Numerical Algorithms for the Linear-Quadratic Optimal Control of Well-Posed Linear Systems

Dissertation zur Erlangung des Doktorgrades
an der Fakultät für Mathematik, Informatik
und Naturwissenschaften

Fachbereich Mathematik
der Universität Hamburg

vorgelegt von

Arash Massoudi

Hamburg, 2016

Datum der Disputation:

12.04.2017

Gutachter der Dissertation:

Prof. Dr. Timo Reis

Prof. Dr. Mark Opmeer

Acknowledgements

As we express our gratitude, we must never forget that the highest appreciation is not to utter words, but to live by them.

—John F. Kennedy

First and foremost I would like to express my deepest gratitude to *Timo Reis* for his great support as my supervisor. His systematic way of thinking and acute insight in the field of mathematical control theory showed me the path of devising proper research plans. I am very thankful for the freedom he gave me to find my own way in mathematics.

I am profoundly grateful to have had the unique opportunity to collaborate with *Mark Opmeer* and *Timo Reis* on two articles which created the main part of this dissertation (Chapter 4). The deep knowledge of *Mark* in the theory of infinite-dimensional systems allowed me to gain a finer understanding in this research area.

I am indebted to *Winnifred Wollner* for providing me with his implementation code of the numerical examples presented in [41, Section 6.1].

I would like to give very special thanks to *Felix Schwenninger* for his exceptional support in the last months before submitting this thesis. His constant suggestions and feedbacks improved the quality of this dissertation considerably. In particular, his substantial comments on Chapter 5 guided me to enhance its presentation. Even in the last days before the deadline, which coincided with Christmas holidays, he offered me his valuable time to proofread essential parts of this work.

I owe particular thanks to *Ines Dorschky*, *Philipp Kunde*, and *Tobias Jordan* for proofreading different chapters of this dissertation.

I am very fortunate to have such invaluable parents and siblings who provided me with enormous source of strength. I shall always be indebted to them for their endless support.

My heartfelt gratitude goes to *Olesya* who encouraged me every single day while I was composing this manuscript. She helped me to enhance the syntactic structure of this dissertation by reading carefully many chapters of this dissertation. Her emotional support was the best gift I could ever receive in the last months of writing my thesis. *Olesya, le parole non bastano per dirti quello che provo ... grazie per quello che sei.*

Contents

List of Figures

List of Notations

$(\alpha_j)_{j=1}^{\infty}$	shift parameters, page 50
$(\psi_j)_{j=1}^{\infty}$	Takenaka–Malmquist system, page 51
$(\varphi_j)_{j=1}^{\infty}$	convolution system, page 50
(A, B, C, D)	generating operators of a weakly regular well-posed linear system Σ, page 24
$\mathbf{G}$	transfer function, page 20
$\mathcal{L}(\mathscr{X}, \mathscr{Y})$	space of bounded linear operators $T : \mathscr{X} \to \mathscr{Y}$, page 7
$\mathbf{P}_\tau$	projection by truncation, page 12
B^*_Λ	(strong) Λ-extension of operator B^*, page 25
$\mathbf{S}^*_\tau$	left shift by some $\tau > 0$, page 12
$\mathbf{S}_\tau$	right shift by some $\tau > 0$, page 12
$B^*_{\Lambda w}$	weak Λ-extension of operator B^*, page 25
$\cdot \underset{\tau}{\Diamond} \cdot$	τ-concatenation, page 13
C_Λ	(strong) Λ-extension of operator C, page 24
$C_{\Lambda w}$	weak Λ-extension of operator C, page 23
ℓ_p	p-summable complex sequences, page 7
$\gamma_{\mathbb{F}}$	growth bound of the input-output map $\mathbb{F}$, page 21
$\hat{u}$	Laplace transform of a function u, page 21
$\mathscr{K}_k(\alpha)$	$\mathscr{K}_k(\alpha) := \operatorname{span}\{\varphi_1, \varphi_2, \dots, \varphi_k\}$, page 51
$\mathscr{R}$	Popov operator, page 32

$\mathscr{R}_c$	complementary Popov operator, page 33
$\mathscr{U}$	input space, page 14
$\mathscr{X}$	state space, page 14
$\mathscr{X}'$	dual space of $\mathscr{X}$, page 8
$\mathscr{Y}$	output space, page 14
$\ker(T)$	kernel of an operator T, page 7
$\mathcal{K}(\mathscr{X},\mathscr{Y})$	space of compact linear operators $T:\mathscr{X}\to\mathscr{Y}$, page 7
$\mathcal{S}_p(\mathscr{X},\mathscr{Y})$	space of p-th Schatten-class operators, page 9
$\mathbb{C}_+$	$\mathbb{C}_+ := \mathbb{C}_0$, page 7
$\mathbb{C}_\omega$	open right half-plane in $\mathbb{C}$, page 7
$\mathbb{F}$	(extended) input-output map, page 14
$\mathbb{T}$	strongly continuous semigroup, page 12
$\lVert\cdot\rVert_{\mathcal{L}(\mathscr{U},\mathscr{Y})}$	operator norm on the space $\mathcal{L}(\mathscr{U},\mathscr{Y})$, page 21
$\lVert\cdot\rVert_{\mathscr{X}}$	norm on the Hilbert space $\mathscr{X}$, page 7
$\lVert\cdot\rVert_{\mathcal{H}^p_\omega(\mathscr{U},\mathscr{Y})}$	norm on the Hardy space $\mathcal{H}^p_\omega(\mathscr{U},\mathscr{Y})$, page 21
$\lVert\cdot\rVert_{gr}$	graph norm, page 7
$\lVert u\rVert^2_\omega$	norm on the Hilbert space $L^2_\omega(0,\infty;\mathscr{U})$, page 21
$\omega_0\,(\mathbb{T})$	growth bound of the semigroup $\mathbb{T}$, page 12
$\otimes$	tensor product, page 54
Φ	input map, page 14
$\Phi\Phi^*$	infinite-time controllability Gramian, page 19
Π	Popov function, page 32
Ψ	(extended) output map, page 14
$\Psi^*\Psi$	infinite-time observability Gramian, page 19
$\rho(T)$	resolvent set of operator T, page 7
$\mathcal{H}^p_\omega(\mathscr{U})$	Hardy space of analytic $\mathscr{U}$-valued functions on $\mathbb{C}_\omega$, page 20
$\mathcal{H}^p_\omega(\mathscr{U},\mathscr{Y})$	Hardy space of analytic $\mathcal{L}(\mathscr{U},\mathscr{Y})$-valued functions on $\mathbb{C}_\omega$, page 21

Σ	well-posed linear system, page 13
$\sigma(T)$	spectrum of operator T, page 7
Σ^a	anticausal interpretation of the dual system Σ^d, page 27
Σ^d	dual of a well-posed linear system Σ, page 25
Σ^K	closed-loop system, page 27
Σ_Ξ	spectral factor system, page 37
$\mathcal{K}(\mathscr{X})$	$\mathcal{K}(\mathscr{X}) := \mathcal{K}(\mathscr{X}, \mathscr{X})$, page 7
$\mathcal{L}(\mathscr{X})$	$\mathcal{L}(\mathscr{X}) := \mathcal{L}(\mathscr{X}, \mathscr{X})$, page 7
$\mathrm{trace}(T)$	trace of an operator T, page 9
Ξ	spectral factor, page 36
A	semigroup generator, page 12
B	control operator, page 17
C	observation operator, page 17
$C(0,\infty;\mathscr{X})$	space of continuous functions $f : [0,\infty) \to \mathscr{X}$, page 7
$C^n(0,\infty;\mathscr{X})$	n times differentiable functions $f : [0,\infty) \to \mathscr{X}$ whose derivatives of order $\leq n$ are in $C(0,\infty;\mathscr{X})$, page 7
D	feedthrough operator, page 23
$D(T)$	domain of an operator T, page 7
F^{opt}	optimal feedback operator, page 36
$g * h$	convolution product of functions g and h, page 50
$H^1(0,\infty;\mathscr{X})$	Sobolev space of locally absolutely continuous functions $f : [0,\infty) \to \mathscr{X}$ for which $\frac{df}{dt} \in L^2(0,\infty;\mathscr{X})$, page 8
$H^1_0(0,\infty;\mathscr{X})$	Sobolev space of functions in $H^1(0,\infty;\mathscr{X})$ with compact support in $[0,\infty)$, page 8
K	output feedback, page 27
$L^p(0,\infty;\mathscr{X})$	Lebesgue space of measurable functions $f : [0,\infty) \to \mathscr{X}$ with the property $\|f(\cdot)\|_{\mathscr{X}} \in L^p(0,\infty)$, page 7
O_k	orthogonal projector onto $\mathscr{K}_k(\alpha) \otimes \mathscr{U}$, page 56
P_k	orthogonal projector onto $\mathscr{K}_k(\alpha) \otimes \mathscr{Y}$, page 56

T^*	adjoint of T, page 8
$T^{1/2}$	square root of a nonnegative operator T, page 9
V_h	the finite element subspace of a vector space V, page 152
X	usually the Riccati operator, page 33
$\langle \cdot, \cdot \rangle_{\mathscr{X}}$	inner product on the Hilbert space $\mathscr{X}$, page 7
$\mathrm{Im}(T)$	image of an operator T, page 7

Chapter 1

Introduction

It ought to be remembered that there is nothing more difficult to take in hand, more perilous to conduct, or more uncertain in its success, than to take the lead in the introduction of a new order of things.

—Niccoló Machiavelli

Contents

1.1 The linear-quadratic optimal control problem in finite dimensions

We start by presenting the central objective of our work. To make the reading of this introduction easier, we avoid using intricate mathematical notations. However, in case of ambiguity of the notation we use, one should consult Section 1.3.

Let us consider the following *finite-dimensional linear time-invariant (LTI) system*

$$\Sigma : \begin{cases} \dot{x}(t) = Ax(t) + Bu(t), & x(0) = x_0 \in \mathbb{C}^n, \\ y(t) = Cx(t), & \forall\, t \geq 0, \end{cases} \tag{1.1}$$

with matrices $A \in \mathbb{C}^{n\times n}$, $B \in \mathbb{C}^{n\times m}$, and $C \in \mathbb{C}^{p\times n}$. We call $x(t) \in \mathbb{C}^n$ the *state* of Σ at time t, $u : (0,\infty) \to \mathbb{C}^m$ the *input function*, and $y : (0,\infty) \to \mathbb{C}^p$ the *output function*. Systems of type (1.1) are usually described by the diagram shown in Figure 1.1: The state of Σ is *controlled* by the input u and *observed* through the output y.

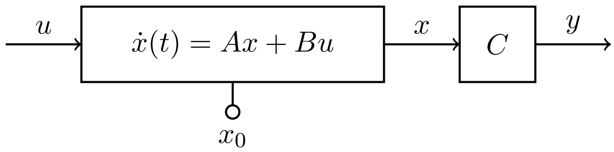


Figure 1.1: The input-state-output description of the system Σ.

The principal focus of this dissertation is on minimizing the cost of *controlling* and *observing* the system given in (1.1). This means that we want to minimize the *cost functional*

$$J(x_0, u) = \int_0^\infty \|u(t)\|^2 + \|y(t)\|^2 \ dt, \tag{1.2}$$

where the output function $y : (0,\infty) \to \mathbb{C}^p$ is subject to the dynamical system (1.1). This optimization problem is called the "linear-quadratic optimal control problem". In order to ensure that the cost functional (1.2) remains finite, we require the input function $u : (0,\infty) \to \mathbb{C}^m$ and the output function $y : (0,\infty) \to \mathbb{C}^p$ to be L^2-integrable (this condition is called *optimizability* of System (1.1)). This means that we minimize the cost functional (1.2) over the set of those *control inputs* $u \in L^2(0,\infty;\mathbb{C}^m)$ which result in $y \in L^2(0,\infty;\mathbb{C}^p)$ (through the dynamical system (1.1)).

The solution to the linear-quadratic optimal control problem can be determined by

$$\inf_{u\in L^2(0,\infty;\mathbb{C}^m)} J(x_0, u) = J(x_0, u^{\mathrm{opt}}) = \langle x_0, Xx_0\rangle\,, \tag{1.3}$$

where $X = X^* \in \mathbb{C}^{n\times n}$ is the smallest positive semidefinite solution of the *algebraic Riccati equation (ARE)*

$$A^*X + XA + C^*C - XBB^*X = 0, \tag{1.4}$$

see for example [38]. The minimizer u^{opt} in (1.3) can be written via the *linear state feedback*

$$u^{\mathrm{opt}}(t) := -B^*Xx(t), \qquad t \geq 0. \tag{1.5}$$

The closed-loop system obtained via the linear state feedback (1.5) (as depicted in Figure 1.2) is called the *optimal closed-loop system*. As shown in Figure 1.2, we have constructed the new input relation $u = v - B^*Xx(t)$ for the system Σ. By substituting this relation in (1.9), we reach the optimal closed-loop system:

$$\begin{cases} \dot{x}(t) = (A - BB^*X)x(t) + Bv(t), & x(0) = x_0 \in \mathbb{C}^n, \\ y(t) = Cx(t), & \forall\, t \geq 0. \end{cases}$$

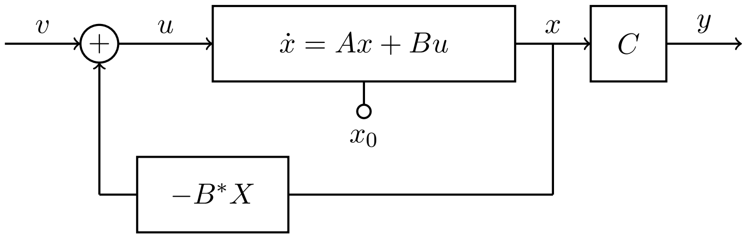


Figure 1.2: The optimal feedback for the system Σ.

If all the eigenvalues of the closed-loop matrix $A_{\text{opt}} := A - BB^*X$ have negative real part, then we call the closed-loop system *stable* and the matrix X is called a *stabilizing* solution of (1.4). Under certain assumptions on the matrices A, B, and C, one can show the existence of a unique stabilizing solution X of the ARE (1.4) (see for example [31]).

As a result of the above observation, one way of solving the linear-quadratic optimal control problem is to solve (1.4) and then the optimal control follows from (1.5). Hence, the numerical solution of the algebraic Riccati equation (1.4) lies in the central focus of the linear-quadratic optimal control problem for systems of type (1.1).

There are many algorithms available for solving (1.4) numerically (see [10] for a review). For problems where the model order is small (e.g., $n = 100$), a direct method based on calculating the eigenvectors of the associated *Hamiltonian* works well [2,33]. However, if the dynamical system (1.1) arises in a finite dimensional approximation of partial differential equations, then n becomes typically very large. In this case, the calculation of eigenvectors turns into a laborious task and a direct method is not suitable anymore.

In this work we focus on two iterative algorithms which provide efficient approximate solutions of (1.4) in *low-rank factored form*. This means that they provide sequences $(X_k)_{k\in\mathbb{N}} \in \mathbb{C}^{n\times n}$ 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 "small" ℓ_k. The main computational cost of these algorithms consists of, at each iteration, solving a linear system of the form $(\alpha_k I - A)x = v$, where $v \in \mathbb{C}^{n\times p}$ and the "shift parameter" $\alpha_k \in \mathbb{C}$ satisfies $\operatorname{Re}(\alpha_k) > 0$. These features make these algorithms attractive for the case where n is large, p is small, and A is sparse. This situation arises for example when considering discretizations of partial differential equations.

The first algorithm that we consider is the recently developed ADI method for solving (1.4). This algorithm was first proposed in [34] without a complete proof of convergence. In [35] a new perspective on this method in terms of the underlying linear-quadratic optimal control problem was introduced. This representation is independent of the Riccati equation and allows a straightforward proof of convergence. Moreover, the setting introduced in [35] allows an extension of the algorithm to *infinite-dimensional systems* [35, Theorem 7.1].

The second algorithm that we discuss is the well-known Newton-Kleinman iteration. This algorithm has received considerable attention in the literature since its introduction by Kleinman [29]. By applying the Newton iteration to the quadratic matrix equation (1.4), one obtains a sequence of *Lyapunov equations*

$$X_{k+1}(A - BB^*X_k) + (A - BB^*X_k)^*X_{k+1} = -X_kBB^*X_k - C^*C. \tag{1.6}$$

Given the current approximation X_k, one needs to solve (1.6) to find the next iterate X_{k+1}. In this dissertation we solve (1.6) by employing the ADI method for solving the Lyapunov equation. This method has been proven to be highly efficient for solving large scale problems when applied in order to compute low-rank factors of the solution (see for example [8, 9]).

1.2 Extension to infinite-dimensional spaces

In this dissertation we focus on the generalization of the setting presented in Section 1.1 to infinite-dimensional Hilbert spaces. This means that we consider systems whose *state space* is not any more $\mathbb{C}^n$, but an infinite-dimensional Hilbert space. This setting arises naturally when considering partial differential or delay differential equations. Our primary goal is to generalize the ADI method and the Newton-Kleinman iteration in order to solve the linear-quadratic optimal control problem for *infinite-dimensional systems*.

The solution theory for systems of type (1.1) with operators A, B, and C acting on infinite-dimensional spaces becomes more complicated. In fact, the local description of the system at a specific time $t \geq 0$ requires the definition of "suitable domains", because one usually encounters differential, trace and other "unbounded" operators. To deal with this issue, we focus on the class of *well-posed linear systems* (Definition 2.4). This class includes many input-state-output systems described by partial differential equations and provides a formal resemblance to finite-dimensional theory. The concept of well-posed linear systems was introduced by Salamon [55, 56] and gives a description of the system using *strongly continuous semigroups* (Definition 2.1) and appropriate *integral operators*. For a complete overview on this class we refer to [58] and the references therein.

The class of well-posed linear systems will be defined in Chapter 2 (see Definition 2.4). In order to make the definition easier to understand, let us take a look at the finite-dimensional LTI system (1.1). Formally, by applying the *variation of constants formula* to (1.1) we obtain

$$\begin{aligned} x(t) &= e^{At}x_0 + \int_0^t e^{A(t-\tau)}Bu(\tau)d\tau, \\ y(t) &= Ce^{At}x_0 + C\int_0^t e^{A(t-\tau)}Bu(\tau)d\tau. \end{aligned} \tag{1.7}$$

Now for all $t \geq 0$, $x_0 \in \mathbb{C}^n$, and $u \in L^2(0, \infty; \mathbb{C}^m)$, we define the following four operators:

$$\begin{aligned}
\mathbb{T}_t x_0 &= e^{At} x_0, \\
\Phi_t u &= \int_0^t e^{A(t-\tau)} Bu(\tau) d\tau, \\
(\Psi x_0)(t) &= Ce^{At} x_0, \\
(\mathbb{F} u)(t) &= C \int_0^t e^{A(t-\tau)} Bu(\tau) d\tau.
\end{aligned} \tag{1.8}$$

With this notation, we write (1.7) as

$$\begin{aligned}
x(t) =& \mathbb{T}_t x_0 + \Phi_t u, \\
y =& \Psi x_0 + \mathbb{F} u.
\end{aligned} \tag{1.9}$$

Here $\mathbb{T}$ is a *strongly continuous semigroup* (Definition 2.1), Φ is called the *input map*, Ψ the *output map*, and $\mathbb{F}$ the *input-output map*. It follows from (1.9) that operators $(\mathbb{T}, \Phi, \Psi, \mathbb{F})$ connect the initial state $x_0 \in \mathbb{C}^n$ and the input $u : [0, \infty) \to \mathbb{C}^m$ to the *state trajectory* $x : [0, \infty) \to \mathbb{C}^n$ and the output function $y : [0, \infty) \to \mathbb{C}^p$. These operators describe the input-state-output behavior of the LTI system (1.1) completely.

Formulas (1.8) and (1.9) hold true when A, B, and C are bounded operators acting on infinite-dimensional system. However, these operators are typically unbounded (in a sense that will be more clear in Section 2.3). For example, this is the case when A is a differential operator, B is a boundary control, and C is a boundary observation. As a result, the nice (classical) structures presented in (1.7)–(1.9) (and also in Section 1.1) do not hold in general. By requiring the operators $(\mathbb{T}, \Phi, \Psi, \mathbb{F})$ to satisfy certain properties (Definition 2.4), we attempt to "imitate" the input-state-output behavior of the LTI system (1.1) also in the infinite-dimensional setting. These properties include in particular *time-invariance* and *causality*. A system for which the operators $(\mathbb{T}, \Phi, \Psi, \mathbb{F})$ satisfy the properties listed in Definition 2.4 is called a *well-posed linear system*.

One can also connect the linear-quadratic optimal control problem to the solution of an *operator algebraic Riccati equation*, i.e., equation of type (1.4) where A, B, and C are not anymore matrices but rather operators acting on infinite-dimensional Hilbert spaces. If A generates a strongly continuous semigroup (Definition 2.3), B and C are bounded, then the *optimal cost operator* X (as in (1.3)) satisfies the following weak form of the algebraic Riccati equation

$$\langle Ax_0, Xz_0 \rangle + \langle Xx_0, Az_0 \rangle + \langle Cx_0, Cz_0 \rangle - \langle B^* X x_0, B^* X z_0 \rangle = 0, \tag{1.10}$$

for all $x_0, z_0 \in D(A)$, where $D(A)$ denotes the domain of the operator A [18, Theorem 6.2.4]. In addition, the minimizer u^{opt} in (1.3) can again be written via the linear state feedback (1.5). However, if the control operator B is unbounded, then it may happen that $Xx \notin D(B^*)$ for some $x \in D(A)$. In this case, the linear feedback (1.5) and the weak form of the algebraic Riccati equation (1.10) does not make sense anymore. We will address this problem in more details in Chapter 3 and present a "Riccati-like" equation from [70] to overcome this problem.

A common technique in the numerical solution of the linear-quadratic optimal control of infinite-dimensional systems is to first apply discretization techniques (e.g., finite element method) to obtain a finite-dimensional system (of the form (1.1)) and then solve the algebraic Riccati equation (1.4) numerically (as explained in Section 1.1). This approach is commonly referred to as "discretizing-then-optimizing". The main drawbacks of this method can be summarized as follows:

- The discretization method may not accurately inherit the properties of the infinite-dimensional operators. Losing these properties may slow down or even destroy the convergence of the resulting approximations.
- Usually, a very fine discretization is needed to acquire correct approximation of the infinite-dimensional operators. This results in high-order state-space systems, which can drastically increase the computational costs of the problem.
- As we have already mentioned, if the control operator B is unbounded, then the algebraic Riccati equation in the form (1.10) (and its matrix version (1.4)) may not be the correct equation to solve, because operator B^*X does not make sense anymore.

Our central motivation in studying numerical algorithms for the optimal control of infinite-dimensional systems is the possibility to apply the so-called "optimizing-then-discretizing" approach. The advantages of this technique over the classical method of "discretizing-then-optimizing" can be summarized as follows:

- In the setting of "optimizing-then-discretizing", one avoids (initial) matrix approximations of the infinite-dimensional operators. These matrices can be difficult to obtain for certain problems (such as linearized fluid flow systems).
- A distinct advantage of "optimizing-then-discretizing" is the possibility of using adaptive refinement techniques: For example, an appropriate mesh adaptation strategy is chosen at each iteration of the optimization method. This leads to significant computational savings (compared to using fixed grid a priori) and increases the convergence accuracy (we refine the mesh where the solution is not smooth).

The main drawback of "optimizing-then-discretizing" is the rather involved convergence theory for infinite-dimensional systems which requires many technical tools from functional analysis and infinite-dimensional control theory.

1.3 Preliminaries and functional analysis notions

In this section we review the functional analytic preliminaries that are important in developing the theory of well-posed linear systems in the next chapter. Our purpose is to unify the notations and make the reading of the forthcoming chapters easier. Since the main focus of this dissertation is on the linear-quadratic optimal control problem, we deal only with operators acting on Hilbert spaces.

Throughout this dissertation $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{R}$, and $\mathbb{C}$ denote the set of natural numbers, integer numbers, real numbers, and complex numbers, respectively. For a complex number $s \in \mathbb{C}$,

$\mathrm{Re}(s)$ and $\mathrm{Im}(s)$ denote the real and the imaginary part of s, respectively. For any $\omega \in [-\infty, \infty)$, $\mathbb{C}_\omega$ denotes the open right half-plane in $\mathbb{C}$ delimited by ω:

$$\mathbb{C}_\omega := \{s \in \mathbb{C} \mid \mathrm{Re}(s) > \omega\},$$

and we define $\mathbb{C}_+ := \mathbb{C}_0$.

In every part of this dissertation, $\mathscr{U}$, $\mathscr{X}$, and $\mathscr{Y}$ denote Hilbert spaces (possibly infinite-dimensional). These spaces are usually referred to as the *input space*, the *state space*, and the *output space*, respectively. Furthermore,

- $\langle \cdot, \cdot \rangle_{\mathscr{X}} : \mathscr{X} \times \mathscr{X} \to \mathbb{C}$ denotes the inner product on the Hilbert space $\mathscr{X}$ and $\| \cdot \|_{\mathscr{X}} : \mathscr{X} \to [0, \infty)$ denotes the norm induced by this inner product. Moreover, $I_{\mathscr{X}}$ denotes the identity operator acting on $\mathscr{X}$. We omit the indices whenever the associated space is clear from the context.

- $\mathcal{L}(\mathscr{X}, \mathscr{Y})$ and $\mathcal{K}(\mathscr{X}, \mathscr{Y})$ denote the spaces of bounded linear and compact linear operators $T : \mathscr{X} \to \mathscr{Y}$, respectively. We define $\mathcal{L}(\mathscr{X}) := \mathcal{L}(\mathscr{X}, \mathscr{X})$ and $\mathcal{K}(\mathscr{X}) := \mathcal{K}(\mathscr{X}, \mathscr{X})$.

- $D(T)$, $\mathrm{Im}(T)$, and $\ker(T)$ denote the domain, the image, and the kernel of an operator T, respectively.

- A linear operator $T : D(T) \subset \mathscr{X} \to \mathscr{Y}$ is called *closed*, if for any sequence $(x_n) \in D(T)$, which satisfies $\lim_{n\to\infty} x_n = x \in \mathscr{X}$ and $\lim_{n\to\infty} T x_n = y \in \mathscr{Y}$, it follows that $x \in D(T)$ and $Tx = y$.

- Let $T : D(T) \subset \mathscr{X} \to \mathscr{Y}$ be a closed linear operator. $D(T)$ equipped with the *graph norm* is a Hilbert space, where the graph norm $\|\cdot\|_{gr}$ is defined as

$$\|x\|_{gr} := \|x\|_{\mathscr{X}}^2 + \|Tx\|_{\mathscr{Y}}^2 . \tag{1.11}$$

- Let $T : D(T) \subset \mathscr{X} \to \mathscr{Y}$ be a closed linear operator. $\rho(T)$ denotes the *resolvent set* of T, which is defined as

$$\rho(T) := \{\lambda \in \mathbb{C} \mid (\lambda I - T) : D(T) \to \mathscr{X} \text{ is bijective}\}.$$

  The *spectrum* of T is defined as the complement of $\rho(T)$, i.e.,

$$\sigma(T) := \mathbb{C} \backslash \rho(T).$$

  The inverse $(\lambda I - T)^{-1}$ is called the *resolvent operator* of T at the point $\lambda \in \rho(T)$.

Let $\mathscr{X}$ be a Hilbert space. $C(0, \infty; \mathscr{X})$ consists of all the continuous functions $f : [0, \infty) \to \mathscr{X}$ and $C^n(0, \infty; \mathscr{X})$, for $n \in \mathbb{N}$, consists of all the n times differentiable functions $f : [0, \infty) \to \mathscr{X}$ whose derivatives of order $\leq n$ are in $C(0, \infty; \mathscr{X})$. For $p \in [1, \infty]$, ℓ_p stands for the p-summable complex sequences and $L^p(0, \infty; \mathscr{X})$ denotes the Lebesgue space of measurable functions $f : [0, \infty) \to \mathscr{X}$ with the property that $\|f(\cdot)\|_{\mathscr{X}} \in L^p(0, \infty)$. Moreover, we define

$$L^p_{loc}(0, \infty; \mathscr{X}) := \{f : [0, \infty) \to \mathscr{X} \mid f \in L^p(0, \tau; \mathscr{X}), \ \forall\, \tau > 0\}.$$

The Sobolev space $H^1(0,\infty;\mathscr{X})$ consists of those locally absolutely continuous functions $f:[0,\infty)\to\mathscr{X}$ for which $\frac{df}{dt}\in L^2(0,\infty;\mathscr{X})$. The space $H^1_0(0,\infty;\mathscr{X})$ consists of those functions in $H^1(0,\infty;\mathscr{X})$ with compact support in $[0,\infty)$ (in particular, functions in $H^1(0,\infty;\mathscr{X})$ which vanish at zero).

Let $\mathscr{X}$ be a Hilbert space. $\mathscr{X}'$ denotes the *dual space* of $\mathscr{X}$, which is the space of bounded linear functionals $\mathscr{X}$. We denote by $\langle x,x'\rangle_{\mathscr{X},\mathscr{X}'}$ the functional $x'\in\mathscr{X}'$ applied to $x\in\mathscr{X}$, so that $\langle x,x'\rangle_{\mathscr{X},\mathscr{X}'}$ is linear in x and antilinear in x' (similarly to the inner product on a Hilbert space). By the Riesz representation theorem [46, p. 48] the operator $J:\mathscr{X}\to\mathscr{X}'$ defined by

$$\langle z,Jx\rangle_{\mathscr{X},\mathscr{X}'}=\langle z,x\rangle_{\mathscr{X}}\,,\qquad\forall\, z,x\in\mathscr{X},\tag{1.12}$$

is an isometric isomorphism. If we do not distinguish between x and Jx in (1.12) (for all $x\in\mathscr{X}$), then we say that $\mathscr{X}$ is *identified* by its dual $\mathscr{X}'$. In this case we call $\mathscr{X}$ a *pivot space*.

Definition 1.1. Let $\mathscr{V}$ and $\mathscr{X}$ be Hilbert spaces such that $\mathscr{V}\subset\mathscr{X}$. The embedding $\mathscr{V}\subset\mathscr{X}$ is called continuous, if there exists an $m\geq 0$ such that

$$\|v\|_{\mathscr{X}}\leq m\,\|v\|_{\mathscr{V}}\,,\qquad\forall\, v\in\mathscr{V}.$$

Proposition 1.2. *[62, Proposition 2.9.2] Let $\mathscr{V}$ and $\mathscr{X}$ be Hilbert spaces with continuous and dense embedding $\mathscr{V}\subset\mathscr{X}$. Then, the function $\|\cdot\|_*:\mathscr{X}\to[0,\infty)$ defined by*

$$\|z\|_*:=\sup_{\varphi\in\mathscr{V},\ \|\varphi\|_{\mathscr{V}}\leq 1}|\langle z,\varphi\rangle_{\mathscr{X}}|,\qquad\forall\, z\in\mathscr{X},$$

is a norm on $\mathscr{X}$. In addition, let $\mathscr{V}^*$ denote the completion of $\mathscr{X}$ with respect to $\|\cdot\|_*$ and define the operator $J:\mathscr{V}^*\to\mathscr{V}'$ as follows: For any $z\in\mathscr{V}^*$,

$$\langle Jz,\varphi\rangle_{\mathscr{V}',\mathscr{V}}=\lim_{n\to\infty}\langle z_n,\varphi\rangle_{\mathscr{X}}\,,\qquad\forall\,\varphi\in\mathscr{V},\tag{1.13}$$

where (z_n) is a sequence in $\mathscr{X}$ such that $\lim_{n\to\infty}z_n=z$ in $\mathscr{V}^*$. Then J is an isomorphism from $\mathscr{V}^*$ to $\mathscr{V}'$.

If $\mathscr{V}$, $\mathscr{X}$, and $\mathscr{V}^*$ are as in Proposition 1.2, then we *identify* $\mathscr{V}^*$ with $\mathscr{V}'$, by not distinguishing between z and Jz (for all $z\in\mathscr{V}^*$). Thus, we have the continuous and dense embedding

$$\mathscr{V}\subset\mathscr{X}\subset\mathscr{V}'.$$

When $\mathscr{V}'$ is identified with $\mathscr{V}^*$ (as above), then we call $\mathscr{V}'$ the dual of $\mathscr{V}$ with respect to the pivot space $\mathscr{X}$. The norm $\|\cdot\|_*$ on $\mathscr{X}$ defined in Proposition 1.2 is called the *dual norm* of $\|\cdot\|_{\mathscr{V}}$ with respect to the pivot space $\mathscr{X}$.

Remark 1.3. $\mathscr{V}$ is uniquely determined by $\mathscr{V}'$: it consists of those $\varphi\in\mathscr{X}$ for which the product $\langle z,\varphi\rangle_{\mathscr{X}}$, regarded as a function of z, has a continuous extension to $\mathscr{V}'$ (see (1.13)).

The adjoint of $T\in\mathcal{L}(\mathscr{X},\mathscr{Y})$ is the operator $T^*\in\mathcal{L}(\mathscr{Y}',\mathscr{X}')$ defined by

$$\left(T^*y'\right)x=y'\left(Tx\right),\qquad\forall\, x\in\mathscr{X},\ \forall\, y'\in\mathscr{Y}'.\tag{1.14}$$

If both $\mathscr{X}$ and $\mathscr{Y}$ are considered to be pivot spaces, then $T^* \in \mathcal{L}(\mathscr{Y}, \mathscr{X})$ and (1.14) becomes

$$\langle x, T^* y\rangle_{\mathscr{Y}} = \langle Tx, y\rangle_{\mathscr{Y}}\,, \qquad \forall\, x \in \mathscr{X},\ \forall\, y \in \mathscr{Y}.$$

An operator $T \in \mathcal{L}(\mathscr{X})$ is called *self-adjoint*, if $T = T^*$. A self-adjoint operator T is called *positive* if

$$\langle x, Tx\rangle_{\mathscr{X}} \geq 0, \qquad \forall\, x \in \mathscr{X}. \tag{1.15}$$

Property (1.15) is written in the form $T \geq 0$. The notion of strict positivity ($T > 0$), negativity ($T \leq 0$), and strict negativity ($T < 0$) can be defined similarly. For two self-adjoint operators $T_1, T_2 \in \mathcal{L}(\mathscr{X})$, we say that $T_1 \geq T_2$, if $T_1 - T_2 \geq 0$. The square root of a nonnegative operator $T \in \mathcal{L}(\mathscr{X})$ is denoted by $T^{1/2}$.

A compact operator $T \in \mathcal{K}(\mathscr{X}, \mathscr{Y})$ admits a singular value decomposition

$$Tx = \sum_{i=1}^{\infty} \sigma_i \langle x, u_i\rangle_{\mathscr{X}} \cdot v_i,$$

where the sequence of singular values $(\sigma_i)_{i\in\mathbb{N}} \in [0, \infty)$ is monotonically decreasing and converges to zero as i tends to infinity. (u_i, v_i) is called the *Schmidt pair* associated to the singular value σ_i, where $(u_i)_{i\in\mathbb{N}}$ and $(v_i)_{i\in\mathbb{N}}$ are orthonormal systems in $\mathscr{X}$ and $\mathscr{Y}$, respectively [46, p. 203].

We close this section, by presenting the operators of *p-th Schatten class* and the associated relations that will be used in the upcoming chapters.

Definition 1.4 (p-th Schatten class)**.** Let $\mathscr{X}$ and $\mathscr{Y}$ be separable Hilbert spaces and let $p \in [1, \infty)$. Then $T \in \mathcal{K}(\mathscr{X}, \mathscr{Y})$ is called a p-th Schatten class operator, if the sequence of its singular values fulfills

$$(\sigma_i)_i \in \ell_p.$$

In this case we write $T \in \mathcal{S}_p(\mathscr{X}, \mathscr{Y})$. Furthermore, we abbreviate $\mathcal{S}_p(\mathscr{X}) := \mathcal{S}_p(\mathscr{X}, \mathscr{X})$. A p-th Schatten class operator is called *nuclear* if $p = 1$, and *Hilbert-Schmidt* if $p = 2$.

The space $\mathcal{S}_p(\mathscr{X}, \mathscr{Y})$ equipped with the norm $\|T\|_{\mathcal{S}_p(\mathscr{X},\mathscr{Y})} = \|(\sigma_i)_i\|_{\ell_p}$ is a Banach space. The trace of a nuclear operator $T \in \mathcal{S}_1(\mathscr{X})$ is given by the expression

$$\operatorname{trace}(T) = \sum_{i=1}^{\infty} \langle e_i, Te_i\rangle_{\mathscr{X}},$$

where (e_i) is an (arbitrary) orthonormal basis of $\mathscr{X}$ [46, p. 206]. For self-adjoint and nonnegative $T \in \mathcal{S}_1(\mathscr{X})$, the spectral theorem implies that

$$\|T\|_{\mathcal{S}_1(\mathscr{X})} = \operatorname{trace}(T).$$

In addition, if $T \in \mathcal{S}_2(\mathscr{X}, \mathscr{Y})$, then $T^*T \in \mathcal{S}_1(\mathscr{X})$ and $TT^* \in \mathcal{S}_1(\mathscr{Y})$ with

$$\|T\|^2_{\mathcal{S}_2(\mathscr{X},\mathscr{Y})} = \|T^*\|^2_{\mathcal{S}_2(\mathscr{Y},\mathscr{X})} = \operatorname{trace}(T^*T) = \operatorname{trace}(TT^*).$$

1.4 Outline of the dissertation

In Chapter 2, we define the class of well-posed linear systems and review those important properties, which we require to solve the linear-quadratic optimal control problem. In particular, we study *admissibility of control and observation operators* (Section 2.3) and *regularity* of the *transfer function* (Section 2.5). Afterwards, we focus on the linear-quadratic optimal control problem in Chapter 3. We start with the *regular optimal control problem* (Section 3.2) and show that the solution to this problem is connected to the *spectral factorization* of the associated *Popov function* (Section 3.3). Subsequently, we present various generalized Riccati equations (Section 3.4). Later on, we turn our focus to the *singular optimal control problem* (Section 3.5), in particular, we deal with the *bounded real* and the *positive real* case (Section 3.6). The main contribution of this work is presented in Chapters 4 and 5. We present two algorithms which provide approximate solutions to the linear-quadratic optimal control problem:

- The ADI method
- The Newton-Kleinman iteration

The ADI method will be applied to stable weakly regular linear systems, whereas for the Newton-Kleinman iteration, we require strong regularity of the well-posed linear system. We will further show that the ADI method can be extended to solve the singular optimal control problem in the bounded real and the positive real case. Finally, in Chapter 6, we apply the algorithms to a heat equation with Robin boundary control and boundary integral observation. We present three numerical examples to show the applicability of our algorithms. The first two examples deal with the application of the ADI method and the Newton-Kleinman iteration to solve the regular optimal control problem. In the last example, we apply the ADI method to solve the positive real optimal control problem.

Chapter 2

Well-posed linear systems

Die Mathematiker sind eine Art Franzosen; redet man mit ihnen, so übersetzen sie es in ihre Sprache, und dann ist es alsobald ganz etwas anderes. (Mathematicians are [like] a sort of Frenchmen; if you talk to them, they translate it into their own language, and then it is immediately something quite different.)

—Johann Wolfgang von Goethe

Contents

We review the theory of well-posed linear systems and (weak) regularity mainly from [58, 63–68, 70]. Since our foremost motive is to study the linear-quadratic optimal control problem for these systems, we gather just the required tools for the upcoming chapters. The notation that we use is mostly from [70] and [15]. For a complete overview on well-posed linear systems we refer to [58].

2.1 Well-posed linear systems: basic definitions

We start this section by introducing the essential notions which are necessary for the definition of well-posed linear systems. We begin with the definition of strongly continuous semigroups of operators on a Hilbert space $\mathscr{X}$. The following definitions are taken from [62].

Definition 2.1 (Strongly continuous semigroup)**.** A family of operators $\mathbb{T} = (\mathbb{T}_t)_{t\geq 0}$ in $\mathcal{L}(\mathscr{X})$ is a *strongly continuous semigroup* on $\mathscr{X}$ if

(a) $\mathbb{T}_{t+\tau} = \mathbb{T}_t\mathbb{T}_\tau$, for every $t, \tau \geq 0$ (the semigroup property),

(b) $\mathbb{T}_0 = I$,

(c) $\lim_{t\downarrow 0} \|\mathbb{T}_t x - x\| = 0$, for all $x \in \mathscr{X}$ (strong continuity).

Definition 2.2 (Growth bound of $\mathbb{T}$)**.** For the strongly continuous semigroup $\mathbb{T}$, the *growth bound* is defined by

$$\omega_0(\mathbb{T}) = \inf\{\omega \in \mathbb{R} \mid \exists\, M \geq 1 \text{ such that } \|\mathbb{T}_t\|_{\mathcal{L}(\mathscr{X})} \leq Me^{\omega t} \quad \forall\, t \geq 0\}.$$

Equivalently, the growth bound is the number defined by (see, e.g., [62, Proposition 2.1.2])

$$\omega_0(\mathbb{T}) = \inf_{t\in(0,\infty)} \frac{1}{t} \log \|\mathbb{T}_t\|.$$

Definition 2.3 (Generator of $\mathbb{T}$)**.** The linear operator $A : D(A) \subset \mathscr{X} \to \mathscr{X}$ defined by

$$\begin{aligned} D(A) &= \left\{ x \in \mathscr{X} \,\middle|\, \lim_{t\downarrow 0} \frac{\mathbb{T}_t x - x}{t} \text{ exists} \right\}, \\ Ax &= \lim_{t\downarrow 0} \frac{\mathbb{T}_t x - x}{t}, \qquad \forall\, x \in D(A), \end{aligned} \tag{2.1}$$

is called the *generator* of the semigroup $\mathbb{T}$.

We introduce some important notations that will be used in defining well-posed linear systems: For a Hilbert space $\mathscr{U}$, $\mathbf{S}_\tau$ denotes the *right shift* by some $\tau > 0$ on $L^2_{loc}(0,\infty;\mathscr{U})$, so that for all $u \in L^2_{loc}(0,\infty;\mathscr{U})$,

$$(\mathbf{S}_\tau u)(t) = \begin{cases} 0, & 0 \leq t < \tau, \\ u(t-\tau), & t \geq \tau. \end{cases}$$

In a similar way, $\mathbf{S}^*_\tau$ denotes the *left shift* by some $\tau > 0$ on $L^2_{loc}(0,\infty;\mathscr{U})$. In fact, if we restrict $\mathbf{S}_\tau$ and $\mathbf{S}^*_\tau$ from L^2_{loc} to L^2, then they are adjoint to each other.

$\mathbf{P}_\tau$ denotes the projection of $L^2_{loc}(0,\infty;\mathscr{U})$ into $L^2(0,\infty;\mathscr{U})$ by truncation, so that for all $u \in L^2_{loc}(0,\infty;\mathscr{U})$

$$(\mathbf{P}_\tau u)(t) = \begin{cases} u(t), & 0 \leq t < \tau, \\ 0, & t \geq \tau. \end{cases}$$

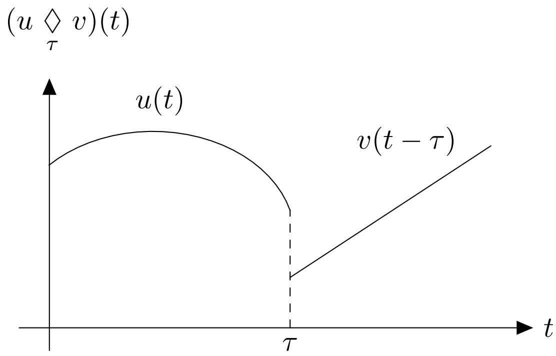


Figure 2.1: An example of the τ-concatenation.

It follows from the definitions of $\mathbf{S}_\tau$ and $\mathbf{P}_\tau$ that

$$\mathbf{S}_\tau^* \mathbf{S}_\tau = I, \qquad \mathbf{S}_\tau \mathbf{S}_\tau^* = I - \mathbf{P}_\tau. \tag{2.2}$$

For all $\tau \geq 0$, we use $\mathbf{S}_\tau$ and $\mathbf{P}_\tau$ to define the τ-*concatenation* of the functions $u, v \in L^2_{loc}(0, \infty; \mathscr{U})$ by

$$(u \underset{\tau}{\diamondsuit} v)(t) = \mathbf{P}_\tau u + \mathbf{S}_\tau v = \begin{cases} u(t), & t \in [0, \tau), \\ v(t-\tau), & \text{else.} \end{cases}$$

An example of the τ-concatenation is depicted in Figure 2.1.

Now we are ready to define the class of well-posed linear systems. We adapt the notation of [70, Definition 4.1] and [15, Definition 2.2].

Definition 2.4 (Well-posed linear system)**.** A well-posed linear system on the Hilbert spaces $\mathscr{U}$, $\mathscr{X}$, and $\mathscr{Y}$ is a quadruple $\Sigma = (\mathbb{T}, \Phi, \Psi, \mathbb{F})$ defined on $L^2(0, \infty; \mathscr{U})$, $\mathscr{X}$, and $L^2(0, \infty; \mathscr{Y})$, such that

(a) $\mathbb{T} = (\mathbb{T}_t)_{t \geq 0}$ is a strongly continuous semigroup of bounded linear operators on $\mathscr{X}$.

(b) $\Phi = (\Phi_t)_{t \geq 0}$ is a family of bounded linear operators from $L^2(0, \infty; \mathscr{U})$ to $\mathscr{X}$ such that

$$\Phi_{\tau+t}(u \underset{\tau}{\diamondsuit} v) = \mathbb{T}_t \Phi_\tau u + \Phi_t v, \tag{2.3}$$

for any $u, v \in L^2(0, \infty; \mathscr{U})$ and all $\tau, t \geq 0$.

(c) Ψ is a continuous linear operator from $\mathscr{X}$ to $L^2_{loc}(0, \infty; \mathscr{Y})$ such that for all $\tau \geq 0$, $\Psi_\tau = \mathbf{P}_\tau \Psi$ is in $\mathcal{L}(\mathscr{X}, L^2(0, \infty; \mathscr{Y}))$ and

$$\Psi x_0 = \Psi x_0 \underset{\tau}{\diamondsuit} \Psi \mathbb{T}_\tau x_0, \tag{2.4}$$

for any $x_0 \in \mathscr{X}$.

