 15A24 · 49N10 · 47J20 · 65F30 · 49M30 · 93B52 · 65K10

1 Introduction

We consider an algorithm for the approximation of the minimal solutions of the bounded real and positive real Lur'e equations. In this introduction we focus on the bounded real case

$$\begin{aligned} A^*X + XA + C^*C &= -K^*K, \\ B^*X + D^*C &= -J^*K, \\ D^*D - I &= -J^*J, \end{aligned} \tag{1}$$

A. Massoudi · T. Reis
Fachbereich Mathematik, Universität Hamburg, Bundesstraße 55, 20146 Hamburg, Germany
E-mail: arash.massoudi@uni-hamburg.de E-mail: timo.reis@uni-hamburg.de

Mark R. Opmeer
Department of Mathematical Sciences, University of Bath, Claverton Down, Bath BA2 7AY, United Kingdom
E-mail: m.opmeer@maths.bath.ac.uk

where $A \in \mathbb{C}^{n\times n}$ is stable (i.e. all its eigenvalues are in the open left half-plane), $B \in \mathbb{C}^{n\times m}$, $C \in \mathbb{C}^{p\times n}$ and $D \in \mathbb{C}^{p\times m}$ are given; the unknowns in this equation are the Hermitian matrix $X \in \mathbb{C}^{n\times n}$ and the further matrices $K \in \mathbb{C}^{q\times n}$, $J \in \mathbb{C}^{q\times m}$ with $q \leq m$. We will call X a *solution of* (1), if there exist $q \in \mathbb{N}_0$ and $K \in \mathbb{C}^{q\times n}$, $J \in \mathbb{C}^{q\times m}$ such that (1) holds true. A solution X is called *minimal*, if $X \leq Y$ (i.e., $Y - X$ is positive semi-definite) for all other solutions Y of (1). Note that if $D^*D - I$ is invertible, then J and K can be eliminated and (1) becomes equivalent to the algebraic Riccati equation

$$A^*X + XA + C^*C + (XB + C^*D)(I - D^*D)^{-1}(B^*X + D^*C) = 0.$$

An important application of the bounded real Lur'e equations is *bounded real balanced truncation* [11,12], a model reduction method which preserves contractivity of a system. In particular in this application there is a need for an efficient numerical method for the large-scale case (i.e., n is large). This large-scale case arises for example when considering discretizations of partial differential equations (see Section 5 for a typical example). In the large scale case it is unfeasible to even store the dense matrix $X \in \mathbb{C}^{n\times n}$. Our algorithm provides a sequence (X_k) of approximate solutions of the form $X_k = R_k^* R_k$ for some $R_k \in \mathbb{C}^{\ell_k \times n}$ with, typically, $\ell_k \ll n$ (i.e., X_k is given in "low-rank factored form"). For a "shift parameter sequence" $(\alpha_j)_{j=1}^k$ with $\alpha_j \in \mathbb{C}$ with $\mathrm{Re}(\alpha_j) > 0$, the main computational cost in the algorithm consists of, for each α_j ($j = 1,\ldots,k$), solving a linear system of the form $(\alpha_j - A)x = v$, where $v \in \mathbb{C}^{n\times p}$. The above features make the proposed algorithm attractive for the case where n is large, p is small and A is sparse. This situation is typical when considering discretizations of partial differential equations.

The proposed algorithm is an extension of the recently developed *ADI method* for algebraic Riccati equations of the type $A^*X + XA + C^*C - XBB^*X = 0$ [8,10], which in turn is an extension of the ADI method for Lyapunov equations [7,9,21].

For the convergence analysis of the algorithm, we use the following connection between the minimal solution of the bounded real Lur'e equation and an optimal control problem. It is well-known that the quadratic form defined by the minimal solution of the bounded real Lur'e equation (1) expresses the *available storage* [26]. Namely, for all $x_0 \in \mathbb{C}^n$ there holds

$$x_0^* X x_0 = \sup_{u\in L^2(0,\infty;\mathbb{C}^m)} \int_0^\infty (\|y(t)\|^2 - \|u(t)\|^2)\,\mathrm{d}t, \tag{2}$$

where

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t), \qquad x(0) = x_0, \\ y(t) &= Cx(t) + Du(t), \end{aligned} \tag{3}$$

see [24–26]. Thereby we follow the ideas in [10], which gives an interpretation of the ADI method for the algebraic Riccati equation [8] in terms of the underlying optimal control problem: The theoretical foundation for our algorithm is a sequence of subspaces

$$\mathscr{K}_k(\alpha) := \mathrm{span}\{\mathrm{e}^{-\alpha_1 t}, \ldots, \mathrm{e}^{-\alpha_k t}\} \subset L^2(0,\infty). \tag{4}$$

In this introduction we assume for notational simplicity that the "shift parameters" α_j are distinct (in the main part of the article we drop this assumption; the definition of $\mathscr{K}_k(\alpha)$ has to be modified in case of non-distinct parameters). Let $P_{k,p}$:

$L^2(0,\infty;\mathbb{C}^p) \to L^2(0,\infty;\mathbb{C}^p)$ denote the orthogonal projection onto $\mathscr{K}_k(\alpha)\otimes\mathbb{C}^p$. The matrix X_k produced by our algorithm is proven to represent the optimal cost for the following control problem (see Theorem 3)

$$x_0^* X_k x_0 = \sup_{u\in L^2(0,\infty;\mathbb{C}^m)} \int_0^\infty (\|(P_{k,p}y)(t)\|^2 - \|u(t)\|^2)\,\mathrm{d}t, \tag{5}$$

subject to (3). Since the spaces $\mathscr{K}_k(\alpha)$ are nested, this representation shows that the sequence (X_k) is monotonically increasing with respect to monotonicity, that is $X_k \geq X_{k-1}$ for all $k \in \mathbb{N}$. In the case where

$$\overline{\bigcup_{k\in\mathbb{N}} \mathscr{K}_k(\alpha)} = L^2(0,\infty), \tag{6}$$

we immediately see that we will have convergence of (X_k) to X. The property (6) is proven in [14] to be equivalent to the *non-Blaschke condition*

$$\sum_{j=1}^{\infty} \frac{\mathrm{Re}(\alpha_j)}{1+|\alpha_j|^2} = \infty. \tag{7}$$

We note that (7) is for example satisfied if the parameters all belong to a fixed compact set contained in the open right half-plane (in particular, if the shift parameters are periodic).

We further consider the ADI method for positive real Lur'e equation

$$\begin{aligned} A^*X + XA &= -K^*K, \\ B^*X - C &= -J^*K, \\ -(D^*+D) &= -J^*J, \end{aligned} \tag{8}$$

where $A \in \mathbb{C}^{n\times n}$ is stable, and $B \in \mathbb{C}^{n\times m}$, $C \in \mathbb{C}^{m\times n}$, $D \in \mathbb{C}^{m\times m}$. Our considerations are based on the fact that the minimal solution expresses the available storage for passivity, that is

$$x_0^* X x_0 = \sup_{u\in L^2(0,\infty;\mathbb{C}^m)} -2\mathrm{Re}\int_0^\infty y(t)^* u(t)\,\mathrm{d}t \tag{9}$$

subject to (3).

At this point, we briefly summarize existing approaches to the solution of bounded real and positive real Lur'e equations. If $I - D^*D$ (resp. $D + D^*$) is invertible, then, of course, the huge variety of existing methods for algebraic Riccati equations (see [2] for an overview) can be used. In the case where this matrix is however singular, there are only few methods available: The *structured doubling algorithm* was recently developed for Lur'e equations [16]. In contrast to our method, the structured doubling algorithm does not provide factorizations of low rank form and is therefore memory consuming in the large-scale case. Another approach to numerical solution was presented in [15], where some "critical part" of the Lur'e equation is extracted such that an algebraic Riccati equation is obtained. The latter is then solved by Newton-Kleinman iteration [2]. This method can be formulated such that approximate low rank factors are obtained. A drawback of this approach is that the extraction of the

critical part consists of successive nullspace computations which may be numerically unstable. We will give a slightly more detailed discussion of the latter method in Remark 9 b).

This article is organized as follows. In the forthcoming Section 2, we introduce the systems theoretic and functional analytic on fundamentals of optimal control and their relations to the minimal solutions of positive real and bounded real Lur'e equations. Section 3 is devoted to the spaces $\mathscr{K}_k(\alpha)$ from (4). We consider an orthonormal basis for these spaces (the Takenaka–Malmquist system). We further consider orthogonal projections of the solution maps of the system (3) to the space $\mathscr{K}_k(\alpha)$ from (4), and we provide matrix representations of these maps with respect to this basis. In Section 4 we apply these findings to the optimal control problem by showing that the matrix representations from Section 3 can be used to determine the solution X_k in (5). This allows to formulate iterative algorithms for the determination of the minimal solutions of the Lur'e equations (1) and (8). We also prove convergence of the algorithm. In Section 5 we consider a numerical example.

At this point, we would like to declare some notation: $L^2(0,\infty;\mathbb{C}^p)$ denotes the Lebesgue space of square integrable $\mathbb{C}^p$-valued functions, which is provided with the standard inner product $\langle f,g\rangle_{L^2} := \int_0^\infty g^*(\tau)f(\tau)\,\mathrm{d}\tau$. In this article, we use the Euclidean inner product in $\mathbb{C}^n$, i.e., $\langle x,y\rangle_{\mathbb{C}^n} := y^*x$. The norm in the inner product space X is $\|x\|_X := \langle x,x\rangle_X^{1/2}$. $X\otimes Y$ denotes the tensor product of the inner product spaces X and Y. We use the inner product in $X\otimes Y$ as introduced in [22, Sec. 4.5]. The tensor product of linear operators A_1 and A_2 is denoted by $A_1\otimes A_2$. We identify $L^2(0,\infty;\mathbb{C}^p) = L^2(0,\infty;\mathbb{C})\otimes\mathbb{C}^p$. A^* is the adjoint of a linear operator A. The identity matrix of size $p\times p$ is denoted by I_p. We omit the subscripts in norms, inner products and identity matrices, if it is clear from context.

2 Linear systems and optimal control

We present the connection between the minimal solutions of the Lur'e equations (1) and (8) to the optimization problems (2) and (9) respectively. We follow the approach in [3] by giving an explicit formula of the minimal solution of the Lur'e equation in terms of operators associated with the linear system (3). This will be the theoretical basis for our algorithms, which are based on discretizations of these operators.

Definition 1 (Output map, input-output map) Assume that $A\in\mathbb{C}^{n\times n}$ is stable, $B\in\mathbb{C}^{n\times m}$, $C\in\mathbb{C}^{p\times n}$ and $D\in\mathbb{C}^{p\times m}$. Consider the following maps associated to the system (3):

a) the *output map* $\Psi:\mathbb{C}^n\to L^2(0,\infty;\mathbb{C}^p)$ which maps the initial state x_0 to the output y (for control $u=0$),
$$\Psi x_0 = t\mapsto Ce^{At}x_0; \tag{10}$$

b) the *input-output map* $\mathbb{F}:L^2(0,\infty;\mathbb{C}^m)\to L^2(0,\infty;\mathbb{C}^p)$ which maps the input u to the output y (for initial condition $x_0=0$);
$$\mathbb{F}u = t\mapsto \int_0^t Ce^{A(t-\tau)}Bu(\tau)\,\mathrm{d}\tau + Du(t). \tag{11}$$

With the above introduced operators, the supremized expression in (2) is $\|\Psi x_0 + \mathbb{F}u\|_{L^2}^2 - \|u\|_{L^2}^2$; the supremized expression in (9) becomes $-2\mathrm{Re}\langle u, \Psi x_0 + \mathbb{F}u\rangle_{L^2}$.