(d) $\mathbb{F}$ is a continuous linear operator from $L^2(0,\infty;\mathscr{U})$ to $L^2_{loc}(0,\infty;\mathscr{Y})$ such that for any $\tau \geq 0$, $\mathbb{F}_\tau = \mathbf{P}_\tau\mathbb{F}$ is in $\mathcal{L}(L^2(0,\infty;\mathscr{U}), L^2(0,\infty;\mathscr{Y}))$ and

$$\mathbb{F}(u \underset{\tau}{\diamondsuit} v) = \mathbb{F}u \underset{\tau}{\diamondsuit} (\Psi\Phi_\tau u + \mathbb{F}v), \tag{2.5}$$

for any $u, v \in L^2(0,\infty;\mathscr{U})$.

The different components of Σ are named as follows: $\mathscr{U}$ is the *input space*, $\mathscr{X}$ the *state space*, $\mathscr{Y}$ the *output space*, Φ the *input map*, Ψ the *(extended) output map*, and $\mathbb{F}$ the *(extended) input-output map*.

Remark 2.5. (a) It follows from (2.3) with $t = 0$ and $v = 0$ that

$$\Phi_\tau \mathbf{P}_\tau = \Phi_\tau, \qquad \forall\ \tau \geq 0. \tag{2.6}$$

Property (2.6) says that Φ is *causal*, i.e., the state does not depend on the future input. It follows from (2.6) and Definition 2.4 that for all $\tau, t \geq 0$,

$$\begin{aligned} \Phi_{\tau+t}\mathbf{P}_\tau &= \mathbb{T}_t\Phi_\tau, \\ \mathbf{P}_\tau\Psi_{\tau+t} &= \Psi_\tau, \\ \mathbf{P}_\tau\mathbb{F}\mathbf{P}_\tau &= \mathbf{P}_\tau\mathbb{F}. \end{aligned} \tag{2.7}$$

The last property in (2.7) says that $\mathbb{F}$ is causal, which means that the past output does not depend on the future input. Moreover, by taking $u = 0$ in (2.5) we obtain

$$\mathbf{S}_\tau\mathbb{F} = \mathbb{F}\mathbf{S}_\tau, \qquad \forall\ \tau \geq 0.$$

This means that $\mathbb{F}$ is *shift-invariant* or *time-invariant*.

(b) If $x_0 \in \mathscr{X}$ is the initial state of $\Sigma = (\mathbb{T}, \Phi, \Psi, \mathbb{F})$ and $u \in L^2_{loc}(0,\infty;\mathscr{U})$ its input function, then the *state trajectory* $x : [0,\infty) \to \mathscr{X}$ and the output function $y \in L^2_{loc}(0,\infty;\mathscr{Y})$ of Σ are defined by

$$\begin{aligned} x(t) &= \mathbb{T}_t x_0 + \Phi_t u, \qquad \forall\ t \geq 0, \\ y &= \Psi x_0 + \mathbb{F}u. \end{aligned} \tag{2.8}$$

(c) It follows from (2.5) that the input-output map $\mathbb{F}$ satisfies the following functional equation for all $\tau \geq 0$:

$$\mathbf{S}^*_\tau\mathbb{F} = \Psi\Phi_\tau + \mathbb{F}\mathbf{S}^*_\tau. \tag{2.9}$$

By combining (2.9) with (2.8) we get

$$\mathbf{S}^*_\tau y = \Psi x(\tau) + \mathbb{F}\mathbf{S}^*_\tau u. \tag{2.10}$$

Definition 2.6 (External, strong, and exponential stability)**.** A well-posed linear system is called

(a) *externally stable*, if it is output stable and input-output stable, i.e.,

$$\begin{aligned} &\Psi \in \mathcal{L}(\mathscr{X}, L^2(0,\infty;\mathscr{Y})), \\ &\mathbb{F} \in \mathcal{L}(L^2(0,\infty;\mathscr{U}), L^2(0,\infty;\mathscr{Y})). \end{aligned} \tag{2.11}$$

(b) *stable*, if the semigroup $\mathbb{T}$ is uniformly bounded, i.e.,

$$\sup_{t\geq 0} \|\mathbb{T}_t\| < \infty. \tag{2.12}$$

(c) *strongly stable*, if the semigroup $\mathbb{T}$ is strongly stable, i.e.,

$$\lim_{t\to\infty} \mathbb{T}_t x_0 = 0, \qquad \forall\, x_0 \in \mathscr{X}. \tag{2.13}$$

(d) *exponentially stable*, if the semigroup $\mathbb{T}$ is exponentially stable, i.e., there exists $M \geq 1$ and $\omega > 0$, such that

$$\|\mathbb{T}_t\| \leq Me^{-\omega t}, \qquad \forall\, t > 0. \tag{2.14}$$

2.2 The rigged spaces $\mathscr{X}_1$ and $\mathscr{X}_{-1}$

In this section we introduce the spaces $\mathscr{X}_1$ and $\mathscr{X}_{-1}$ which are fundamental in the theory of *unbounded control and observation operators* for well-posed linear systems. Most of the results presented in this section are from [62, Section 2.10] and we refer to [58, Section 3.6] for more details.

Definition 2.7. Let A be the generator of a strongly continuous semigroup $\mathbb{T}$ (cf. Definition 2.3) with non-empty resolvent set $\rho(A) \neq \emptyset$. For some $\beta \in \rho(A)$,

(a) the space $\mathscr{X}_1$ is defined as $D(A)$ equipped with the norm

$$\|x\|_1 := \|(\beta I - A)x\|, \qquad \forall\, x \in D(A). \tag{2.15}$$

(b) the space $\mathscr{X}_{-1}$ is defined as the completion of $\mathscr{X}$ with respect to the norm

$$\|x\|_{-1} := \left\|(\beta I - A)^{-1}x\right\|, \qquad \forall\, x \in \mathscr{X}. \tag{2.16}$$

Remark 2.8. It follows from [62, Proposition 2.10.1 & Proposition 2.10.2] that

(a) $\mathscr{X}_1$ and $\mathscr{X}_{-1}$ are Hilbert spaces.

(b) The choice of $\beta \in \rho(A)$ in (2.15) and (2.16) is not important, because different choices lead to equivalent norms on $\mathscr{X}_1$ and $\mathscr{X}_{-1}$.

(c) The norm on $\|\cdot\|_1$ is equivalent to the graph norm on $D(A)$ (as in (1.11)) and $\mathscr{X}_{-1}$ may be regarded as the dual of $D(A^*)$ with respect to the pivot space $\mathscr{X}$ (cf. Section 1.3).

(d) The Hilbert spaces $\mathscr{X}_1$ and $\mathscr{X}_{-1}$ satisfy the continuous and dense embeddings

$$\mathscr{X}_1 \subset \mathscr{X} \subset \mathscr{X}_{-1}.$$

Let $A : D(A) \subset \mathscr{X} \to \mathscr{X}$ be the generator of a strongly continuous semigroup $\mathbb{T}$ on $\mathscr{X}$. It follows from [62, Proposition 2.10.3] that $A \in \mathcal{L}(\mathscr{X}_1, \mathscr{X})$ and A has a unique extension $\widetilde{A} \in \mathcal{L}(\mathscr{X}, \mathscr{X}_{-1})$. Moreover, $(\beta I - A)^{-1} \in \mathcal{L}(\mathscr{X}, \mathscr{X}_1)$ and $(\beta I - \widetilde{A}) \in \mathcal{L}(\mathscr{X}, \mathscr{X}_{-1})$ are two unitary operators (in particular, $\beta \in \rho(\widetilde{A})$).

The semigroup $\mathbb{T}$ can be restricted to $\mathscr{X}_1$ and extended to $\mathscr{X}_{-1}$ using the unitary operators $(\beta I - A)^{-1} \in \mathcal{L}(\mathscr{X}, \mathscr{X}_1)$ and $(\beta I - \widetilde{A}) \in \mathcal{L}(\mathscr{X}, \mathscr{X}_{-1})$, respectively (see [62, Proposition 2.10.4]). The restricted operator $\mathbb{T}$ is a strongly continuous semigroup on $\mathscr{X}_1$, whose generator is the restriction of A to $D(A^2)$. The extended operator $\widetilde{\mathbb{T}}$ is a strongly continuous semigroup on $\mathscr{X}_{-1}$, whose generator is the extended operator $\widetilde{A} \in \mathcal{L}(\mathscr{X}, \mathscr{X}_{-1})$. For the rest of this dissertation we use the same notation as for the original operators, when we restrict or extend the semigroup $\mathbb{T}$ and its generator A to the spaces $\mathscr{X}_1$ and $\mathscr{X}_{-1}$.

In Definition 2.7 we may replace A with A^* and β with $\overline{\beta}$ to obtain the spaces $\mathscr{Z}_1$ and $\mathscr{Z}_{-1}$: The space $\mathscr{Z}_1$ is defined as $D(A^*)$ equipped with the norm $\|z\|_1 := \|(\overline{\beta} I - A^*)z\|$ for all $z \in D(A^*)$. The space $\mathscr{Z}_{-1}$ is defined as the completion of $\mathscr{X}$ with respect to the norm $\|z\|_{-1} := \|(\overline{\beta} I - A^*)^{-1} z\|$ for all $z \in \mathscr{X}$. For these spaces we obtain similar results as for $\mathscr{X}_1$ and $\mathscr{X}_{-1}$. In particular, the following continuous and dense embedding holds

$$\mathscr{Z}_1 \subset \mathscr{X} \subset \mathscr{Z}_{-1}.$$

Moreover, $\mathscr{Z}_1$ and $\mathscr{Z}_{-1}$ are Hilbert spaces. We will use these spaces to define the dual of a well-posed linear system in Section 2.6.

Remark 2.9. Let $\langle \cdot, \cdot \rangle_{\mathscr{X}}$ denote the inner product on the Hilbert space $\mathscr{X}$. It follows from Remark 1.3 that $\langle \cdot, \cdot \rangle_{\mathscr{X}}$ has the continuous extensions $\langle \cdot, \cdot \rangle_{\mathscr{X}_1, \mathscr{Z}_{-1}}$ and $\langle \cdot, \cdot \rangle_{\mathscr{Z}_1, \mathscr{X}_{-1}}$. The Hilbert space $\mathscr{Z}_{-1}$ can be regarded as the dual space of $\mathscr{X}_1$ and the Hilbert space $\mathscr{X}_{-1}$ may be regarded as the dual space of $\mathscr{Z}_1$.

2.3 Control and observation operators

In this section we define the *control and observation operators* of a well-posed linear system. As shown in [64] and [65], the existence of these operators is a consequence of Definition 2.4. We focus in particular on *admissible* control and observation operators, as they are necessary for the linear-quadratic optimal control problem (Chapter 3).

We start with the definition of a *strong solution* of a differential equation of the form $\dot{x}(t) = Ax(t) + f(t)$. The following definition and remark are from [62, Section 4.1]:

Definition 2.10. Consider the differential equation

$$\dot{x}(t) = Ax(t) + f(t), \tag{2.17}$$

where $f \in L^1_{loc}(0, \infty; \mathscr{X}_{-1})$. By a *strong solution* of the differential equation (2.17) in $\mathscr{X}_{-1}$ we mean the function

$$x \in L^1_{loc}(0, \infty; \mathscr{X}) \cap C\left(0, \infty; \mathscr{X}_{-1}\right),$$

which satisfies the following equation in $\mathscr{X}_{-1}$:

$$x(t) - x(0) = \int_0^t [Ax(\tau) + f(\tau)]\, d\tau, \qquad \forall\, t \geq 0. \tag{2.18}$$

Remark 2.11. It follows from Definition 2.10 that the state trajectory x as in (2.18) is absolutely continuous with values in $\mathscr{X}_{-1}$ and (2.17) holds for almost every $t \geq 0$, with the derivative computed with respect to the norm of $\mathscr{X}_{-1}$.

Let $A : D(A) \subset \mathscr{X} \to \mathscr{X}$ be the generator of a strongly continuous semigroup $\mathbb{T}$ on the Hilbert space $\mathscr{X}$ (cf. Definition 2.3). It follows from [64, Theorem 3.9] that assumptions (a) and (b) from Definition 2.4 imply the existence of a unique operator $B \in \mathcal{L}(\mathscr{U}, \mathscr{X}_{-1})$, called the *control operator* of Σ, such that for any $u \in L^2_{loc}(0, \infty; \mathscr{U})$ and all $t \geq 0$,

$$\Phi_t u = \int_0^t \mathbb{T}_{t-\tau} B u(\tau)\, d\tau. \tag{2.19}$$

Moreover, for every $x_0 \in \mathscr{X}$, for any $u \in L^2_{loc}(0, \infty; \mathscr{U})$, and for all $t \geq 0$, the state trajectory

$$x(t) = \mathbb{T}_t x_0 + \Phi_t u \tag{2.20}$$

is continuous in $\mathscr{X}$ and satisfies the differential equation

$$\dot{x}(t) = Ax(t) + Bu(t) \tag{2.21}$$

in the strong sense in $\mathscr{X}_{-1}$ (cf. Definition 2.10). The state trajectory x as in (2.20) is the unique solution of (2.21) which satisfies the initial condition $x(0) = x_0 \in \mathscr{X}$.

In formula (2.19) $\mathbb{T}$ acts on $\mathscr{X}_{-1}$ and the integration is carried out in $\mathscr{X}_{-1}$. We wish to know when the integral (2.19) is in $\mathscr{X}$ (a dense subspace of $\mathscr{X}_{-1}$). This motivates us to define the concept of an *admissible control operator*.

Definition 2.12. An operator $B \in \mathcal{L}(\mathscr{U}, \mathscr{X}_{-1})$ is called an *admissible control operator* for the semigroup $\mathbb{T}$, if for some (hence for any) $t > 0$, there holds

$$\Phi_t u = \int_0^t \mathbb{T}_{t-\tau} B u(\tau)\, d\tau \in \mathscr{X}, \qquad \forall\, u \in L^2(0, \infty; \mathscr{U}).$$

This means that the integral in (2.19) is in $\mathscr{X}$ for all $u \in L^2(0, \infty; \mathscr{U})$. If $B \in \mathcal{L}(\mathscr{U}, \mathscr{X})$, then it is called *bounded*, otherwise it is called *unbounded*.

If $B \in \mathcal{L}(\mathscr{U}, \mathscr{X}_{-1})$ is an admissible control operator for the semigroup $\mathbb{T}$, then by [62, Proposition 4.2.2], for every $t \geq 0$ we have

$$\Phi_t \in \mathcal{L}(L^2(0, \infty; \mathscr{U}), \mathscr{X}).$$

It follows from [65, Theorem 3.3] that assumptions (a) and (c) from Definition 2.4 imply the existence of a unique operator $C \in \mathcal{L}(\mathscr{X}_1, \mathscr{Y})$, called the *observation operator* of $\Sigma = (\mathbb{T}, \Phi, \Psi, \mathbb{F})$, such that for every $x_0 \in \mathscr{X}_1$ and all $t \geq 0$

$$(\Psi x_0)(t) = C\, \mathbb{T}_t x_0. \tag{2.22}$$

Formula (2.22) determines Ψ completely, because $\mathscr{X}_1$ is dense in $\mathscr{X}$.

Definition 2.13. An operator $C \in \mathcal{L}(\mathscr{X}_1, \mathscr{Y})$ is called an *admissible observation operator* for the semigroup $\mathbb{T}$, if for some (hence for every) $t > 0$, there exits a constant $K_t \geq 0$ such that

$$\int_0^t \|C\mathbb{T}_\tau x_0\|_{\mathscr{Y}}^2 \, d\tau \leq K_t^2 \|x_0\|_{\mathscr{X}}^2, \qquad \forall\, x_0 \in \mathscr{X}_1.$$

If $C \in \mathcal{L}(\mathscr{X}, \mathscr{Y})$, then it is called *bounded*, otherwise it is called *unbounded*.

It follows from Definition 2.13 that if $C \in \mathcal{L}(\mathscr{X}_1, \mathscr{Y})$ is the (unique) observation operator of a well-posed linear system Σ (see (2.22)), then C is admissible for the semigroup $\mathbb{T}$ of Σ.

Theorem 2.14. *[62, Theorem 4.3.7] Let $C \in \mathcal{L}(\mathscr{X}_1, \mathscr{Y})$ be an admissible observation operator for the semigroup $\mathbb{T}$ and let $\omega_0(\mathbb{T})$ be the growth bound of $\mathbb{T}$ (cf. Definition 2.2). Then, for every $\alpha > \omega_0(\mathbb{T})$ there exists $K_\alpha \geq 0$ such that*

$$\left\|C(sI - A)^{-1}\right\| \leq \frac{K_\alpha}{\sqrt{\mathrm{Re}(s) - \alpha}}, \qquad \forall\, s \in \mathbb{C}_\alpha. \tag{2.23}$$

We refer to [27, 28] for further details regarding admissible control and observation operators.

Definition 2.15. (see [17, Definition 2.7]) Let $\Sigma = (\mathbb{T}, \Phi, \Psi, \mathbb{F})$ be a well-posed linear system on $L^2(0, \infty; \mathscr{U})$, $\mathscr{X}$, and $L^2(0, \infty; \mathscr{Y})$. If A is the generator of $\mathbb{T}$ (cf. Definition 2.3), B is the control operator of Σ (see 2.19), and C is the observation operator of Σ (see 2.22), then we say that (A, B, C) is the *triple associated with* Σ. A triple of operators (A, B, C) is called *well-posed* if there is a well-posed linear system Σ such that (A, B, C) is the triple associated with Σ

Remark 2.16. So far we have seen how the operators $\mathbb{T}$, Φ, and Ψ of a well-posed linear system Σ can be represented by the triple (A, B, C) (we refer to [17] for more details). The formulas (2.19)–(2.21) and (2.22) resemble those from the finite-dimensional theory (cf. Section 1.1). The representation of $\mathbb{F}$ is more complicated and requires additional tools, mainly the *transfer function* and *regularity* which are defined in Sections 2.4 and 2.5, respectively.

In the last part of this section, we define the notions of *infinite-time admissibility* and *Gramian*. Furthermore, we present an important result on testing admissibility from [62]. We will use these concepts mainly in the development of the Newton-Kleinman iteration given in Chapter 5.

Definition 2.17. Let Φ be the input map given by (2.19). An operator $B \in \mathcal{L}(\mathscr{U}, \mathscr{X}_{-1})$ is called an *infinite-time admissible control operator* for the strongly continuous semigroup $\mathbb{T}$, if there exists a constant $K \geq 0$ such that

$$\|\Phi_t\|_{\mathcal{L}(L^2(0,\infty;\mathscr{U}),\mathscr{X})} \leq K, \qquad \forall\, t \geq 0. \tag{2.24}$$

Equation (2.24) means that for all $u \in L^2(0, \infty; \mathscr{U})$ we have

$$\Phi u := \int_0^\infty \mathbb{T}_t B u(t)\, dt \in \mathscr{X}. \tag{2.25}$$

Obviously, every infinite-time admissible control operator is an admissible control operator (cf. Definition 2.12). It follows from [62, Proposition 4.4.5] that if $\mathbb{T}$ is an exponentially stable semigroup and $B \in \mathcal{L}(\mathscr{U}, \mathscr{X}_{-1})$ is an admissible control operator for $\mathbb{T}$, then B is infinite-time admissible.

Definition 2.18. An operator $C \in \mathcal{L}(\mathscr{X}_1, \mathscr{Y})$ is called an *infinite-time admissible observation operator* for the strongly continuous semigroup $\mathbb{T}$, if there exists a constant $K > 0$ such that

$$\int_0^\infty \|C\mathbb{T}_t x_0\|_{\mathscr{Y}}^2 \, dt \leq K^2 \|x_0\|_{\mathscr{X}}^2, \qquad \forall\, x_0 \in D(A).$$

Equivalently, $C \in \mathcal{L}(\mathscr{X}_1, \mathscr{Y})$ is an *infinite-time admissible observation operator* for the semigroup $\mathbb{T}$, if the output map Ψ has a continuous extension to $\mathscr{X}$, i.e.,

$$\Psi \in \mathcal{L}(\mathscr{X}, L^2(0,\infty;\mathscr{Y})). \tag{2.26}$$

Remark 2.19. The output map of an externally stable well-posed linear system $\Sigma = (\mathbb{T}, \Phi, \Psi, \mathbb{F})$ (cf. Definition 2.6.a) satisfies (2.26). As a result, Σ has a unique observation operator $C \in \mathcal{L}(\mathscr{X}_1, \mathscr{Y})$ which is infinite-time admissible for the strongly continuous semigroup $\mathbb{T}$.

Definition 2.20 (Infinite-time controllability/observability Gramian)**.** Let $A : D(A) \subset \mathscr{X} \to \mathscr{X}$ be the generator of a strongly continuous semigroup $\mathbb{T}$.

(a) Let $B \in \mathcal{L}(\mathscr{U}, \mathscr{X}_{-1})$ be an infinite-time admissible control operator for $\mathbb{T}$ and let Φ be the extended input map defined in (2.25). We define the *infinite-time controllability Gramian* of (A, B) by

$$\Phi\Phi^* \in \mathcal{L}(\mathscr{X}).$$

(b) Let $C \in \mathcal{L}(\mathscr{X}_1, \mathscr{U})$ be an infinite-time admissible observation operator for $\mathbb{T}$ and let Ψ be the extended output map which satisfies (2.26). We define the *infinite-time observability Gramian* of (A, C) by

$$\Psi^*\Psi \in \mathcal{L}(\mathscr{X}).$$

In the following, we present a very important result which allows us to determine if an observation operator $C \in \mathcal{L}(\mathscr{X}_1, \mathscr{Y})$ is infinite-time admissible for the strongly continuous semigroup $\mathbb{T}$. This result links the infinite-time observability Gramian of (A, C) to a solution of the corresponding *observation Lyapunov equation*. The dual result can be also formulated using the infinite-time controllability Gramian and the corresponding *control Lyapunov equation*. For more details we refer to [62, Chapter 5] and [26].

Theorem 2.21. *[62, Theorem 5.1.1] Let $A : D(A) \subset \mathscr{X} \to \mathscr{X}$ be the generator of a strongly continuous semigroup $\mathbb{T}$ and let $C \in \mathcal{L}(\mathscr{X}_1, \mathscr{Y})$. In addition, let Ψ be the output map from (2.22). Then, the following four statements are equivalent:*

(i) C is an infinite-time admissible observation operator for the semigroup $\mathbb{T}$.

(ii) There exists an operator $\Gamma \in \mathcal{L}(\mathscr{X})$ such that

$$\Gamma x_0 = \Psi^* \Psi x_0 = \lim_{t \to \infty} \int_0^t \mathbb{T}_\tau^* C^* C \mathbb{T}_\tau x_0 \, d\tau, \qquad \forall\, x_0 \in D(A). \tag{2.27}$$

(iii) There exists an operator $\Pi \in \mathcal{L}(\mathscr{X})$, $\Pi \geq 0$, which satisfies the following equation

$$\langle A x_0, \Pi z_0 \rangle + \langle \Pi x_0, A z_0 \rangle = - \langle C x_0, C z_0 \rangle\,, \qquad \forall\, x_0, z_0 \in D(A). \tag{2.28}$$

(iv) There exists an operator $\Pi \in \mathcal{L}(\mathscr{X})$, $\Pi \geq 0$, which satisfies the inequality

$$\langle A x_0, \Pi z_0 \rangle + \langle \Pi x_0, A z_0 \rangle \leq - \langle C x_0, C z_0 \rangle\,, \qquad \forall\, x_0, z_0 \in D(A). \tag{2.29}$$

Moreover, if C is infinite-time admissible, then the following statements hold:

(1) $\Gamma = \Psi^*\Psi$ is the infinite-time observability Gramian of (A, C).

(2) $\Gamma = \Psi^*\Psi$ satisfies (2.28).

(3) $\Gamma = \Psi^*\Psi$ is the smallest positive solution of (2.29) [hence, also of (2.28)].

(4) We have

$$\lim_{t \to \infty} \Gamma^{\frac{1}{2}} \mathbb{T}_t x_0 = 0, \qquad \forall\, x_0 \in \mathscr{X}.$$

In particular, if Γ is strictly positive, then $\mathbb{T}$ is strongly stable.

(5) If $\mathbb{T}$ is strongly stable, then $\Gamma = \Psi^*\Psi$ is the unique self-adjoint solution of (2.28).

(6) If $\mathbb{T}$ is uniformly bounded and $\ker(\Gamma) = 0$, then $\mathbb{T}$ is weakly stable.

(2.28) is called a *Lyapunov equation* and (2.29) is called a *Lyapunov inequality*.

2.4 The transfer function

A well-posed linear system $\Sigma = (\mathbb{T}, \Phi, \Psi, \mathbb{F})$ can be described in the *frequency domain* by means of a *transfer function*. When we say “frequency domain description” we mean a description given in terms of *Laplace transformation* (Definition 2.23) of the original functions. In order to define the transfer function of a well-posed linear system, we need the *Hardy space* and the notion of Laplace transformation. These are given in the following definitions:

Definition 2.22 (Hardy spaces)**.** [58, Definition 10.3.1] Let $\mathscr{U}$ and $\mathscr{Y}$ be Hilbert spaces, let $p \in [1, \infty]$ and $\omega \in \mathbb{R}$. Let $\mathbb{C}_\omega$ denote the open right half-plane $\mathbb{C}_\omega := \{s \in \mathbb{C} \mid \mathrm{Re}(s) > \omega\}$.

- The space $\mathcal{H}^p_\omega(\mathscr{U})$ consists of all analytic $\mathscr{U}$-valued functions φ on $\mathbb{C}_\omega$, which satisfy $\|\varphi\|_{\mathcal{H}^p_\omega(\mathscr{U})} < \infty$, where

$$\|\varphi\|_{\mathcal{H}^p_\omega(\mathscr{U})} = \begin{cases} \sup\limits_{\alpha > \omega} \left(\int_{-\infty}^{\infty} \|\varphi(\alpha + i\beta)\|_{\mathscr{U}}^p \, d\beta \right)^{1/p}, & p \in [1, \infty), \\ \sup\limits_{\mathrm{Re}(s) > \omega} \|\varphi(s)\|_{\mathscr{U}}\,, & p = \infty. \end{cases}$$

  In the case $\omega = 0$, we abbreviate $\mathcal{H}^p(\mathscr{U}) := \mathcal{H}^p_0(\mathscr{U})$.

- The space $\mathcal{H}^p_\omega(\mathscr{U},\mathscr{Y})$ consists of all analytic $\mathcal{L}(\mathscr{U},\mathscr{Y})$-valued functions ψ on $\mathbb{C}_\omega$, which satisfy $\|\psi\|_{\mathcal{H}^p_\omega(\mathscr{U},\mathscr{Y})} < \infty$, where

$$\|\psi\|_{\mathcal{H}^p_\omega(\mathscr{U},\mathscr{Y})} = \begin{cases} \sup\limits_{\|u\|_{\mathscr{U}}\leq 1, \alpha>\omega} \left(\int_{-\infty}^{\infty} \|\psi(\alpha+i\beta)u\|^p_{\mathscr{Y}} \ d\beta\right)^{1/p}, & p\in[1,\infty), \\ \sup\limits_{\mathrm{Re}(s)>\omega} \|\psi(s)\|_{\mathcal{L}(\mathscr{U},\mathscr{Y})}, & p=\infty. \end{cases}$$

In the case $\omega = 0$, we abbreviate $\mathcal{H}^p(\mathscr{U},\mathscr{Y}) := \mathcal{H}^p_0(\mathscr{U},\mathscr{Y})$.

Definition 2.23 (Laplace transformation)**.** The *Laplace transform* of a function $u \in L^1_{\mathrm{loc}}(0,\infty)$ is given by

$$\hat{u}(s) = \int_0^\infty e^{-st}u(t)dt, \tag{2.30}$$

for all $s \in \mathbb{C}$ for which the integral (2.30) converges absolutely, i.e.,

$$\int_0^\infty e^{-t\mathrm{Re}(s)}|u(t)|dt < \infty.$$

Notation 2.24. For any Hilbert space $\mathscr{U}$, $L^2_\omega(0,\infty;\mathscr{U})$ is defined as

$$L^2_\omega(0,\infty;\mathscr{U}) := e_\omega L^2(0,\infty;\mathscr{U}), \qquad \text{where} \qquad (e_\omega v)(t) = e^{\omega t}v(t),$$

with the norm

$$\|u\|^2_\omega := \int_0^\infty e^{-2\omega t}\|u(t)\|^2 \ dt.$$

Theorem 2.25 (Paley–Wiener Theorem)**.** *[58, Theorem 10.3.4] Let $\mathscr{U}$ be a Hilbert space, and let $\omega \in \mathbb{R}$. Then the Laplace transform $\hat{u}$ of a function $u \in L^2_\omega(0,\infty;\mathscr{U})$ belongs to $\mathcal{H}^2_\omega(\mathscr{U})$. Conversely, every function $\varphi \in \mathcal{H}^2_\omega(\mathscr{U})$ is the Laplace transform of a function $u \in L^2_\omega(0,\infty;\mathscr{U})$. Moreover,*

$$\|\hat{u}\|_{\mathcal{H}^2_\omega(\mathscr{U})} = \sqrt{2\pi}\,\|u\|_\omega.$$

As shown in [68, Section 3], a shift-invariant operator on L^2 can be represented by a transfer function in the Hardy space $\mathcal{H}^\infty$ (cf. Definition 2.22). For the (shift-invariant) input-output map $\mathbb{F}$, we define the *growth bound* of $\mathbb{F}$ as

$$\gamma_{\mathbb{F}} = \inf\left\{\omega\in\mathbb{R} \mid \mathbb{F}\in\mathcal{L}(L^2_\omega(0,\infty;\mathscr{U}), L^2_\omega(0,\infty;\mathscr{Y}))\right\}. \tag{2.31}$$

If $\mathbb{F}$ has a growth bound which satisfies $\gamma_{\mathbb{F}} < \infty$, then there exists a transfer function representation of $\mathbb{F}$. In fact, the following theorem holds true.

Theorem 2.26. *(see [67, Theorem 3.3] and [68, Section 3]) Let $\mathscr{U}$ and $\mathscr{Y}$ be Hilbert spaces. Suppose $\mathbb{F}$ is a shift-invariant linear operator from $L^2_{loc}(0,\infty;\mathscr{U})$ to $L^2_{loc}(0,\infty;\mathscr{Y})$, with the growth bound $\gamma_{\mathbb{F}} < \infty$. Then, for all $\omega > \gamma_{\mathbb{F}}$, there exists a unique function $\mathbf{G} \in \mathcal{H}^\infty_\omega(\mathcal{L}(\mathscr{U},\mathscr{Y}))$ which satisfies*

$$\widehat{(\mathbb{F}u)}(s) = \mathbf{G}(s)\hat{u}(s),$$

for every $s\in\mathbb{C}_\omega$ and any $u \in L^2_\omega(0,\infty;\mathscr{U})$. Moreover, there holds

$$\|\mathbf{G}\|_{\mathcal{H}^\infty_\omega} = \|\mathbb{F}\|_{\mathcal{L}(L^2_\omega)}. \tag{2.32}$$

Definition 2.27 (Transfer function)**.** If $\mathbb{F}$ and $\mathbf{G}$ are as in Theorem 2.26, then $\mathbf{G}$ is called the transfer function of $\mathbb{F}$.

Let $\Sigma = (\mathbb{T}, \Phi, \Psi, \mathbb{F})$ be a well-posed linear system with the associated triple (A, B, C) (cf. Definition 2.15) and let $\mathbf{G}$ be the transfer function of $\mathbb{F}$. As shown in [56], for any $u \in H_0^1(0, \infty; \mathscr{U})$ and every $\beta \in \rho(A)$, the input-output map $\mathbb{F}$ is given by

$$(\mathbb{F}u)(t) = C\left[\int_0^t \mathbb{T}_{t-\tau} Bu(\tau)\, d\tau - (\beta I - A)^{-1} Bu(t)\right] + \mathbf{G}(\beta)u(t), \qquad \forall\, t \geq 0. \tag{2.33}$$

Equation (2.33) determines $\mathbb{F}$, because $H_0^1(0, \infty; \mathscr{U})$ is dense in $L^2_{loc}(0, \infty; \mathscr{U})$. The representation formula (2.33) shows that (A, B, C) determines $\mathbb{F}$ only up to an additive constant, namely $\mathbf{G}(\beta)$.

By [17, Theorem 4.2] if we substitute $s \in \rho(A)$ instead of β in (2.33), subtract the two equalities side by side, and eventually apply the resolvent identity, then we obtain

$$\begin{aligned} \mathbf{G}(s) - \mathbf{G}(\beta) &= (\beta - s)C(\beta I - A)^{-1}(sI - A)^{-1}B \\ &= C\left[(sI - A)^{-1} - (\beta I - A)^{-1}\right]B, \qquad \forall\, s, \beta \in \rho(A), \end{aligned} \tag{2.34}$$

which results in

$$\mathbf{G}'(s) = -C(sI - A)^{-2}B, \qquad \forall\, s \in \rho(A). \tag{2.35}$$

Equations (2.34) and (2.35) show that $\mathbf{G}$ is determined by the triple (A, B, C) up to an additive constant operator.

Our main goal in this section was solely to introduce the connection between the input-output map $\mathbb{F}$, the transfer function $\mathbf{G}$, and the triple (A, B, C). Hence, we have skipped many important aspects regarding the transfer function and refer to [68] for more details. At this point, we finish this section by providing an important result from [17] on the well-posedness of the triple (A, B, C).

Theorem 2.28. *[17, Theorem 5.1] Let $\mathscr{U}$, $\mathscr{X}$, $\mathscr{Y}$ be Hilbert spaces and let (A, B, C) be a triple of operators such that*

(i) A is the generator of a strongly continuous semigroup $\mathbb{T}$ on $\mathscr{X}$,

(ii) $B \in \mathcal{L}(\mathscr{U}, \mathscr{X}_{-1})$ is an admissible control operator for $\mathbb{T}$,

(iii) $C \in \mathcal{L}(\mathscr{X}_1, \mathscr{Y})$ is an admissible observation operator for $\mathbb{T}$,

(iv) there exists some $\alpha \in \mathbb{R}$ such that some (and hence any) solution $\mathbf{G} : \rho(A) \to \mathcal{L}(\mathscr{U}, \mathscr{Y})$ of equation (2.34) is bounded on $\mathbb{C}_\alpha$,

then (A, B, C) is well-posed (cf. Definition 2.15). This means that there exists a well-posed linear system $\Sigma = (\mathbb{T}, \Phi, \Psi, \mathbb{F})$ such that (A, B, C) is the triple associated with Σ.

2.5 Regular linear systems

Until now we have shown that the transfer function $\mathbf{G}$ can be determined by the triple (A, B, C) up to an additive constant operator. Similarly, the input-output map $\mathbb{F}$ can be represented by the triple (A, B, C) only up to an additive constant. This motivates us to define the concept of *regularity*, which allows us to further define a *feedthrough* operator $D \in \mathcal{L}(\mathscr{U}, \mathscr{Y})$. In addition, we show that the transfer function and the input-output map can be determined uniquely by the *generating operators* (A, B, C, D).

The following definition gives a characterization of *regularity* through the transfer function $\mathbf{G}$. For equivalent characterizations of regularity, we refer to [68].

Definition 2.29 (Regularity)**.** Let $\Sigma = (\mathbb{T}, \Phi, \Psi, \mathbb{F})$ be a well-posed linear system with transfer function $\mathbf{G}$. Let $\gamma_{\mathbb{F}}$ be the growth bound of the input-output map $\mathbb{F}$ as in (2.31) and choose $\omega \in \mathbb{R}$ with $\omega > \gamma_{\mathbb{F}}$. The system Σ (or its transfer function $\mathbf{G}$) is called

(a) *weakly regular*, if there exists an operator $D \in \mathcal{L}(\mathscr{U}, \mathscr{Y})$, such that for all $u \in \mathscr{U}$,

$$\lim_{\lambda \to \infty} \langle y, \mathbf{G}(\lambda) u \rangle_{\mathscr{Y}} = \langle y, Du \rangle_{\mathscr{Y}}, \qquad \lambda \in (\omega, \infty), \quad \forall\, y \in \mathscr{Y}. \tag{2.36}$$

(b) *(strongly) regular*, if there exists an operator $D \in \mathcal{L}(\mathscr{U}, \mathscr{Y})$, such that for all $u \in \mathscr{U}$

$$\lim_{\lambda \to \infty} \mathbf{G}(\lambda) u = Du \quad \text{in } \mathscr{Y}, \qquad \lambda \in (\omega, \infty). \tag{2.37}$$

(c) *line-regular*, if there exists an operator $D \in \mathcal{L}(\mathscr{U}, \mathscr{Y})$, such that for all $u \in \mathscr{U}$

$$\lim_{\mathrm{Re}(\lambda) \to \infty} \mathbf{G}(\lambda) u = Du \quad \text{in } \mathscr{Y}, \qquad \forall\, \lambda \in \mathbb{C}_\omega. \tag{2.38}$$

(d) *uniformly line-regular*, if there exists an operator $D \in \mathcal{L}(\mathscr{U}, \mathscr{Y})$, such that

$$\lim_{\mathrm{Re}(\lambda) \to \infty} \|\mathbf{G}(\lambda) - D\|_{\mathcal{L}(\mathscr{U}, \mathscr{Y})} = 0, \qquad \forall\, \lambda \in \mathbb{C}_\omega. \tag{2.39}$$

In all the above definitions, operator D is called the *feedthrough operator* of Σ.

Uniform line-regularity makes invertibility of the input-output map equivalent to invertibility of its feedthrough operator, as presented in the following proposition.

Proposition 2.30. *[39, Proposition 6.3.1] Let $\Sigma = (\mathbb{T}, \Phi, \Psi, \mathbb{F})$ be a uniformly line-regular linear system. Then the input-output map $\mathbb{F}$ is boundedly invertible if and only if the feedthrough operator $D \in \mathcal{L}(\mathscr{U}, \mathscr{Y})$ is boundedly invertible.*

Definition 2.31 (Λ-extension of $C \in \mathcal{L}(\mathscr{X}_1, \mathscr{Y})$)**.** Let $\mathscr{X}$ and $\mathscr{Y}$ be Hilbert spaces. Let $A : D(A) \subset \mathscr{X} \to \mathscr{X}$ generate a strongly continuous semigroup $\mathbb{T}$ on $\mathscr{X}$ and let $C \in \mathcal{L}(\mathscr{X}_1, \mathscr{Y})$.

(a) The *weak Λ-extension* of C, denoted by $C_{\Lambda w}$, is defined as

$$\langle y, C_{\Lambda w} x \rangle = \lim_{\substack{\lambda \to \infty \\ \lambda \in \mathbb{R}}} \left\langle y, C\lambda\,(\lambda I - A)^{-1}\, x \right\rangle, \qquad \forall\, y \in \mathscr{Y}, \tag{2.40}$$

with its domain $D(C_{\Lambda w})$ consisting of those $x \in \mathscr{X}$ for which the limit in (2.40) exists.

(b) The *(strong)* Λ-*extension* of C, denoted by C_Λ, is defined as

$$C_\Lambda x = \lim_{\substack{\lambda\to\infty \\ \lambda\in\mathbb{R}}} C\lambda\left(\lambda I - A\right)^{-1} x \quad \text{in } \mathscr{Y}, \tag{2.41}$$

with its domain $D(C_\Lambda)$ consisting of those $x \in \mathscr{X}$ for which the limit in (2.41) exists.

For weakly regular linear systems, the formulas for the transfer function $\mathbf{G}$ and the input-output map $\mathbb{F}$ look much the same as those from finite-dimensional theory, with C replaced by $C_{\Lambda w}$ (or C_Λ if the system is strongly regular). In fact, the following theorem gives the desired representations.

Proposition 2.32. *[15, Theorem 2.6]. Let $\mathscr{U}$, $\mathscr{X}$, $\mathscr{Y}$ be Hilbert spaces. Let $\Sigma = (\mathbb{T}, \Phi, \Psi, \mathbb{F})$ be a weakly regular linear system on $L^2(0,\infty;\mathscr{U})$, $\mathscr{X}$, and $L^2(0,\infty;\mathscr{Y})$, with semigroup generator A, control operator B, observation operator C, transfer function $\mathbf{G}$, and feedthrough operator D. Then the following holds:*

1. *For $s \in \mathbb{C}$ with $\mathrm{Re}(s) > \omega_0(\mathbb{T})$, there holds*

$$\mathbf{G}(s) = C_{\Lambda w}\left(sI - A\right)^{-1} B + D.$$

In particular, we have

$$(sI - A)^{-1} B\mathscr{U} \subset D(C_{\Lambda w}). \tag{2.42}$$

2. *$\mathbb{F} : L^2_{loc}(0,\infty;\mathscr{U}) \to L^2_{loc}(0,\infty;\mathscr{Y})$ is given by*

$$\left(\mathbb{F}u\right)(t) = C_{\Lambda w}\int_0^t \mathbb{T}_{t-\sigma} Bu(\sigma) + Du(t), \tag{2.43}$$

for almost all $t \geq 0$.

3. *Let x and y be respectively the state trajectory and the output function of Σ, which are given by (2.8). Then, for almost all $t \geq 0$*

$$x(t) \in D(C_{\Lambda w}), \tag{2.44}$$

and

$$y(t) = C_{\Lambda w} x(t) + Du(t). \tag{2.45}$$

If Σ is strongly regular, then $C_{\Lambda w}$ can be replaced by C_Λ in (2.44) and (2.45).

Remark 2.33. It follows from Proposition 2.32 and equation (2.21) that the weakly regular linear system $\Sigma = (\mathbb{T}, \Phi, \Psi, \mathbb{F})$ is completely determined via

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t), \\ y(t) &= C_{\Lambda w} x(t) + Du(t). \end{aligned} \tag{2.46}$$

Definition 2.34. The quadruple of operators (A, B, C, D) satisfying (2.46) is called the *generator* of Σ (compare with Definition 2.15).

2.6 The dual of a well-posed linear system

In this section we define the dual system and its anticausal version. Dual systems play an important role in the linear-quadratic optimal control problem, which will be treated in the upcoming chapter. An important feature of the duality transformation is that it preserves weak regularity [61]. The following proposition is from [70] (stated without proof). We refer to [61, Section 3] for a proof of this proposition.

Proposition 2.35. *[70, Proposition 6.1] Let $\Sigma = (\mathbb{T}, \Phi, \Psi, \mathbb{F})$ be a well-posed linear system with semigroup generator A, control operator B, observation operator C, and transfer function $\mathbf{G}$. Let $\mathscr{Z}_1$ and $\mathscr{Z}_{-1}$ be the rigged spaces defined in Section 2.2 (by replacing A with A^* and β with $\overline{\beta}$ in Definition 2.7). Then there exists a unique well-posed linear system $\Sigma^d = (\mathbb{T}^*, \Phi^d, \Psi^d, \mathbb{F}^d)$, called the* dual system *of Σ, such that*

(i) $A^* \in \mathcal{L}(\mathscr{Z}_1, \mathscr{X})$ generates the strongly continuous semigroup $\mathbb{T}^*$

(ii) $C^* \in \mathcal{L}(\mathscr{Y}, \mathscr{Z}_{-1})$ is the control operator of Σ^d

(iii) $B^* \in \mathcal{L}(\mathscr{Z}_1, \mathscr{U})$ is the observation operator of Σ^d

(iv) $\mathbf{G}^d(s) = \mathbf{G}(\overline{s})^*$ is the transfer function of Σ^d

Remark 2.36. It follows from [61, Proposition 3.7] that if the well-posed linear system Σ is weakly regular with feedthrough operator D, then its dual system Σ^d is also weakly regular with feedthrough operator D^*.

For the dual system Σ^d, let y^d denote its input function, x_0^d its initial state, $x^d(t)$ its state trajectory at time $t \geq 0$, and u^d its output function. Then, we have

$$\begin{aligned} x^d(t) &= \mathbb{T}_t^* x_0^d + \Phi_t^d y^d, \\ u^d &= \Psi^d x_0^d + \mathbb{F}^d y^d. \end{aligned}$$

If the dual system $\Sigma^d = (\mathbb{T}^*, \Phi^d, \Psi^d, \mathbb{F}^d)$ is weakly regular, then we can determine Σ^d completely via its *generating operators* (A^*, C^*, B^*, D^*). To this end, we define the (weak) Λ-extension of $B^* \in \mathcal{L}(\mathscr{Z}_1, \mathscr{U})$ (compare with Definition 2.31):

Definition 2.37 (Λ-extension of $B^* \in \mathcal{L}(\mathscr{Z}_1, \mathscr{U})$)**.** Let $\mathscr{X}$ and $\mathscr{U}$ be Hilbert spaces. Let $A : D(A) \subset \mathscr{X} \to \mathscr{X}$ generate a strongly continuous semigroup $\mathbb{T}$ on $\mathscr{X}$ and let $B^* \in \mathcal{L}(\mathscr{Z}_1, \mathscr{U})$.

(a) The *weak Λ-extension* of B^*, denoted by $B^*_{\Lambda w}$, is defined as

$$\langle u, B^*_{\Lambda w} x\rangle = \lim_{\lambda\to\infty} \left\langle u, B\lambda \left(\lambda I - A^*\right)^{-1} x \right\rangle, \qquad \forall\, u \in \mathscr{U}, \tag{2.47}$$

with its domain $D(B^*_{\Lambda w})$ consisting of those $x \in \mathscr{X}$ for which the limit in (2.47) exists.

(b) The *(strong) Λ-extension* of B^*, denoted by B^*_{Λ}, is defined as

$$B^*_{\Lambda} x = \lim_{\lambda\to\infty} B\lambda \left(\lambda I - A^*\right)^{-1} x \quad \text{in } \mathscr{U}, \tag{2.48}$$

with its domain $D(B^*_{\Lambda})$ consisting of those $x \in \mathscr{X}$ for which the limit in (2.48) exists.

If the dual system Σ^d is weakly regular, then it follows from Proposition 2.32 that for almost all $t \geq 0$

$$\begin{aligned} \dot{x}^d(t) &= A^* x^d(t) + C^* y^d(t), \\ u^d(t) &= B^*_{\Lambda w} x^d(t) + D^* y^d(t). \end{aligned} \tag{2.49}$$

The transfer function of the weakly regular linear system Σ^d is determined by

$$\mathbf{G}^d(s) = B^*_{\Lambda w}(sI - A^*)^{-1} C^* + D^*.$$

The following proposition from [70] shows the duality between infinite-time admissibility of control and observation operators:

Proposition 2.38. *[70, Proposition 6.2] The following two statements are equivalent:*

(i) C is an infinite-time admissible observation operator for the semigroup $\mathbb{T}$.