Now we study solvability and solutions of general Lur'e equations of the form

$$\begin{aligned} A^*X + XA - C^*QC &= -K^*K, \\ B^*X - (D^*QC + S^*C) &= -J^*K, \\ -(D^*QD + S^*D + D^*S + R) &= -J^*J, \end{aligned} \tag{12}$$

where $A \in \mathbb{C}^{n\times n}$ is stable, $B \in \mathbb{C}^{n\times m}$, $C \in \mathbb{C}^{p\times n}$, $D \in \mathbb{C}^{p\times m}$ and $Q \in \mathbb{C}^{p\times p}$, $S \in \mathbb{C}^{p\times m}$, $R \in \mathbb{C}^{m\times m}$ with $R = R^*$ and $Q = Q^*$. Note that we obtain the bounded real Lur'e equation by setting $Q = -I$, $S = 0$ and $R = I$; the positive real Lur'e equation is given by (12) with $p = m$, $Q = R = 0$ and $S = I$.

The following concepts are crucial for the existence of solutions and their relation to optimization problems.

Definition 2 (Popov function, Popov operator) Assume that $A \in \mathbb{C}^{n\times n}$ is stable, $B \in \mathbb{C}^{n\times m}$, $C \in \mathbb{C}^{p\times n}$, $D \in \mathbb{C}^{p\times m}$ and $Q \in \mathbb{C}^{p\times p}$, $S \in \mathbb{C}^{p\times m}$, $R \in \mathbb{C}^{m\times m}$ with $R = R^*$ and $Q = Q^*$. Then, for $G(s) = C(sI - A)^{-1}B + D$, the *Popov function* $\Pi : \imath\mathbb{R} \to \mathbb{C}^{m\times m}$ is defined by

$$\Pi(\imath\omega) := G(\imath\omega)^* Q G(\imath\omega) + G(\imath\omega)^* S + S^* G(\imath\omega) + R.$$

With $\mathbb{F}$ as defined in Definition 1, the *Popov operator* $\mathscr{R} : L^2(0,\infty;\mathbb{C}^m) \to L^2(0,\infty;\mathbb{C}^m)$ is

$$\mathscr{R} := \mathbb{F}^* Q \mathbb{F} + \mathbb{F}S + S^*\mathbb{F} + R. \tag{13}$$

Next, we give some comments on solvability of Lur'e equations and their specialization to the bounded real and positive real case.

Remark 1 (Popov operator, Popov function, solvability of Lur'e equations)

a) The Popov operator $\mathscr{R}$ is positive semidefinite, if, and only if, the Popov function fulfills $\Pi(\imath\omega) \geq 0$ for all $\omega \in \mathbb{R}$ [3]. If the Lur'e equation (12) is solvable, then the Popov function fulfills $\Pi(\imath\omega) \geq 0$ for all $\omega \in \mathbb{R}$ [17] (and thus $\mathscr{R}$ is positive semidefinite).

b) If the Popov function fulfills $\Pi(\imath\omega) \geq 0$ for all $\omega \in \mathbb{R}$ and the system (3) is controllable, then there exists a minimal solution of the Lur'e equation (12). This follows from the results in [17] and the substitutions

$$\begin{aligned} X &\rightsquigarrow -X, & C^*QC &\rightsquigarrow Q, \\ C^*QD + C^*S &\rightsquigarrow C, & D^*QD + S^*D + D^*S + R &\rightsquigarrow R, \end{aligned} \tag{14}$$

"minimal solution" $\rightsquigarrow$ "maximal solution".

c) In the bounded real case, the Popov operator reads $I - \mathbb{F}^*\mathbb{F}$. Solvability of the bounded real Lur'e equation (1) therefore implies $\|\mathbb{F}\| \leq 1$. This property is called *contractivity* and is equivalent to the $\mathscr{H}_\infty$ norm of $G(s)$ being not larger that one [27, Sec. 4.5].

d) In the positive real case, the Popov operator is given by $\mathscr{R} = \mathbb{F}^* + \mathbb{F}$, whose positive semidefiniteness is called *passivity*. The Popov function reads $\imath\omega \mapsto G^*(\imath\omega) + G(\imath\omega)$. Passivity is equivalent to *positive realness* of $G(s)$. That is, $G(\lambda) + G(\lambda)^* \geq 0$ for all $\lambda \in \mathbb{C}$ with $\mathrm{Re}(\lambda) > 0$ [26].

Now we present the relation between the minimal solutions and optimization problems subject to the linear system (3).

Theorem 1 *Assume that $A \in \mathbb{C}^{n\times n}$ is stable, $B \in \mathbb{C}^{n\times m}$, $C \in \mathbb{C}^{p\times n}$, $D \in \mathbb{C}^{p\times m}$ and $Q \in \mathbb{C}^{p\times p}$, $S \in \mathbb{C}^{p\times m}$, $R \in \mathbb{C}^{m\times m}$ with $R = R^*$ and $Q = Q^*$. Let $\mathbb{F}$ be the input-output map and Ψ be the output map of the system* (3)*. Assume that X is the minimal solution of the Lur'e equations* (12) *and let $K \in \mathbb{C}^{q\times n}$, $J \in \mathbb{C}^{q\times m}$ be such that* (12) *holds true. Then the following hold true:*

a) The system

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t), \quad x(0) = x_0, \\ y_\Xi(t) &= Kx(t) + Ju(t), \end{aligned} \tag{15}$$

with output map $\Psi_\Xi : \mathbb{C}^n \to L^2(0,\infty;\mathbb{C}^q)$ and input-output map $\mathbb{F}_\Xi : L^2(0,\infty;\mathbb{C}^m) \to L^2(0,\infty;\mathbb{C}^q)$ is outer. *That is, $\mathbb{F}_\Xi$ has dense range.*

b) For all $u \in L^2(0,\infty;\mathbb{C}^m)$ and $x_0 \in \mathbb{C}^n$ holds

$$-\left\langle \begin{bmatrix} \mathbb{F}u + \Psi x_0 \\ u \end{bmatrix}, \begin{bmatrix} Q & S \\ S^* & R \end{bmatrix} \begin{bmatrix} \mathbb{F}u + \Psi x_0 \\ u \end{bmatrix} \right\rangle_{L^2} = x_0^* X x_0 - \|\mathbb{F}_\Xi u + \Psi_\Xi x_0\|_{L^2}^2. \tag{16}$$

c) The operator $\mathbb{F}_\Xi$ and the Popov operator (13) *are related by*

$$\mathscr{R} = \mathbb{F}_\Xi^* \mathbb{F}_\Xi. \tag{17}$$

d) The minimal solution fulfills

$$X = \Psi_\Xi^* \Psi_\Xi - \Psi^* Q \Psi. \tag{18}$$

e) The operators $\mathbb{F}_\Xi$, Ψ_Ξ, the output map Ψ, and the input-output map $\mathbb{F}$ of the system (3) *are related by*

$$\mathbb{F}_\Xi^* \Psi_\Xi = (\mathbb{F}^* Q + S^*)\Psi. \tag{19}$$

Proof

a) Let X be the minimal solution. Then, by using the substitutions in (14), it has been shown in [17, Sec. 5] that

$$\mathrm{im} \begin{bmatrix} -\lambda I + A & B \\ K & J \end{bmatrix} = \mathbb{C}^{n+q} \quad \forall \lambda \in \mathbb{C} \text{ with } \mathrm{Re}(\lambda) > 0.$$

Then it follows by a combination of [6, Thm. 3.3 & Thm. 5.1] that (15) is outer.

b) Using [26], we see that for all $t \geq 0$ the solutions of (3) fulfill the *dissipation inequality*

$$x_0^* X x_0 - x(t)^* X x(t) = -\int_0^t \begin{pmatrix} y(\tau) \\ u(\tau) \end{pmatrix}^* \begin{bmatrix} Q & S \\ S^* & R \end{bmatrix} \begin{pmatrix} y(\tau) \\ u(\tau) \end{pmatrix} \mathrm{d}\tau + \int_0^t \|Kx(\tau) + Ju(\tau)\|^2 \mathrm{d}\tau.$$

Using that $u \in L^2(0,\infty;\mathbb{C}^m)$ and A is stable, we obtain that the state trajectory of the system (15) fulfills $\lim_{t\to\infty} x(t) = 0$. Then the result follows by taking the limit $t \to \infty$.

c)-e) Defining the inner product in $\mathbb{C}^m \times L^2$ by the sum of inner products in $\mathbb{C}^m$ and L^2, (16) can be rewritten as

$$\left\langle \begin{pmatrix} x_0 \\ u \end{pmatrix}, \begin{bmatrix} \Psi^* Q \Psi & \Psi^*(Q\mathbb{F}+S) \\ (\mathbb{F}^*Q+S^*)\Psi & \mathscr{R} \end{bmatrix} \begin{pmatrix} x_0 \\ u \end{pmatrix} \right\rangle_{\mathbb{C}^m \times L^2}$$
$$= \left\langle \begin{pmatrix} x_0 \\ u \end{pmatrix}, \begin{bmatrix} \Psi_\Xi^* \Psi_\Xi - X & \Psi_\Xi^* \mathbb{F}_\Xi \\ \mathbb{F}_\Xi^* \Psi_\Xi & \mathbb{F}_\Xi^* \mathbb{F}_\Xi \end{bmatrix} \begin{pmatrix} x_0 \\ u \end{pmatrix} \right\rangle_{\mathbb{C}^m \times L^2} \quad \forall u \in L^2(0,\infty;\mathbb{C}^m), x_0 \in \mathbb{C}^n.$$

This in turn leads to $\mathscr{R} = \mathbb{F}_\Xi^* \mathbb{F}_\Xi$, $X = \Psi_\Xi^* \Psi_\Xi - \Psi^* Q \Psi$ and $\mathbb{F}_\Xi^* \Psi_\Xi = (\mathbb{F}^*Q+S^*)\Psi$. □

Remark 2 (Lur'e equations)

a) Equation (17) is called *spectral factorization* [3, 27].
b) The minimal solution of the bounded real Lur'e equation reads $X = \Psi_\Xi^* \Psi_\Xi + \Psi^* \Psi$. In the positive real case, we have $X = \Psi_\Xi^* \Psi_\Xi$. In both cases, X is positive semidefinite.
c) The property of $\mathbb{F}_\Xi$ being outer implies that for all $x_0 \in \mathbb{C}^n$, $\varepsilon > 0$, there exists some $u \in L^2(0,\infty;\mathbb{C}^m)$ with $\|\mathbb{F}_\Xi u + \Psi_\Xi x_0\|^2 < \varepsilon$. As a consequence, we have, from Theorem 1 e), that for all $x_0 \in \mathbb{C}^n$

$$x_0^* X x_0 = \sup_{u \in L^2(0,\infty;\mathbb{C}^m)} -\left\langle \begin{bmatrix} \mathbb{F}u + \Psi x_0 \\ u \end{bmatrix}, \begin{bmatrix} Q & S \\ S^* & R \end{bmatrix} \begin{bmatrix} \mathbb{F}u + \Psi x_0 \\ u \end{bmatrix} \right\rangle_{L^2}. \tag{20}$$

It follows from Theorem 1 b) that the supremum in the right hand side of the above expression is attained at $u \in L^2(0,\infty;\mathbb{C}^m)$ (i.e. u is an optimal control) if, and only if, $\mathbb{F}_\Xi u + \Psi_\Xi x_0 = 0$. Using Theorem 1 a), this means that there exists some $x : [0,\infty) \to \mathbb{C}^n$ such that the differential-algebraic equation

$$\begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{x}(t) \\ \dot{u}(t) \end{bmatrix} = \begin{bmatrix} A & B \\ K & L \end{bmatrix} \begin{bmatrix} x(t) \\ u(t) \end{bmatrix}, \quad x(0) = x_0 \tag{21}$$

is fulfilled. Then it follows by a transformation of the matrix pencil $\left[\begin{smallmatrix} sI-A & -B \\ -K & -L \end{smallmatrix}\right]$ into Kronecker form [4, Chap. XII, §7] that x and u can be expressed by sums of exponential functions of type $\sum_{k=1}^{\ell} p_k(t) \mathrm{e}^{-\lambda_k t}$, where $p_1, \ldots, p_\ell$ are vector-valued complex polynomials, and the distinct numbers $\lambda_1, \ldots, \lambda_\ell$ are the generalized eigenvalues of the pencil $\left[\begin{smallmatrix} sI-A & -B \\ -K & -L \end{smallmatrix}\right]$. By using the substitutions in (14), the latter are shown in [17] to be the negatives of the stable generalized eigenvalues of the *even matrix pencil*

$$s\mathscr{E} - \mathscr{A} = \begin{bmatrix} 0 & -sI+A & B \\ sI+A^* & -C^*QC & -C^*QD-C^*S \\ B^* & -D^*QC-S^*C & -D^*QD-S^*D-D^*S-R \end{bmatrix}. \tag{22}$$

We will make use of this fact in Section 5 to improve numerical performance by suitable choice of the shift parameters.