(ii) C^* is an infinite-time admissible control operator for the semigroup $\mathbb{T}^*$.

Moreover, if C is infinite-time admissible, then

$$\Psi^* w = \lim_{T \to \infty} \int_0^T \mathbb{T}^*_\tau C^* w(\tau) d\tau =: \Phi^d w, \qquad \forall\, w \in \mathscr{Y},$$

where Φ^d for Σ^d is the analogue of Φ for Σ (see (2.25)).

Theorem 2.39 gives a description of operator $\mathbb{F}^*$, which will be used frequently in Chapters 4 and 5. In addition, this theorem is an important ingredient for the *anticausal interpretation* of dual system Σ^d, which will be presented in the last part of this section.

Theorem 2.39. *[70, Theorem 6.3] Let $\Sigma = (\mathbb{T}, \Phi, \Psi, \mathbb{F})$ be an externally stable weakly regular linear system with generating operators (A, B, C, D) (cf. Definition 2.6.a). Let $w \in L^2(0, \infty; \mathscr{Y})$ and define the function $q : [0, \infty) \to \mathscr{X}$ by*

$$q(t) = \Psi^* \mathbf{S}^*_t w = \lim_{T \to \infty} \int_t^T \mathbb{T}^*_{\tau - t} C^* w(\tau)\, d\tau.$$

Then $q(t) \in D(B^*_{\Lambda w})$ for almost every $t \geq 0$ and there holds

$$(\mathbb{F}^* w)(t) = B^*_{\Lambda w} q(t) + D^* w(t), \quad \textit{for almost every} \quad t \geq 0. \tag{2.50}$$

Remark 2.40. (a) From Remark 2.36 we know that the dual of a weakly regular linear system is also weakly regular. But this is not the case for strongly regular linear systems (see [59, Example 8.1]).

(b) If the output space $\mathscr{Y}$ is finite-dimensional, then weak regularity equals strong regularity. In this case, if Σ is strongly regular, then its dual Σ^d is also strongly regular.

We close this section by giving the *anticausal interpretation* of dual system Σ^d. We need this interpretation to show the feasibility of Newton-Kleinman iteration in Section 5.2.

Definition 2.41. The state trajectory $x^a : [0,\infty) \to \mathscr{X}$ and the output function $u^a \in L^2(0,\infty;\mathscr{U})$ of the *anticausal dual system* Σ^a corresponding to the input function $y^a \in L^2(0,\infty;\mathscr{Y})$ are given by

$$\Sigma^a := \begin{cases} x^a(t) = \Psi^* \mathbf{S}_t^* y^a, & \forall\, t \geq 0. \\ u^a = \mathbb{F}^* y^a. \end{cases} \tag{2.51}$$

Remark 2.42. It follows from (2.51), together with Theorem 2.39 and [70, Proposition 5.2], that

$$\lim_{t\to\infty} x^a(t) = 0,$$

and the functions y^a, x^a, and u^a satisfy

$$\begin{aligned} -\dot{x}^a(t) &= A^* x^a(t) + C^* y^a(t), \\ u^a(t) &= B^*_{\Lambda w} x^a(t) + D^* y^a(t), \end{aligned} \tag{2.52}$$

for almost every $t \geq 0$.

2.7 Linear output feedback theory

In this section we focus on closed-loop systems obtained by imposing the *output feedback law*

$$u = Ky + v,$$

with the feedback operator $K \in \mathcal{L}(\mathscr{Y},\mathscr{U})$ and the new input function v. The closed-loop system, denoted by Σ^K, may not be necessarily well-posed [67]. This motivates us to introduce the concept of *admissible output feedback*:

Definition 2.43. Let $\Sigma = (\mathbb{T},\Phi,\Psi,\mathbb{F})$ be an externally stable well-posed linear system (cf. Definition 2.6.a). Operator $K \in \mathcal{L}(\mathscr{Y},\mathscr{U})$ is called an *admissible output feedback* for Σ, if the time-invariant operator $I - K\mathbb{F} \in \mathcal{L}(L^2(0,\infty;\mathscr{U}))$ has a bounded inverse (equivalently, if the time-invariant operator $I - \mathbb{F}K \in \mathcal{L}(L^2(0,\infty;\mathscr{Y}))$ has a bounded inverse).

Using Laplace transformation of the input-output map $\mathbb{F}$ (cf. Section 2.4), we obtain the following equivalent characterization of admissible output feedback (see also [67]):

Proposition 2.44. *Let Σ be a well-posed linear system with the transfer function $\mathbf{G}$. Operator $K \in \mathcal{L}(\mathscr{Y},\mathscr{U})$ is an* admissible output feedback *for Σ, if the function $I - \mathbf{G}K$ is boundedly invertible on some right half-plane (equivalently, if the function $I - K\mathbf{G}$ is boundedly invertible on some right half-plane).*

An immediate consequence of Definition 2.43 is as follows: By applying an admissible output feedback $u = Ky + v$ to the externally stable well-posed linear system $\Sigma = (\mathbb{T},\Phi,\Psi,\mathbb{F})$, we obtain the following set of equations

$$\begin{aligned} x(t) &= \mathbb{T}_t x_0 + \Phi_t u, \\ y &= \Psi x_0 + \mathbb{F} u, \\ u &= Ky + v. \end{aligned}$$

Since $I - K\mathbb{F}$ is boundedly invertible, the above set of equations is uniquely solvable with respect to the initial state $x_0 \in \mathscr{X}$ and the new input function $v \in L^2(0, \infty; \mathscr{U})$. As a result, we obtain

$$\begin{aligned} x(t) &= \left(\mathbb{T}_t + \Phi_t K (I - \mathbb{F}K)^{-1}\Psi\right) x_0 + \Phi_t \left(I - K\mathbb{F}\right)^{-1} v, \\ y &= \left(I - \mathbb{F}K\right)^{-1} \left(\Psi x_0 + \mathbb{F}v\right), \\ u &= \left(I - K\mathbb{F}\right)^{-1} \left(K\Psi x_0 + v\right). \end{aligned} \tag{2.53}$$

In the above equations we have used the equalities

$$\begin{aligned} \left(I - \mathbb{F}K\right)^{-1} &= I + \mathbb{F}\left(I - K\mathbb{F}\right)^{-1} K, \\ \left(I - K\mathbb{F}\right)^{-1} &= I + K\left(I - \mathbb{F}K\right)^{-1} \mathbb{F}, \\ \left(I - \mathbb{F}K\right)^{-1} \mathbb{F} &= \mathbb{F}\left(I - K\mathbb{F}\right)^{-1}, \\ \left(I - K\mathbb{F}\right)^{-1} K &= K\left(I - \mathbb{F}K\right)^{-1}. \end{aligned}$$

Remark 2.45. It follows from [57, Proposition 20] that $\Sigma^K = (\mathbb{T}^K, \Phi^K, \Psi^K, \mathbb{F}^K)$ with

$$\begin{bmatrix} \mathbb{T}_t^K & \Phi_t^K \\ \Psi^K & \mathbb{F}^K \end{bmatrix} = \begin{bmatrix} \mathbb{T}_t + \Phi_t K (I - \mathbb{F}K)^{-1}\Psi & \Phi_t \left(I - K\mathbb{F}\right)^{-1} \\ (I - \mathbb{F}K)^{-1}\Psi & (I - \mathbb{F}K)^{-1}\mathbb{F} \end{bmatrix}$$

is also an externally stable well-posed linear system. As shown in [57, Lemma 21], stability is preserved under admissible output feedback. In fact, the following result holds:

- Σ^K is strongly stable, if and only if, Σ is strongly stable.
- Σ^K is exponentially stable, if and only if, Σ is exponentially stable.

The following lemma from [39] is an important ingredient in developing the theory of Chapters 3 and 5 (see Remark 3.6.b as well as the proof of Theorems 3.11 and 5.3).

Lemma 2.46. *[39, Lemma 6.6.7] Let $\Sigma = (\mathbb{T}, \Phi, \Psi, \mathbb{F})$ be a well-posed linear system on $L^2(0, \infty; \mathscr{U})$, $\mathscr{X}$, and $L^2(0, \infty; \mathscr{Y})$. Let $\widetilde{\Psi} \in \mathcal{L}(\mathscr{X}, L^2(0, \infty; \mathscr{Y}))$ be another admissible output map such that $\widetilde{\mathbb{T}} := \mathbb{T} + \Phi\widetilde{\Psi}$ is a strongly continuous semigroup.*

- *If $\mathbb{T}$ is strongly stable, then $\widetilde{\mathbb{T}}$ is also strongly stable.*
- *If $\mathbb{T}$ is exponentially stable, then $\widetilde{\mathbb{T}}$ is also exponentially stable.*

For the rest of this section we assume that $\Sigma = (\mathbb{T}, \Phi, \Psi, \mathbb{F})$ is a strongly regular linear system. The main reason why we focus on the notion of strong regularity is the feedback theory from [67]. This theory has essential parts that can not be extended to weakly regular systems. In particular, weak regularity is not preserved under feedback (see for example [58, Remark 7.5.4]).

We denote the generating operators of the closed-loop system $\Sigma^K = (\mathbb{T}^K, \Phi^K, \Psi^K, \mathbb{F}^K)$ by (A^K, B^K, C^K, D^K). The following theorem gives a characterization of these operators:

Theorem 2.47. *[67, Proposition 7.1, Theorem 7.2] Let $K \in \mathcal{L}(\mathscr{Y},\mathscr{U})$ be an admissible output feedback operator and assume that Σ is a strongly regular linear system with feedthrough operator $D \in \mathcal{L}(\mathscr{U},\mathscr{Y})$. Then $I - DK$ (and hence also $I - KD$) is left invertible.*
The closed-loop system Σ^K is strongly regular, if and only if, $I - DK$ (and hence also $I - KD$) is invertible. In this case we have

$$A^K x_0 = \left[A + BK\left(I - DK\right)^{-1} C_\Lambda\right] x_0, \tag{2.54}$$

$$C^K x_0 = \left(I - DK\right)^{-1} C_\Lambda x_0, \tag{2.55}$$

for all $x_0 \in D(A^K)$, where

$$D(A^K) = \left\{x_0 \in D(C_\Lambda) \mid \left(A + BK(I - DK)^{-1}C_\Lambda\right) x_0 \in \mathscr{X}\right\}.$$

Moreover, there holds

$$D(C_\Lambda^K) = D(C_\Lambda), \qquad C_\Lambda^K = (I - DK)^{-1} C_\Lambda.$$

Regarding the operators B^K and D^K we have

$$B^K = B(I - KD)^{-1},$$
$$D^K = D(I - KD)^{-1} = (I - DK)^{-1} D.$$

Remark 2.48. Equation (2.54) can be understood as a perturbation of the semigroup generator A by operator $BK\left(I - DK\right)^{-1} C_\Lambda$. With the assumptions of Theorem 2.47, let $\mathbb{T}^K$ denote the strongly continuous semigroup generated by $A^K = A + BK\left(I - DK\right)^{-1} C_\Lambda$. It follows from [67, Theorem 6.1] that the *perturbation relationship*

$$\mathbb{T}_t^K x_0 = \mathbb{T}_t x_0 + \int_0^\infty \mathbb{T}_{t-\tau} BK\left(I - DK\right)^{-1} C_\Lambda \mathbb{T}_\tau^K x_0 \, d\tau \tag{2.56}$$

holds for every $x_0 \in D(A^K)$ and all $t \geq 0$. We will use the relationship (2.56) in Section 5.3 to construct a connection between the *Newton-Kleinman iteration* and the *Riccati operator*.

We finish this section by showing that (2.54) can be extended to $D(C_\Lambda)$, as proven in [58, Theorem 7.5.1 & Theorem 7.5.3] and [67, Proposition 7.10]. Let $\lambda_0 \in \mathbb{R}$ be such that $[\lambda_0, \infty) \subset \rho(A)$. We define the space $\mathscr{W}_1$ as $D(C_\Lambda)$ equipped with the norm

$$\|x_0\|_{\mathscr{W}_1} := \|x_0\|_{\mathscr{X}} + \sup_{\lambda \geq \lambda_0} \left\|C\lambda(\lambda I - A)^{-1} x_0\right\|_{\mathscr{Y}}. \tag{2.57}$$

It follows from [67, Proposition 5.3] that $\mathscr{W}_1$ is a Hilbert space, $C_\Lambda \in \mathcal{L}(\mathscr{W}_1, \mathscr{Y})$, and there holds

$$\mathscr{X}_1 \subset \mathscr{W}_1 \subset \mathscr{X}, \tag{2.58}$$

with continuous embeddings. Furthermore, for some $\beta \in \rho(A)$, we define the Hilbert space $\mathscr{W}$ by

$$\mathscr{W} = (\beta I - A)\mathscr{W}_1,$$

with the norm

$$\|x_0\|_{\mathscr{W}} := \left\|(\beta I - A)^{-1} x_0\right\|_{\mathscr{W}_1}. \tag{2.59}$$

It follows from the above definition that $(\beta I - A)^{-1}$ is an isomorphism from $\mathscr{W}$ to $\mathscr{W}_1$, just as it is an isomorphism from $\mathscr{X}_{-1}$ to $\mathscr{X}$ and from $\mathscr{X}$ to $\mathscr{X}_1$. As a result, by (2.58) we obtain

$$\mathscr{X} \subset \mathscr{W} \subset \mathscr{X}_{-1},$$

with continuous embeddings. The main advantage of introducing the Hilbert spaces $\mathscr{W}_1$ and $\mathscr{W}$ is that

$$A, A^K \in \mathcal{L}(\mathscr{W}_1, \mathscr{W}),$$

and

$$A^K x_0 = \left[A + BK\,(I - DK)^{-1}\,C_\Lambda\right] x_0, \qquad \forall\; x_0 \in \mathscr{W}_1,$$

as proven in [58, Theorem 7.5.1 & Theorem 7.5.3].

Chapter 3

The linear-quadratic optimal control problem

Crush your fears like stone turns into dust. Then water that dust and make cement. With that cement: build an empire.

—Jack Canfield

Contents

We focus on the linear-quadratic optimal control problem for externally stable well-posed linear systems. We study two cases: the regular optimal control problem, where the associated Popov function is strictly positive (bounded from below) and the singular optimal control problem, where the Popov function is just positive. In the regular case, we show that the solution to the optimal control problem is connected to the spectral factorization of the associated Popov function. Subsequently, we present various generalized Riccati equations. Later on, we turn our focus to the singular optimal control problem, in particular, we deal with the bounded real and positive real case. We note that in the majority of the results presented in this chapter, we do not put any (weak) regularity assumption on the linear systems (in the sense of Section 2.5). The only regularity assumption is on the spectral factorization.

3.1 The quadratic cost functional and the Popov operator

Let $\Sigma = (\mathbb{T}, \Phi, \Psi, \mathbb{F})$ be an externally stable well-posed linear system (cf. Definition 2.6.a). Let $x : [0, \infty) \to \mathscr{X}$ and $y \in L^2(0, \infty; \mathscr{Y})$ be respectively the state trajectory and the output function of Σ corresponding to the initial state $x_0 \in \mathscr{X}$ and the input function $u \in L^2(0, \infty; \mathscr{U})$. This means that there holds

$$\begin{aligned} x(t) &= \mathbb{T}_t x_0 + \Phi_t u, \qquad \forall\, t \geq 0, \\ y &= \Psi x_0 + \mathbb{F} u. \end{aligned} \tag{3.1}$$

We consider the following quadratic *cost functional*

$$J(u, x_0) = \int_0^\infty \left\langle \begin{pmatrix} y(\tau) \\ u(\tau) \end{pmatrix}, \begin{bmatrix} Q & N^* \\ N & R \end{bmatrix} \begin{pmatrix} y(\tau) \\ u(\tau) \end{pmatrix} \right\rangle_{\mathscr{Y} \times \mathscr{U}} d\tau, \tag{3.2}$$

where $R = R^* \in \mathcal{L}(\mathscr{U})$, $Q = Q^* \in \mathcal{L}(\mathscr{Y})$, and $N \in \mathcal{L}(\mathscr{Y}, \mathscr{U})$. This cost functional can be understood as the cost of observing and controlling the system. The linear-quadratic optimal control problem is to find an input function $u^{\text{opt}} \in L^2(0, \infty; \mathscr{U})$ which minimizes (3.2) subject to (3.1).

We can transform our constrained optimization problem to an unconstrained one by substituting $y = \Psi x_0 + \mathbb{F} u$ into (3.2) to obtain

$$J(u, x_0) = \left\langle \begin{pmatrix} x_0 \\ u \end{pmatrix}, \begin{pmatrix} \Psi^* Q \Psi & \Psi^*(Q\mathbb{F} + N^*) \\ (\mathbb{F}^* Q + N)\Psi & \mathscr{R} \end{pmatrix} \begin{pmatrix} x_0 \\ u \end{pmatrix} \right\rangle_{L^2(0,\infty; \mathscr{Y} \times \mathscr{U})}, \tag{3.3}$$

where $\mathscr{R} \in \mathcal{L}(L^2(0, \infty; \mathscr{U}))$ is called the *Popov operator* and is defined by

$$\mathscr{R} = R + N\mathbb{F} + \mathbb{F}^* N^* + \mathbb{F}^* Q \mathbb{F}. \tag{3.4}$$

We consider two main cases: the *regular* and the *singular* optimal control problem. In the regular case, we assume that the Popov operator $\mathscr{R}$ is *coercive*. This means that there exists some $\varepsilon > 0$ such that

$$\langle u, \mathscr{R} u \rangle_{L^2(0,\infty;\mathscr{U})} \geq \varepsilon \, \|u\|_{L^2(0,\infty;\mathscr{U})}, \qquad \forall\, u \in L^2(0, \infty; \mathscr{U}). \tag{3.5}$$

In the singular case we let $\varepsilon = 0$ in (3.5). These cases are treated in Sections 3.2 and 3.5, respectively.

Let $\hat{u}$ denote the Laplace transform of a function $u \in L^2(0, \infty; \mathscr{U})$ (cf. Definition (2.23)) and $\mathbf{G}$ be the transfer function corresponding to the shift-invariant operator $\mathbb{F}$ (cf. Definition 2.27). As shown in [70, Proposition 7.1], $\mathscr{R}$ is a Toeplitz operator, whose unique symbol is the *Popov function* $\Pi \in L^\infty(i\mathbb{R}; \mathcal{L}(\mathscr{U}))$ defined as

$$\Pi(i\omega) = R + NG(i\omega) + G(i\omega)^* N^* + G(i\omega)^* Q G(i\omega). \tag{3.6}$$

By the Paley-Wiener theorem (see Theorem 2.25) there holds

$$\langle u_1, \mathscr{R} u_2 \rangle_{L^2(0,\infty;\mathscr{U})} = \frac{1}{2\pi} \langle \hat{u}_1, \Pi \hat{u}_2 \rangle_{L^2(i\mathbb{R};\mathscr{U})}, \qquad \forall\, u_1, u_2 \in L^2(0, \infty; \mathscr{U}). \tag{3.7}$$

Remark 3.1. Throughout this chapter we assume that $\Sigma = (\mathbb{T}, \Phi, \Psi, \mathbb{F})$ is externally stable (see Definition 2.6.a). For such a system it holds

$$\mathbb{F} \in \mathcal{L}(L^2(0,\infty;\mathscr{U}), L^2(0,\infty;\mathscr{Y})). \tag{3.8}$$

By Theorem 2.26, (3.8) is equivalent to

$$G \in H^\infty\left(\mathcal{L}(\mathscr{U},\mathscr{Y})\right).$$

It follows from [51, Theorem 3.2] that

$$\lim_{\sigma\downarrow 0,\ \sigma\in\mathbb{R}} G(\sigma + i\omega)u = G(i\omega)u,$$

for all $u \in \mathscr{U}$ and almost all $\omega \in \mathbb{R}$. This means that the Popov function (3.6) corresponding to an externally stable well-posed linear system is well-defined.

3.2 The regular linear-quadratic optimal control problem

In the regular optimal control problem we assume that the Popov operator $\mathscr{R}$ from (3.4) is coercive. This means that (3.5) is satisfied for some $\varepsilon > 0$. Condition (3.5) implies in particular that $\mathscr{R}$ is boundedly invertible. The following proposition from [70] gives the *open-loop solution* of the optimal control problem:

Proposition 3.2. *[70, Proposition 7.2.] Let $\Sigma = (\mathbb{T}, \Phi, \Psi, \mathbb{F})$ be an externally stable well-posed linear system (cf. Definition 2.6.a) and let J be the cost functional from (3.2). Then, for every $x_0 \in \mathscr{X}$*

$$\min_{u\in L^2(0,\infty;\mathscr{U})} J(u, x_0) = \langle x_0, Xx_0\rangle_{\mathscr{X}}, \tag{3.9}$$

where operator $X = X^* \in \mathcal{L}(\mathscr{X})$, called the Riccati operator, is defined by

$$X = \Psi^* Q\Psi - \Psi^*(Q\mathbb{F} + N^*)\mathscr{R}^{-1}(\mathbb{F}^* Q + N)\Psi. \tag{3.10}$$

The unique minimizing input function is denoted by u^{opt} and is given by

$$u^{\text{opt}} = -\mathscr{R}^{-1}(\mathbb{F}^* Q + N)\Psi x_0. \tag{3.11}$$

We introduce the *complementary Popov operator* (Definition 3.3), which allows us to present an alternative formulation of the *Riccati operator* (3.10). This alternative formulation is given in Proposition 3.4. In [35], the *Riccati-ADI* algorithm (discussed in Chapter 4) was formulated using the *projected* version of Proposition 3.4.

Definition 3.3. Let $Q = Q^* \in \mathcal{L}(\mathscr{Y})$, $R = R^* \in \mathcal{L}(\mathscr{U})$, and $\widetilde{R} := R - NQ^{-1}N^* \in \mathcal{L}(\mathscr{U})$ be all invertible operators. The *complementary Popov operator* is defined by

$$\mathscr{R}_c := Q^{-1} + (\mathbb{F} + Q^{-1}N^*)\widetilde{R}^{-1}(\mathbb{F}^* + NQ^{-1}) \in \mathcal{L}(\mathscr{Y}). \tag{3.12}$$

Proposition 3.4. *With the assumptions of Definition 3.3, the Riccati operator $X \in \mathcal{L}(\mathscr{X})$ from (3.10) can be represented in the alternative form*

$$X = \Psi^* \mathscr{R}_c^{-1} \Psi. \tag{3.13}$$

Moreover, the optimal input function u^{opt} from (3.11) can be expressed by

$$u^{\text{opt}} = -\widetilde{R}^{-1}(\mathbb{F}^* + NQ^{-1})\mathscr{R}_c^{-1}\Psi x_0. \tag{3.14}$$

Proof. First, we show that the inverse of the complementary Popov operator from (3.12) can be calculated by adapting the generalized Sherman-Morrison-Woodbury formula from Theorem A.1. In fact, by setting

$$A = Q^{-1} \in \mathcal{L}(\mathscr{Y}), \qquad Z^* = Y = \mathbb{F} + Q^{-1}N^* \in \mathcal{L}(\mathscr{U}, \mathscr{Y}), \qquad G = \widetilde{R}^{-1} \in \mathcal{L}(\mathscr{U}),$$

and using (3.12), we obtain that

$$\mathscr{R}_c^{-1} = Q - (Q\mathbb{F} + N^*)\mathscr{R}^{-1}(\mathbb{F}^*Q + N). \tag{3.15}$$

As a result, the Riccati operator X from (3.10) can be written as

$$X = \Psi^* \left(Q - (Q\mathbb{F} + N^*)\mathscr{R}^{-1}(\mathbb{F}^*Q + N)\right)\Psi = \Psi^*\mathscr{R}_c^{-1}\Psi.$$

Now in order to show (3.14), we multiply (3.15) from left by $\mathbb{F}^* + NQ^{-1}$ to get

$$\begin{aligned}
(\mathbb{F}^* + NQ^{-1})\mathscr{R}_c^{-1} &= \mathbb{F}^*Q + N - (\mathbb{F}^*Q\mathbb{F} + \mathbb{F}^*N^* + N\mathbb{F} + NQ^{-1}N^*)\mathscr{R}^{-1}(\mathbb{F}^*Q + N) \\
&= \mathbb{F}^*Q + N - \left(\mathscr{R} - \widetilde{R}\right)\mathscr{R}^{-1}(\mathbb{F}^*Q + N) \\
&= \widetilde{R}\,\mathscr{R}^{-1}(\mathbb{F}^*Q + N).
\end{aligned}$$

By invertibility of $\widetilde{R}$ we obtain

$$\widetilde{R}^{-1}(\mathbb{F}^* + NQ^{-1})\mathscr{R}_c^{-1} = \mathscr{R}^{-1}(\mathbb{F}^*Q + N), \tag{3.16}$$

and hence (3.14) follows by substituting (3.16) in (3.11). □

Let x^{opt} and y^{opt} denote respectively the optimal state trajectory and the optimal output function of the well-posed linear system $\Sigma = (\mathbb{T}, \Phi, \Psi, \mathbb{F})$ with the optimal input function u^{opt} and the initial state $x_0 \in \mathscr{X}$. As a result of (3.1) we have

$$\begin{aligned}
x^{\text{opt}}(t) &= \mathbb{T}_t x_0 + \Phi_t u^{\text{opt}}, \qquad \forall\, t \geq 0, \\
y^{\text{opt}} &= \Psi x_0 + \mathbb{F}u^{\text{opt}},
\end{aligned} \tag{3.17}$$

for all $x_0 \in \mathscr{X}$. By substituting (3.11) in (3.17) we obtain

$$\begin{aligned}
x^{\text{opt}}(t) &= \left[\mathbb{T}_t - \Phi_t\mathscr{R}^{-1}(\mathbb{F}^*Q + N)\Psi\right]x_0, \qquad \forall\, t \geq 0, \\
y^{\text{opt}} &= \left[\Psi - \mathbb{F}\mathscr{R}^{-1}(\mathbb{F}^*Q + N)\Psi\right]x_0.
\end{aligned} \tag{3.18}$$

We denote

$$\begin{aligned}
\mathbb{T}_t^{\text{opt}} &:= \mathbb{T}_t - \Phi_t\mathscr{R}^{-1}(\mathbb{F}^*Q + N)\Psi, \qquad \forall\, t \geq 0, \\
\Psi^{\text{opt}} &:= \Psi - \mathbb{F}\mathscr{R}^{-1}(\mathbb{F}^*Q + N)\Psi, \\
\widetilde{\Psi}^{\text{opt}} &:= -\mathscr{R}^{-1}(\mathbb{F}^*Q + N)\Psi,
\end{aligned} \tag{3.19}$$

and observe that

$$\Psi^{\text{opt}} = \Psi + \mathbb{F}\widetilde{\Psi}^{\text{opt}}. \tag{3.20}$$

Definition 3.5. Let $\mathbb{T}^{\text{opt}}$, Ψ^{opt}, and $\widetilde{\Psi}^{\text{opt}}$ be the operators defined in (3.19). For all $x_0 \in \mathscr{X}$, the *extended (open-loop) optimal well-posed linear system* $\Sigma_{\text{opt,ext}}$ is defined by

$$\Sigma_{\text{opt,ext}} := \begin{cases} x^{\text{opt}}(t) = \mathbb{T}_t^{\text{opt}} x_0, & \forall\, t \geq 0, \\ y^{\text{opt}} = \Psi^{\text{opt}} x_0, \\ u^{\text{opt}} = \widetilde{\Psi}^{\text{opt}} x_0, \end{cases} \tag{3.21}$$

with the associated optimal cost

$$J(u^{\text{opt}}, x_0) = \langle x_0, X x_0 \rangle_{\mathscr{X}}\,.$$

Remark 3.6. (a) Let $\Sigma = (\mathbb{T}, \Phi, \Psi, \mathbb{F})$ be an externally stable well-posed linear system (cf. Definition 2.6.a). It follows from (2.11), together with (3.19), that

$$\begin{aligned} \Psi^{\text{opt}} &\in \mathcal{L}(\mathscr{X}, L^2(0,\infty;\mathscr{Y})), \\ \widetilde{\Psi}^{\text{opt}} &\in \mathcal{L}(\mathscr{X}, L^2(0,\infty;\mathscr{U})), \end{aligned} \tag{3.22}$$

which means that the extended optimal system $\Sigma_{\text{opt,ext}}$ is externally stable.

(b) If $\mathbb{T}$ is strongly [exponentially] stable, then it follows from Lemma 2.46 that $\mathbb{T}^{\text{opt}}$ defined in (3.19) is strongly [exponentially] stable. In this case the extended optimal system $\Sigma_{\text{opt,ext}}$ is called strongly [exponentially] stable.

We would like to determine the generators of the extended optimal system $\Sigma_{\text{opt,ext}}$. Let A^{opt} denote the generator of the strongly continuous semigroup $\mathbb{T}^{\text{opt}}$. $D(A^{\text{opt}})$ is a Hilbert space with the norm

$$\|x\|_1^{\text{opt}} := \|(\beta I - A^{\text{opt}})x\|,$$

for some $\beta \in \rho(A^{\text{opt}})$. It follows from [70, Proposition 9.3] that

$$D(A^{\text{opt}}) \subset D(C_{\Lambda w}),$$

and that the restriction of $C_{\Lambda w}$ to $D(A^{\text{opt}})$ is bounded from $D(A^{\text{opt}})$ to $\mathscr{Y}$. Moreover, for all $x_0 \in D\left(A^{\text{opt}}\right)$ and almost every $t \geq 0$, there holds

$$y^{\text{opt}}(t) = C_{\Lambda w} x^{\text{opt}}(t). \tag{3.23}$$

Let $R = R^* \in \mathcal{L}(\mathscr{U})$ be invertible. It follows from [70, Proposition 8.5] that for every initial state $x_0 \in \mathscr{X}$ and for almost every $t \geq 0$

$$x^{\text{opt}}(t) \in D(C_{\Lambda w}), \qquad X x^{\text{opt}}(t) \in D(B^*_{\Lambda w}), \tag{3.24}$$

and there holds

$$u^{\text{opt}}(t) = -R^{-1}(B^*_{\Lambda w} X + N C_{\Lambda w}) x^{\text{opt}}(t). \tag{3.25}$$

Equation (3.25) shows that the optimal input function u^{opt} of the well-posed linear system Σ can be written in a *feedback* form (familiar from the finite-dimensional theory) which involves the Riccati operator X. Therefore, we define the *optimal feedback operator* $F^{\mathrm{opt}} : D(A^{\mathrm{opt}}) \to \mathscr{U}$ by

$$F^{\mathrm{opt}} x_0 := -R^{-1}(B^*_{\Lambda w} X + N C_{\Lambda w}) x_0. \tag{3.26}$$

Altogether, A^{opt}, $C_{\Lambda w}$, and F^{opt} are the generators (cf. Definition 2.34) of the extended optimal system $\Sigma_{\mathrm{opt,ext}}$. This means that for almost all $t \geq 0$ there holds

$$\Sigma_{\mathrm{opt,ext}} : \begin{cases} \dot{x}^{\mathrm{opt}}(t) = A^{\mathrm{opt}} x^{\mathrm{opt}}(t), \\ y^{\mathrm{opt}}(t) = C_{\Lambda w} x^{\mathrm{opt}}(t), \\ u^{\mathrm{opt}}(t) = F^{\mathrm{opt}} x^{\mathrm{opt}}(t). \end{cases} \tag{3.27}$$

For notational simplicity we use "$\rightsquigarrow$" to mean "generator". With this notation we have

$$\begin{array}{ccc} (A^{\mathrm{opt}}, C_{\Lambda w}) & \rightsquigarrow & \Psi^{\mathrm{opt}}, \\ (A^{\mathrm{opt}}, F^{\mathrm{opt}}) & \rightsquigarrow & \widetilde{\Psi}^{\mathrm{opt}}. \end{array} \tag{3.28}$$

Remark 3.7. (a) As a result of (3.22) we have that F^{opt} and $C_{\Lambda w}$ are infinite-time admissible observation operators for $\mathbb{T}^{\mathrm{opt}}$.

(b) Operator A^{opt} can be also characterized by the Riccati operator X. In fact, it follows from [70, Theorem 9.4] that

$$A^{\mathrm{opt}} x_0 = (A + BF^{\mathrm{opt}}) x_0, \tag{3.29}$$

for all $x_0 \in D(A^{\mathrm{opt}})$.

(c) In all the formulas presented in this section we did not use any regularity assumption. In fact, they hold for all well-posed linear systems. If Σ as well as its dual system Σ^d (cf. Section 2.6) are regular, then we may replace $C_{\Lambda w}$ by C_Λ and $B^*_{\Lambda w}$ by B^*_Λ in all the formulas presented in this section.

3.3 Spectral factorization

A problem with formulas (3.10) and (3.11) is that they contain $\mathscr{R}^{-1}$, which is not easy to compute. In this section we will show that, if a spectral factor of the Popov function (3.6) is known, then $\mathscr{R}^{-1}$ can be expressed in terms of this spectral factor. To this end, we review the concept of *spectral factorization* mainly from [15] and [70].

Definition 3.8 (Spectral factorization, spectral factor system)**.** We say that the Popov function Π from (3.6) has a spectral factorization, if for almost all $\omega \in \mathbb{R}$, there exists an operator $\Xi \in H_\infty(\mathcal{L}(\mathscr{U}))$ such that

$$\Pi(i\omega) = \Xi(i\omega)^* \Xi(i\omega), \qquad \forall\, \omega \in \mathbb{R}. \tag{3.30}$$

Let $\mathbb{F}_\Xi \in \mathcal{L}(L^2(0,\infty;\mathscr{U}))$ denote the shift-invariant operator corresponding to the spectral factor $\Xi \in H_\infty(\mathcal{L}(\mathscr{U}))$ and define the observation operator $\Psi_\Xi : \mathscr{X} \to L^2(0,\infty;\mathscr{U})$ by

$$\Psi_\Xi = (\mathbb{F}_\Xi)^{-*} (\mathbb{F}^* Q + N)\Psi.$$

Then, it follows from [70, Theorem 11.3] that the spectral factor system

$$\Sigma_\Xi = (\mathbb{T}, \Phi, \Psi_\Xi, \mathbb{F}_\Xi)$$

is a well-posed linear system. The spectral factor Ξ is called *outer*, if its range, as a multiplication operator on $H_2(\mathscr{U})$, is dense in $H_2(\mathscr{U})$. If Ξ is outer, then the range of $\mathbb{F}_\Xi$ is also dense.

If the spectral factor Ξ is strongly regular (cf. Definition 2.29), then there exists an operator $D_\Xi \in \mathcal{L}(\mathscr{U})$ such that

$$\lim_{\substack{\lambda\to\infty \\ \lambda\in\mathbb{R}}} \Xi(\lambda)v = D_\Xi v, \qquad \forall\, v \in \mathscr{U}. \tag{3.31}$$

D_Ξ as in (3.31) is called the *feedthrough* operator corresponding to the strongly regular spectral factor Ξ. If additionally, D_Ξ is invertible, then from [70, Theorem 12.4], $-D_\Xi F$ is the observation operator of the regular spectral factor system Σ_Ξ, where

$$Fx_0 = -(D_\Xi^* D_\Xi)^{-1}(B_{\Lambda w}^* X + NC)x_0, \qquad \forall\, x_0 \in D(A).$$

In particular,

$$X \in \mathcal{L}(D(A), B_{\Lambda w}^*).$$

Altogether, the quadruple $(A, B, -D_\Xi F, D_\Xi)$ is the *generator* (cf. Definition 2.34) of the *strongly regular spectral factor system* $\Sigma_\Xi = (\mathbb{T}, \Phi, \Psi_\Xi, \mathbb{F}_\Xi)$ and there holds

$$\Xi(s) = D_\Xi \left(I - F_\Lambda(sI - A)^{-1}B\right).$$

One can observe that F is the optimal state feedback operator for Σ. In fact, it follows from [70, Theorem 12.5] that for all $x_0 \in D(A^{\mathrm{opt}})$,

$$A^{\mathrm{opt}} x_0 = (A + BF_\Lambda)\, x_0,$$

and

$$F^{\mathrm{opt}} x_0 = F_\Lambda x_0,$$

where $D(A^{\mathrm{opt}})$ is defined by

$$D(A^{\mathrm{opt}}) = \{x_0 \in D(F_\Lambda) \mid (A + BF_\Lambda)x_0 \in \mathscr{X}\}.$$

Remark 3.9. It does not always hold that $D_\Xi^* D_\Xi = R$. However, if $\mathscr{Y}$ is finite-dimensional, $B \in \mathcal{L}(\mathscr{U}, \mathscr{X}_{-1})$ is infinite-time admissible (cf. Definition 2.17) for the semigroup $\mathbb{T}$, and C is bounded (i.e., $C \in \mathcal{L}(\mathscr{X}, \mathscr{Y})$), then from [70, Proposition 12.10] any spectral factor Ξ of the Popov function Π is regular and its feedthrough operator D_Ξ satisfies

$$D_\Xi^* D_\Xi = R.$$

3.4 Riccati equations

A classical approach in the finite-dimensional optimal control theory is to solve the Riccati equation to obtain the optimal cost operator (see Section 1.1). This approach can be extended to infinite-dimensional systems with bounded control and observation operators (see for example [18]). This extension becomes more difficult when considering well-posed linear systems with unbounded control and observation operators. Moreover, it is also not easy to write a meaningful Riccati equation in this case.

Let us, for the moment, assume that the control and observation operators are bounded, i.e., $B \in \mathcal{L}(\mathscr{U}, \mathscr{X})$ and $C \in \mathcal{L}(\mathscr{X}, \mathscr{Y})$. A well-posed linear system $\Sigma = (\mathbb{T}, \Phi, \Psi, \mathbb{F})$ with the cost functional (3.2) is called *optimizable*, if for all $x_0 \in \mathscr{X}$, there exists an input function $u \in L^2(0, \infty; \mathscr{U})$ such that this cost functional is finite (see [18, Definition 6.2.1]).

Considering the cost function (3.2) with invertible operator $R = R^* \in \mathcal{L}(\mathscr{U})$, if the well-posed linear system Σ is optimizable, then by [18, Theorem 6.2.4], the Riccati operator X from (3.10) is the minimal nonnegative solution of the following operator algebraic Riccati equation

$$\langle Ax_0, Xz_0\rangle_{\mathscr{X}} + \langle Xx_0, Az_0\rangle_{\mathscr{X}} + \langle Cx_0, QCz_0\rangle_{\mathscr{Y}} = \langle Fx_0, RFz_0\rangle_{\mathscr{U}}, \tag{3.32}$$

for all $x_0, z_0 \in D(A)$, where

$$Fx_0 = -R^{-1}(B^*X + N^*C)x_0. \tag{3.33}$$

By adding the term

$$\langle Fx_0, B^*Xz_0\rangle_{\mathscr{U}} + \langle B^*Xx_0, Fz_0\rangle_{\mathscr{U}}$$

to both sides of (3.32), we obtain the "closed-loop" form of the operator Riccati equation

$$\left\langle A^{\rm opt}x_0, Xz_0\right\rangle_{\mathscr{X}} + \left\langle Xx_0, A^{\rm opt}z_0\right\rangle_{\mathscr{X}} = -\left\langle Cx_0, \widetilde{Q}Cz_0\right\rangle_{\mathscr{Y}} - \left\langle B^*Xx_0, R^{-1}B^*Xz_0\right\rangle_{\mathscr{U}}, \tag{3.34}$$

where

$$A^{\rm opt}x_0 := (A + BF)x_0, \qquad \forall\ x_0 \in D(A^{\rm opt}) = D(A),$$

and

$$\widetilde{Q} := Q - N^*R^{-1}N \ \in \mathcal{L}(\mathscr{Y}).$$

If the control operator is unbounded (i.e., $B \in \mathcal{L}(\mathscr{U}, \mathscr{X}_{-1})$), then it may happen that $Xx \notin D(B^*)$ for some $x \in D(A)$ (see for example [42], [69], and [60]). Therefore, F from (3.33) does not even make sense. To overcome this problem, we consider the Λ-extension of B^* (cf. Definition 2.37) and recall from (3.24) that for all $x_0 \in \mathscr{X}$ and almost all $t \geq 0$ there holds

$$Xx^{\rm opt}(t) \in D(B^*_{\Lambda w}), \qquad \text{with} \qquad x^{\rm opt}(t) = \mathbb{T}^{\rm opt}_t x_0.$$

In addition, we showed that the optimal feedback operator $F^{\rm opt}$ satisfies

$$F^{\rm opt}x_0 = -R^{-1}(B^*_{\Lambda w}X + NC_{\Lambda w})x_0. \tag{3.35}$$

In the following we show that the Riccati operator X from (3.10) satisfies some Riccati equations which can be understood as a generalization of the "closed-loop" Riccati equation (3.34) for unbounded control and observation operators. We refer to [42] for a discussion on several alternative Riccati equations in the case of unbounded control operators.

For the rest of this section we assume that $R = R^* \in \mathcal{L}(\mathscr{U})$ is strictly positive (and hence boundedly invertible). It follows from [70, Proposition 10.4] that the Riccati operator (3.10) satisfies the following "Riccati-like" equation which holds on $D\left(A^{\mathrm{opt}}\right)$: For every $x_0, z_0 \in D(A^{\mathrm{opt}})$,

$$\begin{aligned} &\left\langle A^{\mathrm{opt}}x_0, Xz_0\right\rangle + \left\langle Xx_0, A^{\mathrm{opt}}z_0\right\rangle \\ &\qquad = -\left\langle C_{\Lambda w}x_0, (QC_{\Lambda w} + N^*F^{\mathrm{opt}})z_0\right\rangle + \left\langle F^{\mathrm{opt}}x_0, B^*_{\Lambda w}Xz_0\right\rangle. \end{aligned} \tag{3.36}$$

This equation is called the "Riccati-like" equation, because it looks like the "closed-loop" Riccati equation (3.34) for bounded control and observation operators. The main drawback of (3.36) is that $D(A^{\mathrm{opt}})$ is not a priori known [70]. By substituting (3.35) in the right-hand side of (3.36), we obtain

$$\begin{aligned} &\left\langle A^{\mathrm{opt}}x_0, Xz_0\right\rangle + \left\langle Xx_0, A^{\mathrm{opt}}z_0\right\rangle \\ &\qquad = -\left\langle C_{\Lambda w}x_0, \widetilde{Q}C_{\Lambda w}z_0\right\rangle - \left\langle R^{-1}B^*_{\Lambda w}Xx_0, B^*_{\Lambda w}Xz_0\right\rangle, \end{aligned} \tag{3.37}$$

where $\widetilde{Q} = Q - N^*R^{-1}N$. In Chapter 5, we will use the Riccati-like equation (3.37) to construct an iterative method to find approximate solutions of the linear-quadratic optimal control problem. This method can be understood as an extension of the Newton-Kleinman approach [29] to infinite-dimensional spaces.

If we assume that $\widetilde{Q}$ is positive ($\widetilde{Q} \geq 0$) and R is strictly positive ($R > 0$)[1], then we can write the Riccati-like equation (3.37) as

$$\begin{aligned} &\left\langle A^{\mathrm{opt}}x_0, Xz_0\right\rangle + \left\langle Xx_0, A^{\mathrm{opt}}z_0\right\rangle \\ &\qquad = -\left\langle \widetilde{Q}^{1/2}C_{\Lambda w}x_0, \widetilde{Q}^{1/2}C_{\Lambda w}z_0\right\rangle - \left\langle R^{-1/2}B^*_{\Lambda w}Xx_0, R^{-1/2}B^*_{\Lambda w}Xz_0\right\rangle. \end{aligned} \tag{3.38}$$

Corresponding to (3.38) we consider the system

$$\Sigma_r : \begin{cases} \dot{x}_r(t) = A^{\mathrm{opt}}x_r(t), \\ y_r(t) = C^{\mathrm{opt}}x_r(t), \end{cases} \tag{3.39}$$

for almost all $t \geq 0$, where

$$C^{\mathrm{opt}} := \begin{pmatrix} \widetilde{Q}^{1/2}C_{\Lambda w} \\ -R^{1/2}B^*_{\Lambda w}X \end{pmatrix}. \tag{3.40}$$

The following lemma shows that the Riccati operator (3.10) is the infinite-time observability Gramian of $(A^{\mathrm{opt}}, C^{\mathrm{opt}})$.