3 Convolution systems and matrix representations

In this section we review results from [10] which give matrix representations of the adjoints of the output map Ψ and the input-output map $\mathbb{F}$ with respect to a certain orthonormal basis of L^2. These matrix representations will be crucial in our algorithms.

Definition 3 Let $(\alpha_j)_{j=1}^{\infty}$ be a complex sequence with $\mathrm{Re}(\alpha_j) > 0$ for all $j \in \mathbb{N}$. We define the corresponding *Takenaka–Malmquist system* $(\psi_j)_{j=1}^{\infty}$, $\psi_j \in L^2(0,\infty)$ by

$$\begin{aligned} \phi_1 &= t \mapsto \mathrm{e}^{-\alpha_1 t}, & \psi_1 &= \sqrt{2\mathrm{Re}(\alpha_1)} \cdot \phi_1, \\ \phi_j &= \phi_{j-1} - (\alpha_j + \overline{\alpha_{j-1}}) \cdot (\mathrm{e}^{-\alpha_j \cdot} * \phi_{j-1}), & \psi_j &= \sqrt{2\mathrm{Re}(\alpha_j)} \cdot \phi_j, \quad \text{for } j \geq 2, \end{aligned} \tag{23}$$

where $*$ denotes the convolution product, i.e., $(g*h)(t) = \int_0^t g(t-\tau)h(\tau)\,\mathrm{d}\tau$.
The space generated by the first k Takenaka–Malmquist functions is denoted by $\mathscr{K}_k(\alpha)$.

Remark 3

a) The Takenaka–Malmquist system is orthonormal (see e.g. [14, Appendix B] for a proof).
b) The *convolution system* $(\varphi_j)_{j=1}^{\infty}$, $\varphi_j \in L^2(0,\infty)$, which is defined by

$$\varphi_1 := t \mapsto \mathrm{e}^{-\alpha_1 t}, \qquad \varphi_j := \mathrm{e}^{-\alpha_j \cdot} * \varphi_{j-1}, \tag{24}$$

fulfills $\mathrm{span}\{\varphi_1, \ldots, \varphi_k\} = \mathscr{K}_k(\alpha)$.
c) Consider the distinct numbers $q_1, \ldots, q_J$ with $\{q_1, \ldots, q_J\} = \{\alpha_1, \ldots, \alpha_k\}$. Let ℓ_j be the number of indices in which q_j appears in $(\alpha_j)_{j=1}^{k}$ (thus $k = \ell_1 + \ldots + \ell_J$). Then

$$\mathrm{span}\{\varphi_1, \ldots, \varphi_k\} = \bigoplus_{j=1}^{J} \mathrm{span}\left\{ t \mapsto t^l \mathrm{e}^{-q_j t} \;\middle|\; l = 0, \ldots, \ell_j - 1 \right\},$$

see [10, 14].

The most important property of the above introduced space is that it is $\mathbb{F}^*$-invariant.

Theorem 2 *Let $A \in \mathbb{C}^{n\times n}$ stable and $B \in \mathbb{C}^{n\times m}$, $C \in \mathbb{C}^{p\times n}$, $D \in \mathbb{C}^{p\times m}$. For $\mathbb{F}$ as in* (11) *and $\mathscr{K}_k(\alpha)$ the sequence of subspaces from Definition 3, we have that*

$$\mathbb{F}^* \left(\mathscr{K}_k(\alpha) \otimes \mathbb{C}^p\right) \subset \mathscr{K}_k(\alpha) \otimes \mathbb{C}^m.$$

Proof The proof is contained in [10] for the case $D = 0$. The general result follows by regarding D as a pointwise multiplication operator $D : L^2(0,\infty;\mathbb{C}^m) \to L^2(0,\infty;\mathbb{C}^p)$. The latter obviously fulfills

$$D^* \left(\mathscr{K}_k(\alpha) \otimes \mathbb{C}^p\right) \subset \mathscr{K}_k(\alpha) \otimes \mathbb{C}^m. \qquad \square$$

The above invariance gives rise to the existence of matrix representations of $\mathbb{F}^*$ with respect to the Takenaka–Malmquist systems. These will be explicitly constructed in the following.

Definition 4 Let $(\alpha_j)_{j=1}^{\infty}$ be such that $\mathrm{Re}(\alpha_j) > 0$ for all $j \in \mathbb{N}$. Let $(\psi_j)_{j=1}^{\infty}$, $\psi_j \in L^2(0,\infty)$ be the corresponding Takenaka–Malmquist system (23). For $k \in \mathbb{N}$, the mapping $\iota_k : \mathbb{C}^k \to L^2(0,\infty)$ is defined by

$$\iota_k x = \sum_{j=1}^{k} x_j \cdot \psi_j. \tag{25}$$

Further, for the identity matrix $I_p \in \mathbb{C}^{p\times p}$, we set $\iota_{k,p} := \iota_k \otimes I_p : \mathbb{C}^{kp} \to L^2(0,\infty;\mathbb{C}^p)$.

Orthonormality of the Takenaka–Malmquist system implies that ι_k (and thus also $\iota_{k,p}$) defines an isometric embedding. The orthogonal projector onto $\mathscr{K}_k(\alpha) \otimes \mathbb{C}^p$ is therefore given by

$$P_{k,p} = \iota_{k,p}\iota_{k,p}^* : L^2(0,\infty;\mathbb{C}^p) \to L^2(0,\infty;\mathbb{C}^p). \tag{26}$$

With operators Ψ and $\mathbb{F}$ as in (10) and (11), we define the matrices

$$F_k = \iota_{k,p}^* \mathbb{F} \iota_{k,m} \in \mathbb{C}^{kp\times km}, \quad S_k = \iota_{k,p}^* \Psi \in \mathbb{C}^{kp\times n}. \tag{27}$$

We have

$$P_{k,p}\Psi = \iota_{k,p} S_k, \qquad P_{k,p}\mathbb{F} = P_{k,p}\mathbb{F}P_{k,m} = \iota_{k,p} F_k \iota_{k,m}^*, \tag{28}$$

where the equality $P_{k,p}\mathbb{F} = P_{k,p}\mathbb{F}P_{k,m}$ follows by taking adjoints in $\mathbb{F}^* P_{k,p} = P_{k,m}\mathbb{F}^* P_{k,p}$ and the latter equality follows from Theorem 2.

Alg. 1 from [10] provides a recursive method to compute S_k and F_k. The determination of S_k is based on the fact that the unnormalized Takenaka–Malmquist system system $(\phi_j)_{j=1}^{\infty}$ (23) fulfills

$$\begin{aligned}
&\Psi^*(\phi_1 v) = (\alpha_1 I - A^*)^{-1} C^* v, \\
&\Psi^*(\phi_j v) = \Psi^*(\phi_{j-1} v) - (\alpha_j + \overline{\alpha_{j-1}})(\alpha_j I - A^*)^{-1}\Psi^*(\phi_{j-1} v) \qquad \forall v \in \mathbb{C}^p,
\end{aligned}$$

see [10, Corollary 13]. The determination of F_k relies on the following consideration: Let $\Lambda : L^2(0,\infty;\mathbb{C}^n) \to L^2(0,\infty;\mathbb{C}^p)$ be the input-output map of the system (3) with $B = I$ and $D = 0$. Then $\mathbb{F} = \Lambda B + D$, where $B \in \mathbb{C}^{n\times m}$ and $D \in \mathbb{C}^{p\times m}$ are regarded as constant multiplication operators on $L^2(0,\infty;\mathbb{C}^m)$. Then Λ^* satisfies the recursion (here $(\varphi_j)_{j=1}^{\infty}$ is the convolution system from (24))

$$\begin{aligned}
&\Lambda^*(\varphi_1 v) = (\alpha_1 I - A^*)^{-1} C^* v \varphi_1, \\
&\Lambda^*(\varphi_j v) = (\alpha_j I - A^*)^{-1} C^* v \varphi_j + (\alpha_j I - A^*)^{-1}\Lambda^*(\varphi_{j-1} v) \qquad \forall v \in \mathbb{C}^p,
\end{aligned}$$

see [10, Corollary 14]. A transition from the basis $(\varphi_1,\dots,\varphi_k)$ to the basis $(\psi_1,\dots,\psi_k)$ then gives rise to the construction of F_k. The precise construction is given in Alg. 1 (we refer to [10] for further details).

Algorithm 1 ADI iteration for output and input-output maps.

Input: $A \in \mathbb{C}^{n\times n}$ a stable matrix, $B \in \mathbb{C}^{n\times m}$, $C \in \mathbb{C}^{p\times n}$, $D \in \mathbb{C}^{p\times m}$ and shift parameters $\alpha_1,\dots,\alpha_k \in \mathbb{C}$ with $\mathrm{Re}(\alpha_i) > 0$.
Output: $S_k = \iota_{k,p}^* \Psi \in \mathbb{C}^{kp\times n}$, $F_k = \iota_{k,p}^* \mathbb{F} \iota_{k,m} \in \mathbb{C}^{kp\times km}$

$V_1 = (\alpha_1 I - A^*)^{-1} C^*$
$S_1 = \sqrt{2\mathrm{Re}(\alpha_1)} \cdot V_1^*$
$Q_1 = \sqrt{2\mathrm{Re}(\alpha_1)} \cdot V_1^* B$
$L_1 = \frac{1}{\sqrt{2\mathrm{Re}(\alpha_1)}}$
$F_1 = Q_1 L_1 + D$
for $i = 2,3,\dots,k$ **do**
  $V_i = V_{i-1} - (\alpha_i + \overline{\alpha_{i-1}}) \cdot (\alpha_i I - A^*)^{-1} V_{i-1}$
  $S_i = [S_{i-1}^*,\ \sqrt{2\mathrm{Re}(\alpha_i)} \cdot V_i]^*$
  $Q_i = [Q_{i-1},\ \sqrt{2\mathrm{Re}(\alpha_i)} \cdot V_i^* B]$
  $\gamma_i = \sqrt{\frac{\mathrm{Re}(\alpha_j)}{\mathrm{Re}(\alpha_{j-1})}}$
  $M_{i,1} = \begin{bmatrix} \frac{1}{\sqrt{2\mathrm{Re}(\alpha_1)}} & & \\ & \ddots & \\ & & \frac{1}{\sqrt{2\mathrm{Re}(\alpha_i)}} \end{bmatrix}$, $M_{i,2} = \begin{bmatrix} \overline{\alpha_1} + \alpha_i & & \\ \alpha_1 - \alpha_i & \overline{\alpha_2} + \alpha_i & \\ & \ddots & \\ & \alpha_{i-1} - \alpha_i & \overline{\alpha_i} + \alpha_i \end{bmatrix}$,
  $M_{i,3} = \begin{bmatrix} 1 & \dots & 1 \\ & \ddots & \vdots \\ & & 1 \end{bmatrix}$, $M_{i,4} = \begin{bmatrix} 0 & I \\ 1 & 0 \end{bmatrix}$, $M_{i,5} = \begin{bmatrix} -\sqrt{2\mathrm{Re}(\alpha_1)} & & \\ & \ddots & \\ & & -\sqrt{2\mathrm{Re}(\alpha_{i-1})} \\ & & & 1 \end{bmatrix}$
  $M_i = M_{i,1}^{-1} M_{i,2}^{-1} M_{i,3}^{-1} M_{i,4}^{-1} M_{i,5}^{-1}$
  $L_i = \begin{bmatrix} \gamma_i L_{i-1} & 0 \\ 0 & 0 \end{bmatrix} - M_i \begin{bmatrix} L_{i-1} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \gamma_i(\alpha_i + \overline{\alpha_{i-1}}) I & 0 \\ [0,\gamma_i] & -1 \end{bmatrix}$
  $F_i = \begin{bmatrix} [F_{i-1}, 0] \\ Q_i(\overline{L_i} \otimes I_m) + [0, D] \end{bmatrix}$
end for

4 The projected optimal control problem

In this section we consider the optimal control problems (2) & (3) and (9) & (3), and their relations to the corresponding optimal control problems in which the output y is replaced by $P_{k,p}y$ with the orthogonal projector $P_{k,p}$ as in (26) onto the space $\mathscr{K}_k(\alpha) \otimes \mathbb{C}^p$ generated by the truncated Takenaka–Malmquist system. Thereby we present "discretized versions" of Theorem 1. That is, the input-output map $\mathbb{F}$ and the output map Ψ in (10) and (11) are replaced with F_k and S_k in (27), i.e., the representing matrix of their orthogonal projection onto $\mathscr{K}_k(\alpha) \otimes \mathbb{C}^p$. Then the relations (17) and (19) become matrix equations which have to be solved for "discretized versions" of $\mathbb{F}_\Xi$ and S_Ξ. These are thereafter, by an accordant modification of (18), used to construct X_k. We start with the bounded real case. Towards a better understanding, the reader should compare the following result with Theorem 1 specialized to $Q = -I_p$, $R = I_m$ and $S = 0$.

Theorem 3 *Assume that $A \in \mathbb{C}^{n\times n}$ is stable, $B \in \mathbb{C}^{n\times m}$, $C \in \mathbb{C}^{p\times n}$ and $D \in \mathbb{C}^{p\times m}$. Further assume that bounded real Lur'e equation* (1) *has a minimal solution $X \in$*

$\mathbb{C}^{n\times n}$. *Define* Ψ *and* $\mathbb{F}$ *by* (10) *and* (11). *Let* $(\alpha_j)_{j=1}^{\infty}$ *be a complex sequence with* $\mathrm{Re}(\alpha_j) > 0$ *for all* $j \in \mathbb{N}$, *and let* $F_k \in \mathbb{C}^{kp\times km}$, $S_k \in \mathbb{C}^{kp\times n}$ *be defined as in* (27). *Then the matrix* $I - F_k^* F_k \in \mathbb{C}^{km\times km}$ *is positive semidefinite. In particular, there exists a matrix* $F_{\Xi,k} \in \mathbb{C}^{\ell_k\times km}$ *with full row rank and*

$$I - F_k^* F_k = F_{\Xi,k}^* F_{\Xi,k}. \tag{29}$$

Further, there exists some $S_{\Xi,k} \in \mathbb{C}^{\ell_k\times n}$ *such that*

$$F_{\Xi,k}^* S_{\Xi,k} = -F_k^* S_k. \tag{30}$$

For the orthogonal projector $P_{k,p}$ *as in* (26), *the matrix* X_k *defined by*

$$X_k = S_k^* S_k + S_{\Xi,k}^* S_{\Xi,k}, \tag{31}$$

fulfills

$$x_0^* X_k x_0 = \sup_{u\in L^2(0,\infty;\mathbb{C}^m)} \|P_{k,p}\mathbb{F}u + P_{k,p}\Psi x_0\|^2 - \|u\|^2 \qquad \forall x_0 \in \mathbb{C}^n. \tag{32}$$

Proof Let $X \in \mathbb{C}^{n\times n}$ be a minimal solution of the bounded real Lur'e equation (1). Then Theorem 1 implies that the operator $I - \mathbb{F}^*\mathbb{F}$ is positive semidefinite. Since $P_{k,p} \le I$ we have $\mathbb{F}^* P_{k,p}\mathbb{F} \le \mathbb{F}^*\mathbb{F}$, which implies that $I - \mathbb{F}^* P_{k,p}\mathbb{F} \ge I - \mathbb{F}^*\mathbb{F}$. Hence $I - \mathbb{F}^* P_{k,p}\mathbb{F}$ is as well positive semidefinite. We have

$$I - F_k^* F_k = I - \imath_{k,m}^* \mathbb{F}^* \imath_{k,p} \imath_{k,p}^* \mathbb{F} \imath_{k,m} = \imath_{k,m}^* (I - \mathbb{F}^* P_{k,p}\mathbb{F}) \imath_{k,m} \ge 0,$$

so that $I - F_k^* F_k$ is positive semidefinite. Thus, there exists some $F_{\Xi,k} \in \mathbb{C}^{\ell_k\times km}$ with full row rank and satisfying (29).

We prove that $\mathrm{im}(F_k^* S_k) \subset \mathrm{im}(F_{\Xi,k})$. By taking orthogonal complements, this is equivalent to

$$\ker(F_{\Xi,k}) \subset \ker(S_k^* F_k).$$

Let $x_0 \in \mathbb{C}^n$ and $u \in L^2(0,\infty;\mathbb{C}^m)$. Then, by stability of A, the state $x(t)$ of the system (3) tends to zero, if t tends to infinity. Then (16) together with $R = I_m$ and $Q = -I_p$ yields

$$x_0^* X x_0 \ge \|\mathbb{F}u + \Psi x_0\|^2 - \|u\|^2.$$

By further using (28) and (29), we see that

$$\begin{aligned}
x_0^* X x_0 &\ge \|\mathbb{F}u + \Psi x_0\|^2 - \|u\|^2 \\
&\ge \|P_{k,p}\mathbb{F}u + P_{k,p}\Psi x_0\|^2 - \|P_{k,m}u\|^2 - \|(I - P_{k,m})u\|^2 \\
&= \|\imath_{k,p} F_k \imath_{k,m}^* u + \imath_{k,p} S_k x_0\|^2 - \|\imath_{k,m}^* u\|^2 - \|(I - P_{k,m})u\|^2 \\
&= \|F_k \imath_{k,m}^* u + S_k x_0\|^2 - \|\imath_{k,m}^* u\|^2 - \|(I - P_{k,m})u\|^2 \\
&= \langle \imath_{k,m}^* u, (F_k^* F_k - I)\imath_{k,m}^* u\rangle + 2\mathrm{Re}\langle \imath_{k,m}^* u, F_k^* S_k x_0\rangle + \|S_k x_0\|^2 - \|(I - P_{k,m})u\|^2 \\
&= -\langle \imath_{k,m}^* u, F_{\Xi,k}^* F_{\Xi,k} \imath_{k,m}^* u\rangle + 2\mathrm{Re}\langle \imath_{k,m}^* u, F_k^* S_k x_0\rangle + \|S_k x_0\|^2 - \|(I - P_{k,m})u\|^2 \\
&= -\|F_{\Xi,k}\imath_{k,m}^* u\|^2 + 2\mathrm{Re}\langle \imath_{k,m}^* u, F_k^* S_k x_0\rangle + \|S_k x_0\|^2 - \|(I - P_{k,m})u\|^2.
\end{aligned} \tag{33}$$

Assume that $\ker F_{\Xi,k} \not\subset \ker S_k^* F_k$. Then there exists some $\hat{u} \in \mathbb{C}^{km}$ with $S_k^* F_k \hat{u} \neq 0$ and $F_{\Xi,k}\hat{u} = 0$, and thus we can choose some $x_0 \in \mathbb{C}^n$ such that $x_0^* S_k^* F_k \hat{u} \neq 0$. Then, for $\lambda \in \mathbb{C}$, substituting x_0 and $u := \iota_{k,m}(\lambda \hat{u}) \in L^2(0,\infty;\mathbb{C}^m)$ into (33), we obtain

$$\begin{aligned} x_0^* X x_0 &\geq -\|F_{\Xi,k}\iota_{k,m}^*\iota_{k,m}(\lambda\hat{u})\|^2 + 2\mathrm{Re}\langle \iota_{k,m}^*\iota_{k,m}(\lambda\hat{u}), F_k^* S_k x_0\rangle + \|S_k x_0\|^2 - \|(I-P_{k,m})\iota_{k,m}(\lambda\hat{u})\|^2 \\ &= -\|\lambda F_{\Xi,k}\hat{u}\|^2 + 2\mathrm{Re}(\lambda\langle \hat{u}, F_k^* S_k x_0\rangle) + \|S_k x_0\|^2 \\ &= 2\mathrm{Re}(\lambda\langle \hat{u}, F_k^* S_k x_0\rangle) + \|S_k x_0\|^2. \end{aligned}$$

In particular, by an appropriate choice of $\lambda \in \mathbb{C}$, we can make the expression on the right hand side arbitrarily large, which leads to a contradiction. Hence $\ker(F_{\Xi,k}) \subset \ker(S_k^* F_k)$.

Since $F_{\Xi,k}$ has full row rank, $F_{\Xi,k}F_{\Xi,k}^*$ is invertible and therefore

$$S_{\Xi,k} := (F_{\Xi,k}F_{\Xi,k}^*)^{-1} F_{\Xi,k} F_k S_k \tag{34}$$

is well-defined. We now show that it satisfies (30). Let $x \in \mathbb{C}^n$. By the above established subspace inclusion $\mathrm{im}(F_k^* S_k) \subset \mathrm{im}(F_{\Xi,k}^*)$, there exists a $z \in \mathbb{C}^{km}$ such that $F_k^* S_k x = F_{\Xi,k}^* z$. Then

$$F_{\Xi,k}^* S_{\Xi,k} x = F_{\Xi,k}^*(F_{\Xi,k}F_{\Xi,k}^*)^{-1}F_{\Xi,k}F_k S_k x = F_{\Xi,k}^*(F_{\Xi,k}F_{\Xi,k}^*)^{-1}F_{\Xi,k}F_{\Xi,k}^* z = F_{\Xi,k}^* z = F_k^* S_k x.$$

Since $x \in \mathbb{C}^n$ was arbitrary, this proves that $F_{\Xi,k}^* S_{\Xi,k} = F_k^* S_k$, i.e the above defined $S_{\Xi,k}$ satisfies (30).

It remains to prove that X_k as in (31) fulfills (32). Using (29) and (30), we have for all $x_0 \in \mathbb{C}^n$ and $u \in L^2(0,\infty;\mathbb{C}^m)$ that

$$\begin{aligned} &\|P_{k,p}\mathbb{F}u + P_{k,p}\Psi x_0\|^2 - \|u\|^2 \\ &= -\langle \iota_{k,m}^* u, F_{\Xi,k}^* F_{\Xi,k}\iota_{k,m}^* u\rangle + 2\mathrm{Re}\langle \iota_{k,m}^* u, F_k^* S_k x_0\rangle + \|S_k x_0\|^2 - \|(I-P_{k,m})u\|^2 \\ &= -\langle \iota_{k,m}^* u, F_{\Xi,k}^* F_{\Xi,k}\iota_{k,m}^* u\rangle - 2\mathrm{Re}\langle \iota_{k,m}^* u, F_{\Xi,k}^* S_{\Xi,k} x_0\rangle + \|S_k x_0\|^2 - \|(I-P_{k,m})u\|^2 \\ &= -\|F_{\Xi,k}\iota_{k,m}^* u + S_{\Xi,k}x_0\|^2 + \|S_{\Xi,k}x_0\|^2 + \|S_k x_0\|^2 - \|(I-P_{k,m})u\|^2 \\ &= -\|F_{\Xi,k}\iota_{k,m}^* u + S_{\Xi,k}x_0\|^2 - \|(I-P_{k,m})u\|^2 + x_0^* X_k x_0 \\ &\leq x_0^* X_k x_0. \end{aligned}$$

This gives rise to

$$x_0^* X_k x_0 \geq \sup_{u \in L^2(0,\infty;\mathbb{C}^m)} \|P_{k,p}\mathbb{F}u + P_{k,p}\Psi x_0\|^2 - \|u\|^2.$$

On the other hand, using the surjectivity of $F_{\Xi,k}$, there exists some $\hat{u} \in \mathbb{C}^{km}$ with $F_{\Xi,k}\hat{u} = -S_{\Xi,k}x_0$. Then, for $u = \iota_{k,m}\hat{u}$, we see that equality holds true in the above calculations. This proves (32). □

Remark 4 (Bounded real Lur'e equations and projected optimal control problems) The formula (34) for $S_{\Xi,k}$ shows that X_k equals $S_k^*[I + F_k F_{\Xi,k}^*(F_{\Xi,k}F_{\Xi,k}^*)^{-2}F_{\Xi,k}F_k^*]S_k$. It is easily verified that $F_{\Xi,k}^*(F_{\Xi,k}F_{\Xi,k}^*)^{-2}F_{\Xi,k}$ is the Moore-Penrose pseudo-inverse of $F_{\Xi,k}^* F_{\Xi,k}$. Therefore, $X_k = S_k^*[I + F_k(I - F_k^* F_k)^+ F_k^*]S_k$.