Proposition 3.10. *Let $\Sigma = (\mathbb{T}, \Phi, \Psi, \mathbb{F})$ be an externally stable well-posed linear system such that (A, B, C) is the triple associated with Σ (cf. Definition 2.15). Let J be the cost functional (3.2) such that $R > 0$ and $\widetilde{Q} = Q - N^*R^{-1}N \geq 0$. Moreover, let $\mathbb{T}^{\mathrm{opt}}$, Ψ^{opt},*

[1] This notation has been defined in Section 1.3

and $\widetilde{\Psi}^{\mathrm{opt}}$ *be given by* (3.19)*. Let* A^{opt} *denote the generator of the strongly continuous semigroup* $\mathbb{T}^{\mathrm{opt}}$*. Furthermore, let* X *be the Riccati operator* (3.10) *and let* C^{opt} *be given by* (3.40)*.*
Then the Riccati operator X *is the infinite-time observability Gramian of* $(A^{\mathrm{opt}}, C^{\mathrm{opt}})$*.*

Proof. First, we recall the extended optimal well-posed linear system $\Sigma_{\mathrm{opt,ext}}$ from (3.21). As we showed in (3.27), A^{opt}, $C_{\Lambda w}$, and F^{opt} generate the extended optimal well-posed linear system $\Sigma_{\mathrm{opt,ext}}$. In particular, $C_{\Lambda w}$ and F^{opt} are the infinite-time admissible observation operators for the semigroup $\mathbb{T}^{\mathrm{opt}}$ (cf. Remark 3.7.a). Moreover, $C_{\Lambda w}$ and F^{opt} generate the output maps Ψ^{opt} and $\widetilde{\Psi}^{\mathrm{opt}}$, respectively. By recalling the notation introduced in (3.28), we have

$$\begin{aligned}(A^{\mathrm{opt}}, C_{\Lambda w}) \quad &\rightsquigarrow \quad \Psi^{\mathrm{opt}},\\ (A^{\mathrm{opt}}, F^{\mathrm{opt}}) \quad &\rightsquigarrow \quad \widetilde{\Psi}^{\mathrm{opt}}.\end{aligned}$$

Hence, with $F^{\mathrm{opt}} = -R^{-1}(B^*_{\Lambda w}X + NC_{\Lambda w})$, we observe that

$$\begin{aligned}(A^{\mathrm{opt}}, \widetilde{Q}^{1/2}C_{\Lambda w}) \quad &\rightsquigarrow \quad \widetilde{Q}^{1/2}\Psi^{\mathrm{opt}},\\ (A^{\mathrm{opt}}, -R^{-1}B^*_{\Lambda w}X) \quad &\rightsquigarrow \quad \widetilde{\Psi}^{\mathrm{opt}} + R^{-1}N\Psi^{\mathrm{opt}},\\ (A^{\mathrm{opt}}, -R^{-1/2}B^*_{\Lambda w}X) \quad &\rightsquigarrow \quad R^{1/2}\widetilde{\Psi}^{\mathrm{opt}} + R^{-1/2}N\Psi^{\mathrm{opt}}.\end{aligned} \tag{3.41}$$

As a result, the output map associated with $(A^{\mathrm{opt}}, C^{\mathrm{opt}})$ is given by

$$(A^{\mathrm{opt}}, C^{\mathrm{opt}}) \quad \rightsquigarrow \quad \begin{pmatrix}\Psi_r\\ \widetilde{\Psi}_r\end{pmatrix}, \tag{3.42}$$

where

$$\begin{aligned}\Psi_r &:= \widetilde{Q}^{1/2}\Psi^{\mathrm{opt}} \in \mathcal{L}(\mathscr{X}, L^2(0,\infty;\mathscr{Y})),\\ \widetilde{\Psi}_r &:= R^{1/2}\widetilde{\Psi}^{\mathrm{opt}} + R^{-1/2}N\Psi^{\mathrm{opt}} \in \mathcal{L}(\mathscr{X}, L^2(0,\infty;\mathscr{U})).\end{aligned} \tag{3.43}$$

Now let Γ_r denote the infinite-time observability Gramian of $(A^{\mathrm{opt}}, C^{\mathrm{opt}})$ (cf. Definition 2.20). From (3.42) we conclude that

$$\begin{aligned}\Gamma_r &= \begin{pmatrix}\Psi_r\\ \widetilde{\Psi}_r\end{pmatrix}^* \begin{pmatrix}\Psi_r\\ \widetilde{\Psi}_r\end{pmatrix}\\ &= \begin{pmatrix}\Psi^{\mathrm{opt}}\\ \widetilde{\Psi}^{\mathrm{opt}} + R^{-1}N\Psi^{\mathrm{opt}}\end{pmatrix}^* \begin{bmatrix}\widetilde{Q} & 0\\ 0 & R\end{bmatrix} \begin{pmatrix}\Psi^{\mathrm{opt}}\\ \widetilde{\Psi}^{\mathrm{opt}} + R^{-1}N\Psi^{\mathrm{opt}}\end{pmatrix}.\end{aligned} \tag{3.44}$$

Since R is invertible, we have the following decomposition of the operator matrix $\begin{bmatrix}\widetilde{Q} & 0\\ 0 & R\end{bmatrix}$:

$$\begin{bmatrix}\widetilde{Q} & 0\\ 0 & R\end{bmatrix} = \begin{bmatrix}I & -N^*R^{-1}\\ 0 & I\end{bmatrix} \begin{bmatrix}Q & N^*\\ N & R\end{bmatrix} \begin{bmatrix}I & 0\\ -R^{-1}N & I\end{bmatrix},$$

which can be understood as a generalization of the LDL^T factorization [11] to operators acting on infinite-dimensional spaces. Hence, we can write (3.44) as

$$\Gamma_r = \begin{pmatrix}\Psi^{\mathrm{opt}}\\ \widetilde{\Psi}^{\mathrm{opt}}\end{pmatrix}^* \begin{bmatrix}Q & N^*\\ N & R\end{bmatrix} \begin{pmatrix}\Psi^{\mathrm{opt}}\\ \widetilde{\Psi}^{\mathrm{opt}}\end{pmatrix}. \tag{3.45}$$

From (3.20) we know that $\Psi^{\text{opt}} = \Psi + \mathbb{F}\widetilde{\Psi}^{\text{opt}}$ and therefore

$$\begin{aligned}\Gamma_r &= \begin{pmatrix} \Psi + \mathbb{F}\widetilde{\Psi}^{\text{opt}} \\ \widetilde{\Psi}^{\text{opt}} \end{pmatrix}^* \begin{bmatrix} Q & N^* \\ N & R \end{bmatrix} \begin{pmatrix} \Psi + \mathbb{F}\widetilde{\Psi}^{\text{opt}} \\ \widetilde{\Psi}^{\text{opt}} \end{pmatrix}, \\ &= \begin{pmatrix} \Psi \\ \widetilde{\Psi}^{\text{opt}} \end{pmatrix}^* \begin{bmatrix} I & 0 \\ \mathbb{F}^* & I \end{bmatrix} \begin{bmatrix} Q & N^* \\ N & R \end{bmatrix} \begin{bmatrix} I & \mathbb{F} \\ 0 & I \end{bmatrix} \begin{pmatrix} \Psi \\ \widetilde{\Psi}^{\text{opt}} \end{pmatrix}, \\ &= \begin{pmatrix} \Psi \\ \widetilde{\Psi}^{\text{opt}} \end{pmatrix}^* \begin{bmatrix} Q & Q\mathbb{F} + N^* \\ \mathbb{F}^*Q + N & \mathscr{R} \end{bmatrix} \begin{pmatrix} \Psi \\ \widetilde{\Psi}^{\text{opt}} \end{pmatrix}. \end{aligned} \tag{3.46}$$

Now by substituting $\widetilde{\Psi}^{\text{opt}} = -\mathscr{R}^{-1}(\mathbb{F}^*Q + N)\Psi$ from (3.19) into (3.46) and applying a direct algebraic calculation, we obtain

$$\Gamma_r = \Psi^*Q\Psi - \Psi^*(Q\mathbb{F} + N^*)\mathscr{R}^{-1}(\mathbb{F}^*Q + N)\Psi = X.$$

Hence, the Riccati operator X is the infinite-time observability Gramian of $(A^{\text{opt}}, C^{\text{opt}})$. □

So far we have shown that the Riccati operator (3.10) is the infinite-time observability Gramian of $(A^{\text{opt}}, C^{\text{opt}})$ and satisfies the Riccati-like equation (3.37). If additionally, we assume that $\Sigma = (\mathbb{T}, \Phi, \Psi, \mathbb{F})$ is strongly stable, i.e., the semigroup $\mathbb{T}$ is strongly stable (cf. Definition 2.6.c), then we can show that the Riccati operator is the unique solution of the Riccati-like equation (3.37). In fact, the following theorem holds true:

Theorem 3.11. *Under the assumptions of Proposition 3.10, if the semigroup $\mathbb{T}$ is strongly stable (cf. Definition 2.6.c), then the Riccati operator* (3.10) *is the unique solution of the Riccati-like equation* (3.38).

Proof. By (3.41) – (3.43), we know that

$$\begin{aligned} (A^{\text{opt}}, \widetilde{Q}^{1/2}C_{\Lambda w}) &\quad\rightsquigarrow\quad \Psi_r, \\ (A^{\text{opt}}, -R^{-1/2}B^*_{\Lambda w}X) &\quad\rightsquigarrow\quad \widetilde{\Psi}_r. \end{aligned} \tag{3.47}$$

Now we follow the lines of the proof for [57, Theorem 41]. If $x_0, z_0 \in D(A^{\text{opt}})$, then $t \mapsto \langle \mathbb{T}_t^{\text{opt}} x_0, X\mathbb{T}_t^{\text{opt}} z_0\rangle$ is continuously differentiable on $[0, \infty)$ and we obtain

$$\frac{d}{dt}\left\langle \mathbb{T}_t^{\text{opt}} x_0, X\mathbb{T}_t^{\text{opt}} z_0 \right\rangle = \left\langle A^{\text{opt}}\mathbb{T}_t^{\text{opt}} x_0, X\mathbb{T}_t^{\text{opt}} z_0 \right\rangle + \left\langle \mathbb{T}_t^{\text{opt}} x_0, XA^{\text{opt}}\mathbb{T}_t^{\text{opt}} z_0 \right\rangle.$$

Since the Riccati operator X satisfies the Riccati-like equation (3.38), there holds

$$\begin{aligned} \frac{d}{dt}\left\langle \mathbb{T}_t^{\text{opt}} x_0, X\mathbb{T}_t^{\text{opt}} z_0 \right\rangle_{\mathscr{X}} &= -\left\langle \widetilde{Q}^{1/2}C_{\Lambda w}\mathbb{T}_t^{\text{opt}} x_0, \widetilde{Q}^{1/2}C_{\Lambda w}\mathbb{T}_t^{\text{opt}} z_0 \right\rangle_{\mathscr{Y}} \\ &\quad - \left\langle R^{-1/2}B^*_{\Lambda w}X\mathbb{T}_t^{\text{opt}} x_0, R^{-1/2}B^*_{\Lambda w}X\mathbb{T}_t^{\text{opt}} z_0 \right\rangle_{\mathscr{U}} \\ &= -\langle (\Psi_r x_0)(t), (\Psi_r z_0)(t)\rangle_{\mathscr{Y}} \\ &\quad - \left\langle \left(\widetilde{\Psi}_r x_0\right)(t), \left(\widetilde{\Psi}_r z_0\right)(t) \right\rangle_{\mathscr{U}}, \end{aligned} \tag{3.48}$$

where we have used (3.47). By integrating (3.48), we obtain for each $t > 0$

$$\begin{aligned}\langle x_0, X z_0\rangle_{\mathscr{X}} - \left\langle \mathbb{T}_t^{\mathrm{opt}} x_0, X \mathbb{T}_t^{\mathrm{opt}} z_0 \right\rangle_{\mathscr{X}} = \int_0^t \langle (\Psi_r x_0)(\tau), (\Psi_r z_0)(\tau)\rangle_{\mathscr{Y}} \\ + \int_0^t \left\langle \left(\widetilde{\Psi}_r x_0\right)(\tau), \left(\widetilde{\Psi}_r z_0\right)(\tau)\right\rangle_{\mathscr{U}}.\end{aligned} \tag{3.49}$$

Since $\mathbb{T}$ is a strongly stable semigroup, it follows from Lemma 2.46 that $\mathbb{T}^{\mathrm{opt}}$ is also a strongly stable semigroup (cf. Remark 3.6.b). This means that

$$\lim_{t\to\infty} \mathbb{T}_t^{\mathrm{opt}} x_0 = 0, \qquad \text{in } \mathscr{X}.$$

By letting $t \to \infty$ in (3.49), we conclude that

$$\begin{aligned}\langle x_0, X z_0\rangle_{\mathscr{X}} &= \langle \Psi_r x_0, \Psi_r z_0\rangle_{L^2(0,\infty;\mathscr{Y})} + \left\langle \widetilde{\Psi}_r x_0, \widetilde{\Psi}_r z_0\right\rangle_{L^2(0,\infty;\mathscr{U})} \\ &= \langle x_0, \Psi_r^* \Psi_r z_0\rangle_{\mathscr{X}} + \left\langle x_0, \widetilde{\Psi}_r^* \widetilde{\Psi}_r z_0\right\rangle_{\mathscr{X}}.\end{aligned}$$

This being true for all $x_0, z_0 \in D(A^{\mathrm{opt}})$, we must have

$$\begin{aligned}X &= \Psi_r^* \Psi_r + \widetilde{\Psi}_r^* \widetilde{\Psi}_r \\ &= \Psi^* Q \Psi - \Psi^* (Q\mathbb{F} + N^*) \mathscr{R}^{-1} (\mathbb{F}^* Q + N) \Psi.\end{aligned}$$

□

A problem with the Riccati-like equation (3.36) is that it holds on $D(A^{\mathrm{opt}})$ which is not a priori known. If the Popov function Π has a regular spectral factorization Ξ such that the associated feedthrough operator D_Ξ is invertible, then it follows from [70, Proposition 12.8] that the Riccati operator (3.10) satisfies the "true" Riccati equation

$$A^* X + XA + C^* Q C = \left(B_{\Lambda w}^* X + NC\right)^* \left(D_\Xi^* D_\Xi\right)^{-1} \left(B_{\Lambda w}^* X + NC\right) \tag{3.50}$$

in $\mathcal{L}(\mathscr{X}_1, \mathscr{X}_{-1})$. The term $B_{\Lambda w}^* X + NC$ is regarded as an operator from $\mathscr{X}_1$ to $\mathscr{U}$, so that its adjoint maps $\mathscr{U}$ to $\mathscr{X}_{-1}$. As already mentioned in Remark 3.9, if $\mathscr{Y}$ is finite-dimensional, B is infinite-time admissible, and C is bounded, then any spectral factor Ξ of the Popov function Π is regular and the associated feedthrough operator D_Ξ satisfies $D_\Xi^* D_\Xi = R$. In this case (3.50) looks much the same as the usual Riccati equation known for the case of bounded control and observation operators (see (3.32)).

3.5 The singular linear-quadratic optimal control problem

In this section we weaken the assumption (3.5) on the Popov operator $\mathscr{R}$ by letting $\varepsilon = 0$. This means that for all $u \in L^2(0,\infty;\mathscr{U})$ we assume that

$$\langle u, \mathscr{R} u\rangle_{L^2(0,\infty;\mathscr{U})} \geq 0. \tag{3.51}$$

With the notation introduced after (1.15), this condition is written as $\mathscr{R} \geq 0$.

Following [15], we assume that the well-posed linear system Σ is strongly stable (cf. Definition 2.6.a). We will show that the solution to the singular linear-quadratic optimal control problem is closely related to the spectral factorization of the Popov function (3.6). For the sake of simplicity, we assume throughout this section that Σ has a zero feedthrough operator. This means that for all $u \in \mathscr{U}$

$$\lim_{\lambda \to \infty} \langle y, \mathbf{G}(\lambda) u \rangle_{\mathscr{Y}} = 0, \qquad \lambda \in \mathbb{R}, \quad \forall\, y \in \mathscr{Y}.$$

It follows from (3.7) that positivity of the Popov operator ($\mathscr{R} \geq 0$) is equivalent to positivity of the Popov function ($\Pi \geq 0$). Let $\mathcal{H}^\infty(\mathcal{L}(\mathscr{U}))$ denote the Hardy space of analytic $\mathcal{L}(\mathscr{U})$-valued functions on $\mathbb{C}_+$ (cf. Definition 2.22). Assuming that the Popov function Π has a spectral factorization $\Xi \in \mathcal{H}^\infty(\mathcal{L}(\mathscr{U}))$ (cf. Definition 3.8), we can characterize the solution to the singular optimal control problem by using the associated *spectral factor system* $\Sigma_\Xi = (\mathbb{T}, \Phi, \Psi_\Xi, \mathbb{F}_\Xi)$:

Theorem 3.12. *[15, Theorem 3.2] Let $\Sigma = (\mathbb{T}, \Phi, \Psi, \mathbb{F})$ be an externally stable weakly regular linear system with generating operators $(A, B, C, 0)$. Suppose that the corresponding Popov function Π from (3.6) has a spectral factorization $\Xi(i\omega) \in \mathcal{H}^\infty(\mathcal{L}(\mathscr{U}))$ which is outer. Let $\mathbb{F}_\Xi$ denote the shift-invariant operator corresponding to the transfer function Ξ.*

(a) If there exists $\Psi_\Xi \in \mathcal{L}(\mathscr{X}, L^2(0,\infty;\mathscr{U}))$ that satisfies

$$\mathbb{F}_\Xi^* \Psi_\Xi = (\mathbb{F}^* Q + N)\Psi, \tag{3.52}$$

then Ψ_Ξ is the unique solution to (3.52) and the spectral factor system

$$\Sigma_\Xi = (\mathbb{T}, \Phi, \Psi_\Xi, \mathbb{F}_\Xi)$$

is a well-posed linear system.

(b) Assume the existence of Ψ_Ξ as in part (i), and denote the corresponding infinite-time admissible observation operator by C_Ξ. If we define

$$X = \Psi^* Q \Psi - \Psi_\Xi^* \Psi_\Xi, \tag{3.53}$$

then X satisfies

$$A^* X x + X A x = C_\Xi^* C_\Xi x - C^* Q C x, \qquad \forall\, x \in D(A), \tag{3.54}$$

and

$$J(x_0, u) = \langle x_0, X x_0 \rangle + \|\Psi_\Xi x_0 + \mathbb{F}_\Xi u\|^2. \tag{3.55}$$

In particular,

$$\inf_{u \in L^2(0,\infty;\mathscr{U})} J(x_0, u) = \langle x_0, X x_0 \rangle. \tag{3.56}$$

(c) If additionally, the spectral factor system Σ_Ξ is weakly regular with generating operators (A, B, C_Ξ, D_Ξ), then

$$B_{\Lambda w}^* X x + N C x = D_\Xi^* C_\Xi x, \qquad \forall\, x \in D(A). \tag{3.57}$$

Remark 3.13. (a) It follows from Theorem 3.12.b that (3.56) holds, if and only if

$$\mathbb{F}_\Xi u^{\mathrm{opt}} + \Psi_\Xi x_0 = 0, \tag{3.58}$$

for some $u^{\mathrm{opt}} \in L^2(0,\infty;\mathscr{U})$. The existence of a solution to (3.58) follows from the property of $\mathbb{F}_\Xi$ having a dense range. Let (A, B, C_Ξ, D_Ξ) be the generating operators of the *weakly regular spectral factor system* $\Sigma_\Xi = (\mathbb{T}, \Phi, \Psi_\Xi, \mathbb{F}_\Xi)$. It follows from (3.58) that there exists some $x : [0,\infty) \to \mathscr{X}$ such that the differential-algebraic equation

$$\begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} \begin{pmatrix} \dot{x}(t) \\ \dot{u}(t) \end{pmatrix} = \begin{bmatrix} A & B \\ C_\Xi & D_\Xi \end{bmatrix} \begin{pmatrix} x(t) \\ u(t) \end{pmatrix}, \qquad x(0) = x_0 \in \mathscr{X}, \tag{3.59}$$

is fulfilled. We will use the finite-dimensional version of this remark in Chapter 6 to make a suitable choice of shift parameters for the ADI method and improve numerical performance of the algorithm.

(b) It follows from [15, Proposition 3.3] that the Riccati operator (3.53) is the *maximal self-adjoint solution* of

$$\left\langle \begin{bmatrix} \Psi^* Q \Psi - X & \Psi^*(Q\mathbb{F} + N^*) \\ (\mathbb{F}^* Q + N)\Psi & \mathscr{R} \end{bmatrix} \begin{pmatrix} x_0 \\ u \end{pmatrix}, \begin{pmatrix} x_0 \\ u \end{pmatrix} \right\rangle \geq 0,$$

for all $x_0 \in \mathscr{X}$ and every $u \in L^2(0,\infty;\mathscr{U})$.

(c) Let $\Sigma = (\mathbb{T}, \Phi, \Psi, \mathbb{F})$ be a weakly regular linear system with generating operators $(A, B, C, 0)$, such that its spectral factor system $\Sigma_\Xi = (\mathbb{T}, \Phi, \Psi_\Xi, \mathbb{F}_\Xi)$ is weakly regular with generating operators (A, B, C_Ξ, D_Ξ). Then, it follows from Theorem 3.12 that

$$\left\langle \begin{bmatrix} A^*X + XA + C^*QC & (B^*_{\Lambda w}X + NC)^* \\ B^*_{\Lambda w}X + NC & D^*_\Xi D_\Xi \end{bmatrix} \begin{pmatrix} x_0 \\ v \end{pmatrix}, \begin{pmatrix} x_0 \\ v \end{pmatrix} \right\rangle \geq 0$$

for all $x_0 \in D(A)$ and every $v \in \mathscr{U}$. In fact, it follows from (3.54) and (3.57) that the following factorization holds true on $D(A)$

$$\begin{bmatrix} A^*X + XA + C^*QC & (B^*_{\Lambda w}X + NC)^* \\ B^*_{\Lambda w}X + NC & D^*_\Xi D_\Xi \end{bmatrix} = \begin{pmatrix} C^*_\Xi \\ D^*_\Xi \end{pmatrix} \begin{pmatrix} C_\Xi & D_\Xi \end{pmatrix}. \tag{3.60}$$

(d) (see Remark 3.9) If $\mathscr{Y}$ is finite-dimensional, B is infinite-time admissible for the semigroup $\mathbb{T}$, and C is bounded, then it follows from [70, Proposition 12.10] that any spectral factor Ξ of the Popov function Π is regular and its feedthrough operator D_Ξ satisfies $D^*_\Xi D_\Xi = R$. In this case, (3.60) can be written as

$$\begin{aligned} A^*X + XA + C^*QC &= C^*_\Xi C_\Xi, \\ B^*_{\Lambda w}X + NC &= D^*_\Xi C_\Xi, \\ R &= D^*_\Xi D_\Xi. \end{aligned} \tag{3.61}$$

We refer to (3.61) as the *Lur'e equations* with the unknowns X, C_Ξ, and D_Ξ.

3.6 Bounded real and positive real case

We turn our focus to the bounded real and positive real singular optimal control problems. An important application of these problems is in the bounded real and positive real balanced truncation (see, e.g., [40, 49] for the finite-dimensional case and [24] for the infinite-dimensional case). These are model reduction techniques that preserve *contractivity* and *passivity*, respectively.

In the bounded real case we set $R = I \in \mathcal{L}(\mathscr{U})$, $Q = -I \in \mathcal{L}(\mathscr{Y})$, and $N = 0$. Then, the optimal control problem is to minimize for $x_0 \in \mathscr{X}$

$$J(x_0, u) = \int_0^\infty \|u(\tau)\|^2_{\mathscr{U}} - \|y(\tau)\|^2_{\mathscr{Y}}\, d\tau,$$

over all $u \in L^2(0, \infty; \mathscr{U})$ subject to the externally stable well-posed linear system $\Sigma = (\mathbb{T}, \Phi, \Psi, \mathbb{F})$. We prefer the following formulation of the bounded real optimal control problem:

$$\int_0^\infty \|y(\tau)\|^2_{\mathscr{Y}} - \|u(\tau)\|^2_{\mathscr{U}}\ \ d\tau = \langle x_0, X x_0\rangle_{\mathscr{X}} - \|\Psi_\Xi x_0 + \mathbb{F}_\Xi u\|^2, \tag{3.62}$$

where the Riccati operator (in the bounded real case) is given by

$$X = \Psi_\Xi^* \Psi_\Xi + \Psi\Psi. \tag{3.63}$$

Our motivation is based on the fact that the optimal cost in (3.62) expresses the *available storage* of the system [71]: For all $x_0 \in \mathscr{X}$ there holds

$$\langle x_0, X x_0\rangle_{\mathscr{X}} = \sup_{u \in L^2(0,\infty;\mathscr{U})} \int_0^\infty \|y(\tau)\|^2_{\mathscr{Y}} - \|u(\tau)\|^2_{\mathscr{U}}\, d\tau. \tag{3.64}$$

In the positive real case we set $\mathscr{U} = \mathscr{Y}$, $R = 0$, $Q = 0$, and $N = I \in \mathcal{L}(\mathscr{U})$. Then, the optimal control problem is to minimize for $x_0 \in \mathscr{X}$

$$J(x_0, u) = \int_0^\infty 2\mathrm{Re}\,\langle y(\tau), u(\tau)\rangle_{\mathscr{U}}\ d\tau, \tag{3.65}$$

over all $u \in L^2(0, \infty; \mathscr{U})$ subject to the externally stable well-posed linear system $\Sigma = (\mathbb{T}, \Phi, \Psi, \mathbb{F})$. Similarly to the bounded real case we prefer the following formulation of the positive real optimal control problem:

$$\int_0^\infty -2\mathrm{Re}\,\langle y(\tau), u(\tau)\rangle\ \mathrm{d}\tau = \langle x_0, X x_0\rangle_{\mathscr{X}} - \|\Psi_\Xi x_0 + \mathbb{F}_\Xi u\|^2, \tag{3.66}$$

where the Riccati operator (in the positive real case) is given by

$$X = \Psi_\Xi^* \Psi_\Xi. \tag{3.67}$$

Our consideration is based on the fact that the optimal cost in (3.66) expresses the *available storage for passivity* [71]: For all $x_0 \in \mathscr{X}$ there holds

$$\langle x_0, X x_0\rangle_{\mathscr{X}} = \sup_{u \in L^2(0,\infty;\mathbb{C}^m)} \int_0^\infty -2\mathrm{Re}\,\langle y(\tau), u(\tau)\rangle\ d\tau. \tag{3.68}$$

Remark 3.14. (a) In the bounded real case ($R = I$, $Q = -I$, and $N = 0$) the Popov operator is given by

$$\mathscr{R} = I - \mathbb{F}^*\mathbb{F}.$$

For the singular optimal control problem we assume positivity of the Popov operator ($\mathscr{R} \geq 0$). In the bounded real case, this implies

$$\|\mathbb{F}\|_{\mathcal{L}(L^2)} \leq 1. \tag{3.69}$$

This property is called *contractivity*. It follows from (2.32) that (3.69) is equivalent to

$$\|G\|_{\mathcal{H}^\infty} \leq 1.$$

(b) In the positive real case ($\mathscr{U} = \mathscr{Y}$, $R = 0$, $Q = 0$, $N = I$), the Popov operator is given by

$$\mathscr{R} = \mathbb{F}^* + \mathbb{F},$$

whose positivity is called *passivity*. Passivity is equivalent to *positive realness* of $G(s)$. This means that

$$G(s) + G(s)^* \geq 0,$$

for all $s \in \mathbb{C}$ with $\mathrm{Re}(s) > 0$ [71].

Chapter 4

The ADI method for the optimal control of stable well-posed linear systems

It is in your moments of decision that your destiny is shaped.
—Tony Robbins

Contents

We give an iterative method to solve the linear-quadratic optimal control of stable well-posed linear systems. The algorithm is based on approximating the output map and the input-output map of the well-posed linear system using projections on appropriate subspaces. These projection are determined by the so-called "shift parameters" of the method. We prove that the approximation obtained by this algorithm expresses the optimal cost for a projected optimal control problem. Furthermore, we show that the sequence of approximate solutions obtained by this algorithm is non-decreasing. Under mild assumptions on the shift parameters, we prove convergence of this sequence to the Riccati operator (3.10). Later on, this method is extended to solve the singular optimal control problem in the bounded real and positive real case. The results presented in this chapter are from [35] and [37].

4.1 Introduction

In this chapter we propose an iterative method for solving the linear-quadratic optimal control of stable well-posed linear systems (cf. Definition 2.6.b). In order to convey the main idea behind this algorithm we start by giving a short description of the method and give its important properties. Throughout this chapter the Hilbert spaces $\mathscr{U}$, $\mathscr{X}$, and $\mathscr{Y}$ are assumed to be *pivot spaces* (cf. Section 1.3).

Let $\Sigma = (\mathbb{T}, \Phi, \Psi, \mathbb{F})$ be a stable well-posed linear system. Let $x : [0, \infty) \to \mathscr{X}$ and $y \in L^2(0, \infty; \mathscr{Y})$ be respectively the state trajectory and the output function of Σ corresponding to the initial condition $x_0 \in \mathscr{X}$ and the input function $u \in L^2(0, \infty; \mathscr{U})$. This means that there holds

$$\begin{aligned} x(t) &= \mathbb{T}_t x_0 + \Phi_t u, \qquad \forall\, t \geq 0, \\ y &= \Psi x_0 + \mathbb{F} u. \end{aligned} \tag{4.1}$$

Now let us consider the cost functional (3.2) with $Q = I$, $R = I$, and $N = 0$. This means that we have

$$J(x_0, u) = \int_0^\infty \|u(t)\|^2_{\mathscr{U}} + \|y(t)\|^2_{\mathscr{Y}}\ dt. \tag{4.2}$$

The linear-quadratic optimal control problem is to find for every $x_0 \in \mathscr{X}$ an input function $u^{\text{opt}} \in L^2(0, \infty; \mathscr{U})$, which minimizes (4.2) subject to (4.1). It follows from Proposition 3.2 that for every $x_0 \in \mathscr{X}$

$$\langle x_0, X x_0 \rangle_{\mathscr{X}} = \min_{u \in L^2(0,\infty;\mathscr{U})} J(u, x_0), \tag{4.3}$$

where the Riccati operator X is given by

$$X = \Psi^*\Psi - \Psi^*\mathbb{F}(I + \mathbb{F}^*\mathbb{F})^{-1}\mathbb{F}^*\Psi. \tag{4.4}$$

Moreover, the unique minimizing input function u^{opt} satisfies

$$u^{\text{opt}} = -(I + \mathbb{F}^*\mathbb{F})^{-1}\mathbb{F}^*\Psi x_0. \tag{4.5}$$

By applying the identity $(I + \mathbb{F}^*\mathbb{F})^{-1}\mathbb{F}^* = \mathbb{F}^*(I + \mathbb{F}\mathbb{F}^*)^{-1}$ to (4.4), we can reformulate the Riccati operator as (see Proposition 3.4)

$$X = \Psi^*(I + \mathbb{F}\mathbb{F}^*)^{-1}\Psi. \tag{4.6}$$

The principal idea behind our algorithm is to find a sequence of orthogonal projectors $(P_k)_{k\in\mathbb{N}} : L^2(0, \infty; \mathscr{Y}) \to L^2(0, \infty; \mathscr{Y})$, which allow us to construct approximations of the output map Ψ and the input-output map $\mathbb{F}$. These approximations are defined by

$$\begin{aligned} &\Psi_k :\ \mathscr{X} \to L^2(0, \infty; \mathscr{Y}), && \Psi_k = P_k\Psi, \\ &\mathbb{F}_k :\ L^2(0, \infty; \mathscr{U}) \to L^2(0, \infty; \mathscr{Y}), && \mathbb{F}_k = P_k\mathbb{F}. \end{aligned}$$

Using the operators Ψ_k and $\mathbb{F}_k$, we create a sequence of approximations $(X_k)_{k\in\mathbb{N}}$ for the Riccati operator (4.6). This sequence is given by

$$X_k = \Psi_k^*(I + \mathbb{F}_k\mathbb{F}_k^*)^{-1}\Psi_k. \tag{4.7}$$

We will show in Theorem 4.11 that X_k, given by (4.7), solves the *projected optimal control problem*

$$\langle x_0, X_k x_0\rangle_{\mathscr{X}} = \min_{u\in L^2(0,\infty;\mathscr{X})} \int_0^\infty \|u(t)\|^2 + \|\left(P_k y\right)(t)\|^2 \, dt, \tag{4.8}$$

subject to (4.1).

In order to find an appropriate sequence of orthogonal projectors $(P_k)_{k\in\mathbb{N}}$, we need to introduce suitable subspaces of $L^2(0,\infty)$. These subspaces are defined in Section 4.2. In order to make the definition of these subspaces easier to understand, we explain their key structure in the following: Let $(\alpha_k)_{k=1}^\infty$ be a sequence of *shift parameters* which satisfies $\alpha_k \in \mathbb{C}$ and $\operatorname{Re}(\alpha_k) > 0$. For $k \in \mathbb{N}$ we define the following subspace of $L^2(0,\infty)$

$$\mathscr{V}_k := \operatorname{span}\{t \mapsto \mathrm{e}^{-\alpha_1 t}, \dots, t \mapsto \mathrm{e}^{-\alpha_k t}\}. \tag{4.9}$$

For notational simplicity we assume, for the moment, that the shift parameters α_k are distinct. In Section 4.2 we drop this assumption and the definition of $\mathscr{V}_k$ has to be modified in case of non-distinct shift parameters. Let $P_k : L^2(0,\infty;\mathscr{Y}) \to L^2(0,\infty;\mathscr{Y})$ denote the orthogonal projector onto $\mathscr{V}_k \otimes \mathscr{Y} \subset L^2(0,\infty;\mathscr{Y})$. It follows from $\mathscr{V}_k \subset \mathscr{V}_{k+1}$ and (4.8) that (Theorem 4.21)

$$\langle x_0, X_k x_0\rangle_{\mathscr{X}} \le \langle x_0, X_{k+1} x_0\rangle_{\mathscr{X}} \le \langle x_0, X x_0\rangle_{\mathscr{X}}, \qquad \forall x_0 \in \mathscr{X}.$$

This means that $(X_k)_{k=1}^\infty$ is a non-decreasing sequence bounded from above by X. Hence, $(X_k)_{k=1}^\infty$ converges (see Appendix A.2), but the limit may not necessarily equal X. We will show in Theorem 4.22 that the sequence of approximations X_k converges to X, provided that

$$\overline{\bigcup_{k\in\mathbb{N}} \mathscr{V}_k} = L^2(0,\infty). \tag{4.10}$$

The property (4.10) is proven in [41, Lemma 4.4] to be equivalent to the *non-Blaschke condition*

$$\sum_{j=1}^\infty \frac{\operatorname{Re}(\alpha_j)}{1+|\alpha_j|^2} = \infty. \tag{4.11}$$

We use the following structure to present our method in this chapter: We begin by introducing an orthonormal basis for the space $\mathscr{V}_k$ (the Takenaka–Malmquist system) in Section 4.2. This is used in Section 4.3 to determine the projected output maps Ψ_k and the projected input-output maps $\mathbb{F}_k$. This approximation method, called the *ADI iteration for the output map* Ψ *and the input-output map* $\mathbb{F}$, is presented in Algorithm 1. This algorithm is the crucial ingredient of all the methods that will be presented in this chapter:

- the “Riccati-ADI” method for the regular optimal control problem (Algorithm 2 in Section 4.4)
- the ADI method for the bounded real and positive real singular optimal control problems (Algorithms 3 and 4 in Section 4.6)

We prove that the approximate solutions obtained by all these algorithms express the optimal cost for a projected (regular or singular) linear-quadratic optimal control problem. Furthermore, we show that the sequence of approximate solutions obtained by these algorithms are monotonically non-decreasing. In addition, by assuming the non-Blaschke condition (4.11), we prove convergence of all these algorithms to the respective optimal costs.

Bibliographical notes:

In the finite-dimensional settings, Algorithms 1 and 2 are arithmetically equivalent to [34, Algorithm 2] (ILRSI) for solving the algebraic matrix Riccati equation. Monotonicity of the matrix sequence $(X_k)_{k\in\mathbb{N}}$ produced by ILRSI was proven in [34, Theorem 4.2] (using very different arguments when compared to our setting). An upper bound for the distance between X_k and X in the gap metric was considered in [34]. However, it was left open there, whether or not this upper bound converges to zero. In [35] a new perspective on this method in terms of the underlying linear-quadratic optimal control problem was introduced. We use this perspective throughout this chapter. This representation is independent of the Riccati equation and allows a straightforward proof of convergence. Moreover, the setting introduced in [35] allows an extension of the algorithm to infinite-dimensional systems, as shown in [35, Theorem 7.1].

In [37] the application of the ADI method for the bounded real and positive real Lur'e equations is considered. This application is based on Algorithm 1 and allows an extension to the singular optimal control problem of strongly stable well-posed systems in the bounded real and positive real case (see Algorithms 3 and 4).

4.2 Convolution and Takenaka–Malmquist systems

In this section we consider two special subspaces of $L^2(0,\infty)$, namely the *convolution system* and the *Takenaka–Malmquist system*. The application of the Takenaka–Malmquist system to approximate the input map Φ (or in the dual form the output map Ψ) was first introduced in [41]. In that paper the authors used this approximation to solve the operator Lyapunov equation for controllability (or the operator Lyapunov equation for observability in the dual form). The application of the Takenaka–Malmquist system to approximate the input-output map $\mathbb{F}$ was first introduced in [35].

The convolution system and the Takenaka–Malmquist system (Definition 4.1) are constructed by the *convolution product* of the exponential functions of the form $t \mapsto \mathrm{e}^{-\alpha_i t}$, where $\alpha_i \in \mathbb{C}$ and $\mathrm{Re}(\alpha_j) > 0$. The convolution product of two functions g and h is defined by

$$(g * h)(t) = \int_0^t g(t-\tau)h(\tau)\,d\tau.$$

Definition 4.1 (Convolution system, Takenaka–Malmquist system)**.** Let $(\alpha_j)_{j=1}^{\infty} \in \mathbb{C}$ be the *shift parameters* such that $\mathrm{Re}(\alpha_j) > 0$ for all $j \in \mathbb{N}$. We define

- the *convolution system* $(\varphi_j)_{j=1}^{\infty}$, $\varphi_j \in L^2(0,\infty)$ by

$$\begin{aligned} \varphi_1 &:= t \mapsto \mathrm{e}^{-\alpha_1 t}, \\ \varphi_j &:= \mathrm{e}^{-\alpha_j \cdot} * \varphi_{j-1}, \end{aligned} \tag{4.12}$$

and set

$$\mathscr{K}_k(\alpha) := \operatorname{span}\{\varphi_1, \varphi_2, \ldots, \varphi_k\}. \tag{4.13}$$

- the *Takenaka–Malmquist system* $(\psi_j)_{j=1}^{\infty}$, $\psi_j \in L^2(0, \infty)$ by

$$\begin{aligned} \phi_1 &= t \mapsto \mathrm{e}^{-\alpha_1 t}, & \psi_1 &= \sqrt{2\mathrm{Re}(\alpha_1)} \cdot \phi_1, \\ \phi_j &= \phi_{j-1} - (\alpha_j + \overline{\alpha_{j-1}}) \cdot (\mathrm{e}^{-\alpha_j \cdot} * \phi_{j-1}), & \psi_j &= \sqrt{2\mathrm{Re}(\alpha_j)} \cdot \phi_j. \end{aligned} \tag{4.14}$$

Remark 4.2. If we denote by $\widehat{\varphi}_j$ and $\widehat{\psi}_j$ respectively the Laplace transform of φ_j and ψ_j, then

a) For all $s \in \mathbb{C}$ with $\mathrm{Re}(s) > 0$, the Laplace transformation of (4.12) results in

$$\widehat{\varphi}_1(s) = \frac{1}{s + \alpha_1}, \qquad \widehat{\varphi}_j(s) = \frac{1}{s + \alpha_j} \cdot \widehat{\varphi}_{j-1}(s),$$

and therefore by induction

$$\widehat{\varphi}_j(s) = \prod_{\ell=1}^{j} \frac{1}{s + \alpha_\ell}. \tag{4.15}$$

b) Assume that the numbers $q_1, \ldots, q_J$ are pairwise different and there holds

$$\{q_1, \ldots, q_J\} = \{\alpha_1, \ldots, \alpha_k\}, \qquad J \le k.$$

Further, let ℓ_j be the number of times in which q_j appears in $(\alpha_j)_{j=1}^{k}$ (thus $k = \ell_1 + \ldots + \ell_J$). Then

$$\mathscr{K}_k(\alpha) = \operatorname{span}\{\varphi_1, \ldots, \varphi_k\} = \bigoplus_{j=1}^{J} \operatorname{span}\left\{ \, t \mapsto t^l \mathrm{e}^{-q_j t} \,\middle|\, l = 0, \ldots, \ell_j - 1 \, \right\}. \tag{4.16}$$

The relation (4.16) can be obtained by using the partial fraction expansions of (4.15) and eventually applying the inverse Laplace transformation.

If the numbers $\alpha_1, \ldots, \alpha_k$ are distinct, then

$$\operatorname{span}\{\varphi_1, \ldots, \varphi_k\} = \operatorname{span}\{\mathrm{e}^{-\alpha_1 \cdot}, \ldots, \mathrm{e}^{-\alpha_k \cdot}\}.$$

c) If $(\widetilde{\alpha}_j)_{j=1}^{k}$ is a permutation of $(\alpha_j)_{j=1}^{k}$ and the corresponding convolution systems are denoted respectively by $(\widetilde{\varphi}_j)_{j=1}^{k}$ and $(\varphi_j)_{j=1}^{k}$, then it follows from b that

$$\operatorname{span}\{\widetilde{\varphi}_1, \ldots, \widetilde{\varphi}_k\} = \operatorname{span}\{\varphi_1, \ldots, \varphi_k\}.$$

d) The Takenaka–Malmquist system is orthonormal (see, e.g., [41, Appendix B] for a proof).

e) For all $s \in \mathbb{C}$ with $\mathrm{Re}(s) > 0$, the Laplace transformation of (4.14) yields

$$\begin{aligned} \widehat{\phi}_1(s) &= \frac{1}{s+\alpha_1}, & \widehat{\psi}_1(s) &= \sqrt{2\mathrm{Re}(\alpha_1)} \cdot \widehat{\phi}_1(s), \\ \widehat{\phi}_j(s) &= \widehat{\phi}_{j-1}(s) - (\alpha_j + \overline{\alpha_{j-1}}) \cdot \frac{1}{s+\alpha_j} \cdot \widehat{\phi}_{j-1}(s), & \widehat{\psi}_j(s) &= \sqrt{2\mathrm{Re}(\alpha_j)} \cdot \widehat{\phi}_j(s), \end{aligned} \tag{4.17}$$

and therefore by induction

$$\widehat{\psi}_j(s) = \frac{\sqrt{2\mathrm{Re}(\alpha_j)}}{(s+\alpha_j)} \cdot \prod_{\ell=1}^{j-1} \frac{s-\overline{\alpha_\ell}}{s+\alpha_\ell}. \tag{4.18}$$

f) By using partial fraction expansions of (4.15) and (4.18), we obtain

$$\mathscr{K}_k(\alpha) = \mathrm{span}\{\psi_1, \ldots, \psi_k\}.$$

As a result, $\{\psi_1, \ldots, \psi_k\}$ is an orthonormal basis of $\mathscr{K}_k(\alpha)$.

For the rest of this section, we assume that $\Sigma = (\mathbb{T}, \Phi, \Psi, \mathbb{F})$ is a stable weakly regular linear system (cf. Definition 2.6.c). As a result, the semigroup $\mathbb{T}$ is uniformly bounded and hence by the theorem of Hille–Yosida (see for example [58, Theorem 3.4.1]), there holds $\mathbb{C}_+ \subset \rho(A)$.

Now let us determine how the operators Ψ^* and $\mathbb{F}^*$ act on the considered bases of $\mathscr{K}_k(\alpha)$. To this end, we define the following two operators for $t \geq 0$

$$\Upsilon_t : L^2(0,\infty;\mathscr{Y}) \to \mathscr{X}, \quad \Upsilon_t y := \int_t^\infty \mathbb{T}^*_{\tau-t} C^* y(\tau)\, d\tau, \tag{4.19}$$

$$\Lambda : L^2(0,\infty;\mathscr{Y}) \to L^2(0,\infty;\mathscr{X}), \quad \Lambda y := t \mapsto \int_t^\infty \mathbb{T}^*_{\tau-t} C^* y(\tau)\, d\tau. \tag{4.20}$$

We have $\Lambda y = t \mapsto \Upsilon_t y$. In addition, by Theorem 2.39, Definition 2.41, and Remark 2.42, we observe that

$$x^a(t) = \Upsilon_t y^a = \Psi^* \mathbf{S}^*_t y^a, \tag{4.21}$$

and

$$u^a = (B^*_{\Lambda w}\Lambda + D^*)\, y^a = \mathbb{F}^* y^a, \tag{4.22}$$

where y^a, x^a, and u^a are the input, state, and output of the anticausal dual system Σ^a. At this point, we are ready to give a crucial technical result which is used to show how the operators Ψ^* and $\mathbb{F}^*$ act on the convolution and Takenaka–Malmquist systems:

Lemma 4.3. *Let Υ_t be the operator defined in (4.19) for $t \geq 0$. Then for $\mu \in \rho(A)$, $v \in \mathscr{Y}$ and $y \in L^2(0,\infty;\mathscr{Y})$ there holds*

$$\Upsilon_t(\mathrm{e}^{-\mu\cdot} v) = (\mu I - A^*)^{-1} C^* v \mathrm{e}^{-\mu t}, \tag{4.23}$$

and

$$\Upsilon_t(\mathrm{e}^{-\mu\cdot} * y) = (\mu I - A^*)^{-1} C^* (\mathrm{e}^{-\mu\cdot} * y)(t) + (\mu I - A^*)^{-1} \Upsilon_t(y). \tag{4.24}$$

Proof. We follow the proof of [35, Lemma 3.5]. First, we consider (4.23). By the change of variables $\theta := \tau - t$, we obtain

$$\begin{aligned}\Upsilon_t(\mathrm{e}^{-\mu\cdot}v) = \int_t^\infty \mathbb{T}^*_{\tau-t}C^*v\mathrm{e}^{-\mu\tau}\,d\tau &= \int_0^\infty \mathbb{T}^*_\theta C^*v\mathrm{e}^{-\mu\theta}\mathrm{e}^{-\mu t}\,d\theta\\ &= \mathrm{e}^{-\mu t}\int_0^\infty \mathrm{e}^{-\mu\theta}\mathbb{T}^*_\theta C^*v\,d\theta.\end{aligned}$$

The result follows then by the definition of the Laplace transform.