Next we prove that the sequence (X_k) is monotonically increasing with respect to definiteness. We further present a criterion on the shift parameters such that convergence to the minimal solutions is achieved.

Theorem 4 *Assume that $A \in \mathbb{C}^{n\times n}$ is stable, $B \in \mathbb{C}^{n\times m}$, $C \in \mathbb{C}^{p\times n}$ and $D \in \mathbb{C}^{p\times m}$. Further assume that the bounded real Lur'e equation* (1) *has a minimal solution $X \in \mathbb{C}^{n\times n}$. Define Ψ and $\mathbb{F}$ by* (10) *and* (11).
Let $(\alpha_j)_{j=1}^{\infty}$ be a complex sequence with $\mathrm{Re}(\alpha_j) > 0$ for all $j \in \mathbb{N}$, and let $F_k \in \mathbb{C}^{kp\times km}$, $S_k \in \mathbb{C}^{kp\times n}$ be defined as in (27)*; let X_k be defined as in Theorem 3.*
Then

$$X_k \le X_{k+1}, \quad X_k \le X \qquad \forall k \in \mathbb{N},$$

and the sequence (X_k) converges. If, additionally, $(\alpha_j)_{j=1}^{\infty}$ satisfies the non-Blaschke condition (7)*, then (X_k) converges to X.*

Proof For $x_0 \in \mathbb{C}^n$ and $u \in L^2(0,\infty;\mathbb{C}^m)$ we have

$$\|P_{k,p}\mathbb{F}u + P_{k,p}\Psi x_0\|_{L^2}^2 \le \|P_{k+1,p}\mathbb{F}u + P_{k+1,p}\Psi x_0\|_{L^2}^2,$$

since $\mathscr{K}_k(\alpha) \subset \mathscr{K}_{k+1}(\alpha)$. It follows that

$$\begin{aligned} x_0^* X_k x_0 &= \sup_{u\in L^2(0,\infty;\mathbb{C}^m)} \|P_{k,p}\mathbb{F}u + P_{k,p}\Psi x_0\|^2 - \|u\|^2 \\ &\le \sup_{u\in L^2(0,\infty;\mathbb{C}^m)} \|P_{k+1,p}\mathbb{F}u + P_{k+1,p}\Psi x_0\|^2 - \|u\|^2 = x_0^* X_{k+1} x_0. \end{aligned}$$

Similarly, using that

$$\|P_{k,p}\mathbb{F}u + P_{k,p}\Psi x_0\|_{L^2}^2 \le \|\mathbb{F}u + \Psi x_0\|_{L^2}^2,$$

we obtain

$$x_0^* X_k x_0 \le x_0^* X x_0 \qquad \forall x_0 \in \mathbb{C}^n.$$

Convergence of the sequence (X_k) follows by the fact that it is non-decreasing and bounded from above by X with respect to definiteness.
In the case where the non-Blaschke condition (7) is fulfilled, the union of the spaces $\mathscr{K}_k(\alpha)$ over all $k \in \mathbb{N}$ is dense in $L^2(0,\infty;\mathbb{C}^p)$ [14]. The sequence $(P_{k,p})$ therefore converges to the identity in the strong operator topology, that is

$$\lim_{k\to\infty} P_{k,p} y = y \qquad \forall y \in L^2(0,\infty;\mathbb{C}^p). \tag{35}$$

Let $x_0 \in \mathbb{C}^n$ and $\varepsilon > 0$. By (20) there exists some $u \in L^2(0,\infty;\mathbb{C}^m)$ with

$$x_0^* X x_0 < \|\mathbb{F}u + \Psi x_0\|^2 - \|u\|^2 + \tfrac{\varepsilon}{2}.$$

By (35), there exists some $N \in \mathbb{N}$ with $\|(\mathbb{F}u + \Psi x_0) - P_{k,p}(\mathbb{F}u + \Psi x_0)\|^2 \le \frac{\varepsilon}{2}$ for all $k \ge N$. Then we obtain that for all $k \ge N$ there holds

$$\begin{aligned} x_0^* X x_0 &< \|\mathbb{F}u + \Psi x_0\|^2 - \|u\|^2 + \tfrac{\varepsilon}{2} \\ &\le \|P_{k,p}\mathbb{F}u + P_{k,p}\Psi x_0\|^2 + \|(\mathbb{F}u + \Psi x_0) - P_{k,p}(\mathbb{F}u + \Psi x_0)\|^2 - \|u\|^2 + \tfrac{\varepsilon}{2} \\ &\le \|P_{k,p}\mathbb{F}u + P_{k,p}\Psi x_0\|^2 - \|u\|^2 + \varepsilon \le x_0^* X_k x_0 + \varepsilon. \end{aligned}$$

Using further that $X_k \leq X$, we obtain

$$|x_0^*(X - X_k)x_0| = x_0^* X x_0 - x_0^* X_k x_0 < \varepsilon \qquad \forall k \geq N.$$

It follows that the sequence (X_k) converges to X. □

Next we introduce a slighty different, numerically more advantageous, representation for the matrix X_k as in (31).

Theorem 5 *Assume that $A \in \mathbb{C}^{n\times n}$ is stable, $B \in \mathbb{C}^{n\times m}$, $C \in \mathbb{C}^{p\times n}$ and $D \in \mathbb{C}^{p\times m}$. Further assume that the bounded real Lur'e equation* (1) *has a minimal solution $X \in \mathbb{C}^{n\times n}$. Define Ψ and $\mathbb{F}$ by* (10) *and* (11).
Let $(\alpha_j)_{j=1}^{\infty}$ be a complex sequence with $\mathrm{Re}(\alpha_j) > 0$ for all $j \in \mathbb{N}$, and let $F_k \in \mathbb{C}^{kp\times km}$, $S_k \in \mathbb{C}^{kp\times n}$ be defined as in (27).
Then there exists a matrix $G_k \in \mathbb{C}^{\tilde{\ell}_k \times kp}$ with full row rank and

$$I - F_k F_k^* = G_k^* G_k. \tag{36}$$

Further, there exists some $R_k \in \mathbb{C}^{\tilde{\ell}_k \times n}$ such that

$$G_k^* R_k = S_k. \tag{37}$$

The matrix X_k as in (31) *fulfills*

$$X_k = R_k^* R_k. \tag{38}$$

Proof The matrix $I - F_k F_k^* \in \mathbb{C}^{kp\times kp}$ is positive semidefinite by Theorem 3. Therefore, $I - F_k^* F_k \in \mathbb{C}^{km\times km}$ is positive semidefinite as well. This implies the existence of some matrix $G_k \in \mathbb{C}^{\tilde{\ell}_k \times kp}$ with full row rank such that (36) holds.

By (29) we have $\ker(I - F_k^* F_k) = \ker(F_{\Xi,k})$. From (30) we obtain $\ker(F_{\Xi,k}) \subset \ker(S_k^* F_k)$, whence $\ker(I - F_k^* F_k) \subset \ker(S_k^* F_k)$.

We now prove $\mathrm{im}(S_k) \subset \mathrm{im}(I - F_k F_k^*)$. This is equivalent to $\ker(I - F_k F_k^*) \subset \ker(S_k^*)$. Let $y \in \ker(I - F_k F_k^*)$. Then $y = F_k F_k^* y$. Therefore

$$S_k^* y = S_k^* F_k F_k^* y \tag{39}$$

and $F_k^* y = F_k^* F_k F_k^* y$. The latter is equivalent to $(I - F_k^* F_k) F_k^* y = 0$. Thereby we obtain that $F_k^* y \in \ker(I - F_k^* F_k)$, which by the inclusion of nullspaces established in the previous paragraph gives $F_k^* y \in \ker(S_k^* F_k)$. Hence $S_k^* F_k F_k^* y = 0$. From (39) we then obtain $S_k^* y = 0$. We conclude that $\ker(I - F_k F_k^*) \subset \ker(S_k^*)$, as desired.

From (36) we obtain $\ker(I - F_k F_k^*) = \ker(G_k)$, so that $\mathrm{im}(I - F_k F_k^*) = \mathrm{im}(G_k^*)$. Together with the already established subspace inclusion $\mathrm{im}(S_k) \subset \mathrm{im}(I - F_k F_k^*)$, this shows that $\mathrm{im}(S_k) \subset \mathrm{im}(G_k^*)$. Since G_k has full row rank, $G_k G_k^*$ is invertible and therefore

$$R_k := (G_k G_k^*)^{-1} G_k S_k \tag{40}$$

is well-defined. We now show that it satisfies (37). Let $x \in \mathbb{C}^n$. By the above established subspace inclusion $\mathrm{im}(S_k) \subset \mathrm{im}(G_k^*)$, there exists a $z \in \mathbb{C}^{kp}$ such that $S_k x = G_k^* z$. Then

$$G_k^* R_k x = G_k^* (G_k G_k^*)^{-1} G_k S_k x = G_k^* (G_k G_k^*)^{-1} G_k G_k^* z = G_k^* z = S_k x.$$

Since $x \in \mathbb{C}^n$ was arbitrary this proves that $G_k^* R_k = S_k$, i.e the above defined R_k satisfies (37).

By Remark 4 we have $X_k = S_k^*[I + F_k(I - F_k^* F_k)^+ F_k^*]S_k$. Using the above established subspace inclusion $\mathrm{im}(S_k) \subset \mathrm{im}(I - F_k F_k^*)$ and the fact that $(I - F_k F_k^*)^+(I - F_k F_k^*)$ is the orthogonal projection onto $\mathrm{im}(I - F_k F_k^*)$ we may alternatively write this as

$$X_k = S_k^*[(I - F_k F_k^*)^+(I - F_k F_k^*) + F_k(I - F_k^* F_k)^+ F_k^*]S_k.$$

The following identity for Moore-Penrose pseudo-inverses is most easily proven by verifying the Moore-Penrose conditions [5, Sec. 5.5.4]:

$$(I - F_k F_k^*)^+ = (I - F_k F_k^*)^+(I - F_k F_k^*) + F_k(I - F_k^* F_k)^+ F_k^*.$$

From this we see that

$$X_k = S_k^*(I - F_k F_k^*)^+ S_k. \tag{41}$$

On the other hand we have, using (40),

$$R_k^* R_k = S_k^* G_k^* (G_k G_k^*)^{-2} G_k S_k,$$

and it is easily verified that $G_k^*(G_k G_k^*)^{-2} G_k$ is the Moore-Penrose pseudo-inverse of $G_k^* G_k$. Since $G_k^* G_k = I - F_k F_k^*$ by (36), it follows that $R_k^* R_k = X_k$. □

Remark 5 (Bounded real Lur'e equations)

a) It follows from (28) that $\imath_{k,p}^*(I - \mathbb{F}_k \mathbb{F}_k^*)\imath_{k,p} = I - F_k F_k^*$.
b) Observing the lower triangular block structure of matrix F_i in Alg. 1, that is

$$F_i = \begin{bmatrix} [F_{i-1}, 0] \\ Q_i(\overline{L_i} \otimes I_m) + [0, D] \end{bmatrix}, \tag{42}$$

we can determine the matrices $G_i \in \mathbb{C}^{\tilde{\ell}_i \times ip}$ and $R_i \in \mathbb{C}^{\tilde{\ell}_i \times n}$ recursively as follows: We have

$$\begin{aligned} & I - F_i F_i^* \\ &= \begin{bmatrix} I - F_{i-1} F_{i-1}^* & -\begin{bmatrix} F_{i-1} & 0 \end{bmatrix}\left(Q_i(\overline{L_i} \otimes I_m)\right)^* \\ -\left(Q_i(\overline{L_i} \otimes I_m)\right)\begin{bmatrix} F_{i-1} & 0 \end{bmatrix}^* & I - \left(Q_i(\overline{L_i} \otimes I_m) + [0, D]\right)\left(Q_i(\overline{L_i} \otimes I_m) + [0, D]\right)^* \end{bmatrix}. \end{aligned}$$