Now we consider (4.24). We have

$$\begin{aligned}\Upsilon_t(\mathrm{e}^{-\mu\cdot} * y) &= \int_t^\infty \mathbb{T}^*_{\tau-t}C^*\int_0^\tau \mathrm{e}^{-\mu(\tau-\sigma)}y(\sigma)\,d\sigma d\tau\\ &= \int_t^\infty\int_0^\tau \mathrm{e}^{-\mu(\tau-t)}\mathbb{T}^*_{\tau-t}C^*\mathrm{e}^{-\mu(t-\sigma)}y(\sigma)\,d\sigma d\tau.\end{aligned}$$

By interchanging the order of integration we obtain

$$\begin{aligned}\Upsilon_t(\mathrm{e}^{-\mu\cdot} * y) &= \int_0^t\int_t^\infty \mathrm{e}^{-\mu(\tau-t)}\mathbb{T}^*_{\tau-t}C^*\mathrm{e}^{-\mu(t-\sigma)}y(\sigma)\,d\tau d\sigma\\ &\qquad + \int_t^\infty\int_\sigma^\infty \mathrm{e}^{-\mu(\tau-t)}\mathbb{T}^*_{\tau-t}C^*\mathrm{e}^{-\mu(t-\sigma)}y(\sigma)\,d\tau d\sigma\\ &= \int_0^t \Big[-(\mu I - A^*)^{-1}\mathrm{e}^{-\mu(\tau-t)}\mathbb{T}^*_{\tau-t}C^*\mathrm{e}^{-\mu(t-\sigma)}y(\sigma)\Big]_{\tau=t}^\infty\,d\sigma\\ &\qquad + \int_t^\infty \Big[-(\mu I - A^*)^{-1}\mathrm{e}^{-\mu(\tau-t)}\mathbb{T}^*_{\tau-t}C^*\mathrm{e}^{-\mu(t-\sigma)}y(\sigma)\Big]_{\tau=\sigma}^\infty\,d\sigma\\ &= \int_0^t (\mu I - A^*)^{-1}C^*\mathrm{e}^{-\mu(t-\sigma)}y(\sigma)\,d\sigma\\ &\qquad + \int_t^\infty (\mu I - A^*)^{-1}\mathrm{e}^{-\mu(\sigma-t)}\mathbb{T}^*_{\sigma-t}C^*\mathrm{e}^{-\mu(t-\sigma)}y(\sigma)\,d\sigma\\ &= (\mu I - A^*)^{-1}C^*\int_0^t \mathrm{e}^{-\mu(t-\sigma)}y(\sigma)\,d\sigma + (\mu I - A^*)^{-1}\int_t^\infty \mathbb{T}^*_{\sigma-t}C^*y(\sigma)\,d\sigma\\ &= (\mu I - A^*)^{-1}C^*\,(\mathrm{e}^{-\mu\cdot} * y)(t) + (\mu I - A^*)^{-1}\Upsilon_t(y),\end{aligned}$$

as claimed. □

As a consequence of Lemma 4.3, we obtain the following result on the action of Υ_t, Ψ^* and Λ on the convolution system.

Proposition 4.4. *Let $(\alpha_j)_{j=1}^\infty$ such that $\mathrm{Re}(\alpha_j) > 0$ for all $j \in \mathbb{N}$, $(\varphi_j)_{j=1}^\infty$ as in Definition 4.1 and $v \in \mathscr{Y}$.*

1. *With Υ_t as in (4.19) there holds*

$$\begin{aligned}&\Upsilon_t(\varphi_1 v) = (\alpha_1 I - A^*)^{-1}C^*v\varphi_1(t),\\ &\Upsilon_t(\varphi_j v) = (\alpha_j I - A^*)^{-1}C^*v\varphi_j(t) + (\alpha_j I - A^*)^{-1}\Upsilon_t(\varphi_{j-1}v).\end{aligned}$$

2. *With* Ψ *as in* (2.22) *there holds*

$$\Psi^*(\varphi_1 v) = (\alpha_1 I - A^*)^{-1} C^* v,$$
$$\Psi^*(\varphi_j v) = (\alpha_j I - A^*)^{-1} \Psi^*(\varphi_{j-1} v).$$

3. *With* Λ *as in* (4.20) *there holds*

$$\Lambda(\varphi_1 v) = (\alpha_1 I - A^*)^{-1} C^* v \varphi_1,$$
$$\Lambda(\varphi_j v) = (\alpha_j I - A^*)^{-1} C^* v \varphi_j + (\alpha_j I - A^*)^{-1} \Lambda(\varphi_{j-1} v).$$

Proof. We first prove part 1. The first formula follows directly from (4.23) with $\mu := \alpha_1$. The second formula follows from multiplying the iterative definition of $(\varphi_j)_{j=1}^{\infty}$ in (4.12) by v, applying Υ_t to the result and using that by Lemma 4.3,

$$\Upsilon_t(\mathrm{e}^{-\alpha_j \cdot} * \varphi_{j-1} v) = (\alpha_j I - A^*)^{-1} C^* v \varphi_j(t) + (\alpha_j I - A^*)^{-1} \Upsilon_t(\varphi_{j-1} v).$$

To prove part 2, we consider (4.21) for $t = 0$ and obtain that $\Psi^* = \Upsilon_0$. The result then follows from part 1 by using $\varphi_1(0) = 1$ and $\varphi_j(0) = 0$ for $j > 1$. Part 3 follows from part 1 using that $\Lambda z = t \mapsto \Upsilon_t z$. □

From Proposition 4.4 and the fact that $\mathbb{F}^* = B^*_{\Lambda w}\Lambda + D^*$ we can conclude that $\mathscr{K}_k(\alpha)$ is, in a certain sense, an invariant subspace of the adjoint input-output map. To this end, we define the following subspaces of $L^2(0,\infty;\mathscr{U})$ and $L^2(0,\infty;\mathscr{Y})$

$$\mathscr{K}_k(\alpha) \otimes \mathscr{U} := \{f(\cdot)\cdot v \mid f \in \mathscr{K}_k(\alpha),\ v \in \mathscr{U}\} \subset L^2(0,\infty;\mathscr{U}),$$
$$\mathscr{K}_k(\alpha) \otimes \mathscr{Y} := \{f(\cdot)\cdot v \mid f \in \mathscr{K}_k(\alpha),\ v \in \mathscr{Y}\} \subset L^2(0,\infty;\mathscr{Y}).$$

Corollary 4.5 shows the invariance of $\mathscr{K}_k(\alpha)$ with respect to $\mathbb{F}^*$.

Corollary 4.5. *Let* $(\alpha_j)_{j=1}^{\infty}$ *such that* $\mathrm{Re}(\alpha_j) > 0$ *for all* $j \in \mathbb{N}$*. Then for* $\mathscr{K}_k(\alpha)$ *as in* (4.13) *there holds*

$$\mathbb{F}^*(\mathscr{K}_k(\alpha) \otimes \mathscr{Y}) \subset \mathscr{K}_k(\alpha) \otimes \mathscr{U}. \tag{4.25}$$

Proof. By (2.50), we know that $\mathbb{F}^* = B^*_{\Lambda w}\Lambda + D^*$. The proof follows from Proposition 3 and regarding D as a pointwise operator $D : L^2(0,\infty;\mathscr{U}) \to L^2(0,\infty;\mathscr{Y})$ which fulfills

$$D^*(\mathscr{K}_k(\alpha) \otimes \mathscr{Y}) \subset \mathscr{K}_k(\alpha) \otimes \mathscr{U}.$$

□

Now we describe the action of Υ_t, Ψ^* and Λ on the Takenaka–Malmquist system.

Proposition 4.6. *Let* $(\alpha_j)_{j=1}^{\infty}$ *such that* $\mathrm{Re}(\alpha_j) > 0$ *for all* $j \in \mathbb{N}$, $(\phi_j)_{j=1}^{\infty}$ *and* $(\psi_j)_{j=1}^{\infty}$ *as in Definition 4.1 and* $v \in \mathscr{Y}$.

1. *For* $j > 1$ *and with* Ψ *as in* (2.22) *there holds*

$$\Psi^*(\phi_1 v) = (\alpha_1 I - A^*)^{-1} C^* v,$$
$$\Psi^*(\phi_j v) = \Psi^*(\phi_{j-1} v) - (\alpha_j + \overline{\alpha_{j-1}})(\alpha_j I - A^*)^{-1} \Psi^*(\phi_{j-1} v).$$

2. For $j > 1$ and with Λ as in (4.20) and $\gamma_j := \sqrt{\frac{\mathrm{Re}(\alpha_j)}{\mathrm{Re}(\alpha_{j-1})}}$ there holds

$$\begin{aligned}\Lambda(\psi_1 v) &= (\alpha_1 I - A^*)^{-1} C^* v \psi_1, \\ \Lambda(\psi_j v) &= \gamma_j \Lambda(\psi_{j-1} v) - \gamma_j (\alpha_j + \overline{\alpha_{j-1}}) \big[\quad (\alpha_j I - A^*)^{-1} C^* v \mathrm{e}^{-\alpha_j \cdot} * \psi_{j-1} \\ &\qquad + (\alpha_j I - A^*)^{-1} \Lambda(\psi_{j-1} v) \big].\end{aligned}$$

Proof. We first prove part 1. The first equation follows from (4.23) with $\mu := \alpha_1$ using that $\Psi^* = \Upsilon_0$. The second equation is obtained by multiplying (4.14) by v, applying Ψ^* to the result and using that by Lemma 4.3 (using that $\Psi^* = \Upsilon_0$),

$$\Psi^*(\mathrm{e}^{-\alpha_j \cdot} * \phi_{j-1} v) = (\alpha_j I - A^*)^{-1} \Psi^*(\phi_{j-1} v). \tag{4.26}$$

To prove part 2, we observe from (4.14) that there holds

$$\Lambda(\psi_j v) = \gamma_j \Lambda(\psi_{j-1} v) - \gamma_j (\alpha_j + \overline{\alpha_{j-1}}) \Lambda(\mathrm{e}^{-\alpha_j \cdot} * \psi_{j-1} v).$$

From Lemma 4.3 we get that

$$\Lambda(\mathrm{e}^{-\alpha_j \cdot} * \psi_{j-1} v) = (\alpha_j I - A^*)^{-1} C^* v \mathrm{e}^{-\alpha_j \cdot} * \psi_{j-1} + (\alpha_j I - A^*)^{-1} \Lambda(\psi_{j-1} v),$$

and the desired result follows. □

4.3 The ADI iteration for the output map and the input-output map

Using the convolution system and the Takenaka-Malmquist system presented in Section 4.2, we present an iterative method to approximate the output map Ψ and the input-output map $\mathbb{F}$. The approximation is obtained by projecting Ψ and $\mathbb{F}$ on the subspace $\mathscr{K}_k(\alpha) \otimes \mathscr{Y}$ defined in Section 4.2.

We consider the Laplace transformation of $\mathscr{K}_k(\alpha)$ (from (4.16)), $\mathscr{K}_k(\alpha) \otimes \mathscr{U}$, and $\mathscr{K}_k(\alpha) \otimes \mathscr{Y}$, namely

$$\widehat{\mathscr{K}_k}(\alpha) = \bigoplus_{j=1}^{J} \mathrm{span} \left\{ \frac{1}{(s - q_j)^l} \;\middle|\; l = 0, \dots, \ell_j - 1 \right\} \subset \mathcal{H}_2.$$

$$\begin{aligned}\widehat{\mathscr{K}_k}(\alpha) \otimes \mathscr{U} &:= \left\{ \widehat{f}(\cdot) \cdot v \mid \widehat{f} \in \widehat{\mathscr{K}_k}(\alpha),\ v \in \mathscr{U} \right\} \subset \mathcal{H}_2(\mathscr{U}), \\ \widehat{\mathscr{K}_k}(\alpha) \otimes \mathscr{Y} &:= \left\{ \widehat{f}(\cdot) \cdot v \mid \widehat{f} \in \widehat{\mathscr{K}_k}(\alpha),\ v \in \mathscr{Y} \right\} \subset \mathcal{H}_2(\mathscr{Y}),\end{aligned}$$

where $\mathcal{H}_2$ denotes the Hardy space (cf. Definition 2.22).

We define the mapping $\iota_{\mathscr{Y}_k}$ as

$$\begin{aligned}\iota_{\mathscr{Y}_k} : \qquad \mathscr{Y}^k &\to \mathscr{K}_k(\alpha) \otimes \mathscr{Y} \subset L^2(0, \infty; \mathscr{Y}), \\ (y_1, \dots, y_k) &\mapsto \sum_{j=1}^{k} \psi_j \cdot y_j,\end{aligned} \tag{4.27}$$

and similarly the mapping $\iota_{\mathscr{U}_k}$ is defined as

$$\begin{aligned} \iota_{\mathscr{U}_k}: \qquad\qquad & \mathscr{U}^k \to \mathscr{K}_k(\alpha) \otimes \mathscr{U} \subset L^2(0,\infty;\mathscr{U}), \\ & (u_1,\ldots,u_k) \mapsto \sum_{j=1}^{k} \psi_j \cdot u_j. \end{aligned} \tag{4.28}$$

It follows from orthonormality of the Takenaka-Malmquist system that $\iota_{\mathscr{Y}_k}$ and $\iota_{\mathscr{U}_k}$ define isometric embeddings. In particular,

$$\begin{aligned} P_k &:= \iota_{\mathscr{Y}_k}\iota_{\mathscr{Y}_k}^* : L^2(0,\infty;\mathscr{Y}) \mapsto L^2(0,\infty;\mathscr{Y}), \\ O_k &:= \iota_{\mathscr{U}_k}\iota_{\mathscr{U}_k}^* : L^2(0,\infty;\mathscr{U}) \mapsto L^2(0,\infty;\mathscr{U}), \end{aligned} \tag{4.29}$$

are orthogonal projectors onto $\mathscr{K}_k(\alpha) \otimes \mathscr{Y}$ and $\mathscr{K}_k(\alpha) \otimes \mathscr{U}$, respectively.

Using the orthogonal projector P_k, we define the projected output map Ψ_k and the projected input-output map $\mathbb{F}_k$ as follows

$$\Psi_k : \ \mathscr{X} \to L^2(0,\infty;\mathscr{Y}), \qquad \Psi_k = P_k\Psi, \tag{4.30}$$

$$\mathbb{F}_k : \ L^2(0,\infty;\mathscr{U}) \to L^2(0,\infty;\mathscr{Y}), \qquad \mathbb{F}_k = P_k\mathbb{F}. \tag{4.31}$$

Using the mappings $\iota_{\mathscr{Y}_k}$ and $\iota_{\mathscr{U}_k}$, we introduce the operators

$$S_k = \iota_{\mathscr{Y}_k}^*\Psi \in \mathcal{L}(\mathscr{X},\mathscr{Y}^k), \tag{4.32}$$

$$F_k = \iota_{\mathscr{Y}_k}^*\mathbb{F}\iota_{\mathscr{U}_k} \in \mathcal{L}(\mathscr{U}^k,\mathscr{Y}^k). \tag{4.33}$$

It follows from (4.30) and (4.31) that

$$\begin{aligned} \Psi_k = P_k\Psi = \iota_{\mathscr{Y}_k}\iota_{\mathscr{Y}_k}^*\Psi = \iota_{\mathscr{Y}_k}S_k, \\ \iota_{\mathscr{Y}_k}F_k = P_k\mathbb{F}\iota_{\mathscr{U}_k} = \mathbb{F}_k\iota_{\mathscr{U}_k}. \end{aligned} \tag{4.34}$$

Remark 4.7. If $\mathscr{U}$, $\mathscr{X}$, and $\mathscr{Y}$ are finite-dimensional, i.e., $\mathscr{U} = \mathbb{C}^m$, $\mathscr{X} = \mathbb{C}^n$, and $\mathscr{Y} = \mathbb{C}^p$ for some $m,n,p \in \mathbb{N}$, then S_k as in (4.32) is the matrix representation of $\Psi_k : \mathscr{X} \to \mathscr{K}_k(\alpha) \otimes \mathscr{Y}$ with respect to the basis given by the tensor product of $\{\psi_1,\ldots,\psi_k\}$ and the canonical basis of $\mathscr{Y}$. Similarly, F_k as in (4.33) is the matrix representation of $\mathbb{F}_k|_{\mathscr{K}_k(\alpha)\otimes\mathscr{U}} : \mathscr{K}_k(\alpha) \otimes \mathscr{U} \to \mathscr{K}_k(\alpha) \otimes \mathscr{Y}$ with respect to the basis given by the tensor product of $\{\psi_1,\ldots,\psi_k\}$ and the canonical basis of $\mathscr{U}$ (respectively $\mathscr{Y}$).

Algorithm 1 presents a recursive method to determine S_k and F_k. Note that determination of S_k is basically the same as in [41]. We call Algorithm 1 "*the ADI iteration for the output map* Ψ *and the input-output map* $\mathbb{F}$". Theorems 4.9 and 4.10 show that S_k and F_k in (4.32) and (4.33) are indeed the operators computed in Algorithm 1.

Remark 4.8. (a) If the input and output spaces of Σ are finite-dimensional, i.e., $\mathscr{U} = \mathbb{C}^m$ and $\mathscr{Y} = \mathbb{C}^p$ for some $m,n \in \mathbb{N}$, then Algorithm 1 provides finite rank approximations of the output map Ψ and input-output map $\mathbb{F}$. That is

$$S_k \in \mathcal{L}(\mathscr{X},\mathbb{C}^{kp}), \qquad F_k \in \mathcal{L}(\mathbb{C}^{km},\mathbb{C}^{kp}).$$

Algorithm 1 ADI iteration for the output map Ψ and the input-output map $\mathbb{F}$.

Input: $\Sigma = (\mathbb{T}, \Phi, \Psi, \mathbb{F})$ a stable weakly regular well-posed linear system with generating operators (A, B, C, D) (cf. Definition 2.6.b). Shift parameters $\alpha_1, \ldots, \alpha_k \in \mathbb{C}$ with $\mathrm{Re}(\alpha_i) > 0$.

Output: $S_k = \iota^*_{\mathscr{Y}_k}\Psi \in \mathcal{L}(\mathscr{X}, \mathscr{Y}^k)$, $F_k = \iota^*_{\mathscr{Y}_k}\mathbb{F}\iota_{\mathscr{U}_k} \in \mathcal{L}(\mathscr{U}^k, \mathscr{Y}^k)$ such that $S_k \approx \Psi$ and $F_k \approx \mathbb{F}$.

$V_1 = (\alpha_1 I - A^*)^{-1} C^*$

$S_1 = \sqrt{2\mathrm{Re}(\alpha_1)} \cdot V_1^*$

$Q_1 = \sqrt{2\mathrm{Re}(\alpha_1)} \cdot V_1^* B$

$L_1 = \frac{1}{\sqrt{2\mathrm{Re}(\alpha_1)}}$

$F_1 = Q_1 L_1 + D$

for $i = 2, 3, \ldots, k$ **do**

$\quad V_i = V_{i-1} - (\alpha_i + \overline{\alpha_{i-1}}) \cdot (\alpha_i I - A^*)^{-1} V_{i-1}$

$\quad S_i = [\, S_{i-1}^* \,,\, \sqrt{2\mathrm{Re}(\alpha_i)} \cdot V_i \,]^*$

$\quad Q_i = [\, Q_{i-1} \,,\, \sqrt{2\mathrm{Re}(\alpha_i)} \cdot V_i^* B \,]$

$\quad \gamma_i := \sqrt{\frac{\mathrm{Re}(\alpha_j)}{\mathrm{Re}(\alpha_{j-1})}}$

$$M_{i,1} := \begin{bmatrix} \frac{1}{\sqrt{2\mathrm{Re}(\alpha_1)}} & & \\ & \ddots & \\ & & \frac{1}{\sqrt{2\mathrm{Re}(\alpha_i)}} \end{bmatrix}, \quad M_{i,2} = \begin{bmatrix} \overline{\alpha_1} + \alpha_i & & & \\ \alpha_1 - \alpha_i & \overline{\alpha_2} + \alpha_i & & \\ & & \ddots & \\ & & \alpha_{i-1} - \alpha_i & \overline{\alpha_i} + \alpha_i \end{bmatrix},$$

$$M_{i,3} = \begin{bmatrix} 1 & \ldots & 1 \\ & \ddots & \vdots \\ & & 1 \end{bmatrix}, \quad M_{i,4} = \begin{bmatrix} 0 & I \\ 1 & 0 \end{bmatrix}, \quad M_{i,5} = \begin{bmatrix} -\sqrt{2\mathrm{Re}(\alpha_1)} & & & \\ & \ddots & & \\ & & -\sqrt{2\mathrm{Re}(\alpha_{i-1})} & \\ & & & 1 \end{bmatrix}$$

$\quad M_i = M_{i,1}^{-1} M_{i,2}^{-1} M_{i,3}^{-1} M_{i,4}^{-1} M_{i,5}^{-1}$

$$L_i = \begin{bmatrix} \gamma_i L_{i-1} & 0 \\ 0 & 0 \end{bmatrix} - M_i \begin{bmatrix} L_{i-1} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \gamma_i(\alpha_i + \overline{\alpha_{i-1}}) I & 0 \\ [0, \gamma_i] & -1 \end{bmatrix}$$

$$F_i = \begin{bmatrix} [F_{i-1}, 0] \\ Q_i \left(\overline{L_i} \otimes I_{\mathscr{U}}\right) + [0, D] \end{bmatrix}$$

end for

(b) In Algorithm 1 we assume that $\Sigma = (\mathbb{T}, \Phi, \Psi, \mathbb{F})$ is a stable well-posed linear system. This means that the semigroup $\mathbb{T}$ is uniformly bounded (Definition 2.6.b), i.e.,

$$\sup_{t\geq 0} \|\mathbb{T}_t\| < \infty.$$

Therefore, we have the growth bound $\omega_0(\mathbb{T}) = 0$ and the shift parameters $\alpha_1, \ldots, \alpha_k \in \mathbb{C}$ must satisfy $\mathrm{Re}(\alpha_i) > 0$. One could extend the application of Algorithm 1 to externally stable well-posed linear systems for which the semigroup $\mathbb{T}$ is not necessarily bounded. In this case the shift parameters must satisfy

$$\mathrm{Re}(\alpha_i) > \omega_0(\mathbb{T}).$$

(c) The main computational cost in Algorithm 1 consists of solving Steps 1 and 7, namely

$$\begin{aligned} V_1 &= (\alpha_1 I - A^*)^{-1} C^*, \\ V_i &= V_{i-1} - (\alpha_i + \overline{\alpha_{i-1}}) \cdot (\alpha_i I - A^*)^{-1} V_{i-1}. \end{aligned}$$

If the underlying infinite-dimensional system arises from the abstract formulation of partial differential equations, then the above equations are equivalent to solving a PDE system. This can be done efficiently by applying the adaptive refinement techniques, as will be discussed in Chapter 6.

Theorem 4.9. *Let $(\alpha_j)_{j=1}^{\infty}$ with $\mathrm{Re}(\alpha_j) > 0$ for all $j \in \mathbb{N}$. Then for all $k \in \mathbb{N}$, S_k determined by Algorithm 1 fulfills (4.32).*

Proof. By Algorithm 1, $S_k \in \mathcal{L}(\mathscr{X}, \mathscr{Y}^k)$ is the operator row matrix

$$S_k = \left[\sqrt{2\mathrm{Re}(\alpha_1)} \cdot V_1 \quad \ldots \quad \sqrt{2\mathrm{Re}(\alpha_k)} \cdot V_k\right]^*,$$

where the sequence $V_k \in \mathcal{L}(\mathscr{Y}, \mathscr{X})$ is recursively defined by

$$V_1 = (\alpha_1 I - A^*)^{-1} C^*, \qquad V_k = V_{k-1} - (\alpha_k + \overline{\alpha_{k-1}}) \cdot (\alpha_k I - A^*)^{-1} V_{k-1}. \tag{4.35}$$

The result then follows from Proposition 4.6.1 together with the definition of the Takenaka-Malmquist system in (4.14). We note that Theorem 4.9 was already established in [41], where the case $B = 0$ (for which the Riccati equation becomes a Lyapunov equation) was considered. □

Theorem 4.10. *Let $(\alpha_j)_{j=1}^{\infty}$ such that $\mathrm{Re}(\alpha_j) > 0$ for all $j \in \mathbb{N}$. Then for all $k \in \mathbb{N}$, the operator F_k determined by Algorithm 1 fulfills (4.33).*

Proof. By (4.22), the input-output map $\mathbb{F}$ fulfills

$$\mathbb{F}^* = B_{\Lambda w}^* \Lambda + D^*.$$

As a result, it follows from Proposition 4.6.2 that

$$\mathbb{F}^* \psi_1 v = \left(B_{\Lambda w}^* (\alpha_1 I - A^*)^{-1} C^* + D^*\right) v \psi_1, \qquad \forall\, v \in \mathscr{Y}.$$

By (4.33), we have for $k=1$

$$F_1^* = \iota_{\mathscr{U}_1}^* \mathbb{F}^* \iota_{\mathscr{Y}_1} \in \mathcal{L}(\mathscr{Y},\mathscr{U}),$$

and therefore

$$F_1^* = B_{\Lambda w}^*(\alpha_1 I - A^*)^{-1}C^* + D^*. \tag{4.36}$$

The adjoint of the expression in (4.36) can be calculated using the definition of $B_{\Lambda w}^*$ (cf. Definition 2.37). In fact, for $\lambda \in (\omega_0(\mathbb{T}),\infty)$, for all $v \in \mathscr{Y}$, and every $w \in \mathscr{U}$, we have

$$\begin{aligned}
\left\langle Bw(\alpha_1 I - A^*)^{-1}C^*v, w\right\rangle_{\mathscr{U}} &= \lim_{\lambda\to\infty} \left\langle B^*\lambda(\lambda I - A^*)^{-1}(\alpha_1 I - A^*)^{-1}C^*v, w\right\rangle_{\mathscr{U}} \\
&= \lim_{\lambda\to\infty} \left\langle \lambda(\lambda I - A^*)^{-1}(\alpha_1 I - A^*)^{-1}C^*v, Bw\right\rangle_{\mathscr{X}} \\
&= \lim_{\lambda\to\infty} \left\langle C^*v, \lambda(\lambda I - A)^{-1}(\overline{\alpha_1} I - A)^{-1}Bw\right\rangle_{\mathscr{X}} \\
&= \lim_{\lambda\to\infty} \left\langle v, C\lambda(\lambda I - A)^{-1}(\overline{\alpha_1} I - A)^{-1}Bw\right\rangle_{\mathscr{Y}} \\
&= C_{\Lambda w}(\overline{\alpha_1} I - A)^{-1}Bw.
\end{aligned}$$

Thus we obtain

$$F_1 = C_{\Lambda w}(\overline{\alpha_1} I - A)^{-1}B + D. \tag{4.37}$$

To obtain the structure of the operator F_k, we observe that by (4.33) there holds

$$\iota_{\mathscr{U}_{k-1}}^* \mathbb{F}^* \iota_{\mathscr{Y}_{k-1}} = F_{k-1}^* \in \mathcal{L}(\mathscr{Y}^{k-1},\mathscr{U}^{k-1}).$$

By Corollary 4.5, the following invariance holds true

$$\mathbb{F}^*\left(\mathscr{K}_{k-1}(\alpha)\otimes\mathscr{Y}\right) \subset \mathscr{K}_{k-1}(\alpha)\otimes\mathscr{U}.$$

As a result, we have

$$\iota_{\mathscr{U}_k}^* \mathbb{F}^* \iota_{\mathscr{Y}_{k-1}} = \begin{bmatrix} F_{k-1}^* \\ 0 \end{bmatrix}.$$

Therefore, we obtain that F_k has the form

$$F_k = \begin{bmatrix} [F_{k-1}, 0] \\ N_k \end{bmatrix}, \tag{4.38}$$

for some $N_k \in \mathcal{L}(\mathscr{U}^k,\mathscr{Y})$. The operator N_k has the following operator column matrix structure

$$N_k = \begin{bmatrix} N_{k1} & \cdots & N_{kk} \end{bmatrix}, \tag{4.39}$$

where $N_{ki} \in \mathcal{L}(\mathscr{U},\mathscr{Y})$. It follows from (4.34), together with (4.38) and (4.39), that

$$B_{\Lambda w}^*\Lambda(\psi_k v) + D^*\psi_k v = \mathbb{F}^*(\psi_k v) = \sum_{i=1}^{k} N_{ki}^* v \cdot \psi_i, \qquad \forall\, v \in \mathscr{Y}. \tag{4.40}$$

We set the following ansatz for $\Lambda(\psi_k v)$ (compare Proposition 4.6.1 with Proposition 4.6.2)

$$\Lambda(\psi_k v) = \sum_{j=1}^{k} \Psi^*(\psi_j v) \cdot \sum_{\ell=1}^{j} l_{j\ell,k} \cdot \psi_\ell, \qquad \forall\, v \in \mathscr{Y}, \tag{4.41}$$

for some coefficients $l_{j\ell,k} \in \mathbb{C}$.

Applying the resolvent $(\alpha_k I - A^*)^{-1}$ to (4.41) and using (4.26), we obtain that for all $v \in \mathscr{Y}$

$$\begin{aligned} (\alpha_k I - A^*)^{-1}\Lambda(\psi_k v) &= \sum_{j=1}^{k} (\alpha_k I - A^*)^{-1}\Psi^*(\psi_j v) \sum_{\ell=1}^{j} l_{j\ell,k} \cdot \psi_\ell \\ &= \sum_{j=1}^{k} \Psi^*(\mathrm{e}^{-\alpha_k \cdot} * \psi_j v) \sum_{\ell=1}^{j} l_{j\ell,k} \cdot \psi_\ell. \end{aligned} \tag{4.42}$$

In addition, by Proposition 4.6.1, we have that

$$(\alpha_k I - A^*)^{-1} C^* v = \Psi^*(\mathrm{e}^{-\alpha_k \cdot} v), \qquad \forall\, v \in \mathscr{Y}. \tag{4.43}$$

By substituting (4.42) and (4.43) in the second equation of Proposition 4.6.2, we obtain

$$\begin{aligned} \Lambda(\psi_k v) = \gamma_k \Lambda(\psi_{k-1} v) - \gamma_k(\alpha_k + \overline{\alpha_{k-1}}) \Big[&\Psi^*(\mathrm{e}^{-\alpha_k \cdot} v)\mathrm{e}^{-\alpha_k \cdot} * \psi_{k-1} \\ &+ \sum_{j=1}^{k-1} \Psi^*(\mathrm{e}^{-\alpha_k \cdot} * \psi_j v) \sum_{\ell=1}^{j} l_{j\ell,k} \cdot \psi_\ell \Big]. \end{aligned} \tag{4.44}$$

In order to write $\Lambda(\psi_k v)$ in (4.44) as a linear combination of the Takenaka-Malmquist basis $\{\psi_1, \dots, \psi_k\}$, we need a change of coordinates between the bases $\{\psi_1, \dots, \psi_{k-1}, \mathrm{e}^{-\alpha_k \cdot} * \psi_{k-1}\}$ and $\{\psi_1, \dots, \psi_k\}$, as well as a transformation between the bases $\{\mathrm{e}^{-\alpha_k \cdot} * \psi_1, \dots, \mathrm{e}^{-\alpha_k \cdot} * \psi_{k-1}, \mathrm{e}^{-\alpha_k \cdot}\}$ and $\{\psi_1, \dots, \psi_k\}$. This means that we will find invertible matrices $T_k, M_k \in \mathbb{C}^{k \times k}$ such that

$$\begin{bmatrix} \psi_1 \\ \vdots \\ \psi_{k-1} \\ \mathrm{e}^{-\alpha_k \cdot} * \psi_{k-1} \end{bmatrix} = T_k \begin{bmatrix} \psi_1 \\ \vdots \\ \psi_{k-1} \\ \psi_k \end{bmatrix}, \tag{4.45}$$

$$\begin{bmatrix} \mathrm{e}^{-\alpha_k \cdot} * \psi_1 \\ \vdots \\ \mathrm{e}^{-\alpha_k \cdot} * \psi_{k-1} \\ \mathrm{e}^{-\alpha_k \cdot} \end{bmatrix} = M_k^\top \begin{bmatrix} \psi_1 \\ \vdots \\ \psi_{k-1} \\ \psi_k \end{bmatrix}. \tag{4.46}$$

To find the matrix T_k, we use the recursive definition of the Takenaka–Malmquist basis in (4.17) to obtain

$$\begin{aligned} \mathrm{e}^{-\alpha_k \cdot} * \psi_{k-1} &= \sqrt{2\mathrm{Re}(\alpha_{k-1})}\ \mathrm{e}^{-\alpha_k \cdot} * \phi_{k-1} = \sqrt{2\mathrm{Re}(\alpha_{k-1})}\ \frac{\phi_{k-1} - \phi_k}{\alpha_k + \overline{\alpha_{k-1}}} \\ &= \frac{1}{\alpha_k + \overline{\alpha_{k-1}}} \left(\psi_{k-1} - \frac{\sqrt{2\mathrm{Re}(\alpha_{k-1})}}{\sqrt{2\mathrm{Re}(\alpha_k)}} \psi_k \right) = \frac{1}{\alpha_k + \overline{\alpha_{k-1}}} \psi_{k-1} + \frac{1}{\gamma_k(\alpha_k + \overline{\alpha_{k-1}})} \psi_k. \end{aligned}$$

As a result, we have

$$T_k = \begin{bmatrix} I_{k-1} & 0 \\ \begin{bmatrix} 0 & \frac{1}{\alpha_k + \overline{\alpha_{k-1}}} \end{bmatrix} & \frac{-1}{\gamma_k(\alpha_k + \overline{\alpha_{k-1}})} \end{bmatrix}. \tag{4.47}$$

To determine M_k, we apply the Laplace transform to (4.46) and obtain

$$\begin{bmatrix} \frac{\widehat{\psi}_1(s)}{s+\alpha_k} \\ \vdots \\ \frac{\widehat{\psi}_{k-1}(s)}{s+\alpha_k} \\ \frac{1}{s+\alpha_k} \end{bmatrix} = M_k^\top \begin{bmatrix} \widehat{\psi}_1(s) \\ \vdots \\ \widehat{\psi}_{k-1}(s) \\ \widehat{\psi}_k(s) \end{bmatrix}.$$

In the sequel, we show that

$$M_k = (M_{k,5}M_{k,4}M_{k,3}M_{k,2}M_{k,1})^{-1},$$

where

$$M_{k,1} := \begin{bmatrix} \frac{1}{\sqrt{2\mathrm{Re}(\alpha_1)}} & & \\ & \ddots & \\ & & \frac{1}{\sqrt{2\mathrm{Re}(\alpha_k)}} \end{bmatrix}, \quad M_{k,2} := \begin{bmatrix} \overline{\alpha_1} + \alpha_k & & & \\ \alpha_1 - \alpha_k & \overline{\alpha_2} + \alpha_k & & \\ & & \ddots & \\ & & \alpha_{k-1} - \alpha_k & \overline{\alpha_k} + \alpha_k \end{bmatrix},$$

$$M_{k,3} := \begin{bmatrix} 1 & \dots & 1 \\ & \ddots & \vdots \\ & & 1 \end{bmatrix}, \qquad M_{k,4} := \begin{bmatrix} 0 & I \\ 1 & 0 \end{bmatrix},$$

$$M_{k,5} := \begin{bmatrix} -\sqrt{2\mathrm{Re}(\alpha_1)} & & & \\ & \ddots & & \\ & & -\sqrt{2\mathrm{Re}(\alpha_{k-1})} & \\ & & & 1 \end{bmatrix}.$$

It follows from (4.18) that

$$\begin{aligned} E_k(s) &:= \begin{bmatrix} \frac{\widehat{\psi}_1(s)}{s+\alpha_k} & \dots & \frac{\widehat{\psi}_{k-1}(s)}{s+\alpha_k} & \frac{1}{s+\alpha_k} \end{bmatrix} \\ &= \left[\frac{\sqrt{\mathrm{Re}(\alpha_1)}}{(s+\alpha_k)(s+\alpha_1)}, \quad \dots, \quad \frac{\sqrt{2\mathrm{Re}(\alpha_{k-1})}}{(s+\alpha_k)(s+\alpha_{k-1})} \prod_{\ell=1}^{k-2} \frac{s-\overline{\alpha_\ell}}{s+\alpha_\ell}, \quad \frac{1}{s+\alpha_k} \right]. \end{aligned}$$

Consecutive application of the matrices $(M_{k,j})_{j=1}^5$ to $E_k(s)$ results in

$$
\begin{aligned}
E_k(s)M_{k,5} &= \left[\tfrac{-2\mathrm{Re}(\alpha_1)}{(s+\alpha_k)(s+\alpha_1)}, \quad \dots, \quad \tfrac{-2\mathrm{Re}(\alpha_{k-1})}{(s+\alpha_k)(s+\alpha_{k-1})}\prod_{\ell=1}^{k-2}\tfrac{s-\overline{\alpha_\ell}}{s+\alpha_\ell}, \quad \tfrac{1}{s+\alpha_k}\right].\\
E_k(s)M_{k,5}M_{k,4} &= \left[\tfrac{1}{s+\alpha_k}, \quad \tfrac{-2\mathrm{Re}(\alpha_1)}{(s+\alpha_k)(s+\alpha_1)}, \quad \dots, \quad \tfrac{-2\mathrm{Re}(\alpha_{k-1})}{(s+\alpha_k)(s+\alpha_{k-1})}\prod_{\ell=1}^{k-2}\tfrac{s-\overline{\alpha_\ell}}{s+\alpha_\ell}\right].\\
E_k(s)M_{k,5}M_{k,4}M_{k,3} &= \left[\tfrac{1}{s+\alpha_k}, \quad \tfrac{s-\overline{\alpha_1}}{(s+\alpha_k)(s+\alpha_1)}, \quad \dots, \quad \tfrac{1}{(s+\alpha_k)}\prod_{\ell=1}^{k-1}\tfrac{s-\overline{\alpha_\ell}}{s+\alpha_\ell}\right].\\
E_k(s)M_{k,5}M_{k,4}M_{k,3}M_{k,2} &= \left[\tfrac{2\mathrm{Re}(\alpha_1)}{s+\alpha_1}, \quad \tfrac{2\mathrm{Re}(\alpha_2)(s-\overline{\alpha_1})}{(s+\alpha_2)(s+\alpha_1)}, \quad \dots, \quad \tfrac{2\mathrm{Re}(\alpha_k)}{(s+\alpha_k)}\prod_{\ell=1}^{k-1}\tfrac{s-\overline{\alpha_\ell}}{s+\alpha_\ell}\right].\\
E_k(s)M_{k,5}M_{k,4}M_{k,3}M_{k,2}M_{k,1} &= \left[\tfrac{\sqrt{2\mathrm{Re}(\alpha_1)}}{s+\alpha_1}, \quad \tfrac{\sqrt{2\mathrm{Re}(\alpha_2)}(s-\overline{\alpha_1})}{(s+\alpha_2)(s+\alpha_1)}, \quad \dots, \quad \tfrac{\sqrt{2\mathrm{Re}(\alpha_k)}}{(s+\alpha_k)}\prod_{\ell=1}^{k-1}\tfrac{s-\overline{\alpha_\ell}}{s+\alpha_\ell}\right].\\
E_k(s)M_k^{-1} &= \left[\widehat{\psi}_1(s) \quad \dots \quad \widehat{\psi}_{k-1}(s) \quad \widehat{\psi}_k(s)\right].
\end{aligned}
$$

We note that the above procedure is similar to the proof of [34, Proposition 3.2].

With the help of the transformation matrices T_k and M_k, we can determine a recursive formula for the determination of the coefficients $l_{j\ell,k} \in \mathbb{C}$ in (4.41). To this end, for $v \in \mathscr{Y}$, we define the matrices

$$
\begin{aligned}
L_k &:= \begin{bmatrix} l_{11,k} & & \\ \vdots & \ddots & \\ l_{k1,k} & \dots & l_{kk,k} \end{bmatrix} \in \mathbb{C}^{k\times k},\\
\widetilde{S}_k(v) &:= \begin{bmatrix} \Psi(\psi_1 v) \\ \vdots \\ \Psi(\psi_{k-1}v) \\ \Psi(\psi_k v) \end{bmatrix} \in \mathbb{C}^{k\times n},\\
\widehat{S}_k(v) &:= \begin{bmatrix} \Psi(\mathrm{e}^{-\alpha_k \cdot} * \psi_1 v) \\ \vdots \\ \Psi(\mathrm{e}^{-\alpha_k \cdot} * \psi_{k-1}v) \\ \Psi(\mathrm{e}^{-\alpha_k \cdot} v) \end{bmatrix} \in \mathbb{C}^{k\times n},
\end{aligned}
\tag{4.48}
$$

and the formal expressions

$$
\psi := \begin{bmatrix} \psi_1 \\ \vdots \\ \psi_{k-1} \\ \psi_k \end{bmatrix}, \qquad \widetilde{\psi} := \begin{bmatrix} \psi_1 \\ \vdots \\ \psi_{k-1} \\ \mathrm{e}^{-\alpha_k \cdot} * \psi_{k-1} \end{bmatrix}.
$$

Using the matrices $\widetilde{S}_k(v)$ and L_k we can reformulate (4.41) as

$$
\Lambda(\psi_k v) = \widetilde{S}_k^*(v) L_k \psi. \tag{4.49}
$$

Further, by the definition of $\widehat{S}_k(v)$, we obtain

$$\Psi^*(\mathrm{e}^{-\alpha_k \cdot} v)\mathrm{e}^{-\alpha_k \cdot} * \psi_{k-1} + \sum_{j=1}^{k-1} \Psi^*(\mathrm{e}^{-\alpha_k \cdot} * \psi_j v) \sum_{\ell=1}^{j} l_{j\ell,k} \cdot \psi_\ell = \widehat{S}_k^*(v) \begin{bmatrix} L_{k-1} & 0 \\ 0 & 1 \end{bmatrix} \widetilde{\psi}. \quad (4.50)$$

By substituting (4.49) and (4.50) in (4.44) we obtain

$$\widetilde{S}_k^*(v) L_k \psi = \gamma_k \widetilde{S}_k^*(v) \begin{bmatrix} L_{k-1} & 0 \\ 0 & 0 \end{bmatrix} \psi + \gamma_k(\alpha_k + \overline{\alpha_{k-1}}) \widehat{S}_k^*(v) \begin{bmatrix} L_{k-1} & 0 \\ 0 & 1 \end{bmatrix} \widetilde{\psi}. \quad (4.51)$$

It follows from (4.45) and (4.46), together with the definitions of ψ, $\widetilde{\psi}$, $\widehat{S}_k^*(v)$ and $\widetilde{S}_k^*(v)$, that

$$T_k \psi = \widetilde{\psi}, \qquad \widetilde{S}_k^*(v) M_k = \widehat{S}_k^*(v). \quad (4.52)$$

Now, by substituting (4.52) in (4.51), we obtain

$$\widetilde{S}_k^*(v) L_k \psi = \gamma_k \widetilde{S}_k^*(v) \begin{bmatrix} L_{k-1} & 0 \\ 0 & 0 \end{bmatrix} \psi + \gamma_k(\alpha_k + \overline{\alpha_{k-1}}) \widetilde{S}_k^*(v) M_k \begin{bmatrix} L_{k-1} & 0 \\ 0 & 1 \end{bmatrix} T_k \psi.$$

The matrix L_k can therefore be recursively determined by

$$\begin{aligned} L_1 &= \frac{1}{\sqrt{2\mathrm{Re}(\alpha_1)}}, \\ L_k &= \begin{bmatrix} \gamma_k L_{k-1} & 0 \\ 0 & 0 \end{bmatrix} + \gamma_k(\alpha_k + \overline{\alpha_{k-1}}) M_k \begin{bmatrix} L_{k-1} & 0 \\ 0 & 1 \end{bmatrix} T_k, \qquad k = 2, 3, \dots, \end{aligned} \quad (4.53)$$

where we can determine L_1 by considering (4.49) for $k = 1$

$$\Lambda(\psi_1 v) = \widetilde{S}_1^*(v) L_1 \psi = \Psi^*(\psi_1 v) L_1 \psi_1,$$

and using that by Proposition 4.6, we have

$$\Lambda(\psi_1 v) = (\alpha_1 I - A^*)^{-1} C^* v \psi_1, \qquad \Psi^*(\psi_1 v) = \sqrt{2\mathrm{Re}(\alpha_1)}\, (\alpha_1 I - A^*)^{-1} C^* v.$$

Note that (4.53) includes Steps 4 and 13 in Algorithm 1.

It follows from (4.40) and (4.49) that

$$B_{\Lambda w}^* \widetilde{S}_k^*(v) L_k \psi + [0, D^* v] \psi = \begin{bmatrix} N_{k1}^* v & \cdots & N_{kk}^* v \end{bmatrix} \psi, \qquad \forall\, v \in \mathscr{Y},$$

where $[0, D^*] \in \mathcal{L}(\mathscr{Y}^k, \mathscr{U})$. By equating the coefficients in the above relation and denoting the i-th unit vector by $e_i \in \mathbb{C}^k$, we obtain

$$\begin{aligned} N_{ki}^* v &= B_{\Lambda w}^* \widetilde{S}_k^*(v) L_k e_i, \qquad \text{for} \quad i = 1, 2, \dots, k-1, \\ N_{kk}^* v &= B_{\Lambda w}^* \widetilde{S}_k^*(v) L_k e_k + D^* v, \qquad \forall\, v \in \mathscr{Y}. \end{aligned}$$

By substituting the definition of the matrices $\widetilde{S}_k(v)$ and L_k from (4.48), the above equations result in

$$\begin{aligned} N_{ki}^* v &= \sum_{j=i}^{k} B_{\Lambda w}^* \Psi^*(\psi_j v) \cdot l_{ji,k}, \qquad \text{for} \quad i = 1, 2, \dots, k-1, \\ N_{kk}^* v &= B_{\Lambda w}^* \Psi^*(\psi_k v) \cdot l_{kk,k} + D^* v, \qquad \forall\, v \in \mathscr{Y}. \end{aligned} \tag{4.54}$$

From Proposition 4.6.1 and the definition of the operators $(V_j)_{j=1}^k$ in (4.35), we obtain that

$$\Psi^*(\psi_j v) = \sqrt{2\mathrm{Re}(\alpha_j)} \cdot V_j\ v, \qquad \forall\, v \in \mathscr{Y}.$$

We substitute this into (4.54) to get

$$\begin{aligned} N_{ki}^* v &= \sum_{j=i}^{k} \sqrt{2\mathrm{Re}(\alpha_j)} \cdot B_{\Lambda w}^* V_j\ l_{ji,k}\ v, \qquad \text{for} \quad i = 1, 2, \dots, k-1, \\ N_{kk}^* v &= \left(\sqrt{2\mathrm{Re}(\alpha_k)} \cdot B_{\Lambda w}^* V_k\ l_{kk,k}\ + D^* \right) v, \qquad \forall\, v \in \mathscr{Y}. \end{aligned} \tag{4.55}$$

The adjoint of the expressions in (4.55) can be calculated by using the definition of $B_{\Lambda w}^*$ (cf. Definition 2.37). In fact, for $\lambda \in (\omega_0(\mathbb{T}), \infty)$, for all $v \in \mathscr{Y}$, and every $w \in \mathscr{U}$ we have

$$\begin{aligned} \left\langle \sqrt{2\mathrm{Re}(\alpha_k)} \cdot B_{\Lambda w}^* V_k\ l_{kk,k}\ v, w \right\rangle_{\mathscr{U}} &= \sqrt{2\mathrm{Re}(\alpha_k)} \lim_{\lambda \to \infty} \left\langle B^* \lambda (\lambda I - A^*)^{-1} V_k\ l_{kk,k}\ v, w \right\rangle_{\mathscr{U}} \\ &= \sqrt{2\mathrm{Re}(\alpha_k)} \lim_{\lambda \to \infty} \left\langle \lambda (\lambda I - A^*)^{-1} V_k\ l_{kk,k}\ v, Bw \right\rangle_{\mathscr{X}} \\ &= \sqrt{2\mathrm{Re}(\alpha_k)} \lim_{\lambda \to \infty} \left\langle V_k\ l_{kk,k}\ v, \lambda (\lambda I - A)^{-1} Bw \right\rangle_{\mathscr{X}} \\ &= \sqrt{2\mathrm{Re}(\alpha_k)} \left\langle V_k\ l_{kk,k}\ v, Bw \right\rangle_{\mathscr{X}} \\ &= \sqrt{2\mathrm{Re}(\alpha_k)} \left\langle v, \overline{l_{kk,k}}\ V_k^* Bw \right\rangle_{\mathscr{Y}}. \end{aligned}$$

As a result, by taking the adjoint of the expressions in (4.55) we obtain

$$\begin{aligned} \langle v, N_{ki} w \rangle_{\mathscr{Y}} &= \sum_{j=i}^{k} \sqrt{2\mathrm{Re}(\alpha_j)} \left\langle v, \overline{l_{ji,k}}\ V_j^* Bw \right\rangle_{\mathscr{Y}}, \qquad \text{for} \quad i = 1, 2, \dots, k-1, \\ \langle v, N_{kk} w \rangle_{\mathscr{Y}} &= \sqrt{2\mathrm{Re}(\alpha_k)} \left\langle v, \overline{l_{kk,k}}\ V_k^* Bw \right\rangle_{\mathscr{Y}} + \langle v, D^* w \rangle_{\mathscr{Y}}, \qquad \forall\, v \in \mathscr{Y},\ \forall\, w \in \mathscr{U}. \end{aligned}$$

Hence, we can conclude that for $i = 1, 2, \dots, k-1$

$$\begin{aligned} N_{ki} &= \sum_{j=i}^{k} \sqrt{2\mathrm{Re}(\alpha_j)} \cdot V_j^* B \cdot \overline{l_{ji,k}}\ I_{\mathscr{U}} \\ N_{kk} &= \sqrt{2\mathrm{Re}(\alpha_k)} \cdot V_k^* B \cdot \overline{l_{kk,k}}\ I_{\mathscr{U}} + D. \end{aligned}$$

Using the definition of L_k, this is equivalent to

$$\begin{aligned} N_{ki} &= \left[\sqrt{2\mathrm{Re}(\alpha_1)} \cdot V_1^* B \quad \cdots \quad \sqrt{2\mathrm{Re}(\alpha_k)} \cdot V_k^* B \right] \left(\left(\overline{L_k} e_i \right) \otimes I_{\mathscr{U}} \right), \\ N_{kk} &= \left[\sqrt{2\mathrm{Re}(\alpha_1)} \cdot V_1^* B \quad \cdots \quad \sqrt{2\mathrm{Re}(\alpha_k)} \cdot V_k^* B \right] \left(\left(\overline{L_k} e_k \right) \otimes I_{\mathscr{U}} \right) + D. \end{aligned}$$

In other words, for $N_k = \begin{bmatrix} N_{k1} & \cdots & N_{kk} \end{bmatrix}$ we can conclude that

$$N_k = \begin{bmatrix} \sqrt{2\mathrm{Re}(\alpha_1)} \cdot V_1^* B & \cdots & \sqrt{2\mathrm{Re}(\alpha_k)} \cdot V_k^* B \end{bmatrix} \left(\overline{L_k} \otimes I_{\mathscr{U}} \right) + [0, D].$$

As a result, the operator N_k can be determined recursively as

$$\begin{aligned} Q_1 &:= \sqrt{2\mathrm{Re}(\alpha_1)} \cdot V_1^* B, \\ Q_k &:= \begin{bmatrix} Q_{k-1} & \sqrt{2\mathrm{Re}(\alpha_k)} \cdot V_k^* B \end{bmatrix}, \\ N_k &= Q_k \left(\overline{L_k} \otimes I_{\mathscr{U}} \right) + [0, D], \end{aligned}$$

with $[0, D] \in \mathcal{L}(\mathscr{U}^k, \mathscr{Y})$. It follows from (4.37) and (4.38) that F_k is recursively defined by

$$F_1 = C(\overline{\alpha_1} I - A)^{-1} B + D = \frac{1}{\sqrt{2\mathrm{Re}(\alpha_1)}} V_1^* B + D = Q_1 L_1 + D,$$

$$F_k = \begin{bmatrix} [F_{k-1}, 0] \\ N_k \end{bmatrix} = \begin{bmatrix} [F_{k-1}, 0] \\ Q_k \left(\overline{L_k} \otimes I_{\mathscr{U}} \right) + [0, D] \end{bmatrix}.$$

Note that these correspond to Steps 3, 5, 9, and 14 in Algorithm 1. □

4.4 The projected regular optimal control problem (Riccati-ADI)

In this section we consider the projected optimal control problem and show that the solution to this problem can be formulated using the operators F_k and S_k, which are calculated by Algorithm 1. Following [36], we call this method the *Riccati-ADI* algorithm.

Recall the definition of the orthogonal projector P_k, the projected output map Ψ_k, and the projected input-output map $\mathbb{F}_k$ from (4.29), (4.30), and (4.31), respectively. By applying P_k to the cost functional (3.2), we can define the projected optimal control problem by

$$\langle X_k x_0, x_0 \rangle_{\mathscr{X}} = \min_{u \in L^2(0,\infty;\mathscr{U})} \int_0^\infty \left\langle \begin{pmatrix} (P_k y)(\tau) \\ u(\tau) \end{pmatrix}, \begin{bmatrix} Q & N^* \\ N & R \end{bmatrix} \begin{pmatrix} (P_k y)(\tau) \\ u(\tau) \end{pmatrix} \right\rangle_{\mathscr{Y} \times \mathscr{U}} d\tau, \quad (4.56)$$

where u, y, and x_0 are the input function, output function, and the initial state of the stable weakly regular linear system Σ. $X_k \in \mathcal{L}(\mathscr{X})$ is called the *projected Riccati operator*.

We define the *projected Popov operator* by $\mathscr{R}_k : L^2(0,\infty;\mathscr{U}) \to L^2(0,\infty;\mathscr{U})$,

$$\mathscr{R}_k = R + N\mathbb{F}_k + \mathbb{F}_k^* N^* + \mathbb{F}_k^* Q \mathbb{F}_k. \quad (4.57)$$

Since $\mathbb{F}_k|_{\mathscr{K}_k(\alpha)\otimes\mathscr{U}} \in \mathcal{L}(\mathscr{K}_k(\alpha) \otimes \mathscr{U}, \mathscr{K}_k(\alpha) \otimes \mathscr{Y})$, we have that

$$\mathscr{R}_k|_{\mathscr{K}_k(\alpha)\otimes\mathscr{U}} \in \mathcal{L}(\mathscr{K}_k(\alpha) \otimes \mathscr{U}).$$

It follows from $\mathscr{R} \geq \varepsilon \cdot I_{\mathscr{U}}$ that $\mathscr{R}_k \geq \varepsilon \cdot I_{\mathscr{U}}$. Hence, $\mathscr{R}_k$ is boundedly invertible and we have

$$\mathscr{R}_k^{-1}|_{\mathscr{K}_k(\alpha)\otimes\mathscr{U}} \in \mathcal{L}(\mathscr{K}_k(\alpha) \otimes \mathscr{U}). \quad (4.58)$$

In the following, we consider the projected version of Proposition 3.2. Later, we will show that the operators F_k and S_k in Algorithm 1 indeed provide the solution of (4.56).

Theorem 4.11. *Let $(\alpha_j)_{j=1}^{\infty}$ such that $\operatorname{Re}(\alpha_j) > 0$ for all $j \in \mathbb{N}$. Define Ψ_k and $\mathbb{F}_k$ by (4.30) and (4.31). Then, the unique minimizer of the optimal control problem (4.56) is given by*

$$u_k^{\text{opt}} = -\mathscr{R}_k^{-1} \left(\mathbb{F}_k^* Q + N\right) \Psi_k x_0, \tag{4.59}$$

and the optimal cost is given by $\langle X_k x_0, x_0 \rangle_{\mathscr{X}}$ with the projected Riccati operator

$$X_k = \Psi_k^* Q \Psi_k - \Psi_k^* \left(Q \mathbb{F}_k + N^*\right) \mathscr{R}_k^{-1} \left(\mathbb{F}_k^* Q + N\right) \Psi_k. \tag{4.60}$$

Proof. Noting that

$$P_k y = \Psi_k x_0 + \mathbb{F}_k u,$$

we use a "completion of the square" formula similar to [70, Proposition 7.2]. This means that

$$\begin{aligned}
J(x_0, u) &= \left\langle \begin{pmatrix} P_k y \\ u \end{pmatrix}, \begin{bmatrix} Q & N^* \\ N & R \end{bmatrix} \begin{pmatrix} P_k y \\ u \end{pmatrix} \right\rangle_{L^2(0,\infty;\mathscr{Y}\times\mathscr{U})} \\
&= \langle \Psi_k x_0 + \mathbb{F}_k u, Q\left(\Psi_k x_0 + \mathbb{F}_k u\right)\rangle + \langle \Psi_k x_0 + \mathbb{F}_k u, N^* u\rangle \\
&\quad + \langle u, N\left(\Psi_k x_0 + \mathbb{F}_k u\right)\rangle + \langle u, Ru\rangle \\
&= \left\langle \left[\Psi_k^* Q \Psi_k - \Psi_k^* \left(Q\mathbb{F}_k + N^*\right) \mathscr{R}_k^{-1} \left(\mathbb{F}_k^* Q + N\right) \Psi_k\right] x_0, x_0 \right\rangle_{\mathscr{X}} \\
&\quad + \left\langle \mathscr{R}_k \left[u + \mathscr{R}_k^{-1} \left(\mathbb{F}_k^* Q + N\right) \Psi_k x_0\right], u + \mathscr{R}_k^{-1} \left(\mathbb{F}_k^* Q + N\right) \Psi_k x_0 \right\rangle_{L^2(0,\infty;\mathscr{U})}.
\end{aligned}$$

In particular, for

$$X_k = \Psi_k^* Q \Psi_k - \Psi_k^* \left(Q \mathbb{F}_k + N^*\right) \mathscr{R}_k^{-1} \left(\mathbb{F}_k^* Q + N\right) \Psi_k,$$

we have that $J(x_0, u) \geq \langle X_k x_0, x_0\rangle$. In the case where the input reads

$$u = -\mathscr{R}_k^{-1} \left(\mathbb{F}_k^* Q + N\right) \Psi_k x_0,$$

the second summand vanishes. Thus, we have equality between $J(x_0, u)$ and the quadratic form $\langle X_k x_0, x_0\rangle$ in this case. ☐

Corollary 4.12. *Under the assumptions and with the notation of Theorem 4.11, we have*

$$u_k^{\text{opt}} \in \mathscr{K}_k(\alpha) \otimes \mathscr{U}.$$

Proof. By (4.31) we know that $\mathbb{F}_k = P_k \mathbb{F}$. As a result, (4.59) can be written as

$$u_k^{\text{opt}} = \mathscr{R}_k^{-1} \left(\mathbb{F}^* z + w\right),$$

where

$$z := -P_k Q \Psi_k x_0 \in \mathscr{K}_k(\alpha) \otimes \mathscr{Y}, \qquad w := -N \Psi_k x_0 \in \mathscr{K}_k(\alpha) \otimes \mathscr{U}.$$

From Corollary 4.5 we see that $\mathbb{F}^*$ maps $\mathscr{K}_k(\alpha) \otimes \mathscr{Y}$ into $\mathscr{K}_k(\alpha) \otimes \mathscr{U}$. Therefore

$$\mathbb{F}^* z + w \in \mathscr{K}_k(\alpha) \otimes \mathscr{U}.$$

From (4.58) we know that $\mathscr{R}_k^{-1}|_{\mathscr{K}_k(\alpha)\otimes\mathscr{U}} \in \mathcal{L}(\mathscr{K}_k(\alpha) \otimes \mathscr{U})$. As a consequence, $u_k^{\text{opt}} \in \mathscr{K}_k(\alpha) \otimes \mathscr{U}$, as desired. ☐

In the sequel we show that the projected Riccati operator X_k as in (4.60), can be indeed calculated using the operators S_k and F_k determined by Algorithm 1. We start by presenting two lemmas which will be helpful in proving the last theorem of this section.

For Hilbert spaces H_1, H_2, an operator $T \in \mathcal{L}(H_1, H_2)$, and the identity matrix $I_k \in \mathbb{R}^{k\times k}$, we define the operator matrix $I_k \otimes T \in \mathcal{L}(H_1^k, H_2^k)$ by

$$I_k \otimes T := \begin{bmatrix} T & & & \\ & T & & \\ & & \ddots & \\ & & & T \end{bmatrix} \in \mathcal{L}(H_1^k, H_2^k). \tag{4.61}$$

Lemma 4.13. *Let $\iota_{\mathscr{Y}_k}$ and $\iota_{\mathscr{U}_k}$ be the mappings defined in (4.27) and (4.28), respectively. Then for operators $Q \in \mathcal{L}(\mathscr{Y})$, $R \in \mathcal{L}(\mathscr{U})$, and $N \in \mathcal{L}(\mathscr{Y}, \mathscr{U})$, there holds*

$$\begin{aligned} Q\iota_{\mathscr{Y}_k} &= \iota_{\mathscr{Y}_k}(I_k \otimes Q), & R\iota_{\mathscr{U}_k} &= \iota_{\mathscr{U}_k}(I_k \otimes R),\\ N^*\iota_{\mathscr{U}_k} &= \iota_{\mathscr{Y}_k}(I_k \otimes N^*), & N\iota_{\mathscr{Y}_k} &= \iota_{\mathscr{U}_k}(I_k \otimes N^*). \end{aligned}$$

Proof. For $\mathbf{y} := (y_1, y_2, \dots, y_k) \in \mathscr{Y}^k$, we use the definition of $\iota_{\mathscr{Y}_k}$ and $\iota_{\mathscr{U}_k}$ in (4.27) and (4.28) to obtain

$$Q\iota_{\mathscr{Y}_k}\mathbf{y} = Q\sum_{i=1}^k \psi_i y_i = \sum_{i=1}^k Q\psi_i y_i = \sum_{i=1}^k \psi_i Q y_i = \iota_{\mathscr{Y}_k}(I_k \otimes Q)\mathbf{y}.$$

The relationships for $R\iota_{\mathscr{U}_k}$ and $N^*\iota_{\mathscr{U}_k}$ can be proven analogously. $\square$

Lemma 4.14. *Let P_k and O_k be the orthogonal projectors onto $\mathscr{K}_k(\alpha)\otimes\mathscr{Y}$ and $\mathscr{K}_k(\alpha)\otimes\mathscr{U}$, respectively (as defined in (4.29)). Let the operator $\mathcal{R}_k \in \mathcal{L}(\mathscr{U}^k)$ be defined by*

$$\mathcal{R}_k = \iota_{\mathscr{U}_k}^* \mathscr{R}_k \iota_{\mathscr{U}_k}. \tag{4.62}$$

Then the following relations hold true:

$$\mathcal{R}_k = (I_k \otimes R) + (I_k \otimes N)F_k + F_k^*(I_k \otimes N^*) + F_k^*(I_k \otimes Q)F_k, \tag{4.63}$$

$$\iota_{\mathscr{U}_k}\mathcal{R}_k^{-1}\iota_{\mathscr{U}_k}^* = O_k\mathscr{R}_k^{-1}O_k. \tag{4.64}$$

Proof. To show (4.63), we use Lemma 4.13 and the definition of $\mathbb{F}_k$ in (4.33) to obtain that

$$\begin{aligned} \mathcal{R}_k &= \iota_{\mathscr{U}_k}^* \mathscr{R}_k \iota_{\mathscr{U}_k}\\ &= \iota_{\mathscr{U}_k}^* \left(R + N\mathbb{F}_k + \mathbb{F}_k^* N^* + \mathbb{F}_k^* Q \mathbb{F}_k\right)\iota_{\mathscr{U}_k}\\ &= \iota_{\mathscr{U}_k}^* R\iota_{\mathscr{U}_k} + \iota_{\mathscr{U}_k}^* N P_k \mathbb{F}\iota_{\mathscr{U}_k} + \iota_{\mathscr{U}_k}^* \mathbb{F} P_k N^* \iota_{\mathscr{U}_k} + \iota_{\mathscr{U}_k}^* \mathbb{F}^* P_k Q P_k \mathbb{F}\iota_{\mathscr{U}_k}\\ &= \iota_{\mathscr{U}_k}^* R\iota_{\mathscr{U}_k} + \iota_{\mathscr{U}_k}^* N \iota_{\mathscr{Y}_k}\iota_{\mathscr{Y}_k}^* \mathbb{F}\iota_{\mathscr{U}_k} + \iota_{\mathscr{U}_k}^* \mathbb{F}_k \iota_{\mathscr{Y}_k}\iota_{\mathscr{Y}_k}^* N^* \iota_{\mathscr{U}_k} + \iota_{\mathscr{U}_k}^* \mathbb{F}^* \iota_{\mathscr{Y}_k}\iota_{\mathscr{Y}_k}^* Q \iota_{\mathscr{Y}_k}\iota_{\mathscr{Y}_k}^* \mathbb{F}\iota_{\mathscr{U}_k}\\ &= (I_k \otimes R) + (I_k \otimes N)\iota_{\mathscr{Y}_k}^* \mathbb{F}\iota_{\mathscr{U}_k} + \iota_{\mathscr{U}_k}^* \mathbb{F}^* \iota_{\mathscr{Y}_k}(I_k \otimes N^*) + \iota_{\mathscr{U}_k}^* \mathbb{F}^* \iota_{\mathscr{Y}_k}\iota_{\mathscr{Y}_k}^* Q \iota_{\mathscr{Y}_k}\iota_{\mathscr{Y}_k}^* \mathbb{F}\iota_{\mathscr{U}_k}\\ &= (I_k \otimes R) + (I_k \otimes N)\iota_{\mathscr{Y}_k}^* \mathbb{F}\iota_{\mathscr{U}_k} + \iota_{\mathscr{U}_k}^* \mathbb{F}^* \iota_{\mathscr{Y}_k}(I_k \otimes N^*) + \iota_{\mathscr{U}_k}^* \mathbb{F}^* \iota_{\mathscr{Y}_k}(I_k \otimes Q)\iota_{\mathscr{Y}_k}^* \mathbb{F}\iota_{\mathscr{U}_k}\\ &= (I_k \otimes R) + (I_k \otimes N)F_k + F_k^*(I_k \otimes N^*) + F_k^*(I_k \otimes Q)F_k. \end{aligned} \tag{4.65}$$

To prove (4.64), we use the definition of $\mathcal{R}_k$ in (4.62) to obtain that

$$\iota_{\mathscr{U}_k}\mathcal{R}_k\iota_{\mathscr{U}_k}^* = \iota_{\mathscr{U}_k}\iota_{\mathscr{U}_k}^*\mathscr{R}_k\iota_{\mathscr{U}_k}\iota_{\mathscr{U}_k}^* = O_k\mathscr{R}_kO_k.$$

Hence, we have

$$\iota_{\mathscr{U}_k}\mathcal{R}_k^{-1}\iota_{\mathscr{U}_k}^* = O_k\mathscr{R}_k^{-1}O_k.$$

□

Remark 4.15. If the input space $\mathscr{U}$ is finite-dimensional, i.e., $\mathscr{U} = \mathbb{C}^m$ for some $m \in \mathbb{N}$, then $\mathcal{R}_k \in \mathcal{L}(\mathscr{U}^k)$ is the matrix representation of the Popov operator $\mathscr{R}$.

Now we are ready to provide an approximation of the Riccati operator (3.10) using the ADI method, as presented in Algorithm 2. We call this algorithm the "Riccati-ADI" method.

Algorithm 2 ADI iteration for the Riccati operator (Riccati-ADI).

Input: $\Sigma = (\mathbb{T}, \Phi, \Psi, \mathbb{F})$ a stable weakly regular well-posed linear system with generating operators (A, B, C, D) (cf. Definition 2.6.b). Cost functional (3.2) such that the associated Popov operator $\mathscr{R}$ is coercive. Shift parameters $\alpha_1, \dots, \alpha_k \in \mathbb{C}$ with $\mathrm{Re}(\alpha_i) > 0$.
Output: $X_k \in \mathcal{L}(\mathscr{X})$ such that $X_k \approx X$, where X is the Riccati operator (3.10).

1: Perform Algorithm 1 to obtain the operators F_k and S_k for some $k \in \mathbb{N}$

2: $\mathcal{R}_k = (I_k \otimes R) + (I_k \otimes N)F_k + F_k^*(I_k \otimes N^*) + F_k^*(I_k \otimes Q)F_k$

3: $X_k = S_k^*(I_k \otimes Q)S_k - S_k^*\left[(I_k \otimes Q)F_k + (I_k \otimes N^*)\right]\mathcal{R}_k^{-1}\left[F_k^*(I_k \otimes Q) + (I_k \otimes N)\right]S_k$

Theorem 4.16 (Riccati-ADI)**.** *Let $\Sigma = (\mathbb{T}, \Phi, \Psi, \mathbb{F})$ be a stable weakly regular linear system (cf. Definition 2.6.b) and $(\alpha_j)_{j=1}^\infty$ such that $\mathrm{Re}(\alpha_j) > 0$ for all $j \in \mathbb{N}$. Then for all $k \in \mathbb{N}$, the operator X_k in (4.60) can be determined by*

$$X_k = S_k^*(I_k \otimes Q)S_k - S_k^*\left[(I_k \otimes Q)F_k + (I_k \otimes N^*)\right]\mathcal{R}_k^{-1}\left[F_k^*(I_k \otimes Q) + (I_k \otimes N)\right]S_k, \quad (4.66)$$

where the operators F_k and S_k are determined by Algorithm 1 and $\mathcal{R}_k$ is calculated by

$$\mathcal{R}_k = (I_k \otimes R) + (I_k \otimes N)F_k + F_k^*(I_k \otimes N^*) + F_k^*(I_k \otimes Q)F_k.$$

Proof. Using Lemma 4.13 and Lemma 4.14, together with the definition of S_k and F_k in

(4.32) and (4.33), we obtain

$$
\begin{aligned}
X_k &= \Psi_k^* Q \Psi_k - \Psi_k^* \left(Q\mathbb{F}_k + N^*\right) \mathscr{R}_k^{-1} \left(\mathbb{F}_k^* Q + N\right) \Psi_k \\
&= S_k^* \iota_{\mathscr{Y}_k}^* Q \iota_{\mathscr{Y}_k} S_k - S_k^* \iota_{\mathscr{Y}_k}^* \left(Q\mathbb{F}_k + N^*\right) \mathscr{R}_k^{-1} \left(\mathbb{F}_k^* Q + N\right) \iota_{\mathscr{Y}_k} S_k \\
&= S_k^* \iota_{\mathscr{Y}_k}^* \iota_{\mathscr{Y}_k} (I_k \otimes Q) S_k \\
&\qquad - S_k^* \left[(I_k \otimes Q)\iota_{\mathscr{Y}_k}^* \mathbb{F}_k + (I_k \otimes N^*)\iota_{\mathscr{U}_k}^*\right] \mathscr{R}_k^{-1} \left[\mathbb{F}_k^* \iota_{\mathscr{Y}_k}(I_k \otimes Q) + \iota_{\mathscr{U}_k}(I_k \otimes N)\right] S_k \\
&= S_k^* (I_k \otimes Q) S_k \\
&\qquad - S_k^* \left[(I_k \otimes Q)\iota_{\mathscr{Y}_k}^* \mathbb{F}_k + (I_k \otimes N^*)\iota_{\mathscr{U}_k}^*\right] O_k \mathscr{R}_k^{-1} O_k \left[\mathbb{F}_k^* \iota_{\mathscr{Y}_k}(I_k \otimes Q) + \iota_{\mathscr{U}_k}(I_k \otimes N)\right] S_k \\
&= S_k^* (I_k \otimes Q) S_k \\
&\qquad - S_k^* \left[(I_k \otimes Q)\iota_{\mathscr{Y}_k}^* \mathbb{F}_k + (I_k \otimes N^*)\iota_{\mathscr{U}_k}^*\right] \iota_{\mathscr{U}_k} \mathcal{R}_k^{-1} \iota_{\mathscr{U}_k}^* \left[\mathbb{F}_k^* \iota_{\mathscr{Y}_k}(I_k \otimes Q) + \iota_{\mathscr{U}_k}(I_k \otimes N)\right] S_k \\
&= S_k^* (I_k \otimes Q) S_k \\
&\qquad - S_k^* \left[(I_k \otimes Q) F_k + (I_k \otimes N^*)\right] \mathcal{R}_k^{-1} \left[F_k^* (I_k \otimes Q) + (I_k \otimes N)\right] S_k.
\end{aligned}
$$

□

In [35] the projected optimal control problem is formulated using the projected version of Proposition 3.4 from Chapter 3. All the results presented in this section (with analogous proofs) can be reformulated using the *complementary Popov operator*.

Theorem 4.17. *Let $Q = Q^* \in \mathcal{L}(\mathscr{Y})$, $R = R^* \in \mathcal{L}(\mathscr{U})$, and $\widetilde{R} := R - NQ^{-1}N^* \in \mathcal{L}(\mathscr{U})$ be all invertible operators. Let $(\alpha_j)_{j=1}^{\infty}$ such that $\mathrm{Re}(\alpha_j) > 0$ for all $j \in \mathbb{N}$. Define Ψ_k and $\mathbb{F}_k$ by (4.30) and (4.31), where $P_k : L^2(0,\infty;\mathscr{Y}) \to L^2(0,\infty;\mathscr{Y})$ is the orthogonal projector onto $\mathscr{K}_k(\alpha) \otimes \mathscr{Y}$ with $\mathscr{K}_k(\alpha)$ as in Definition 4.1. In addition, define the* projected complementary Popov operator *by $\mathscr{R}_{c,k} :\ L^2(0,\infty;\mathscr{Y}) \to L^2(0,\infty;\mathscr{Y})$*

$$
\mathscr{R}_{c,k} = Q^{-1} + \left(\mathbb{F}_k + Q^{-1}N^*\right) \left(R - NQ^{-1}N^*\right)^{-1} \left(\mathbb{F}_k^* + NQ^{-1}\right). \tag{4.67}
$$

The unique minimizer of the optimal control problem (4.56) is given by

$$
u_k^{\mathrm{opt}} = -\widetilde{R}^{-1} \left(\mathbb{F}_k^* + NQ^{-1}\right) \mathscr{R}_{c,k}^{-1} \Psi_k x_0.
$$

The optimal cost is given by $\langle X_k x_0, x_0 \rangle$ with

$$
X_k = \Psi_k^* \mathscr{R}_{c,k}^{-1} \Psi_k. \tag{4.68}
$$

Lemma 4.18. *Let $Q = Q^* \in \mathcal{L}(\mathscr{Y})$, $R = R^* \in \mathcal{L}(\mathscr{U})$, and $\widetilde{R} := R - NQ^{-1}N^* \in \mathcal{L}(\mathscr{U})$ be all invertible operators. Let P_k be the orthogonal projector onto $\mathscr{K}_k(\alpha) \otimes \mathscr{Y}$ as defined in (4.29). Let the operator $\mathcal{R}_{c,k} \in \mathcal{L}(\mathscr{Y}^k)$ be defined by*

$$
\mathcal{R}_{c,k} = \iota_{\mathscr{Y}_k}^* \mathscr{R}_{c,k} \iota_{\mathscr{Y}_k}. \tag{4.69}
$$

Then the following relations hold true:

$$
\mathcal{R}_{c,k} = (I_k \otimes Q^{-1}) + \left(F_k + I_k \otimes (Q^{-1}N^*)\right) (I_k \otimes \widetilde{R}^{-1}) \left(F_k^* + I_k \otimes (NQ^{-1})\right), \tag{4.70}
$$

$$
\iota_{\mathscr{Y}_k} \mathcal{R}_{c,k}^{-1} \iota_{\mathscr{Y}_k}^* = P_k \mathscr{R}_{c,k}^{-1} P_k. \tag{4.71}
$$

Remark 4.19. If the output space $\mathscr{Y}$ is finite-dimensional, i.e., $\mathscr{Y} = \mathbb{C}^p$ for some $p \in \mathbb{N}$, then $\mathcal{R}_{c,k} \in \mathcal{L}(\mathscr{Y}^k)$ is the matrix representation of the complementary Popov operator $\mathscr{R}_c$.

Theorem 4.20 (Riccati-ADI)**.** *Let $Q = Q^* \in \mathcal{L}(\mathscr{Y})$, $R = R^* \in \mathcal{L}(\mathscr{U})$, and $\widetilde{R} := R - NQ^{-1}N^* \in \mathcal{L}(\mathscr{U})$ be all invertible operators. Let $\Sigma = (\mathbb{T}, \Phi, \Psi, \mathbb{F})$ be a stable weakly regular linear system (cf. Definition 2.6.b) and $(\alpha_j)_{j=1}^{\infty}$ such that $\operatorname{Re}(\alpha_j) > 0$ for all $j \in \mathbb{N}$. Then for all $k \in \mathbb{N}$, the operator X_k in (4.60) can be determined by*

$$X_k = S_k^* \mathcal{R}_{c,k}^{-1} S_k, \tag{4.72}$$

where $\mathcal{R}_{c,k}$ is calculated by

$$\mathcal{R}_{c,k} = (I_k \otimes Q^{-1}) + \left(F_k + I_k \otimes (Q^{-1}N^*)\right)(I_k \otimes \widetilde{R}^{-1})\left(F_k^* + I_k \otimes (NQ^{-1})\right),$$

and the operators F_k and S_k are determined by Algorithm 1.

4.5 Monotonicity and convergence of Riccati-ADI

In this section we prove the monotonicity and convergence of the Riccati-ADI algorithm (cf. Theorem 4.16) under the following assumption:

$$\left\langle \begin{pmatrix} w \\ v \end{pmatrix}, \begin{bmatrix} Q & N^* \\ N & R \end{bmatrix} \begin{pmatrix} w \\ v \end{pmatrix} \right\rangle \geq 0, \qquad \forall\, v \in \mathscr{U}, \quad \forall\, w \in \mathscr{Y}. \tag{4.73}$$

Theorem 4.11 and Theorem 4.16 imply that the operator X_k computed by the Riccati-ADI method (Algorithm 2) expresses the optimal cost (4.60) of the projected optimal control problem (4.56). Since the ranges of projectors P_k are nested, we can easily deduce that the sequence $(X_k)_{k\in\mathbb{N}}$ is monotone and bounded from above by X, as shown in the following theorem.

Theorem 4.21. *Let $\Sigma = (\mathbb{T}, \Phi, \Psi, \mathbb{F})$ be a stable weakly regular linear system (cf. Definition 2.6.b) and $(\alpha_j)_{j=1}^{\infty}$ be the shift parameters such that $\operatorname{Re}(\alpha_j) > 0$ for all $j \in \mathbb{N}$. Let $X \in \mathcal{L}(\mathscr{X})$ be the Riccati operator (3.10). Let P_k and O_k be the orthogonal projectors defined as in (4.29). Further, let Ψ_k and $\mathbb{F}_k$ be given by (4.30) and (4.31), respectively. Let X be the Riccati operator (3.10) and X_k be the projected Riccati operator defined by (4.60). Moreover, let the assumption (4.73) hold true.*
Then the sequence $(X_k)_{k\in\mathbb{N}}$ as in (4.60) satisfies

$$\langle x_0, X_k x_0 \rangle_{\mathscr{X}} \leq \langle x_0, X_{k+1} x_0 \rangle_{\mathscr{X}} \leq \langle x_0, X x_0 \rangle_{\mathscr{X}}, \qquad \forall x_0 \in \mathscr{X}.$$

Proof. Since $\mathscr{K}_k(\alpha) \subset \mathscr{K}_{k+1}(\alpha)$, we have

$$P_k \leq P_{k+1}, \qquad O_k \leq O_{k+1}. \tag{4.74}$$

Let $u \in L^2(0,\infty;\mathscr{U})$, $x_0 \in \mathscr{X}$, and $y \in L^2(0,\infty;\mathscr{Y})$ be respectively the input, initial state, and output of the well-posed linear system Σ. It follows from (4.74) that

$$\begin{aligned}
&\left\langle \begin{pmatrix} P_{k+1}y \\ u \end{pmatrix}, \begin{bmatrix} Q & N^* \\ N & R \end{bmatrix} \begin{pmatrix} P_{k+1}y \\ u \end{pmatrix} \right\rangle_{L^2(0,\infty;\mathscr{Y}\times\mathscr{U})} \\
&= \left\langle \begin{pmatrix} P_k y + (P_{k+1}-P_k)y \\ O_k u + (O_{k+1}-O_k)u \end{pmatrix}, \begin{bmatrix} Q & N^* \\ N & R \end{bmatrix} \begin{pmatrix} P_k y + (P_{k+1}-P_k)y \\ O_k u + (O_{k+1}-O_k)u \end{pmatrix} \right\rangle_{L^2(0,\infty;\mathscr{Y}\times\mathscr{U})} \\
&= \left\langle \begin{pmatrix} P_k y \\ O_k u \end{pmatrix}, \begin{bmatrix} Q & N^* \\ N & R \end{bmatrix} \begin{pmatrix} P_k y \\ O_k u \end{pmatrix} \right\rangle_{L^2(0,\infty;\mathscr{Y}\times\mathscr{U})} \\
&\quad + \left\langle \begin{pmatrix} (P_{k+1}-P_k)y \\ (O_{k+1}-O_k)u \end{pmatrix}, \begin{bmatrix} Q & N^* \\ N & R \end{bmatrix} \begin{pmatrix} (P_{k+1}-P_k)y \\ (O_{k+1}-O_k)u \end{pmatrix} \right\rangle_{L^2(0,\infty;\mathscr{Y}\times\mathscr{U})},
\end{aligned} \tag{4.75}$$

where we have used

$$\begin{aligned}
\langle P_k y, Q(P_{k+1}-P_k)y\rangle &= 0, & \langle O_k u, R(O_{k+1}-O_k)u\rangle &= 0, \\
\langle P_k y, N^*(O_{k+1}-O_k)u\rangle &= 0, & \langle O_k u, N(P_{k+1}-P_k)y\rangle &= 0.
\end{aligned}$$

Now it follows from (4.75), together with the assumption (4.73), that

$$\begin{aligned}
&\min_{u\in L^2(0,\infty;\mathscr{U})} \left\langle \begin{pmatrix} P_{k+1}y \\ u \end{pmatrix}, \begin{bmatrix} Q & N^* \\ N & R \end{bmatrix} \begin{pmatrix} P_{k+1}y \\ u \end{pmatrix} \right\rangle_{L^2(0,\infty;\mathscr{Y}\times\mathscr{U})} \\
&\qquad \geq \min_{u\in L^2(0,\infty;\mathscr{U})} \left\langle \begin{pmatrix} P_k y \\ O_k u \end{pmatrix}, \begin{bmatrix} Q & N^* \\ N & R \end{bmatrix} \begin{pmatrix} P_k y \\ O_k u \end{pmatrix} \right\rangle_{L^2(0,\infty;\mathscr{Y}\times\mathscr{U})}.
\end{aligned}$$

From Corollary 4.12 we know that

$$u_k^{\mathrm{opt}} \in \mathscr{K}_k(\alpha) \otimes \mathscr{U}. \tag{4.76}$$

Consequently, it follows from Theorem 4.11 that

$$\begin{aligned}
\langle x_0, X_{k+1}x_0\rangle &= \min_{u\in L^2(0,\infty;\mathscr{U})} \left\langle \begin{pmatrix} P_{k+1}y \\ u \end{pmatrix}, \begin{bmatrix} Q & N^* \\ N & R \end{bmatrix} \begin{pmatrix} P_{k+1}y \\ u \end{pmatrix} \right\rangle_{L^2(0,\infty;\mathscr{Y}\times\mathscr{U})} \\
&\geq \min_{u\in L^2(0,\infty;\mathscr{U})} \left\langle \begin{pmatrix} P_k y \\ O_k u \end{pmatrix}, \begin{bmatrix} Q & N^* \\ N & R \end{bmatrix} \begin{pmatrix} P_k y \\ O_k u \end{pmatrix} \right\rangle_{L^2(0,\infty;\mathscr{Y}\times\mathscr{U})} \\
&= \min_{u\in L^2(0,\infty;\mathscr{U})} \left\langle \begin{pmatrix} P_k y \\ u \end{pmatrix}, \begin{bmatrix} Q & N^* \\ N & R \end{bmatrix} \begin{pmatrix} P_k y \\ u \end{pmatrix} \right\rangle_{L^2(0,\infty;\mathscr{Y}\times\mathscr{U})} \\
&= \langle x_0, X_k x_0\rangle\,.
\end{aligned}$$

Since P_k and O_k are orthogonal projectors, we have

$$P_k \leq I, \qquad O_k \leq I.$$

Analogously to the first part of this proof we see that

$$\begin{aligned}
&\left\langle \begin{pmatrix} y \\ u \end{pmatrix}, \begin{bmatrix} Q & N^* \\ N & R \end{bmatrix} \begin{pmatrix} y \\ u \end{pmatrix} \right\rangle_{L^2(0,\infty;\mathscr{Y}\times\mathscr{U})} \\
&= \left\langle \begin{pmatrix} P_k y + (I-P_k)y \\ O_k u + (I-O_k)u \end{pmatrix}, \begin{bmatrix} Q & N^* \\ N & R \end{bmatrix} \begin{pmatrix} P_k y + (I-P_k)y \\ O_k u + (I-O_k)u \end{pmatrix} \right\rangle_{L^2(0,\infty;\mathscr{Y}\times\mathscr{U})} \\
&= \left\langle \begin{pmatrix} P_k y \\ O_k u \end{pmatrix}, \begin{bmatrix} Q & N^* \\ N & R \end{bmatrix} \begin{pmatrix} P_k y \\ O_k u \end{pmatrix} \right\rangle_{L^2(0,\infty;\mathscr{Y}\times\mathscr{U})} \\
&\quad + \left\langle \begin{pmatrix} (I-P_k)y \\ (I-O_k)u \end{pmatrix}, \begin{bmatrix} Q & N^* \\ N & R \end{bmatrix} \begin{pmatrix} (I-P_k)y \\ (I-O_k)u \end{pmatrix} \right\rangle_{L^2(0,\infty;\mathscr{Y}\times\mathscr{U})},
\end{aligned} \tag{4.77}$$

where we have used

$$\begin{aligned}
\langle P_k y, Q(I-P_k)y\rangle &= 0, \qquad & \langle O_k u, R(I-O_k)u\rangle &= 0, \\
\langle P_k y, N^*(I-O_k)u\rangle &= 0, \qquad & \langle O_k u, N(I-P_k)y\rangle &= 0.
\end{aligned}$$

Now it follows from (4.77), together with the assumption (4.73), that

$$\begin{aligned}
&\min_{u\in L^2(0,\infty;\mathscr{U})} \left\langle \begin{pmatrix} y \\ u \end{pmatrix}, \begin{bmatrix} Q & N^* \\ N & R \end{bmatrix} \begin{pmatrix} y \\ u \end{pmatrix} \right\rangle_{L^2(0,\infty;\mathscr{Y}\times\mathscr{U})} \\
&\qquad \geq \min_{u\in L^2(0,\infty;\mathscr{U})} \left\langle \begin{pmatrix} P_k y \\ O_k u \end{pmatrix}, \begin{bmatrix} Q & N^* \\ N & R \end{bmatrix} \begin{pmatrix} P_k y \\ O_k u \end{pmatrix} \right\rangle_{L^2(0,\infty;\mathscr{Y}\times\mathscr{U})}.
\end{aligned}$$

Consequently, we can conclude from Proposition 3.2 and Theorem 4.11, together with (4.76), that

$$\begin{aligned}
\langle x_0, X x_0\rangle &= \min_{u\in L^2(0,\infty;\mathscr{U})} \left\langle \begin{pmatrix} y \\ u \end{pmatrix}, \begin{bmatrix} Q & N^* \\ N & R \end{bmatrix} \begin{pmatrix} y \\ u \end{pmatrix} \right\rangle_{L^2(0,\infty;\mathscr{Y}\times\mathscr{U})} \\
&\geq \min_{u\in L^2(0,\infty;\mathscr{U})} \left\langle \begin{pmatrix} P_k y \\ O_k u \end{pmatrix}, \begin{bmatrix} Q & N^* \\ N & R \end{bmatrix} \begin{pmatrix} P_k y \\ O_k u \end{pmatrix} \right\rangle_{L^2(0,\infty;\mathscr{Y}\times\mathscr{U})} \\
&= \min_{u\in L^2(0,\infty;\mathscr{U})} \left\langle \begin{pmatrix} P_k y \\ u \end{pmatrix}, \begin{bmatrix} Q & N^* \\ N & R \end{bmatrix} \begin{pmatrix} P_k y \\ u \end{pmatrix} \right\rangle_{L^2(0,\infty;\mathscr{Y}\times\mathscr{U})} \\
&= \langle x_0, X_k x_0\rangle\,.
\end{aligned}$$

□

So far we have shown that the sequence $(X_k)_{k\in\mathbb{N}}$ generated by the Riccati-ADI method (Algorithm 2) is non-decreasing and bounded from above by the Riccati operator X (Theorem 4.21). Therefore, it follows from Theorem A.2 that there exists some self-adjoint operator $\widetilde{X} \in \mathcal{L}(\mathscr{X})$ such that

$$X_k \leq \widetilde{X} \leq X, \qquad \forall k \in \mathbb{N},$$

and $(X_k)_{k\in\mathbb{N}}$ converges to $\widetilde{X}$ in the strong operator topology, i.e.,

$$\lim_{k\to\infty} X_k x = \widetilde{X} x, \qquad \forall\, x \in \mathscr{X}\,.$$

We note that it does not necessarily hold that $\widetilde{X} = X$. To deal with this issue, we require the additional assumption

$$\overline{\bigcup_{k\in\mathbb{N}} \mathscr{K}_k(\alpha) \otimes \mathscr{Y}} = L^2(0,\infty;\mathscr{Y}). \tag{4.78}$$

It follows from [41, Lemma 4.4] that (4.78) is equivalent to the *non-Blaschke condition*

$$\sum_{j=1}^{\infty} \frac{\mathrm{Re}(\alpha_j)}{1+|\alpha_j|^2} = \infty, \tag{4.79}$$

where $(\alpha_k)_{k=1}^{\infty}$ is the sequence of shift parameters associated with the Takenaka-Malmquist system. This non-Blaschke condition is satisfied, for example, if the parameters all belong to a fixed compact set contained in the open right half-plane. With the help of a numerical example, we will show in Chapter 6 that if the shift parameters do not satisfy the non-Blaschke (4.79), then the sequence $(X_k)_{k\in\mathbb{N}}$ may converge to an operator which is not the Riccati operator (3.10).

The following theorem gives the convergence of the Riccati-ADI method to the Riccati operator (3.10). This convergence result was previously obtained for the special case of operator Lyapunov equation in [41].

Theorem 4.22. *Under the assumptions of Theorem 4.21, the sequence $(X_k)_{k\in\mathbb{N}}$ defined by (4.60) converges in the strong operator topology. If additionally, $(\alpha_j)_{j=1}^{\infty}$ satisfies the non-Blaschke condition (4.79), then X_k converges even to X. This means that*

$$\lim_{k\to\infty} X_k x = Xx, \qquad \forall\, x \in \mathscr{X}.$$

Furthermore, under the additional assumption that $Q = Q^* \in \mathcal{L}(\mathscr{Y})$, $R = R^* \in \mathcal{L}(\mathscr{U})$, and $\widetilde{R} := R - NQ^{-1}N^* \in \mathcal{L}(\mathscr{U})$ are all invertible operators,

(i) if X is compact, then X_k converges to X in the uniform operator topology. This means that if $X \in \mathcal{K}(\mathscr{X})$, then

$$\lim_{k\to\infty} \|X_k - X\|_{\mathcal{L}(\mathscr{X})} = 0.$$

(ii) if X is in the Schatten class $S_p(\mathscr{X})$ for $p \in [1,\infty]$, then X_k converges to X in the topology of $S_p(\mathscr{X})$. This means that

$$\lim_{k\to\infty} \|X_k - X\|_{\mathcal{S}_p(\mathscr{X})} = 0.$$

Proof. Since, by Theorem 4.21, $(X_k)_{k\in\mathbb{N}}$ is a non-decreasing sequence which is bounded from above by X, we obtain convergence in the strong operator topology to some operator $\widetilde{X} \in \mathcal{L}(\mathscr{X})$ with $\widetilde{X} \le X$ (see Theorem A.2). If X is compact, then the non-decreasing sequence $(X_k)_{k\in\mathbb{N}}$ is bounded from above by a compact operator and therefore converges in the uniform operator topology. If X is in the Schatten class, then the non-decreasing sequence $(X_k)_{k\in\mathbb{N}}$ is bounded from above by a Schatten class operator and therefore converges in the Schatten class topology.

Since $\mathscr{K}_k(\alpha) \subset \mathscr{K}_{k+1}(\alpha)$ we have $P_k \leq P_{k+1}$ and since P_k is an orthogonal projector, we have $P_k \leq I$. It follows from [50, p.263] that $(P_k)_{k\in\mathbb{N}}$ converges in the strong operator topology to some orthogonal projector $P \leq I$. It was shown in [41, Lemma 4.4] that $P = I$, if and only if, the non-Blaschke condition is satisfied (this result is shown there actually only for the case $L^2(0,\infty)$, but the behavior of tensor products under the strong operator topology [30, Theorem 1 part b] gives the general case for $L^2(0,\infty;\mathscr{Y})$).