By making the ansatz $G_i = \begin{bmatrix} G_{i-1} & G_{12,i} \\ 0 & G_{22,i} \end{bmatrix}$, we obtain

$$\begin{aligned} & \begin{bmatrix} G_{i-1}^* G_{i-1} & G_{i-1}^* G_{12,i} \\ G_{12,i}^* G_{i-1} & G_{12,i}^* G_{12,i} + G_{22,i}^* G_{22,i} \end{bmatrix} \\ &= G_i^* G_i = I - F_i F_i^* \\ &= \begin{bmatrix} I - F_{i-1} F_{i-1}^* & -\begin{bmatrix} F_{i-1} & 0 \end{bmatrix}\left(Q_i(\overline{L_i} \otimes I_m)\right)^* \\ -\left(Q_i(\overline{L_i} \otimes I_m)\right)\begin{bmatrix} F_{i-1} & 0 \end{bmatrix}^* & I - \left(Q_i(\overline{L_i} \otimes I_m) + [0, D]\right)\left(Q_i(\overline{L_i} \otimes I_m) + [0, D]\right)^* \end{bmatrix}. \end{aligned}$$

Thus, the matrix $G_{12,i}$ is the unique solution of the linear equation

$$G_{i-1}^* G_{12,i} = -\begin{bmatrix} F_{i-1} & 0 \end{bmatrix}\left(Q_i(\overline{L_i} \otimes I_m)\right)^*.$$

Thereafter, the matrix $G_{22,i}$ can be obtained by a factorization

$$G_{22,i}^* G_{22,i} = I - \left(Q_i(\overline{L_i} \otimes I_m) + \begin{bmatrix} 0, D \end{bmatrix}\right)\left(Q_i(\overline{L_i} \otimes I_m) + \begin{bmatrix} 0, D \end{bmatrix}\right)^* - G_{12,i}^* G_{12,i}.$$

Since, by Alg. 1, S_i is obtained from S_{i-1} by

$$S_i = \begin{bmatrix} S_{i-1} \\ \sqrt{2\mathrm{Re}(\alpha_i)} \cdot V_i^* \end{bmatrix}, \tag{43}$$

we can, by making the ansatz $R_i = \begin{bmatrix} R_{i-1} \\ R_{2,i} \end{bmatrix}$, rewrite equation (37) as

$$\begin{bmatrix} G_{i-1}^* & 0 \\ G_{12,i}^* & G_{22,i}^* \end{bmatrix} \begin{bmatrix} R_{i-1} \\ R_{2,i} \end{bmatrix} = \begin{bmatrix} S_{i-1} \\ \sqrt{2\mathrm{Re}(\alpha_i)} \cdot V_i^* \end{bmatrix}.$$

Hence, $R_{2,i}$ is the solution of the linear equation

$$G_{22,i}^* R_{2,i} = \sqrt{2\mathrm{Re}(\alpha_i)} \cdot V_i^* - G_{12,i}^* R_{i-1}.$$

By Theorem 5 and Remark 5 b), we can set up the following algorithm for the determination of the minimal solution of bounded real Lur'e equations.

Algorithm 2 ADI iteration for the bounded real Lur'e equation.

Input: a stable matrix $A \in \mathbb{C}^{n\times n}$, and $B \in \mathbb{C}^{n\times m}$, $C \in \mathbb{C}^{p\times n}$, $D \in \mathbb{C}^{p\times m}$ such that the bounded real Lur'e equation (1) has the minimal solution $X \in \mathbb{C}^{n\times n}$, and shift parameters $\alpha_1, \ldots, \alpha_k \in \mathbb{C}$ with $\mathrm{Re}(\alpha_i) > 0$.
Output: $R_k \in \mathbb{C}^{\tilde{\ell}_k \times n}$ such that $R_k^* R_k = X_k \approx X$.

1: Perform steps 1–5 in Alg. 1
2: Determine a matrix G_1 with full row rank and $G_1^* G_1 = I - F_1 F_1^*$
3: Determine a matrix R_1 with $G_1^* R_1 = S_1$
4: **for** $i = 2, 3, \ldots, k$ **do**
5: Perform steps 7–14 in Alg. 1.
6: Determine a matrix $G_{12,i}$ with $G_{i-1}^* G_{12,i} = -\begin{bmatrix} F_{i-1} & 0 \end{bmatrix}\left(Q_i(\overline{L_i} \otimes I_m)\right)^*$
7: Determine a matrix $G_{22,i}$ with full row rank and
$G_{22,i}^* G_{22,i} = I - \left(Q_i(\overline{L_i} \otimes I_m) + \begin{bmatrix} 0, D \end{bmatrix}\right)\left(Q_i(\overline{L_i} \otimes I_m) + \begin{bmatrix} 0, D \end{bmatrix}\right)^* - G_{12,i}^* G_{12,i}$
8: $G_i = \begin{bmatrix} G_{i-1} & G_{12,i} \\ 0 & G_{22,i} \end{bmatrix}$
9: Determine a matrix $R_{2,i}$ with $G_{22,i}^* R_{2,i} = \sqrt{2\mathrm{Re}(\alpha_i)} \cdot V_i^* - G_{12,i}^* R_{i-1}$
10: $R_i = \begin{bmatrix} R_{i-1} \\ R_{2,i} \end{bmatrix}$
11: **end for**

Remark 6 If $A \in \mathbb{C}^{n\times n}$ is stable, $B = 0 \in \mathbb{C}^{n\times m}$, $C \in \mathbb{C}^{p\times n}$ and $D = 0 \in \mathbb{C}^{p\times m}$, then the bounded real Lur'e equations reduce to the Lyapunov equation

$$A^* X + XA + C^* C = 0.$$

In this case, the matrices in Alg. 2 read $F_i = 0$, $G_i = I$ and $S_i = R_i$. Then Alg. 2 reduces to the well-known and established ADI iteration for Lyapunov equations [7, 9, 21].

Now we consider positive real Lur'e equations. First we present a version of Theorem 3 for positive real systems. The proof can be done by adapting the lines of the proof of Theorem 3. Again, it may help to compare the following result with Theorem 1 specialized to $R = Q = I_m$ and $S = 0$.

Theorem 6 *Assume that $A \in \mathbb{C}^{n\times n}$ is stable, $B \in \mathbb{C}^{n\times m}$, $C \in \mathbb{C}^{m\times n}$ and $D \in \mathbb{C}^{m\times m}$. Further assume that the positive real Lur'e equation* (8) *has a minimal solution $X \in \mathbb{C}^{n\times n}$.*
Define Ψ and $\mathbb{F}$ by (10) *and* (11)*. Let $(\alpha_j)_{j=1}^{\infty}$ be a complex sequence with $\mathrm{Re}(\alpha_j) > 0$ for all $j \in \mathbb{N}$, and let $F_k \in \mathbb{C}^{kp\times km}$, $S_k \in \mathbb{C}^{kp\times n}$ be defined as in* (27)*.*
Then the matrix $F_k^* + F_k \in \mathbb{C}^{km\times km}$ is positive semidefinite. In particular, there exists some $F_{\Xi,k} \in \mathbb{C}^{\ell_k\times km}$ with full row rank and

$$F_k^* + F_k = F_{\Xi,k}^* F_{\Xi,k}. \tag{44}$$

Further, there exists some $S_{\Xi,k} \in \mathbb{C}^{\ell_k\times n}$ such that

$$F_{\Xi,k}^* S_{\Xi,k} = S_k. \tag{45}$$

For the orthogonal projector $P_{k,m}$ as in (26)*, the matrix X_k defined by*

$$X_k = S_{\Xi,k}^* S_{\Xi,k}. \tag{46}$$

fulfills,

$$x_0^* X_k x_0 = \sup_{u\in L^2(0,\infty;\mathbb{C}^m)} -2\mathrm{Re}\langle u, P_{k,m}\mathbb{F}u + P_{k,m}\Psi x_0\rangle \qquad \forall x_0 \in \mathbb{C}^n. \tag{47}$$

Again, we can formulate a convergence result. The proof is analogous to that of Theorem 4 and therefore omitted.

Theorem 7 *Assume that $A \in \mathbb{C}^{n\times n}$ is stable, $B \in \mathbb{C}^{n\times m}$, $C \in \mathbb{C}^{m\times n}$ and $D \in \mathbb{C}^{m\times m}$. Further assume that the positive real Lur'e equation* (8) *has a minimal solution $X \in \mathbb{C}^{n\times n}$. Define Ψ and $\mathbb{F}$ by* (10) *and* (11)*.*
Let $(\alpha_j)_{j=1}^{\infty}$ be a complex sequence with $\mathrm{Re}(\alpha_j) > 0$ for all $j \in \mathbb{N}$, and let $F_k \in \mathbb{C}^{kp\times km}$, $S_k \in \mathbb{C}^{kp\times n}$ be defined as in (27)*; let X_k be defined as in Theorem 6.*
Then

$$X_k \le X_{k+1}, \quad X_k \le X \qquad \forall k \in \mathbb{N},$$

and the sequence (X_k) converges. If, additionally, $(\alpha_j)_{j=1}^{\infty}$ satisfies the non-Blaschke condition (7)*, then (X_k) converges to X.*

Remark 7 (Positive real Lur'e equations and projected optimal control problems) In the following we show that, by using the fact that the matrix F_i has the lower triangular block structure as in (42), the matrices $F_{\Xi,i} \in \mathbb{C}^{\ell_i\times im}$ and $S_{\Xi,i} \in \mathbb{C}^{\ell_i\times n}$ can be recursively determined (cf. Remark 5 b):
We have

$$F_i + F_i^* = \begin{bmatrix} F_{i-1} + F_{i-1}^* & \begin{bmatrix} I_{(i-1)m} & 0 \end{bmatrix} (Q_i(\overline{L_i}\otimes I_m))^* \\ (Q_i(\overline{L_i}\otimes I_m)) \begin{bmatrix} I_{(i-1)m} \\ 0 \end{bmatrix} & D + D^* + \begin{bmatrix} 0 & I_m \end{bmatrix} (Q_i(\overline{L_i}\otimes I_m))^* + (Q_i(\overline{L_i}\otimes I_m)) \begin{bmatrix} 0 \\ I_m \end{bmatrix} \end{bmatrix}.$$

By making the ansatz $F_{\Xi,i} = \begin{bmatrix} F_{\Xi,i-1} & F_{\Xi 12,i} \\ 0 & F_{\Xi 22,i} \end{bmatrix}$, we obtain

$$\begin{bmatrix} F_{\Xi,i-1}^* F_{\Xi,i-1} & F_{\Xi,i-1}^* F_{\Xi 12,i} \\ F_{\Xi 12,i}^* F_{\Xi,i-1} & F_{\Xi 12,i}^* F_{\Xi 12,i} + F_{\Xi 22,i}^* F_{\Xi 22,i} \end{bmatrix} = F_{\Xi,i}^* F_{\Xi,i} = F_i + F_i^*$$
$$= \begin{bmatrix} F_{i-1} + F_{i-1}^* & \begin{bmatrix} I_{(i-1)m} & 0 \end{bmatrix} (Q_i(\overline{L_i} \otimes I_m))^* \\ (Q_i(\overline{L_i} \otimes I_m)) \begin{bmatrix} I_{(i-1)m} \\ 0 \end{bmatrix} & D + D^* + \begin{bmatrix} 0 & I_m \end{bmatrix} (Q_i(\overline{L_i} \otimes I_m))^* + (Q_i(\overline{L_i} \otimes I_m)) \begin{bmatrix} 0 \\ I_m \end{bmatrix} \end{bmatrix}.$$

Thus, the matrix $F_{\Xi 12,i}$ is the unique solution of the linear equation

$$F_{\Xi,i-1}^* F_{\Xi 12,i} = \begin{bmatrix} I_{(i-1)m} & 0 \end{bmatrix} (Q_i(\overline{L_i} \otimes I_m))^*.$$

Thereafter, the matrix $F_{\Xi 22,i}$ can be obtained by a factorization

$$F_{\Xi 22,i}^* F_{\Xi 22,i} = D + D^* + \begin{bmatrix} 0 & I_m \end{bmatrix} (Q_i(\overline{L_i} \otimes I_m))^* + (Q_i(\overline{L_i} \otimes I_m)) \begin{bmatrix} 0 \\ I_m \end{bmatrix} - F_{\Xi 12,i}^* F_{\Xi 12,i}.$$

Since, by Alg. 1, the matrices S_i and S_{i-1} are related by (43) we see, by making the ansatz $S_{\Xi,i} = \begin{bmatrix} S_{\Xi,i-1} \\ S_{\Xi 2,i} \end{bmatrix}$, that equation (45) now reads

$$\begin{bmatrix} F_{\Xi,i-1}^* & 0 \\ F_{\Xi 12,i}^* & F_{\Xi 22,i}^* \end{bmatrix} \begin{bmatrix} S_{\Xi,i-1} \\ S_{\Xi 2,i} \end{bmatrix} = \begin{bmatrix} S_{i-1} \\ \sqrt{2\mathrm{Re}(\alpha_i)} \cdot V_i^* \end{bmatrix}.$$

Hence, $S_{\Xi 2,i}$ is the solution of the linear equation

$$F_{\Xi 22,i}^* S_{\Xi 2,i} = \sqrt{2\mathrm{Re}(\alpha_i)} \cdot V_i^* - F_{\Xi 12,i}^* S_{\Xi,i-1}.$$

Algorithm 3 ADI iteration for the positive real Lur'e equation.

Input: $A \in \mathbb{C}^{n\times n}$ a stable matrix, $B \in \mathbb{C}^{n\times m}$, $C \in \mathbb{C}^{p\times n}$, $D \in \mathbb{C}^{p\times m}$ such that the positive real Lur'e equation (8) has the minimal solution $X \in \mathbb{C}^{n\times n}$, and shift parameters $\alpha_1, \ldots, \alpha_k \in \mathbb{C}$ with $\mathrm{Re}(\alpha_i) > 0$.
Output: $S_{\Xi,k} \in \mathbb{C}^{\ell_k \times n}$ such that $S_{\Xi,k}^* S_{\Xi,k} = X_k \approx X$.

1: Perform steps 1–5 in Alg. 1
2: Determine a matrix $F_{\Xi,1}$ with full row rank and $F_{\Xi,1}^* F_{\Xi,1} = F_1 + F_1^*$
3: Determine a matrix $S_{\Xi,1}$ with $F_{\Xi,1}^* S_{\Xi,1} = S_1$
4: **for** $i = 2, 3, \ldots, k$ **do**
5: Perform steps 7–14 in Alg. 1.
6: Determine a matrix $F_{\Xi 12,i}$ with $F_{\Xi,i-1}^* F_{\Xi 12,i} = \begin{bmatrix} I_{(i-1)m} & 0 \end{bmatrix} (Q_i(\overline{L_i} \otimes I_m))^*$
7: Determine a matrix $F_{\Xi 22,i}$ with full row rank and
$F_{\Xi 22,i}^* F_{\Xi 22,i} = D + D^* + \begin{bmatrix} 0 & I_m \end{bmatrix} (Q_i(\overline{L_i} \otimes I_m))^* + (Q_i(\overline{L_i} \otimes I_m)) \begin{bmatrix} 0 \\ I_m \end{bmatrix} - F_{\Xi 12,i}^* F_{\Xi 12,i}$
8: $F_{\Xi,i} = \begin{bmatrix} F_{\Xi,i-1} & F_{\Xi 12,i} \\ 0 & F_{\Xi 22,i} \end{bmatrix}$
9: Determine a matrix $S_{\Xi 2,i}$ with $F_{\Xi 22,i}^* S_{\Xi 2,i} = \sqrt{2\mathrm{Re}(\alpha_i)} \cdot V_i - F_{\Xi 12,i}^* S_{\Xi,i-1}$
10: $S_{\Xi,i} = \begin{bmatrix} S_{\Xi,i-1} \\ S_{\Xi 2,i} \end{bmatrix}$
11: **end for**

Remark 8 We note that Alg. 2 reduces to well-known ADI iteration for Lyapunov equations [7, 9, 21] (cf. Remark 6): If $A \in \mathbb{C}^{n\times n}$ is stable, $B = 0 \in \mathbb{C}^{n\times m}$, $C \in \mathbb{C}^{m\times n}$ and $D = \frac{1}{2}I_m \in \mathbb{C}^{m\times m}$, then the positive real Lur'e equation reduces to the Lyapunov equation

$$A^*X + XA + C^*C = 0.$$

The matrices in Alg. 3 then read $F_i = 0$, $F_{\Xi,i} = \frac{1}{2}I$ and $S_{\Xi,i} = S_i$, whence Alg. 2 then again reduces to ADI iteration for Lyapunov equations.

Remark 9 (Numerical effort for ADI iteration)

a) At this point we would like to address the computational cost in Alg. 2 and Alg. 3: As in ADI iteration for Lyapunov equations [7, 9, 21], our algorithms for Lur'e equations require two properties:
   (i) The numerical rank of X is small. That is, X has only few eigenvalues which exceed a small number $\varepsilon > 0$.
   (ii) The output dimension is small (i.e., $p \ll n$).

   Property (i) guarantees that X can be well approximated by a product $R_k^*R_k$ for some $R_k \in \mathbb{C}^{\ell_k\times n}$ with $\ell_k \ll n$ (note that $p = m$ in the positive real case). This enables that ADI iteration gives a good approximation after only a few steps. If Property (ii) is fulfilled, then the numerical effort for all steps in Alg. 2 and Alg. 3, except for the computation of the matrix $V_i = V_{i-1} - (\alpha_i + \overline{\alpha_{i-1}}) \cdot (\alpha_i I - A^*)^{-1} V_{i-1}$, are relatively negligible. The computation of V_i requires the solution of p linear systems with n degrees of freedom. In particular, possible sparsity of A can be exploited as in ADI iteration for Lyapunov equations [2, 13].

b) Our algorithm is of a totally different nature than the one presented in [15], which is based on a determination of the $\mathscr{E}$-neutral deflating space corresponding to the infinite eigenvalues of the even matrix pencil $s\mathscr{E} - \mathscr{A}$ as in (22). That is, the approach in [15] relies on a determination of a matrix sequence (V_i) with $\operatorname{im} V_0 = \ker \mathscr{E}$, $\operatorname{im} \mathscr{E}V_{i+1} = \operatorname{im} \mathscr{A}V_i$ and $V_{i+1}^*\mathscr{E}V_{i+1} = 0$. This sequence is shown to be stagnating. Thereafter, a projector $\Pi \in \mathbb{C}^{n\times n}$ is determined from the matrices V_i. The minimal solution X of the Lur'e equations is decomposed as

$$X = \Pi^*X\Pi + (I-\Pi)^*X\Pi + \Pi^*X(I-\Pi) + (I-\Pi)^*X(I-\Pi).$$

   The matrix $X(I-\Pi)$ can be directly computed from the matrices V_i. The remaining expression $(I-\Pi)^*X(I-\Pi)$ is shown to be the stabilizing solution of a certain projected Riccati equation, which is thereafter solved by Newton-Kleinman iteration. The bottleneck is indeed the determination of the matrices V_i, which requires in turn a successive computation of nullspaces. This may be a numerically ill-posed problem, if matrices (whose nullspaces have to be determined) with small singular values occur. We note that the algorithms in this article are working with the "original coordinates" and do not require any nullspace computations.

c) The choice of the shift parameters has a tremendous influence on the speed of convergence of ADI. By Remark 2 c), it might be reasonable to choose the shift parameters according to the generalized eigenvalues of the even matrix pencil (22). Selection of (sub-)optimal shift parameters however remains to be an open

problem. Furthermore, adaptive shift parameter selection methods, such as those in [1,20] for Lyapunov equations, are worthwhile to investigate for our algorithms.

5 Numerical Example

We present a numerical example to show the applicability of our algorithm and to demonstrate the expected performance of the ADI iteration for the positive real Lur'e equation in terms of monotonicity and convergence behavior. All the calculations were done using MATLAB 8.5 (R2015a) on a 64-bit server with 24 CPU cores of type Intel Xeon X5650 at 2.67GHz and 48 GB main memory available.

We consider a convection-diffusion equation on the unit square $\Omega := [0,1] \times [0,1]$, namely

$$\frac{\partial x}{\partial t}(\xi,t) = k\Delta x(\xi,t) + b^\top \nabla x(\xi,t), \quad (\xi,t) \in \Omega \times \mathbb{R}_{\geq 0}. \tag{48}$$

The input is a scalar function formed by the Robin boundary condition

$$u(t) = \nu(\xi)^\top \nabla x(\xi,t) + \alpha x(\xi,t), \quad (\xi,t) \in \partial\Omega \times \mathbb{R}_{\geq 0},$$

and the output consists of the integral of Dirichlet boundary values, i.e.

$$y(t) = \int_{\partial\Omega} x(\xi,t) d\sigma_\xi,$$

where $\partial\Omega$ denotes the boundary of Ω, σ_ξ denotes the surface measure, and $\nu(\xi)$ denotes the outward unit normal.

To discretize the PDE (48), we apply a finite element discretization with uniform triangular elements of fixed size $h = \frac{1}{N-1}$, where $N \in \mathbb{N}$ is the number of points in each coordinate direction. An example of the grid (for $N = 6$) that we used in our computations is shown in Fig. 1. In addition, we define the subspace $V_h \subset H^1(\Omega)$ using piecewise-linear basis functions. As a result, we obtain a finite dimensional dynamical system

$$\begin{aligned} E\dot{x}(t) &= Ax(t) + Bu(t), \\ y(t) &= Cx(t) \end{aligned} \tag{49}$$

with state space dimension $n = N^2$. $E \in \mathbb{R}^{n\times n}$ is a symmetric positive definite mass matrix, $A \in \mathbb{R}^{n\times n}$ is a non-symmetric stiffness matrix, $B \in \mathbb{R}^{n\times 1}$ is the input matrix, and $C \in \mathbb{R}^{1\times n}$ is the output matrix.

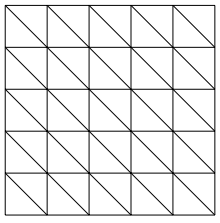

Fig. 1 An example of the chosen triangular element for $N = 6$

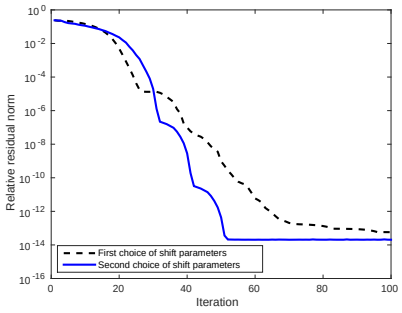

Fig. 2 Comparison of different shift parameters for ADI iteration: convection-diffusion equation with $n = 4900$, $b^\top = [10 \;\; 10]$, $k = 0.45$, and $\alpha = 3$

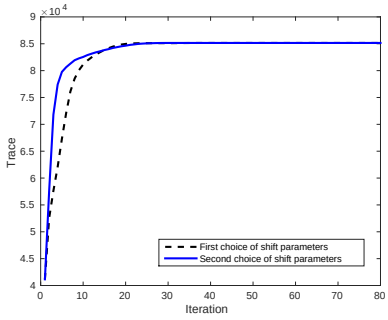

Fig. 3 Monotonicity of ADI iteration: convection-diffusion equation with $n = 4900$, $b^\top = [10 \;\; 10]$, $k = 0.45$, and $\alpha = 3$

The system is asymptotically stable and the matrix $A + A^*$ is negative definite. Furthermore, we have $B = k \cdot C^*$. A simple calculation then shows that the system is passive. Since $D = 0$, the corresponding positive real Lur'e equations cannot be turned into an algebraic Riccati equation.