From now on we assume the non-Blaschke condition (4.79), so that $P = I$. Then, since the sequence $(P_k)_{k\in\mathbb{N}}$ is uniformly bounded by identity ($P_k \leq I$), we obtain that

$$\begin{aligned}\mathscr{R}_k &= R + N\mathbb{F}_k + \mathbb{F}^*N^* + \mathbb{F}_k^*Q\mathbb{F}_k \\ &= R + NP_k\mathbb{F} + \mathbb{F}^*P_kN^* + \mathbb{F}^*P_kQP_k\mathbb{F}\end{aligned}$$

converges to

$$\mathscr{R} = R + N\mathbb{F} + \mathbb{F}^*N^* + \mathbb{F}^*Q\mathbb{F}$$

in the strong operator topology. As a result, $\mathscr{R}_k^{-1}$ converges to $\mathscr{R}^{-1}$ in the strong operator topology (e.g., by [20, Theorem 7.6.1]). It follows from the uniform boundedness principle that $\mathscr{R}_k^{-1}$ is uniformly bounded. Since $\mathscr{R}_k^{-1}$ converges to $\mathscr{R}^{-1}$ in the strong operator topology, we have that

$$(QP_k\mathbb{F} + N^*)\mathscr{R}_k^{-1}(\mathbb{F}^*P_kQ + N)$$

converges in the strong operator topology to

$$(Q\mathbb{F} + N^*)\mathscr{R}^{-1}(\mathbb{F}^*Q + N).$$

By sequential continuity, we conclude that

$$\begin{aligned}X_k &= \Psi_k^*Q\Psi_k - \Psi_k^*(Q\mathbb{F}_k + N^*)\mathscr{R}_k^{-1}(\mathbb{F}_k^*Q + N)\Psi_k \\ &= \Psi^*P_kQP_k\Psi - \Psi^*P_k(QP_k\mathbb{F} + N^*)\mathscr{R}_k^{-1}(\mathbb{F}^*P_kQ + N)P_k\Psi\end{aligned}$$

converges to

$$X = \Psi^*Q\Psi - \Psi^*(Q\mathbb{F} + N^*)\mathscr{R}^{-1}(\mathbb{F}^*Q + N)\Psi$$

in the strong operator topology.

To prove the last part of this theorem, let us assume additionally that $Q = Q^* \in \mathcal{L}(\mathscr{Y})$, $R = R^* \in \mathcal{L}(\mathscr{U})$, and $\widetilde{R} := R - NQ^{-1}N^* \in \mathcal{L}(\mathscr{U})$ are all invertible operators. Then, it follows from Proposition 3.4 that the Riccati operator X can be represented in the alternative form

$$X = \Psi^*\mathscr{R}_c^{-1}\Psi,$$

where $\mathscr{R}_c$ is the complementary Popov operator given by

$$\mathscr{R}_c = Q^{-1} + (\mathbb{F} + Q^{-1}N^*)\widetilde{R}^{-1}(\mathbb{F}^* + NQ^{-1}).$$

If X is compact, then (since $\mathscr{R}_c$ is self-adjoint and invertible) Ψ is compact. Now it follows from Theorem A.3.a that

$$\begin{array}{rcl}\Psi^*P_kQP_k\Psi & \longrightarrow & \Psi^*Q\Psi \\ (QP_k\mathbb{F} + N^*)\mathscr{R}_k^{-1}(\mathbb{F}^*P_kQ + N)P_k\Psi & \longrightarrow & (Q\mathbb{F} + N^*)\mathscr{R}^{-1}(\mathbb{F}^*Q + N)\Psi\end{array}$$

in the uniform operator topology. Hence, we conclude that X_k converges to X in the uniform operator topology. The convergence in the Schatten norm $\mathcal{S}_p(\mathscr{X})$ follows analogously by applying Theorem A.3.b. $\square$

4.6 The ADI iteration for the bounded real and positive real optimal control problems

In this section we show how the ADI method developed in Section 4.3 can be applied to find approximate solutions of the singular optimal control problem in the bounded real and positive real case subject to a strongly stable weakly regular linear system (cf. Definition 2.6.c). We prove that the sequence of approximate solutions is monotonically increasing. If the shift parameters satisfy the non-Blaschke condition (4.79), then the sequence is proven to be convergent to the optimal cost of the singular optimal control problem in the bounded real and positive real case.

Here we assume finite-dimensionality of the input and output spaces (i.e., $\mathscr{U} = \mathbb{C}^m$ and $\mathscr{Y} = \mathbb{C}^p$ for some $m, p \in \mathbb{N}$). This implies that we allow only finitely many variables to control and observe the system (which is justified in the actual applications). In this case the ADI algorithm provides approximate solutions in *low-rank factored form*. This means that we provide sequences $(X_k)_{k\in\mathbb{N}} \in \mathcal{L}(\mathscr{X})$ of approximate solutions of the form $X_k = R_k^* R_k$ for some $R_k \in \mathcal{L}(\mathscr{X}, \mathbb{C}^{\ell_k})$, with typically "small" ℓ_k.

Throughout this section the Riccati operator X is the solution of the singular optimal control problem (3.62) or (3.66). We construct the projected versions of these singular optimal control problems by replacing the output function y with $P_k y$, where P_k is the orthogonal projector given by (4.29). Thereby we present the projected versions of Theorem 3.12 (for the bounded real or positive real case). In this regard, the output map Ψ and the input-output map $\mathbb{F}$ are replaced with S_k and F_k (generated by Algorithm 1), respectively. Then the relation (3.52) has to be solved for projected versions of $\mathbb{F}_\Xi$ and Ψ_Ξ. These are thereafter, by an accordant modification of (3.53), used to construct X_k.

4.6.1 The ADI method for the bounded real singular optimal control problem

Let the input space $\mathscr{U}$ and the output space $\mathscr{Y}$ be finite-dimensional, i.e., $\mathscr{U} = \mathbb{C}^m$ and $\mathscr{Y} = \mathbb{C}^p$ for some $m, p \in \mathbb{N}$. In the bounded real case we consider the cost functional (3.2) with $Q = -I$, $R = I$ and $N = 0$ (Section 3.6). In the singular optimal control problem we assume that the Popov operator (3.4) satisfies (3.51). In the bounded real case this means that

$$\langle u, (I - \mathbb{F}^*\mathbb{F})u\rangle_{L^2(0,\infty;\mathbb{C}^m)} \geq 0. \tag{4.80}$$

With the notation introduced after (1.15), this condition is written as $I - \mathbb{F}^*\mathbb{F} \geq 0$. The following theorem is the projected version of the bounded real singular optimal control problem.

Theorem 4.23. *Let $\Sigma = (\mathbb{T}, \Phi, \Psi, \mathbb{F})$ be a strongly stable weakly regular linear system (cf. Definition 2.6.c) with generating operators (A, B, C, D). Assume that the input space $\mathscr{U}$ and the output space $\mathscr{Y}$ are finite-dimensional (i.e., $\mathscr{U} = \mathbb{C}^m$ and $\mathscr{Y} = \mathbb{C}^p$). Further assume that the Popov operator $\mathscr{R} = I - \mathbb{F}^*\mathbb{F}$ satisfies (4.80). Let $(\alpha_j)_{j=1}^\infty$ be a complex sequence with $\mathrm{Re}(\alpha_j) > 0$ for all $j \in \mathbb{N}$. Moreover, let $S_k \in \mathcal{L}(\mathscr{X}, \mathbb{C}^{kp})$ and $F_k \in \mathbb{C}^{kp\times km}$ be defined as in (4.32) and (4.33), respectively.*
Then, the matrix $I - F_k^* F_k \in \mathbb{C}^{km\times km}$ is positive semidefinite (i.e., $I - F_k^* F_k \geq 0$). In

particular, there exists some matrix $F_{\Xi,k} \in \mathbb{C}^{\ell_k \times km}$ with full row rank such that

$$I - F_k^* F_k = F_{\Xi,k}^* F_{\Xi,k}. \tag{4.81}$$

Furthermore, there exists some operator $S_{\Xi,k} \in \mathcal{L}(\mathscr{X}, \mathbb{C}^{\ell_k})$ such that

$$F_{\Xi,k}^* S_{\Xi,k} = -F_k^* S_k. \tag{4.82}$$

For the orthogonal projector P_k as in (4.29), the operator X_k defined by

$$X_k = S_k^* S_k + S_{\Xi,k}^* S_{\Xi,k} \tag{4.83}$$

fulfills

$$\langle x_0, X_k x_0 \rangle = \sup_{u \in L^2(0,\infty;\mathbb{C}^m)} \|P_k \mathbb{F} u + P_k \Psi x_0\|^2 - \|u\|^2, \qquad \forall\, x_0 \in \mathscr{X}. \tag{4.84}$$

Proof. Recall that definition of the matrix $F_k \in \mathbb{C}^{kp \times km}$ and the operator $S_k \in \mathcal{L}(\mathscr{X}, \mathbb{C}^{kp})$ from (4.33) and (4.32). Since $P_k \leq I$, we have that $\mathbb{F}^* P_k \mathbb{F} \leq \mathbb{F}^* \mathbb{F}$, which implies that $I - \mathbb{F}^* P_k \mathbb{F} \geq I - \mathbb{F}^* \mathbb{F} \geq 0$. Hence we obtain

$$\begin{aligned} I - F_k^* F_k &= \iota_{\mathscr{U}_k}^* \iota_{\mathscr{U}_k} - \iota_{\mathscr{U}_k}^* \mathbb{F}^* \iota_{\mathscr{Y}_k} \iota_{\mathscr{Y}_k}^* \mathbb{F} \iota_{\mathscr{U}_k} \\ &= \iota_{\mathscr{U}_k}^* \iota_{\mathscr{U}_k} - \iota_{\mathscr{U}_k}^* \mathbb{F}^* P_k \mathbb{F} \iota_{\mathscr{U}_k} \\ &= \iota_{\mathscr{U}_k}^* (I - \mathbb{F}^* P_k \mathbb{F}) \iota_{\mathscr{U}_k} \geq 0. \end{aligned}$$

Next, we prove that

$$\operatorname{Im}(F_k^* S_k) \subset \operatorname{Im}(F_{\Xi,k}^*) \tag{4.85}$$

By taking orthogonal complements, this is equivalent to

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

It follows from (3.62) that for all $u \in L^2(0, \infty; \mathbb{C}^m)$ and $x_0 \in \mathscr{X}$,

$$\langle x_0, X x_0 \rangle \geq \|\mathbb{F} u + \Psi x_0\|^2 - \|u\|^2.$$

By further using (4.32), (4.33), and (4.81), we observe that

$$\begin{aligned} \langle x_0, X x_0 \rangle &\geq \|\mathbb{F} u + \Psi x_0\|^2 - \|u\|^2 \\ &\geq \|P_k \mathbb{F} u + P_k \Psi x_0\|^2 - \|O_k u\|^2 - \|(I - O_k) u\|^2 \\ &= \|\iota_{\mathscr{Y}_k} F_k \iota_{\mathscr{U}_k}^* u + \iota_{\mathscr{Y}_k} S_k x_0\|^2 - \|\iota_{\mathscr{U}_k}^* u\|^2 - \|(I - O_k) u\|^2 \\ &= \|F_k \iota_{\mathscr{U}_k}^* u + S_k x_0\|^2 - \|\iota_{\mathscr{U}_k}^* u\|^2 - \|(I - O_k) u\|^2 \\ &= \left\langle \iota_{\mathscr{U}_k}^* u, (F_k^* F_k - I) \iota_{\mathscr{U}_k}^* u \right\rangle + 2\mathrm{Re} \left\langle \iota_{\mathscr{U}_k}^* u, F_k^* S_k x_0 \right\rangle \\ &\qquad + \|S_k x_0\|^2 - \|(I - O_k) u\|^2 \\ &= -\left\langle \iota_{\mathscr{U}_k}^* u, F_{\Xi,k}^* F_{\Xi,k}) \iota_{\mathscr{U}_k}^* u \right\rangle + 2\mathrm{Re} \left\langle \iota_{\mathscr{U}_k}^* u, F_k^* S_k x_0 \right\rangle \\ &\qquad + \|S_k x_0\|^2 - \|(I - O_k) u\|^2 \\ &= -\|F_{\Xi,k} \iota_{\mathscr{U}_k}^* u\|^2 + 2\mathrm{Re} \left\langle \iota_{\mathscr{U}_k}^* u, F_k^* S_k x_0 \right\rangle + \|S_k x_0\|^2 - \|(I - O_k) u\|^2. \end{aligned} \tag{4.86}$$

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 non-trivial $x_0 \in \mathscr{X}$ such that $\langle S_k x_0, F_k \hat{u}\rangle \neq 0$. Then, for $\lambda \in \mathbb{C}$, by substituting x_0 and $u := \iota_{\mathscr{U}_k}(\lambda\hat{u}) \in L^2(0,\infty;\mathbb{C}^m)$ into (4.86), we obtain

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

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

$$S_{\Xi,k} := (F_{\Xi,k}F_{\Xi,k}^*)^{-1}F_{\Xi,k}F_k S_k \tag{4.87}$$

is well-defined. We now show that $S_{\Xi,k}$ as in (4.87) satisfies (4.82). From (4.85) we have that for all $x \in \mathscr{X}$ 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 $x \in \mathscr{X}$ was arbitrary, this proves that $F_{\Xi,k}^* S_{\Xi,k} = F_k^* S_k$. Hence, $S_{\Xi,k} \in \mathcal{L}(\mathscr{X}, \mathbb{C}^{\ell_k})$ satisfies (4.82).

It remains to prove that X_k as in (4.83) fulfills (4.84). Using (4.81) and (4.82), we have for all $x_0 \in \mathscr{X}$ and $u \in L^2(0,\infty;\mathbb{C}^m)$ that

$$\begin{aligned}&\|P_k \mathbb{F} u + P_k \Psi x_0\|^2 - \|u\|^2 \\ &= -\left\langle \iota_{\mathscr{U}_k}^* u, F_{\Xi,k}^* F_{\Xi,k}\iota_{\mathscr{U}_k}^* u\right\rangle + 2\mathrm{Re}\left\langle \iota_{\mathscr{U}_k}^* u, F_k^* S_k x_0\right\rangle + \|S_k x_0\|^2 - \|(I-O_k)u\|^2 \\ &= -\left\langle \iota_{\mathscr{U}_k}^* u, F_{\Xi,k}^* F_{\Xi,k}\iota_{\mathscr{U}_k}^* u\right\rangle - 2\mathrm{Re}\left\langle \iota_{\mathscr{U}_k}^* u, F_{\Xi,k}^* S_{\Xi,k} x_0\right\rangle + \|S_k x_0\|^2 - \|(I-O_k)u\|^2 \\ &= -\|F_{\Xi,k}\iota_{\mathscr{U}_k}^* u + S_{\Xi,k}x_0\|^2 + \|S_{\Xi,k}x_0\|^2 + \|S_k x_0\|^2 - \|(I-O_k)u\|^2 \\ &= -\|F_{\Xi,k}\iota_{\mathscr{U}_k}^* u + S_{\Xi,k}x_0\|^2 - \|(I-O_k)u\|^2 + \langle x_0, X_k x_0\rangle \\ &\leq \langle x_0, X_k x_0\rangle.\end{aligned}$$

This gives rise to

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

On the other hand, using the surjectivity of $F_{\Xi,k} \in \mathbb{C}^{\ell_k \times km}$, there exists some $\hat{u} \in \mathbb{C}^{km}$ with $F_{\Xi,k}\hat{u} = -S_{\Xi,k}x_0$. Then, for $u = \iota_{\mathscr{U}_k}\hat{u}$ we see that equality holds true in the above calculations. This proves (4.84). □

Remark 4.24. The formula (4.87) for $S_{\Xi,k} \in \mathcal{L}(\mathscr{X}, \mathbb{C}^{\ell_k})$ shows that

$$X_k = S_k^*[I + F_k F_{\Xi,k}^*(F_{\Xi,k}F_{\Xi,k}^*)^{-2}F_{\Xi,k}F_k^*]S_k.$$

It is easily verified that $F_{\Xi,k}^*(F_{\Xi,k}F_{\Xi,k}^*)^{-2}F_{\Xi,k}$ is the Moore-Penrose inverse of $F_{\Xi,k}^*F_{\Xi,k}$. Therefore, we have

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

Next we prove that the sequence $(X_k)_{k\in\mathbb{N}}$ is monotonically increasing with respect to definiteness. We further present a criterion on the shift parameters such that convergence to the solution of the bounded real singular optimal control problem is achieved.

Theorem 4.25. *Let $\Sigma = (\mathbb{T}, \Phi, \Psi, \mathbb{F})$ be a strongly stable weakly regular linear system (cf. Definition 2.6.c) with generating operators (A, B, C, D). Assume that the input space $\mathscr{U}$ and the output space $\mathscr{Y}$ are finite-dimensional (i.e., $\mathscr{U} = \mathbb{C}^m$ and $\mathscr{Y} = \mathbb{C}^p$). Further assume that the Popov operator $\mathscr{R} = I - \mathbb{F}^*\mathbb{F}$ satisfies (4.80). Let $(\alpha_j)_{j=1}^{\infty}$ be a complex sequence with $\mathrm{Re}(\alpha_j) > 0$ for all $j \in \mathbb{N}$. Moreover, let $S_k \in \mathcal{L}(\mathscr{X}, \mathbb{C}^{kp})$ and $F_k \in \mathbb{C}^{kp\times km}$ be defined as in (4.32) and (4.33), respectively. Let X_k be defined as in Theorem 4.23. Then,*

$$\langle x_0, X_k x_0\rangle \le \langle x_0, X_{k+1} x_0\rangle \le \langle x_0, X x_0\rangle, \qquad \forall\ x_0 \in \mathscr{X}, \quad \forall\ k \in \mathbb{N},$$

and the sequence $(X_k)_{k\in\mathbb{N}}$ converges in the strong operator topology. Additionally, if $(\alpha_j)_{j=1}^{\infty}$ satisfies the non-Blaschke condition (4.79), then $(X_k)_{k\in\mathbb{N}}$ converges even to X. This means that for all $x_0 \in \mathscr{X}$

$$\lim_{k\to\infty} X_k x_0 = X x_0.$$

Moreover, if X is compact, then $(X_k)_{k\in\mathbb{N}}$ converges to X in the uniform operator topology. If X is in the Schatten class $S_p(\mathscr{X})$ for $p \in [1, \infty]$, then $(X_k)_{k\in\mathbb{N}}$ converges to X in the topology of $S_p(\mathscr{X})$.

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

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

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

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

Similarly, using that

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

we obtain

$$\langle x_0, X_k x_0\rangle \le \langle x_0, X x_0\rangle, \qquad \forall\ x_0 \in \mathscr{X}.$$

Since the sequence $(X_k)_{k\in\mathbb{N}}$ is non-decreasing and bounded from above by X, it follows from Theorem A.2 that $(X_k)_{k\in\mathbb{N}}$ converges in the strong operator topology to some operator $\widetilde{X} \in \mathcal{L}(\mathscr{X})$, such that $\widetilde{X} \le X$.
In the case where the non-Blaschke condition (4.79) is fulfilled, by [41, Lemma 4.4] there holds

$$\overline{\bigcup_{k\in\mathbb{N}} \mathscr{K}_k(\alpha) \otimes \mathbb{C}^p} = L^2(0, \infty; \mathbb{C}^p).$$

Therefore, the sequence $(P_k)_{k\in\mathbb{N}}$ converges to the identity in the strong operator topology, that is

$$\lim_{k\to\infty} P_k y = y, \qquad \forall\, y \in L^2(0,\infty;\mathbb{C}^p). \tag{4.88}$$

Let $x_0 \in \mathscr{X}$ and $\varepsilon > 0$. By (3.62) there exists some $u \in L^2(0,\infty;\mathbb{C}^m)$ with

$$\langle x_0, Xx_0\rangle < \|\mathbb{F}u + \Psi x_0\|^2 - \|u\|^2 + \tfrac{\varepsilon}{2}.$$

By (4.88), there exists some $N \in \mathbb{N}$ with $\|(\mathbb{F}u+\Psi x_0) - P_k(\mathbb{F}u+\Psi x_0)\|^2 \le \frac{\varepsilon}{2}$ for all $k \ge N$. Then we obtain that for all $k \ge N$ there holds

$$\begin{aligned}\langle x_0, Xx_0\rangle &< \|\mathbb{F}u + \Psi x_0\|^2 - \|u\|^2 + \tfrac{\varepsilon}{2}\\ &\le \|P_k\mathbb{F}u + P_k\Psi x_0\|^2 + \|(\mathbb{F}u + \Psi x_0) - P_k(\mathbb{F}u + \Psi x_0)\|^2 - \|u\|^2 + \tfrac{\varepsilon}{2}\\ &\le \|P_k\mathbb{F}u + P_k\Psi x_0\|^2 - \|u\|^2 + \varepsilon \le \langle x_0, X_k x_0\rangle + \varepsilon.\end{aligned}$$

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

$$|\,\langle x_0, (X - X_k)x_0\rangle\,| = \langle x_0, Xx_0\rangle - \langle x_0, X_k x_0\rangle < \varepsilon, \qquad \forall\, k \ge N.$$

Hence, it follows that the sequence $(X_k)_{k\in\mathbb{N}}$ converges to X.

If additionally, we assume that X is compact, then the non-decreasing sequence $(X_k)_{k\in\mathbb{N}}$ is bounded from above by a compact operator and therefore converges in the uniform operator topology. If X is in the Schatten class, then the non-decreasing sequence $(X_k)_{k\in\mathbb{N}}$ is bounded from above by a Schatten class operator and therefore converges in the Schatten class topology. □

Next, we introduce a slightly different representation for the operator X_k as in (4.83), which is numerically more advantageous.

Theorem 4.26. *Let $\Sigma = (\mathbb{T}, \Phi, \Psi, \mathbb{F})$ be a strongly stable weakly regular linear system (cf. Definition 2.6.c) with generating operators (A, B, C, D). Assume that the input space $\mathscr{U}$ and output space $\mathscr{Y}$ are finite-dimensional (i.e., $\mathscr{U} = \mathbb{C}^m$ and $\mathscr{Y} = \mathbb{C}^p$). Further assume that the Popov operator $\mathscr{R} = I - \mathbb{F}^*\mathbb{F}$ satisfies (4.80). Let $(\alpha_j)_{j=1}^{\infty}$ be a complex sequence with $\mathrm{Re}(\alpha_j) > 0$ for all $j \in \mathbb{N}$. Moreover, let $S_k \in \mathcal{L}(\mathscr{X}, \mathbb{C}^{kp})$ and $F_k \in \mathbb{C}^{mp\times km}$ be defined as in (4.32) and (4.33), respectively.*
Then there exists some matrix $G_k \in \mathbb{C}^{\ell_k\times kp}$ with finite rank such that

$$I - F_kF_k^* = G_k^*G_k. \tag{4.89}$$

Further, there exists some $R_k \in \mathcal{L}(\mathscr{X}, \mathbb{C}^{\ell_k})$ such that

$$G_k^*R_k = S_k. \tag{4.90}$$

The operator X_k as in (4.83) fulfills

$$X_k = R_k^*R_k. \tag{4.91}$$

Proof. By Theorem 4.23, the matrix $I - F_k^* F_k \in \mathbb{C}^{km \times km}$ is positive semidefinite. Therefore, $I - F_k F_k^* \in \mathbb{C}^{kp \times kp}$ is positive semidefinite as well. Hence, there exists some matrix $G_k \in \mathbb{C}^{\ell_k \times kp}$ with full row rank such that (4.89) holds.

By (4.81) we have $\ker(I - F_k^* F_k) = \ker(F_{\Xi,k})$. It follows from (4.82) that $\ker(F_{\Xi,k}) \subset \ker(S_k^* F_k)$ and hence

$$\ker(I - F_k^* F_k) \subset \ker(S_k^* F_k). \tag{4.92}$$

We now prove

$$\operatorname{Im}(S_k) \subset \operatorname{Im}(I - F_k F_k^*). \tag{4.93}$$

By taking orthogonal complements, this is equivalent to

$$\ker(I - F_k F_k^*) \subset \ker(S_k^*).$$

Taking $y \in \ker(I - F_k F_k^*)$, it follows that $y = F_k F_k^* y$. Therefore

$$S_k^* y = S_k^* F_k F_k^* y \tag{4.94}$$

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 (4.92) gives $F_k^* y \in \ker(S_k^* F_k)$. Hence $S_k^* F_k F_k^* y = 0$. From (4.94) we then obtain $S_k^* y = 0$. We conclude that $\ker(I - F_k F_k^*) \subset \ker(S_k^*)$, as desired.

From (4.89) we obtain $\ker(I - F_k F_k^*) = \ker(G_k)$, so that $\operatorname{Im}(I - F_k F_k^*) = \operatorname{Im}(G_k^*)$. Together with (4.93), this shows that $\operatorname{Im}(S_k) \subset \operatorname{Im}(G_k^*)$. Since G_k has full row rank, $G_k G_k^*$ is invertible and therefore

$$R_k := (G_k G_k^*)^{-1} G_k S_k \tag{4.95}$$

is well-defined. We now show that the operator R_k satisfies (4.90). It follows from $\operatorname{Im}(S_k) \subset \operatorname{Im}(G_k^*)$ that for all $x \in \mathscr{X}$, 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 $x \in \mathscr{X}$ was arbitrary this proves that $G_k^* R_k = S_k$, i.e., R_k as defined in (4.95) satisfies (4.90).

By Remark 4.24 we have $X_k = S_k^*[I + F_k(I - F_k^* F_k)^+ F_k^*] S_k$. Using (4.93) and the fact that $(I - F_k F_k^*)^+ (I - F_k F_k^*)$ is the orthogonal projection onto $\operatorname{Im}(I - F_k F_k^*)$ we may alternatively write X_k 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-inverse is most easily proven by verifying the Moore-Penrose conditions [23, 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

$$X_k = S_k^* (I - F_k F_k^*)^+ S_k. \tag{4.96}$$

On the other hand, by using (4.95), we obtain

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

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 (4.89), it follows that $R_k^* R_k = X_k$. $\square$

Remark 4.27 (Projected bounded real singular optimal control problem). It follows from Algorithm 1 that the matrix $F_i \in \mathbb{C}^{ip \times im}$ has the lower triangular block structure

$$F_i = \begin{bmatrix} [F_{i-1}, 0] \\ Q_i(\overline{L_i} \otimes I_m) + \big[0, D\big] \end{bmatrix}. \tag{4.97}$$

As a result, we can determine the matrix $G_i \in \mathbb{C}^{\ell_i \times pi}$ and the operator $R_i \in \mathcal{L}(\mathscr{X}, \mathbb{C}^{\ell_i})$ recursively as follows: We have

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

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

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

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

$$G_{i-1}^* G_{12,i} = -\big[F_{i-1} \quad 0\big]\big(Q_i(\overline{L_i} \otimes I_m)\big)^*.$$

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

$$G_{22,i}^* G_{22,i} = I - \big(Q_i(\overline{L_i} \otimes I_m) + \big[0, D\big]\big)\big(Q_i(\overline{L_i} \otimes I_m) + \big[0, D\big]\big)^* - G_{12,i}^* G_{12,i}.$$

It follows from Algorithm 1 that S_i is obtained from S_{i-1} by

$$S_i = \begin{bmatrix} S_{i-1} \\ \sqrt{2\mathrm{Re}(\alpha_i)} \cdot V_i^* \end{bmatrix}. \tag{4.98}$$

Hence, by making the ansatz $R_i = \begin{bmatrix} R_{i-1} \\ R_{2,i} \end{bmatrix}$, we can rewrite equation (4.90) 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

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

Algorithm 3 ADI iteration for the bounded real singular optimal control problem.

Input: $\Sigma = (\mathbb{T}, \Phi, \Psi, \mathbb{F})$ a strongly stable weakly regular linear system with generating operators (A, B, C, D), such that $I - \mathbb{F}^*\mathbb{F} \geq 0$. Finite-dimensional input space $\mathscr{U} = \mathbb{C}^m$ and output space $\mathscr{Y} = \mathbb{C}^p$. Shift parameters $\alpha_1, \ldots, \alpha_k \in \mathbb{C}$ with $\mathrm{Re}(\alpha_i) > 0$.
Output: $R_k \in \mathcal{L}(\mathscr{X}, \mathbb{C}^{\ell_k})$ such that $R_k^* R_k = X_k \approx X$, where X is given by (3.63)

1: Perform Steps 1–5 in Algorithm 1

2: Determine a matrix $G_1 \in \mathbb{C}^{\ell_1 \times p}$ with full row rank such that

$$G_1^* G_1 = I - F_1 F_1^*$$

3: Determine an operator $R_1 \in \mathcal{L}(\mathscr{X}, \mathbb{C}^{\ell_1})$ such that

$$G_1^* R_1 = S_1$$

4: **for** $i = 2, 3, \ldots, k$ **do**

5: Perform Steps 7–14 in Algorithm 1

6: Determine a matrix $G_{12,i}$ such that

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

7: Determine a matrix $G_{22,i}$ with full row rank such that

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

8: $G_i = \begin{bmatrix} G_{i-1} & G_{12,i} \\ 0 & G_{22,i} \end{bmatrix}$

9: Determine an operator $R_{2,i}$ such that

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

10: $R_i = \begin{bmatrix} R_{i-1} \\ R_{2,i} \end{bmatrix}$

11: **end for**

4.6.2 The ADI method for the positive real singular optimal control problem

In the positive case we assume that $\mathscr{Y} = \mathscr{U} = \mathbb{C}^m$ and consider the cost functional (3.2) with $Q = 0$, $R = 0$, and $N = I$ (Section 3.6). In the singular optimal control problem we have that the Popov operator (3.4) satisfies (3.51). In the positive real case this means that

$$\langle u, (\mathbb{F}^* + \mathbb{F})u\rangle_{L^2(0,\infty;\mathbb{C}^m)} \geq 0. \tag{4.99}$$

With the notation introduced after (1.15), this condition is written in the form $\mathbb{F}^* + \mathbb{F} \geq 0$. The following theorem is the projected version of the positive real singular optimal control problem. The proof can be done by adapting the lines of the proof of Theorem 4.23 and therefore is omitted.

Theorem 4.28. *Let $\Sigma = (\mathbb{T}, \Phi, \Psi, \mathbb{F})$ be a strongly stable weakly regular linear system (cf. Definition 2.6.c) with generating operators (A, B, C, D). Let $\mathscr{Y} = \mathscr{U} = \mathbb{C}^m$. Further assume that the Popov operator $\mathscr{R} = \mathbb{F}^* + \mathbb{F}$ satisfies (4.99). Let $(\alpha_j)_{j=1}^{\infty}$ be a complex sequence with $\mathrm{Re}(\alpha_j) > 0$ for all $j \in \mathbb{N}$. Moreover, let $S_k \in \mathcal{L}(\mathscr{X}, \mathbb{C}^{km})$ and $F_k \in \mathbb{C}^{km\times km}$ be defined as in (4.32) and (4.33), respectively.*
Then, the matrix $F_k^* + F_k \in \mathbb{C}^{km\times km}$ is positive semidefinite. In particular, there exists some matrix $F_{\Xi,k} \in \mathbb{C}^{\ell_k\times km}$ such that

$$F_k^* + F_k = F_{\Xi,k}^* F_{\Xi,k}. \tag{4.100}$$

Further, there exists some operator $S_{\Xi,k} \in \mathcal{L}(\mathscr{X}, \mathbb{C}^{\ell_k})$ such that

$$F_{\Xi,k}^* S_{\Xi,k} = S_k. \tag{4.101}$$

For the orthogonal projector P_k as in (4.29), the operator X_k defined by

$$X_k = S_{\Xi,k}^* S_{\Xi,k} \tag{4.102}$$

fulfills

$$x_0^* X_k x_0 = \sup_{u\in L^2(0,\infty;\mathbb{C}^m)} -2\mathrm{Re}\,\langle u, P_k\mathbb{F}u + P_k\Psi x_0\rangle, \qquad \forall\, x_0 \in \mathscr{X}. \tag{4.103}$$

Again, we can formulate a convergence result. The proof is analogous to that of Theorem 4.25 and therefore omitted.

Theorem 4.29. *Let $\Sigma = (\mathbb{T}, \Phi, \Psi, \mathbb{F})$ be a strongly stable weakly regular linear system (cf. Definition 2.6.c) with generating operators (A, B, C, D). Let $\mathscr{Y} = \mathscr{U} = \mathbb{C}^m$. Further assume that the Popov operator $\mathscr{R} = \mathbb{F}^* + \mathbb{F}$ satisfies (4.99). Let $(\alpha_j)_{j=1}^{\infty}$ be a complex sequence with $\mathrm{Re}(\alpha_j) > 0$ for all $j \in \mathbb{N}$. Moreover, let $S_k \in \mathcal{L}(\mathscr{X}, \mathbb{C}^{km})$ and $F_k \in \mathbb{C}^{km\times km}$ be defined as in (4.32) and (4.33), respectively. Let X_k be defined as in Theorem 4.28. Then*

$$\langle x_0, X_k x_0\rangle \leq \langle x_0, X_{k+1}x_0\rangle \leq \langle x_0, X x_0\rangle, \qquad \forall\, x_0 \in \mathscr{X}, \quad \forall\, k \in \mathbb{N},$$

and the sequence $(X_k)_{k\in\mathbb{N}}$ *converges in the strong operator topology. Additionally, if* $(\alpha_j)_{j=1}^{\infty}$ *satisfies the non-Blaschke condition* (4.79)*, then* $(X_k)_{k\in\mathbb{N}}$ *converges even to* X*. This means that for all* $x_0 \in \mathscr{X}$

$$\lim_{k\to\infty} X_k x_0 = X x_0.$$

Moreover, if X *is compact, then* X_k *converges to* X *in the uniform operator topology and if* X *is in the Schatten class* $S_p(\mathscr{X})$ *for* $p \in [1,\infty]$*, then* X_k *converges to* X *in the topology of* $S_p(\mathscr{X})$*.*

Remark 4.30 (Projected positive real singular optimal control problem). We know that $F_i \in \mathbb{C}^{im\times im}$ has the lower triangular block structure (4.97). This allows us to determine $F_{\Xi,i} \in \mathbb{C}^{\ell_i\times mi}$ and $S_{\Xi,i} \in \mathcal{L}(\mathscr{X},\mathbb{C}^{\ell_i})$ recursively as follows (cf. Remark 4.27): We have

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

By making the ansatz $F_{\Xi,i} = \left[\begin{smallmatrix} F_{\Xi,i-1} & F_{\Xi 12,i} \\ 0 & F_{\Xi 22,i}\end{smallmatrix}\right]$, we obtain

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

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

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

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

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

It follows from Algorithm 1 that the operators S_i and S_{i-1} are related by (4.98). Hence, by making the ansatz $S_{\Xi,i} = \left[\begin{smallmatrix} S_{\Xi,i-1} \\ S_{\Xi 2,i}\end{smallmatrix}\right]$, we see that (4.101) can be written as

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

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

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

Remark 4.31 (Numerical effort for ADI iteration). Consider the discretized state space $V_h \subset \mathscr{X}$ with dimension n such that $p \ll n$ (p is the dimension of the output space $\mathscr{Y} = \mathbb{C}^p$). In this case the numerical effort for all steps in Algorithm 3 and Algorithm 4, except for the computation of the operator $V_i = V_{i-1} - (\alpha_i + \overline{\alpha_{i-1}})\cdot(\alpha_i I - A^*)^{-1}V_{i-1}$, are relatively negligible. The computation of V_i requires the solution of a PDE equation, which can be done by applying adaptive finite element methods, as shown by the help of a numerical example in Chapter 6.

Algorithm 4 ADI iteration for the positive real singular optimal control problem.

Input: $\Sigma = (\mathbb{T}, \Phi, \Psi, \mathbb{F})$ a strongly stable weakly regular linear system with generating operators (A, B, C, D) such that $\mathbb{F} + \mathbb{F}^* \geq 0$ and $\mathscr{Y} = \mathscr{U} = \mathbb{C}^m$. Shift parameters $(\alpha_i)_{i=1}^k$ with $\alpha_i \in \mathbb{C}$ and $\mathrm{Re}(\alpha_i) > 0$.

Output: $S_{\Xi,k} \in \mathcal{L}(\mathscr{X}, \mathbb{C}^{\ell_k})$ such that $S_{\Xi,k}^* S_{\Xi,k} = X_k \approx X$, where X is given by (3.67).

1: Perform Steps 1–5 in Algorithm 1

2: Determine a matrix $F_{\Xi,1} \in \mathbb{C}^{\ell_1 \times m}$ with full row rank such that

$$F_{\Xi,1}^* F_{\Xi,1} = F_1 + F_1^*$$

3: Determine an operator $S_{\Xi,1} \in \mathcal{L}(\mathscr{X}, \mathbb{C}^{\ell_1})$ such that

$$F_{\Xi,1}^* S_{\Xi,1} = S_1$$

4: **for** $i = 2, 3, \ldots, k$ **do**

5: Perform Steps 7–14 in Algorithm 1

6: Determine a matrix $F_{\Xi 12,i}$ such that

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

7: Determine a matrix $F_{\Xi 22,i}$ with full row rank such that

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

8: $F_{\Xi,i} = \begin{bmatrix} F_{\Xi,i-1} & F_{\Xi 12,i} \\ 0 & F_{\Xi 22,i} \end{bmatrix}$

9: Determine an operator $S_{\Xi 2,i}$ such that

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

10: $S_{\Xi,i} = \begin{bmatrix} S_{\Xi,i-1} \\ S_{\Xi 2,i} \end{bmatrix}$

11: **end for**

Chapter 5

The Newton-Kleinman method for the optimal control of regular linear systems

Truth is ever to be found in simplicity, and not in the multiplicity and confusion of things.

—Isaac Newton

Contents

We give an algorithmic approach to find the approximate solution of the linear-quadratic optimal control problem for regular linear systems whose dual system is also regular. The algorithm is an extension of the *Newton-Kleinman method* [29] to the infinite-dimensional spaces. This extension was studied in [12] for exponentially stable well-posed linear systems with bounded control and observation operators. We propose an extension to externally stable regular linear systems with unbounded control and observation operators. We construct a sequence of infinite-time observability Gramians to approximate the Riccati operator (3.10). We show the feasibility of the iterations by establishing an interconnection of the system with its anticausal dual at each Newton's iteration. To prove monotonicity and convergence of our algorithm, we set the additional assumption of strong stability on the system and require the control operator to be bounded.

5.1 The Newton-Kleinman iteration

Throughout this chapter the Hilbert spaces $\mathscr{U}$, $\mathscr{X}$, and $\mathscr{Y}$ are assumed to be *pivot spaces* (cf. Section 1.3). Moreover, we suppose that $\Sigma = (\mathbb{T}, \Phi, \Psi, \mathbb{F})$ is an externally stable regular linear system with generating operators $(A, B, C, 0)$, such that its dual system Σ^d is also regular (cf. Section 2.6). Moreover, we consider the cost functional (3.2) such that

$$R > 0, \qquad \widetilde{Q} := Q - N^* R^{-1} N \geq 0.$$

Our setting is mostly based on the results presented in Chapter 3. Note that we did not make any regularity assumption on $\Sigma = (\mathbb{T}, \Phi, \Psi, \mathbb{F})$ in Chapter 3 and the results presented there hold for all well-posed linear systems. In this chapter we mainly assume that Σ and its dual Σ^d are regular. Hence, we may replace $C_{\Lambda w}$ by C_Λ and $B^*_{\Lambda w}$ by B^*_Λ in all the formulas from Chapter 3.

We start by recalling the "Riccati-like" equation (3.37) from Chapter 3. This equation is our main inspiration for proposing the Newton-Kleinman iteration. We know from Section 3.2 that F^{opt} and A^{opt} can be characterized by the Riccati operator (3.10) as follows:

$$F^{\mathrm{opt}} x_0 = -R^{-1} \left(B^*_\Lambda X + N C_\Lambda \right) x_0, \tag{5.1}$$

$$A^{\mathrm{opt}} x_0 = (A + B F^{\mathrm{opt}}) x_0, \qquad \forall\, x_0 \in D(A^{\mathrm{opt}}). \tag{5.2}$$

The Riccati operator (3.10) satisfies the following Riccati-like equation, which holds on $D\left(A^{\mathrm{opt}}\right)$: For every $x_0, z_0 \in D(A^{\mathrm{opt}})$,

$$\begin{aligned} &\left\langle A^{\mathrm{opt}} x_0, X z_0 \right\rangle + \left\langle X x_0, A^{\mathrm{opt}} z_0 \right\rangle = \\ &\qquad - \left\langle C_\Lambda x_0, (Q C_\Lambda + N^* F^{\mathrm{opt}}) z_0 \right\rangle + \left\langle F^{\mathrm{opt}} x_0, B^*_\Lambda X z_0 \right\rangle. \end{aligned} \tag{5.3}$$

By substituting (5.1) in the right-hand side of (5.3), we obtain

$$\left\langle A^{\mathrm{opt}} x_0, X z_0 \right\rangle + \left\langle X x_0, A^{\mathrm{opt}} z_0 \right\rangle = - \left\langle C_\Lambda x_0, \widetilde{Q} C_\Lambda z_0 \right\rangle - \left\langle R^{-1} B^*_\Lambda X x_0, B^*_\Lambda X z_0 \right\rangle, \tag{5.4}$$

where $\widetilde{Q} = Q - N^* R^{-1} N$. The Riccati-like equation (5.4) motivates us to propose an iterative method, presented in Algorithm 5, to find an approximation of the Riccati operator (3.10). Algorithm 5 can be understood as a generalization of the Newton-Kleinman method [29]: Given a self-adjoint operator $X_k \in \mathcal{L}(\mathscr{X})$, $k = 0, 1, 2, \ldots$, define $A_k : D(A_k) \subset \mathscr{X} \to \mathscr{X}$ and $F_k : D(A_k) \subset \mathscr{X} \to \mathscr{U}$ as

$$\begin{aligned} A_k x_0 &:= (A + B F_k) x_0, \\ F_k x_0 &:= -R^{-1} \left(B^*_\Lambda X_k + N C_\Lambda \right) x_0, \end{aligned}$$

for all $x_0 \in D(A_k)$, where $D(A_k)$ is defined by

$$D(A_k) := \left\{ x_0 \in D(F_k) \mid (A + B F_k) x_0 \in \mathscr{X} \right\}.$$

The Newton-Kleinman iteration is to find a self-adjoint operator $X_{k+1} \in \mathcal{L}(\mathscr{X})$, for $k = 0, 1, 2, \ldots$, which satisfies the Lyapunov equation

$$\begin{aligned} &\langle A_k x_0, X_{k+1} z_0 \rangle + \langle X_{k+1} x_0, A_k z_0 \rangle = \\ &\qquad - \left\langle C_\Lambda x_0, \widetilde{Q} C_\Lambda z_0 \right\rangle - \left\langle R^{-1} B^*_\Lambda X_k x_0, B^*_\Lambda X_k z_0 \right\rangle, \end{aligned} \tag{5.5}$$

for all $x_0, z_0 \in D(A_k)$. In Algorithm 5 we give the Newton-Kleinman method to solve the linear-quadratic optimal control problem for externally stable regular linear systems whose dual system is also regular. Without loss of generality, we assume that the system has a zero feedthrough operator.

Algorithm 5 The Newton-Kleinman method for the optimal control of externally stable regular linear systems.

Input: $\Sigma = (\mathbb{T}, \Phi, \Psi, \mathbb{F})$ an externally stable regular linear system with generating operators $(A, B, C, 0)$ such that its dual system is also regular. Cost functional (3.2) such that $R > 0$ and $\widetilde{Q} := Q - N^* R^{-1} N \geq 0$.

Output: $\Psi_n^{\mathrm{aug}} \in \mathcal{L}\left(\mathscr{X}, L^2\left(0, \infty; \mathscr{Y} \times \mathscr{U}\right)\right)$, such that $\left(\Psi_n^{\mathrm{aug}}\right)^* \Psi_n^{\mathrm{aug}} \approx X$, where X is the Riccati operator (3.10).

1: $\Psi_0^{\mathrm{aug}} := \begin{pmatrix} \widetilde{Q}^{1/2}\Psi \\ 0 \end{pmatrix}$

2: **for** $k = 1, 2, \ldots, n$ **do**

3: $X_k = \left(\Psi_{k-1}^{\mathrm{aug}}\right)^* \Psi_{k-1}^{\mathrm{aug}}, \quad F_k = -R^{-1}\left(B_\Lambda^* X_k + N C_\Lambda\right), \quad A_k = A + B F_k$

4: $C_k^{\mathrm{aug}} = \begin{pmatrix} \widetilde{Q}^{1/2} C_\Lambda \\ R^{-1/2} B_\Lambda^* X_k \end{pmatrix} \in \mathcal{L}(D(C_\Lambda), \mathscr{Y} \times \mathscr{U}), \quad \Sigma_k^{\mathrm{aug}} := \begin{cases} \dot{x}(t) = A_k x(t) \\ y(t) = C_k^{\mathrm{aug}} x(t) \end{cases}$

5: Find the output map Ψ_k^{aug} corresponding to the pair $(A_k, C_k^{\mathrm{aug}})$

6: $k = k + 1$

7: **end for**

In the upcoming section we will show that

(i) The Lyapunov equation (5.5) is well-defined.

(ii) A_k generates a strongly continuous semigroup $\mathbb{T}^k$ on $\mathscr{X}$.

(iii) C_k^{aug} defined as

$$C_k^{\mathrm{aug}} := \begin{pmatrix} \widetilde{Q}^{1/2} C_\Lambda \\ R^{-1/2} B_\Lambda^* X_k \end{pmatrix} \in \mathcal{L}(D(C_\Lambda), \mathscr{Y} \times \mathscr{U}),$$

is an infinite-time admissible observation operator for the semigroup $\mathbb{T}^k$.

(iv) There exists a solution to (5.5) given by

$$X_{k+1} = \left(\Psi_k^{\mathrm{aug}}\right)^* \Psi_k^{\mathrm{aug}} \in \mathcal{L}(\mathscr{X}),$$

where
$$\Psi_k^{\text{aug}} \in \mathcal{L}\left(\mathscr{X}, L^2\left(0,\infty;\mathscr{Y}\times\mathscr{U}\right)\right)$$
is defined as
$$\left(\Psi_k^{\text{aug}} x_0\right)(t) = C_k^{\text{aug}} \mathbb{T}_t^k x_0, \qquad \forall\, x_0 \in \mathscr{X}, \quad t \geq 0.$$

Remark 5.1. If the input and output spaces are finite-dimensional (i.e., $\mathscr{U} = \mathbb{C}^m$ and $\mathscr{Y} = \mathbb{C}^p$, for some $m, p \in \mathbb{N}$), then Algorithm 5 provides approximative solutions of finite rank.