We consider $N = 70$ ($n = N^2 = 4900$), $b = \left[\begin{smallmatrix} 10 \\ 10 \end{smallmatrix}\right]$, $k = 0.45$, and set $\alpha = 3$. We find an approximate solution $X \in \mathbb{C}^{n \times n}$ of the positive real Lur'e equation (8) by applying Alg. 3. Thereby, we use the modifications proposed in [10, Remark 7.1] & [8, Remark 3.3] which allow computations without explicit inversion of E. In addition, in steps 6 and 7 of Alg. 3, we do not need to compute the expression $Q_i(\overline{L_i} \otimes I_m)$, because we compute it once in step 14 of Alg. 1. In fact, we need to just access the last p rows of the matrix F_i in order to obtain the value of this expression (cf. Remark 7).

The choice of shift parameters has a major effect on the convergence speed of the ADI algorithm. In our example, we choose the following two different sets of shift parameters.

1. As a first set of shift parameters, we generate 30 parameters using the Wachspress method [21] on the basis of 4900 eigenvalues of the Dirichlet Laplacian given by $\pi^2(i^2 + j^2)$, $i, j = 1, 2, \ldots, 70$. To obtain a smooth convergence of the ADI method, we sort these shift parameters in an increasing order with respect to the values of their real part. We use the obtained shift parameters in the first 30 iterations. Afterwards, we select a subset of these parameters which provided the highest reduction in the value of residual norm. In our case, we choose 13 shift parameters and re-use them every 13 iterations. The computation time of these shift parameters for state space dimension $n = 4900$ is about 0.0025 seconds.
2. The second set of shift parameters is motivated by the statements in Remark 2 c). Specifically, we generate a set of 30 shift parameters using Penzl's heuristic procedure [13] on negatives of the stable eigenvalues of the even matrix pencil (22) with $Q = R = 0$ and $S = I_m$. In order to approximate the spectrum of this even matrix pencil, we calculate 450 Ritz values using the shift-and-invert Arnoldi process [19] with the shift $\sigma = 1$. The Arnoldi process is initialized with a random vector in $\mathbb{R}^n$. The computation time of these shift parameters for state space dimension $n = 4900$ is about 80 seconds.

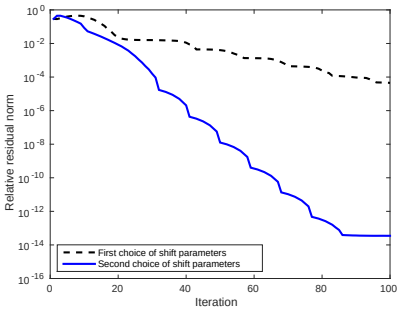

Fig. 4 Comparison of different shift parameters for ADI iteration: convection-diffusion equation with $n = 4900$, $b^\top = [45 \;\; 45]$, $k = 0.45$, and $\alpha = 3$

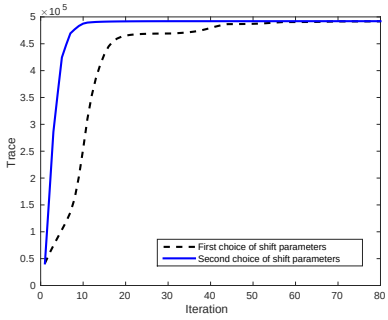

Fig. 5 Monotonicity of ADI iteration: convection-diffusion equation with $n = 4900$, $b^\top = [45 \;\; 45]$, $k = 0.45$, and $\alpha = 3$

We sort the obtained 30 parameters in an increasing order with respect to the values of their real part in order to obtain a smooth convergence. We perform 30 iterations of Alg. 3 using these shift parameters. Subsequently, we extract a subset of these parameters which provided the highest reduction in the value of residual norm. From this set of shift parameters, we extract 10 parameters to re-use every 10 iterations.

We add a large real shift parameter of order 10^{12} to the above two sets of shift parameters and consider it to be the first parameter in the set. We use this large shift parameter just in the first iteration of Alg. 3 and do not repeat it in the further iterations. The reason for adding a very big shift parameter can be explained as follows. Since in the positive real case the Popov function has a zero at infinity, a delta impulse will occur in the optimal control. The Takenaka-Malmquist basis function corresponding to a big shift parameter should suitably approximate the behavior of this delta impulse.

At each iteration i, we observe the relative residual norm of the positive real Lur'e equation using the approach proposed in [15, Sec. 6]. Fig. 2 shows the relative residual norm with respect to the iteration for the space dimension $n = 4900$ and for the two different choices of shift parameters which we have introduced earlier. We can conclude from this figure that the second set of shift parameters provides a faster convergence to the solution of positive real Lur'e equation corresponding to the system (49). In fact, with a tolerance of 10^{-13} on the relative residual norm for the problem with the space dimension $n = 4900$, the second choice of shift parameters leads to convergence in 51 iterations whereas the first set of parameters requires more than 80 iterations for the desired convergence.

In order to illustrate the monotonicity of the ADI iteration, we observe the trace of X_i, denoted by $\mathrm{trace}\,(X_i)$, at each iteration of Alg. 3. The trace of X_i can be computed efficiently as

$$\mathrm{trace}\,(X_i) = \mathrm{trace}\,\left(S^*_{\Xi,i} S_{\Xi,i}\right) = \|S_{\Xi,i}\|_F^2 \;,$$

where $\|\cdot\|_F$ denotes the Frobenius norm. Fig. 3 shows the trace of solutions X_i generated by Alg. 3 with the two sets of shift parameters introduced earlier in this example.

From this figure we observe that trace $(X_i) \leq$ trace (X_{i+1}), for all $i \in \mathbb{N}$, which is consistent with Theorem 7.

The execution time of the ADI algorithm for this example (including the computation of relative residual norm and trace of the solution at each iteration) is about 257 seconds for 100 iterations for the first choice of shift parameters and about 470 seconds for 100 iterations for the second choice of shift parameters.

We finish our numerical example by showing the effect of increasing the convection term on the convergence of the ADI algorithm. To this end, we increase the convection coefficients to $b = \left[\begin{smallmatrix} 45 \\ 45 \end{smallmatrix}\right]$ and keep the other parameters unchanged. In this case, the spectrum of the associated even matrix pencil gets more complicated. As a result, the selection of shift parameters becomes a more delicate task. We recompute the two sets of shift parameters in exactly the same way as for the case $b = \left[\begin{smallmatrix} 10 \\ 10 \end{smallmatrix}\right]$. The computation time for the first set of shift parameters is about 0.13 seconds and for the second set of shift parameters is about 78 seconds.

Fig. 4 shows that the first set of shift parameters fails to provide a fast convergence when the problem is convection dominated. The second choice of shift parameters still provides a convenient convergence, but the speed of convergence is rather slower compared to Fig. 2. The trace of X_i for both sets of shift parameters is depicted in Fig. 5. The execution time of the ADI algorithm in this case (including the computation of relative residual norm and trace of the solution at each iteration) is about 240 seconds for 100 iterations for the first choice of shift parameters and about 443 seconds for 100 iterations for the second choice of shift parameters.

6 Conclusions

We have introduced new numerical methods for the solutions of bounded real and positive real Lur'e equations. Thereby, only solvability of these equations together with stability of the underlying system have been assumed. Our methods generalize the well-known ADI iteration for Lyapunov equations and provide low-rank factors of the solution. Each iteration step basically consists of the solution of a linear system $(\alpha_i I - A^*)x = b$, where $\alpha_i \in \mathbb{C}$ is a so-called *shift parameter*. This enables the application of our algorithms to large-scale systems.

The theoretical basis for our convergence analysis is the fact that solutions of the Lur'e equations express the *available storage* of a system, which is a particular linear-quadratic optimal control problem. The matrices obtained by iteration are shown to correspond to a certain projected optimal control problem, where the projections are determined by the shift parameters. This gives rise to a simple sufficient criterion on the shift parameters for convergence of our method.

References

1. Peter Benner, Patrick Kürschner and Jens Saak. Self-generating and efficient shift parameters in ADI methods for large Lyapunov and Sylvester equations. Electron. Trans. Numer. Anal., **43**:142–162, 2014.
2. Peter Benner and Jens Saak. Numerical solution of large and sparse continuous time algebraic matrix Riccati and Lyapunov equations: a state of the art survey. GAMM-Mitt., **36**(1):32–52, 2013.

3. Ruth F. Curtain. Linear operator inequalities for strongly stable weakly regular linear systems. Math. Control Signals Systems, **14**(4):299–338, 1997.
4. Felix R. Gantmacher. The Theory of Matrices (Vol. II). Chelsea, New York, 1959.
5. Gene H. Golub and Charles F. Van Loan. Matrix Computations. The John Hopkins University Press, Baltimore/London, third edition, 1996.
6. Achim Ilchmann and Timo Reis. Outer transfer functions of differential-algebraic systems. ESAIM Control Optim. Calc. Var., 2016. Article first published online, DOI:10.1051/cocv/2015051.
7. Jing-Rebecca Li and Jacob White. Low rank solution of Lyapunov equations. SIAM J. Matrix Anal. Appl., **24**(1):260–280, 2002.
8. Yiding Lin and Valeria Simoncini. A new subspace iteration method for the algebraic Riccati equation. Numer. Linear Algebra Appl., **22**(1):26-47, 2015.
9. An Lu and Eugene L. Wachspress, Solution of Lyapunov equations by alternating direction implicit iteration, Comput. Math. Appl., **21**(9):43–58, 1991.
10. Arash Massoudi and Mark R. Opmeer and Timo Reis. Analysis of an iteration method for the algebraic Riccati equation. SIAM. J. Matrix Anal. Appl., accepted for publication, 2016. Preprint available at `https://preprint.math.uni-hamburg.de/public/papers/hbam/hbam2014-16.pdf`.
11. Philippe C. Opdenacker and Edmond A. Jonckheere. A contraction mapping preserving balanced reduction scheme and its infinity norm error bounds. IEEE Trans. Circuits Syst. I Regul. Pap., **35**(2):184–189, 1988.
12. Chris Guiver and Mark R. Opmeer. Error bounds in the gap metric for dissipative balanced approximations. Lin. Alg. Appl., **439**(12):3659–3698, 2013.
13. Thilo Penzl. A cyclic low-rank Smith method for large sparse Lyapunov equations. SIAM J. Sci. Comput., **21**(4):1401–1418, 1999/00.
14. Mark R. Opmeer, Timo Reis, and Winnifried Wollner. Finite-rank ADI iteration for operator Lyapunov equations. SIAM J. Control Optim., **51**(5):4084–4117, 2013.
15. Federico Poloni and Timo Reis. A deflation approach for large-scale Lur'e equations. SIAM J. Matrix Anal. Appl., **33**(4):1339–1368, 2012.
16. Federico Poloni and Timo Reis. A structured doubling algorithm for Lur'e equations. Numer. Linear Algebra Appl., **23**(1):169–186, 2016.
17. Timo Reis. Lur'e equations and even matrix pencils. Lin. Alg. Appl., **434**:152–173, 2011.
18. Timo Reis and Tatjana Stykel. Positive real and bounded real balancing for model reduction of descriptor systems. Int. J. Control, **83**(1):74–88, 2010.
19. Yousef Saad. Numerical Methods for Large Eigenvalue Problems. Manchester University Press, Manchester, UK, 1992.
20. John Sabino. Solution of Large-Scale Lyapunov Equations via the Block Modified Smith Method. Phd thesis, Rice University, 2006.
21. Eugene L. Wachspress. Iterative solution of the Lyapunov matrix equation. Appl. Math. Lett., **1**:87–90, 1988.
22. Joachim Weidmann. Linear Operators in Hilbert Spaces. Springer, New York, Heidelberg, Berlin, 1980.
23. Martin Weiss and George Weiss. Optimal control of stable weakly regular linear systems. Math. Control Signals Systems, **10**(4):287–330, 1997.
24. Jan C. Willems. Least squares stationary optimal control and the algebraic Riccati equation. IEEE Trans. Automat. Control, **16**:621–634, 1971.
25. Jan C. Willems. Dissipative dynamical systems. Part I: General theory. Arch. Ration. Mech. Anal., **45**:321–351, 1972.
26. Jan C. Willems. Dissipative dynamical systems. Part II: Linear systems with quadratic supply rates. Arch. Ration. Mech. Anal., **45**:352–393, 1972.
27. Kemin Zhou and John D. Doyle and Keith Glover. Robust and Optimal Control. Prentice-Hall, Princeton, 1996.