Remark 5.2. (a) Each iteration of the Newton-Kleinman method consists of a state feedback and a subsequent solution of the corresponding Lyapunov equation. In general, weak regularity is not preserved under feedback (see for example [58, Remark 7.5.4]). This is the main reason why we restrict ourselves to regular linear systems. Moreover, the dual of a regular linear system is not necessarily regular (see [61, Example 8.1]). Hence, it is necessary to further assume that the dual system is also regular.

(b) If the output space is finite-dimensional (i.e., $\mathscr{Y} = \mathbb{C}^p$ for some $p \in \mathbb{N}$), then the dual of a regular linear system is also regular. The reason is that weak regularity is equivalent to (strong) regularity, if the output space is finite-dimensional. Note that weak regularity is preserved under duality transformation [61].

5.2 Feasibility of the algorithm

In this section we show that equation (5.5) is well-defined for all $k = 0, 1, 2, \ldots$. To this end, we first recall the anticausal interpretation of the dual system Σ^d on the interval $[0,\infty)$ (cf. Section 2.6). This is an important ingredient in proving feasibility of Algorithm 5. In fact, we will show that at each iteration of the Newton-Kleinman algorithm, the infinite-time observability Gramian
$$X_{k+1} = \left(\Psi_k^{\text{aug}}\right)^* \Psi_k^{\text{aug}}$$
can be interpreted in terms of an interconnection of Σ_k^{aug} (Step 4 of Algorithm 5) with its anticausal dual. This idea is inspired by [70, Section 8] and we refer to this paper for more details.

Let $y^a \in L^2(0,\infty;\mathscr{Y})$, $x^a : [0,\infty) \to \mathscr{X}$, and $u^a \in L^2(0,\infty;\mathscr{U})$ be respectively the input function, the state trajectory, and the output function of the anticausal dual system Σ^a. As a result, we have (see Section 2.6)
$$\begin{aligned} x^a(t) &= \Psi^* \mathbf{S}_t^* y^a, \qquad \forall\, t \geq 0, \\ u^a &= \mathbb{F}^* y^a. \end{aligned} \tag{5.6}$$

It follows from [70, Theorem 6.3 and Proposition 5.2] that the state trajectories $x^a : [0,\infty) \to \mathscr{X}$ vanish at infinity. This means that
$$\lim_{t\to\infty} x^a(t) = 0.$$
For almost all $t \geq 0$, the functions y^a, x^a, and u^a satisfy
$$\begin{aligned} -\dot{x}^a(t) &= A^* x^a(t) + C^* y^a(t), \\ u^a(t) &= B_\Lambda^* x^a(t). \end{aligned} \tag{5.7}$$

Now we are ready to show feasibility of Algorithm 5 for externally stable regular linear systems whose dual system is also regular.

Theorem 5.3. *Let $\Sigma = (\mathbb{T}, \Phi, \Psi, \mathbb{F})$ be an externally stable regular linear system with generating operators $(A, B, C, 0)$, such that its dual system is also regular. Let J be the cost functional (3.2) such that $R > 0$ and $\widetilde{Q} := Q - N^* R^{-1} N \geq 0$. For all $k \in \mathbb{N}$, let X_k, F_k, A_k, C_k^{aug}, and Ψ_k^{aug} be the operators generated by Algorithm 5. Then, for all $k \in \mathbb{N}$*

(a) A_k generates a strongly continuous semigroup $\mathbb{T}^k$ on $\mathscr{X}$.

(b) $D(A_k) \subset D(C_\Lambda)$.

(c) C_Λ, $B_\Lambda^* X_k$, and F_k are infinite-time admissible observation operators for the semigroup $\mathbb{T}^k$. In particular,

$$C_k^{\mathrm{aug}} = \begin{pmatrix} \widetilde{Q}^{1/2} C_\Lambda \\ R^{-1/2} B_\Lambda^* X_k \end{pmatrix} \in \mathcal{L}(D(C_\Lambda), \mathscr{Y} \times \mathscr{U})$$

is an infinite-time admissible observation operator for $\mathbb{T}^k$.

(d) There exists a solution to (5.5) given by

$$X_{k+1} = \left(\Psi_k^{\mathrm{aug}}\right)^* \Psi_k^{\mathrm{aug}} \in \mathcal{L}(\mathscr{X}).$$

Proof. The proof is done by induction. Inspired by [70, Section 8], we interpret the sequence of infinite-time observability Gramians

$$X_{k+1} = \left(\Psi_k^{\mathrm{aug}}\right)^* \Psi_k^{\mathrm{aug}}$$

in terms of an interconnection of the respective systems with their anticausal dual systems.

Step 1 (base case). We start the induction by showing that the first iteration of the Newton-Kleinman algorithm is well-defined. Let $X_0 = 0$ and consider (5.5) for $k = 0$, i.e.,

$$\langle A x_0, X_1 z_0 \rangle + \langle X_1 x_0, A z_0 \rangle = -\left\langle C_\Lambda x_0, \widetilde{Q} C_\Lambda z_0 \right\rangle, \qquad \forall\, x_0, z_0 \in D(A). \tag{5.8}$$

The external stability of the regular linear system Σ (cf. Definition 2.6.a) means in particular that C_Λ is an infinite-time admissible observation operator for $\mathbb{T}$. Hence, it follows from Theorem 2.21 that X_1 given by

$$X_1 = \Psi^* \widetilde{Q} \Psi, \tag{5.9}$$

is a solution of the Lyapunov equation (5.8).

Next, we show that $B_\Lambda^* X_1$ and F_1 are infinite-time admissible observation operators for the semigroup $\mathbb{T}$. We follow a procedure similar to [70, Section 8]. Let $x : [0, \infty) \to \mathscr{X}$ and $y \in L^2(0, \infty; \mathscr{Y})$ be the state trajectory and output function of $\Sigma = (\mathbb{T}, \Phi, \Psi, \mathbb{F})$ corresponding to the input function $u \in L^2(0, \infty; \mathscr{U})$ and initial state $x_0 \in \mathscr{X}$. As a result, we have

$$y = \Psi x_0 + \mathbb{F} u.$$

Since Σ is time-invariant, we have for all $t \geq 0$ (see (2.10))

$$\mathbf{S}_t^* y = \Psi x(t) + \mathbb{F}\mathbf{S}_t^* u.$$

By setting $u = 0$ we obtain

$$\mathbf{S}_t^* y = \Psi x(t). \tag{5.10}$$

Now let y^a, x^a, and u^a be respectively the input function, the state trajectory, and the output function of the anticausal dual system Σ^a. We interconnect Σ with its anticausal dual system Σ^a by setting

$$y^a = \widetilde{Q}y, \tag{5.11}$$

as depicted in Figure 5.1.

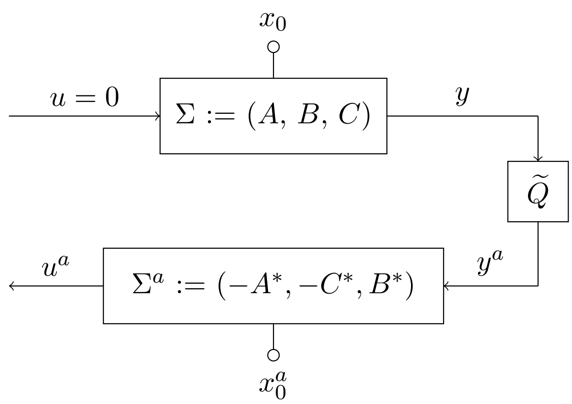


Figure 5.1: Interconnection of Σ and Σ^a.

At this point, by using the relations (5.6), (5.9) – (5.11), we obtain for all $t \geq 0$

$$x^a(t) = \Psi^* \mathbf{S}_t^* y^a = \Psi^* \widetilde{Q}\mathbf{S}_t^* y = \Psi^* \widetilde{Q}\Psi x(t) = X_1 x(t). \tag{5.12}$$

Now it follows from Proposition 2.32 and Theorem 2.39 that $x(t) \in D(C_\Lambda)$ and $x^a(t) \in D(B_\Lambda^*)$. Hence, we conclude from (5.12) that

$$X_1 x(t) \in D(B_\Lambda^*).$$

The output of the anticausal dual system is an L^2 function (cf. Section 2.6). In this regard, we obtain that

$$u^a(\cdot) = B_\Lambda^* x^a(\cdot) = B_\Lambda^* X_1 x(\cdot) \in L^2(0, \infty; \mathscr{U}).$$

Therefore, $B_\Lambda^* X_1$ is an infinite-time admissible observation operator for the semigroup $\mathbb{T}$. Since C_Λ and $B_\Lambda^* X_1$ are infinite-time admissible observation operators for $\mathbb{T}$, we have that

$$F_1 := -R^{-1}\left(B_\Lambda^* X_1 + NC_\Lambda\right)$$

is also an infinite-time admissible observation operator for $\mathbb{T}$. As a result, it follows from Theorem 2.47 that $A_1 := A + BF_1$ generates a strongly continuous semigroup $\mathbb{T}^1$ (set

$K = I$ and $D = 0$ in Theorem 2.47). Moreover, C_Λ, $B_\Lambda^* X_1$, and F_1 are infinite-time admissible observation operators for $\mathbb{T}^1$. Hence,

$$C_1^{\text{aug}} = \begin{pmatrix} \widetilde{Q}^{1/2} C_\Lambda \\ R^{-1/2} B_\Lambda^* X_1 \end{pmatrix} \in \mathcal{L}(D(C_\Lambda), \mathscr{Y} \times \mathscr{U})$$

is also an infinite-time admissible observation operator for $\mathbb{T}^1$. It remains to prove that

$$D(A_1) \subset D(C_\Lambda).$$

Let $x_0 \in D(A_1)$ and $s \in \mathbb{C}$ with $\operatorname{Re}(s) > \max\{\omega_0(\mathbb{T}), \omega_0(\mathbb{T}^1)\}$. Then, there exists some $w_0 \in \mathscr{X}$ such that $x_0 = (sI - A_1)^{-1} w_0$. By using the resolvent identity for $A_1 = A + BF_1$, we obtain

$$x_0 = (sI - A)^{-1} w_0 + (sI - A)^{-1} BF_1 x_0.$$

We know that $(sI - A)^{-1} w_0 \in D(A) \subset D(C_\Lambda)$. In addition, since Σ is regular, there holds (see (2.42) in Theorem 2.32)

$$(sI - A)^{-1} BF_1 x_0 \in D(C_\Lambda).$$

As a result, we obtain that $x_0 \in D(C_\Lambda)$ and therefore $D(A_1) \subset D(C_\Lambda)$.

Step 2 (inductive step). In order to show the main step of the induction, let us make the following induction hypothesis: For some $k \in \mathbb{N}$

(i) $X_k \in \mathcal{L}(\mathscr{X})$ is the k-th operator generated by Algorithm 5.

(ii) $A_k := A + BF_k$ generates a strongly continuous semigroup $\mathbb{T}^k$.

(iii) $C_k^{\text{aug}} \in \mathcal{L}(D(C_\Lambda), \mathscr{Y} \times \mathscr{U})$ is an infinite-time admissible observation operator for $\mathbb{T}^k$.

(iv) $D(A_k) \subset D(C_\Lambda)$.

Let us consider the *augmented observation operator* $C_k^{\text{aug}} \in \mathcal{L}(D(C_\Lambda), \mathscr{Y} \times \mathscr{U})$, which is defined by

$$C_k^{\text{aug}} = \begin{pmatrix} \widetilde{Q}^{1/2} C_\Lambda \\ R^{-1/2} B_\Lambda^* X_k \end{pmatrix}.$$

Furthermore, for almost all $t \geq 0$, let us define the *augmented system*

$$\Sigma_k^{\text{aug}} := \begin{cases} \dot{x}(t) = Ax(t) + Bu(t), \\ y(t) = C_k^{\text{aug}} x(t). \end{cases} \tag{5.13}$$

At this point we set up the state-feedback relationship $u(t) := F_k x(t) + v(t)$ (see Figure 5.2) and obtain the *augmented closed-loop system*

$$\Sigma_k^{\text{aug,cl}} := \begin{cases} \dot{x}(t) = (A + BF_k)x(t) + Bv(t), \\ y(t) = C_k^{\text{aug}} x(t). \end{cases} \tag{5.14}$$

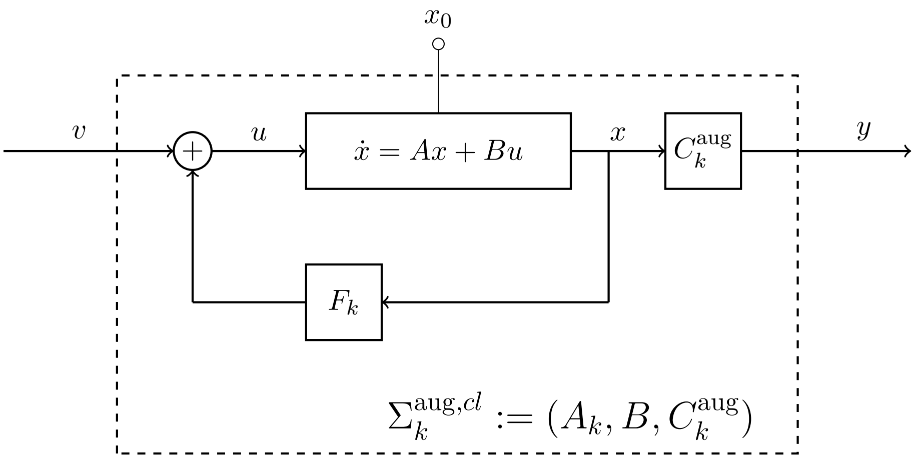


Figure 5.2: The augmented closed-loop system $\Sigma_k^{\text{aug,cl}}$.

Now let Ψ_k^{aug} and $\mathbb{F}_k^{\text{aug}}$ denote the output map and the input-output map of $\Sigma_k^{\text{aug,cl}}$, respectively. As a result, we have

$$y = \Psi_k^{\text{aug}} x_0 + \mathbb{F}_k^{\text{aug}} v.$$

Since $\Sigma_k^{\text{aug,cl}}$ is time-invariant, we have for all $t \geq 0$

$$\mathbf{S}_t^* y = \Psi_k^{\text{aug}} x(t) + \mathbb{F}_k^{\text{aug}} \mathbf{S}_t^* v.$$

By setting $v = 0$ we obtain

$$\mathbf{S}_t^* y = \Psi_k^{\text{aug}} x(t). \tag{5.15}$$

At the beginning of our inductive step we assumed that $A_k := A + BF_k$ generates a strongly continuous semigroup $\mathbb{T}_k$ and C_k^{aug} is an infinite-admissible observation operator for $\mathbb{T}_k$. Hence, it follows from Theorem 2.21 that the infinite-time observability Gramian of the pair (A_k, C_k^{aug}) is a solution of the Lyapunov equation (5.5). This means that

$$X_{k+1} = \left(\Psi_k^{\text{aug}}\right)^* \Psi_k^{\text{aug}} \tag{5.16}$$

is a solution of (5.5). At this point we show that X_{k+1}, as in (5.16), can be constructed using an interconnection of $\Sigma_k^{\text{aug,cl}}$ with its anticausal dual system:

(1) Let $\Sigma_k^{a,\text{aug}}$ denote the anticausal dual of the augmented closed-loop system $\Sigma_k^{\text{aug,cl}}$.

(2) Let y^a, x^a, and u^a be the input function, the state trajectory, and the output function of $\Sigma_k^{a,\text{aug}}$, respectively.

Then, for all $t \geq 0$ we have

$$\begin{cases} x^a(t) = \left(\Psi_k^{\text{aug}}\right)^* \mathbf{S}_t^* y^a, \\ \quad u^a = \left(\mathbb{F}_k^{\text{aug}}\right)^* y^a. \end{cases} \tag{5.17}$$

Moreover, for almost all $t \geq 0$ there holds (see (5.6) and (5.7))

$$\Sigma_k^{a,\text{aug}} := \begin{cases} -\dot{x}^a(t) = A_k^* x^a(t) + \left(C_k^{\text{aug}}\right)^* y^a(t), \\ \quad u^a(t) = B_\Lambda^* x^a(t). \end{cases} \tag{5.18}$$

We interconnect $\Sigma_k^{\text{aug,cl}}$ with $\Sigma_k^{a,\text{aug}}$ by setting

$$y^a = y, \tag{5.19}$$

as shown in Figure 5.3.

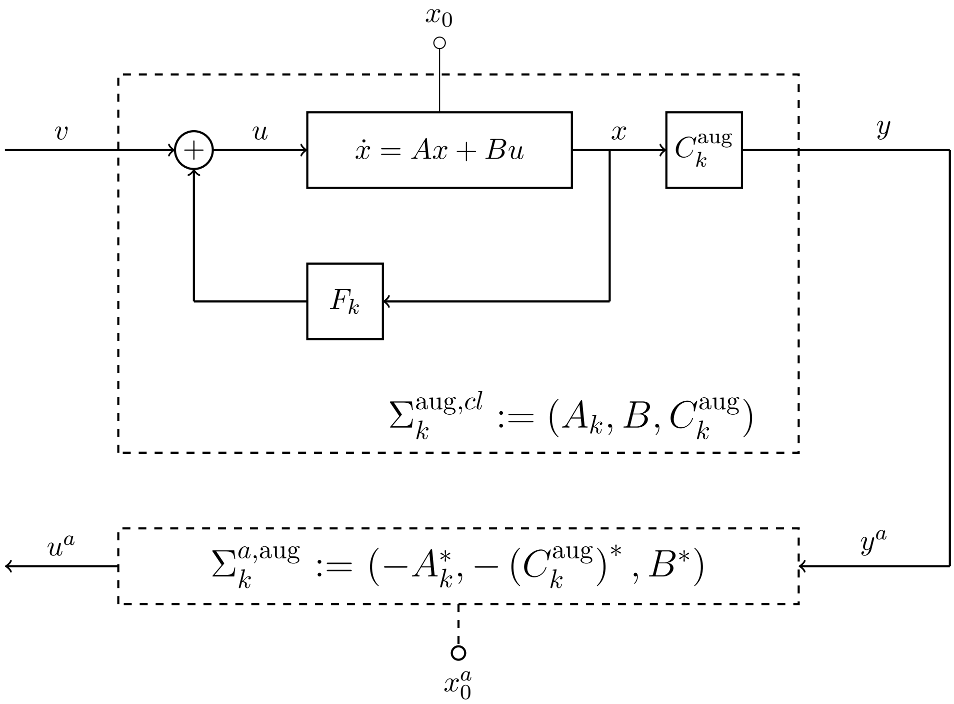


Figure 5.3: Interconnection of Σ_k^{aug} and $\Sigma_k^{a,\text{aug,cl}}$.

Using the relations (5.15) – (5.17) and (5.19), we obtain for all $t \geq 0$

$$\begin{aligned} x^a(t) &= \left(\Psi_k^{\text{aug}}\right)^* \mathbf{S}_t^* y^a = \left(\Psi_k^{\text{aug}}\right)^* \mathbf{S}_t^* y \\ &= \left(\Psi_k^{\text{aug}}\right)^* \Psi_k^{\text{aug}} x(t) = X_{k+1} x(t). \end{aligned} \tag{5.20}$$

Now it follows from Proposition 2.32 and Theorem 2.39 that $x(t) \in D(C_\Lambda)$ and $x^a(t) \in D(B_\Lambda^*)$, for almost all $t \geq 0$. Hence, we conclude from (5.20) that

$$X_{k+1} x(t) \in D(B_\Lambda^*).$$

The output of the anticausal dual system is an L^2 function (cf. Section 2.6). In this regard, we obtain

$$u^a(\cdot) = B_\Lambda^* X_{k+1} x(\cdot) \in L^2(0, \infty; \mathscr{U}).$$

Therefore, $B_\Lambda^* X_{k+1}$ is an infinite-time admissible observation operator for the semigroup $\mathbb{T}^k$. At this point, since C_Λ and $B_\Lambda^* X_{k+1}$ are infinite-time admissible observation operators for the semigroup $\mathbb{T}^k$, we obtain that

$$F_{k+1} := -R^{-1}\left(B_\Lambda^* X_{k+1} + NC_\Lambda\right) \tag{5.21}$$

is also an infinite-time admissible observation operator for $\mathbb{T}^k$. It follows from Theorem 2.47 that

$$A_{k+1} := A + BF_{k+1} = A_k + B(F_{k+1} - F_k)$$

generates a strongly continuous semigroup $\mathbb{T}^{k+1}$ on $\mathscr{X}$ (set $K = I$ and $D = 0$ in Theorem 2.47). Moreover, C_Λ and $B_\Lambda^* X_{k+1}$ are infinite-time admissible observation operators for $\mathbb{T}^{k+1}$. Hence,

$$C_{k+1}^{\mathrm{aug}} = \begin{pmatrix} \widetilde{Q}^{1/2} C_\Lambda \\ R^{-1/2} B_\Lambda^* X_{k+1} \end{pmatrix} \in \mathcal{L}(D(C_\Lambda), \mathscr{Y} \times \mathscr{U})$$

is also an infinite-time admissible observation operator for $\mathbb{T}^{k+1}$. Finally, by the resolvent identity (see the proof of $D(A_1) \subset D(C_\Lambda)$ in Step 1), we can show that

$$D(A_{k+1}) \subset D(C_\Lambda).$$

Its proof is analogous to Step 1 and therefore omitted. □

Remark 5.4. In the proof of Theorem 5.3, we showed that $B_\Lambda^* X_k$ is a bounded operator from $D(C_\Lambda)$ to $\mathscr{U}$, for all $k \in \mathbb{N}$. In addition, $D(A_k) \subset D(C_\Lambda)$. Hence, the right-hand side of the Lyapunov equation (5.5) is well-defined for all iterations $k \in \mathbb{N}$.

5.3 Connection between observability Gramians and the Riccati operator

In this section we establish a direct connection between the Riccati operator (3.10) and the infinite-time observability Gramians generated by Algorithm 5. Recall the augmented observation operator

$$C_k^{\mathrm{aug}} = \begin{pmatrix} \widetilde{Q}^{1/2} C_\Lambda \\ R^{-1/2} B_\Lambda^* X_k \end{pmatrix} \in \mathcal{L}(D(C_\Lambda), \mathscr{Y} \times \mathscr{U}),$$

and the augmented closed-loop system

$$\Sigma_k^{\mathrm{aug}} := \begin{cases} \dot{x}(t) = A_k x(t) + Bv(t), \\ y(t) = C_k^{\mathrm{aug}} x(t). \end{cases}$$

Let $\Psi_{C,k}$ and $\Psi_{B,k}$ denote the output maps corresponding to the pairs (A_k, C_Λ) and $(A_k, B_\Lambda^* X_k)$, respectively. With the "generator notation" (3.28) we have

$$\begin{aligned} (A_k, C_\Lambda) \quad &\rightsquigarrow \quad \Psi_{C,k}, \\ (A_k, B_\Lambda^* X_k) \quad &\rightsquigarrow \quad \Psi_{B,k}. \end{aligned} \tag{5.22}$$

As a result, we conclude that

$$(A_k, C_k^{\mathrm{aug}}) \qquad \rightsquigarrow \qquad \Psi_k^{\mathrm{aug}} = \begin{bmatrix} \widetilde{Q}^{1/2}\Psi_{C,k} \\ R^{-1/2}\Psi_{B,k} \end{bmatrix}.$$

As we have already shown, X_{k+1} (generated by Algorithm 5) is the infinite-time observability Gramian of the pair $(A_k, C_k^{\mathrm{aug}})$. In this regard, we obtain that

$$X_{k+1} = \left(\Psi_k^{\mathrm{aug}}\right)^* \Psi_k^{\mathrm{aug}} = \begin{bmatrix} \Psi_{C,k}^* & \Psi_{B,k}^* \end{bmatrix} \begin{bmatrix} \widetilde{Q} & 0 \\ 0 & R^{-1} \end{bmatrix} \begin{bmatrix} \Psi_{C,k} \\ \Psi_{B,k} \end{bmatrix}. \tag{5.23}$$

Now we present an important relation between the Riccati operator (3.10) and the observability Gramian (5.23).

Theorem 5.5. *Let $\Sigma = (\mathbb{T}, \Phi, \Psi, \mathbb{F})$ be an externally stable regular linear system with generating operators $(A, B, C, 0)$, such that its dual system is also regular. Let J be the cost functional (3.2) such that $R > 0$ and $\widetilde{Q} := Q - N^* R^{-1} N \geq 0$. Let $\mathscr{R}$ be the Popov operator (3.4) and assume that $\mathscr{R}$ is coercive, i.e., there exists some $\varepsilon > 0$ such that*

$$\langle u, \mathscr{R} u \rangle_{L^2(0,\infty;\mathscr{U})} \geq \varepsilon \, \|u\|_{L^2(0,\infty;\mathscr{U})}, \qquad \forall\; u \in L^2(0,\infty;\mathscr{U}).$$

Furthermore, let X be the Riccati operator (3.10). Let X_k, F_k, and A_k be the operators generated by Algorithm 5.
Then, with the notation (5.22), there holds

$$X_{k+1} = X + \Pi_k^* \mathscr{R} \Pi_k, \qquad \forall\; k \in \mathbb{N}, \tag{5.24}$$

where

$$\Pi_k := R^{-1}\Psi_{B,k} + R^{-1} N \Psi_{C,k} - \mathscr{R}^{-1}(\mathbb{F}^* Q + N)\Psi.$$

In particular, for all $x_0 \in \mathscr{X}$ and every $k \in \mathbb{N}$, it follows that

$$\langle x_0, X_k x_0 \rangle \geq \langle x_0, X x_0 \rangle. \tag{5.25}$$

Proof. Let $\mathbb{T}$ and $\mathbb{T}^k$ be strongly continuous semigroups with generators A and $A_k = A + BF_k$, respectively. We apply the perturbation relationship (2.56) to $A_k = A + BF_k$ and obtain that for every $x_0 \in D(A_k)$ and all $t \geq 0$ there holds

$$\begin{aligned} \mathbb{T}_t^k x_0 &= \mathbb{T}_t x_0 + \int_0^\infty \mathbb{T}_{t-\tau} B F_k \mathbb{T}_\tau^k x_0 \, d\tau, \\ &= \mathbb{T}_t x_0 - \int_0^\infty \mathbb{T}_{t-\tau} B R^{-1} (B_\Lambda^* X_k + N C_\Lambda) \mathbb{T}_\tau^k x_0 \, d\tau. \end{aligned} \tag{5.26}$$

We multiply (5.26) from left by C_Λ and use the notation (5.22) to obtain

$$\Psi_{C,k} = \Psi - \mathbb{F} R^{-1} \left(\Psi_{B,k} + N \Psi_{C,k}\right). \tag{5.27}$$

We further consider the following decomposition of the operator matrix $\left[\begin{smallmatrix} \widetilde{Q} & 0 \\ 0 & R \end{smallmatrix}\right]$:

$$\begin{bmatrix} \widetilde{Q} & 0 \\ 0 & R \end{bmatrix} = \begin{bmatrix} I & -N^* R^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} Q & N^* \\ N & R \end{bmatrix} \begin{bmatrix} I & 0 \\ -R^{-1} N & I \end{bmatrix}. \tag{5.28}$$

Using (5.28), we write X_{k+1} (given by (5.23)) as

$$\begin{aligned} X_{k+1} &= \begin{bmatrix} \Psi^*_{C,k} & \Psi^*_{B,k} R^{-1} \end{bmatrix} \begin{bmatrix} \widetilde{Q} & 0 \\ 0 & R \end{bmatrix} \begin{bmatrix} \Psi_{C,k} \\ R^{-1}\Psi_{B,k} \end{bmatrix} \\ &= \begin{bmatrix} \Psi^*_{C,k} & \Psi^*_{B,k} R^{-1} \end{bmatrix} \begin{bmatrix} I & -N^* R^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} Q & N^* \\ N & R \end{bmatrix} \begin{bmatrix} I & 0 \\ -R^{-1}N & I \end{bmatrix} \begin{bmatrix} \Psi_{C,k} \\ R^{-1}\Psi_{B,k} \end{bmatrix} \\ &= \begin{bmatrix} \Psi^*_{C,k} & \left(\Psi^*_{B,k} - \Psi^*_{C,k} N^*\right) R^{-1} \end{bmatrix} \begin{bmatrix} Q & N^* \\ N & R \end{bmatrix} \begin{bmatrix} \Psi_{C,k} \\ R^{-1}\left(\Psi_{B,k} - N\Psi_{C,k}\right) \end{bmatrix}. \end{aligned} \tag{5.29}$$

By substituting (5.27) in (5.29), we obtain

$$\begin{aligned} X_{k+1} &= \begin{bmatrix} \Psi - \mathbb{F}R^{-1}\left(\Psi_{B,k} + N\Psi_{C,k}\right) \\ R^{-1}\left(\Psi_{B,k} - N\Psi_{C,k}\right) \end{bmatrix}^* \begin{bmatrix} Q & N^* \\ N & R \end{bmatrix} \begin{bmatrix} \Psi - \mathbb{F}R^{-1}\left(\Psi_{B,k} + N\Psi_{C,k}\right) \\ R^{-1}\left(\Psi_{B,k} - N\Psi_{C,k}\right) \end{bmatrix} \\ &= P_k^* \begin{bmatrix} Q & N^* \\ N & R \end{bmatrix} P_k, \end{aligned} \tag{5.30}$$

where

$$P_k := \begin{bmatrix} \Psi \\ 0 \end{bmatrix} + \begin{bmatrix} -\mathbb{F} \\ I \end{bmatrix} R^{-1}\Psi_{B,k} - \begin{bmatrix} \mathbb{F} \\ I \end{bmatrix} R^{-1}N\Psi_{C,k}.$$

Now we observe that the Riccati operator (3.10) can be reformed as

$$\begin{aligned} X &= \Psi^* Q\Psi - \Psi^*(Q\mathbb{F} + N^*)\mathscr{R}^{-1}(\mathbb{F}^*Q + N)\Psi \\ &= \Psi^* Q\Psi - \begin{bmatrix} \Psi^* & 0 \end{bmatrix} \begin{bmatrix} Q & N^* \\ N & R \end{bmatrix} \begin{bmatrix} \mathbb{F} \\ I \end{bmatrix} \mathscr{R}^{-1} \begin{bmatrix} \mathbb{F}^* & I \end{bmatrix} \begin{bmatrix} Q & N^* \\ N & R \end{bmatrix} \begin{bmatrix} \Psi \\ 0 \end{bmatrix}. \end{aligned} \tag{5.31}$$

Using (5.30) and (5.31), we can apply a direct algebraic calculation to obtain

$$X_{k+1} = X + \Pi_k^* \mathscr{R} \Pi_k, \tag{5.32}$$

where

$$\begin{aligned} \Pi_k &= R^{-1}\Psi_{B,k} + R^{-1}N\Psi_{C,k} - \mathscr{R}^{-1} \begin{bmatrix} \mathbb{F}^* & I \end{bmatrix} \begin{bmatrix} Q & N^* \\ N & R \end{bmatrix} \begin{bmatrix} \Psi \\ 0 \end{bmatrix}, \\ &= R^{-1}\Psi_{B,k} + R^{-1}N\Psi_{C,k} - \mathscr{R}^{-1}\left(\mathbb{F}^*Q + N\right)\Psi. \end{aligned}$$

Since the Popov operator $\mathscr{R}$ is coercive, we have that

$$\langle x_0, \Pi_k^* \mathscr{R} \Pi_k x_0 \rangle \geq 0, \qquad \forall\, x_0 \in \mathscr{X}.$$

As a result, it follows from (5.32) that

$$\langle x_0, X_k x_0 \rangle \geq \langle x_0, X x_0 \rangle, \qquad \forall\, x_0 \in \mathscr{X}.$$

□

5.3.1 Open problem: calculation of X_k in terms of Ψ and $\mathbb{F}$

Note: Since this section is left as an open problem, the reader may skip this part and continue with Section 5.4.

The convergence analysis of Algorithm 5 would be much simpler if we could compute $\Psi_{B,k}$ and $\Psi_{C,k}$ in terms of Ψ and $\mathbb{F}$. In such case, we could determine X_k in terms of Ψ and $\mathbb{F}$ via (5.23). In order to make our claim easier to perceive, we present our idea by considering the first two iterations of the Newton-Kleinman algorithm (Algorithm 5). If we could generalize these calculations to all iterations, then we would be able to prove convergence of our algorithm independently of any Lyapunov equation. At this moment of time, we leave this part as an open problem.

The first iteration of Algorithm 5 produces the infinite-time observability Gramian X_1, which is given by

$$X_1 = \Psi^* \widetilde{Q} \Psi. \tag{5.33}$$

Let $\Sigma = (\mathbb{T}, \Phi, \Psi, \mathbb{F})$ be an externally stable regular linear system with generating operators $(A, B, C, 0)$, whose dual system is also regular. For almost all $t \geq 0$, let us construct the feedback relation $u(t) = F_1 x(t) + v(t)$, where

$$F_1 := -R^{-1} \left(B_\Lambda^* X_1 + N C_\Lambda \right) \tag{5.34}$$

and $u, v \in L^2(0, \infty; \mathscr{U})$ (as depicted in Figure 5.4). This means that for all $x_0 \in \mathscr{X}$ and for almost all $t \geq 0$, we have the following closed-loop system

$$\begin{cases} \dot{x}(t) = Ax(t) + Bu(t), \\ y(t) = C_\Lambda x(t), \\ z(t) = F_1 x(t), \\ u(t) = z(t) + v(t). \end{cases} \tag{5.35}$$

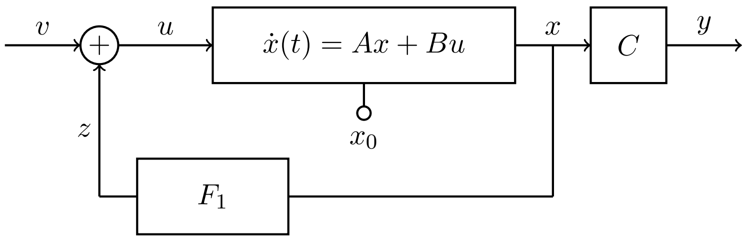


Figure 5.4: The system Σ with the state feedback $u(t) = F_1 x(t) + v(t)$.

Let $\widetilde{\Psi}_1$ and $\widetilde{\mathbb{F}}_1$ denote respectively the output map and the input-output map corresponding to the observation operator F_1 in (5.35). As a result, for every $x_0 \in \mathscr{X}$ and all $t \geq 0$ we have

$$\begin{cases} x(t) = \mathbb{T}_t x_0 + \Phi_t u, \\ y = \Psi x_0 + \mathbb{F} u, \\ z = \widetilde{\Psi}_1 x_0 + \widetilde{\mathbb{F}}_1 u, \\ u = z + v = \widetilde{\Psi}_1 x_0 + \widetilde{\mathbb{F}}_1 u + v, \end{cases} \tag{5.36}$$

where

$$\begin{aligned} \left(\widetilde{\Psi}_1 x_0\right)(t) &:= F_1 \mathbb{T}_t x_0, \\ \left(\widetilde{\mathbb{F}}_1 u\right)(t) &:= F_1 \Phi_t u. \end{aligned} \tag{5.37}$$

The following proposition gives a representation of $\widetilde{\Psi}_1$ and $\widetilde{\mathbb{F}}_1$ in terms of Ψ and $\mathbb{F}$.

Proposition 5.6. *Let $\Sigma = (\mathbb{T}, \Phi, \Psi, \mathbb{F})$ be an externally stable regular linear system with generating operators $(A, B, C, 0)$, whose dual system is also regular. Let X_1 be as in (5.33) and let F_1 be given by (5.34). Moreover, let J be the cost functional (3.2) such that $R > 0$ and $\widetilde{Q} = Q - N^* R^{-1} N \geq 0$. Furthermore, let $\widetilde{\Psi}_1$ and $\widetilde{\mathbb{F}}_1$ be defined by (5.37). Then,*

$$\begin{aligned} \widetilde{\Psi}_1 &= -R^{-1}\left(\mathbb{F}^* \widetilde{Q} + N\right)\Psi, \\ \left(\widetilde{\mathbb{F}}_1 u\right)(t) &= -R^{-1}\left(\mathbb{F}^* \widetilde{Q} \mathbb{F} \mathbf{P}_t u\right)(t) - R^{-1}\left(N \mathbb{F} u\right)(t), \qquad \forall\, t \geq 0. \end{aligned} \tag{5.38}$$

Proof. Recall the anticausal dual system Σ^a (Section 2.6) with generating operators $(A^*, C^*, B^*, 0)$: For almost all $t \geq 0$

$$\begin{cases} -\dot{x}^a(t) = A^* x^a(t) + C^* y^a(t), \\ u^a(t) = B^*_\Lambda x^a(t). \end{cases}$$

Moreover, we have

$$\begin{cases} x^a(t) = \Psi^* \mathbf{S}_t^* y^a, \qquad \forall\, t \geq 0, \\ u^a = \mathbb{F}^* y^a. \end{cases}$$

As a result, for almost all $t \geq 0$ there holds

$$B^*_\Lambda \Psi^* \mathbf{S}_t^* y^a = B^*_\Lambda x^a(t) = u^a(t) = \left(\mathbb{F}^* y^a\right)(t). \tag{5.39}$$

To determine $\widetilde{\Psi}_1$, we choose the input y^a of the anticausal dual system Σ^a to be

$$y^a = \widetilde{Q} \Psi x_0$$

(see Figure 5.1), and observe that for all $t \geq 0$

$$y^a = \widetilde{Q}\Psi x_0 = \widetilde{Q}\Psi \mathbb{T}_0 x_0 = \widetilde{Q}\Psi \mathbf{S}_t \mathbb{T}_t x_0 = \mathbf{S}_t \widetilde{Q} \Psi \mathbb{T}_t x_0.$$

By the property $\mathbf{S}_t^* \mathbf{S}_t = I$ we obtain

$$\mathbf{S}_t^* y^a = \widetilde{Q} \Psi \mathbb{T}_t x_0. \tag{5.40}$$

It follows from (5.33), together with (5.39) and (5.40), that

$$B^*_\Lambda X_1 \mathbb{T}_t x_0 = B^*_\Lambda \Psi^* \widetilde{Q} \Psi \mathbb{T}_t x_0 = B^*_\Lambda \Psi^* \mathbf{S}_t^* y^a = \left(\mathbb{F}^* y^a\right)(t) = \left(\mathbb{F}^* \widetilde{Q} \Psi x_0\right)(t).$$

Hence, we conclude that

$$\left(\widetilde{\Psi}_1 x_0\right)(t) = F_1 \mathbb{T}_t x_0 = -R^{-1}(B^*_\Lambda X_1 + N C_\Lambda)\mathbb{T}_t x_0 = -R^{-1}\left(\mathbb{F}^* \widetilde{Q} \Psi x_0 + N \Psi x_0\right)(t).$$

Therefore,

$$\widetilde{\Psi}_1 = -R^{-1}\left(\mathbb{F}^*\widetilde{Q} + N\right)\Psi.$$

To compute $\widetilde{\mathbb{F}}_1$, we recall from (2.9) that for all $t \geq 0$ there holds

$$\Psi\Phi_t = \mathbf{S}_t^*\mathbb{F} - \mathbb{F}\mathbf{S}_t^*. \tag{5.41}$$

From (5.33) together with (5.41) we see that

$$B_\Lambda^* X_1 \Phi_t u = B_\Lambda^* \Psi^* \widetilde{Q} \Psi \Phi_t u = B_\Lambda^* \Psi^* \widetilde{Q}(\mathbf{S}_t^*\mathbb{F}u - \mathbb{F}\mathbf{S}_t^* u).$$

Using the properties $\mathbf{S}_t^*\mathbf{S}_t = I$, $\mathbf{S}_t\mathbf{S}_t^* = I - \mathbf{P}_t$ (see (2.2)), and the shift-invariance of $\mathbb{F}$ (i.e., $\mathbf{S}_t\mathbb{F} = \mathbb{F}\mathbf{S}_t$), we obtain

$$\begin{aligned}
B_\Lambda^* X_1 \Phi_t u &= B_\Lambda^* \Psi^* \mathbf{S}_t^* \widetilde{Q}(\mathbb{F}u - \mathbf{S}_t\mathbb{F}\mathbf{S}_t^* u) \\
&= B_\Lambda^* \Psi^* \mathbf{S}_t^* \widetilde{Q}(\mathbb{F}u - \mathbb{F}\mathbf{S}_t\mathbf{S}_t^* u) \\
&= B_\Lambda^* \Psi^* \mathbf{S}_t^* \widetilde{Q}(\mathbb{F}u - \mathbb{F}(I - \mathbf{P}_t)u) \\
&= B_\Lambda^* \Psi^* \mathbf{S}_t^* \widetilde{Q}\mathbb{F}\mathbf{P}_t u \\
&= \left(\mathbb{F}^*\widetilde{Q}\mathbb{F}\mathbf{P}_t u\right)(t).
\end{aligned}$$

As a result of the above observation, we conclude that

$$\begin{aligned}
\left(\widetilde{\mathbb{F}}_1 u\right)(t) = F_1\Phi_t u &= -R^{-1}(B_\Lambda^* X_1 + NC_\Lambda)\Phi_t u \\
&= -R^{-1}\left(\mathbb{F}^*\widetilde{Q}\mathbb{F}\mathbf{P}_t u\right)(t) - R^{-1}\left(N\mathbb{F}u\right)(t).
\end{aligned}$$

□

At this point, it seems to be intricate (if not impossible) to show that $I - \widetilde{\mathbb{F}}_1$ is invertible for unbounded control operators $B \in \mathcal{L}(\mathscr{U}, \mathscr{X}_{-1})$ (a proof for bounded control operators $B \in \mathcal{L}(\mathscr{U}, \mathscr{X})$ is given as part of the proof of Proposition 5.7). If we could show that $I - \widetilde{\mathbb{F}}_1$ is boundedly invertible, then we would be able to calculate X_2 (generated by the second iteration of Algorithm 5) by

$$X_2 = X + \Pi_1^* \mathscr{R} \Pi_1, \tag{5.42}$$

where

$$\Pi_1 = \left[(I - \widetilde{\mathbb{F}}_1)^{-1}R^{-1} - \mathscr{R}^{-1}\right](\mathbb{F}^*Q + N)\Psi.$$

Calculation of X_2:

In order to obtain (5.42), we observe that if $I - \widetilde{\mathbb{F}}_1$ is boundedly invertible, then we get from the last equation in (5.36) that

$$u = (I - \widetilde{\mathbb{F}}_1)^{-1}\widetilde{\Psi}_1 x_0 + (I - \widetilde{\mathbb{F}}_1)^{-1} v. \tag{5.43}$$

By substituting (5.43) in the other equations of (5.36), we would obtain for all $t \geq 0$

$$\begin{cases} x(t) = \left(\mathbb{T}_t + \Phi_t(I - \widetilde{\mathbb{F}}_1)^{-1}\widetilde{\Psi}_1\right) x_0 + \Phi_t(I - \widetilde{\mathbb{F}}_1)^{-1} v(t), \\ y = \left(\Psi + \mathbb{F}(I - \widetilde{\mathbb{F}}_1)^{-1}\widetilde{\Psi}_1\right) x_0 + \mathbb{F}(I - \widetilde{\mathbb{F}}_1)^{-1} v, \\ z = \left(\widetilde{\Psi}_1 + \widetilde{\mathbb{F}}_1(I - \widetilde{\mathbb{F}}_1)^{-1}\widetilde{\Psi}_1\right) x_0 + \widetilde{\mathbb{F}}_1(I - \widetilde{\mathbb{F}}_1)^{-1} v, \\ u = (I - \widetilde{\mathbb{F}}_1)^{-1}\widetilde{\Psi}_1 x_0 + (I - \widetilde{\mathbb{F}}_1)^{-1} v. \end{cases} \tag{5.44}$$

Let us consider the closed-loop system (5.35). By substituting $u(t) = F_1 x(t) + v(t)$ in $\dot{x}(t) = Ax(t) + Bu(t)$, we obtain for almost all $t \geq 0$

$$\dot{x}(t) = Ax(t) + B\left(F_1 x(t) + v(t)\right) = (A + BF_1)x(t) + Bv(t) = A_1 x(t) + Bv(t).$$

Therefore, (5.35) can be reformed as

$$\begin{cases} \dot{x}(t) = A_1 x(t) + Bv(t), \\ y(t) = C_\Lambda x(t), \\ z(t) = F_1 x(t), \\ u(t) = z(t) + v(t). \end{cases} \tag{5.45}$$

By comparing (5.45) with (5.44) and using the "generator notation" (5.22), we conclude that

$$\begin{aligned} (A_1, C_\Lambda) \quad &\rightsquigarrow \quad \Psi + \mathbb{F}(I - \widetilde{\mathbb{F}}_1)^{-1}\widetilde{\Psi}_1, \\ (A_1, F_1) \quad &\rightsquigarrow \quad \widetilde{\Psi}_1 + \widetilde{\mathbb{F}}_1(I - \widetilde{\mathbb{F}}_1)^{-1}\widetilde{\Psi}_1 = (I - \widetilde{\mathbb{F}}_1)^{-1}\widetilde{\Psi}_1. \end{aligned}$$

Moreover, with $F_1 = -R^{-1}(B_\Lambda^* X_1 + NC_\Lambda)$ we obtain

$$\begin{aligned} \Psi_{C,1} &= \Psi + \mathbb{F}(I - \widetilde{\mathbb{F}}_1)^{-1}\widetilde{\Psi}_1, \\ -R^{-1}\Psi_{B,1} - R^{-1}N\Psi_{C,1} &= (I - \widetilde{\mathbb{F}}_1)^{-1}\widetilde{\Psi}_1. \end{aligned} \tag{5.46}$$

At this point, we can determine X_2 by setting $k = 1$ in (5.24) and applying (5.46) to obtain

$$X_2 = X + \Pi_1^* \mathscr{R} \Pi_1,$$

where

$$\begin{aligned} \Pi_1 &= R^{-1}\Psi_{B,1} + R^{-1}N\Psi_{C,1} - \mathscr{R}^{-1}(\mathbb{F}^* Q + N)\Psi \\ &= -(I - \widetilde{\mathbb{F}}_1)^{-1}\widetilde{\Psi}_1 - \mathscr{R}^{-1}(\mathbb{F}^* Q + N)\Psi \\ &= (I - \widetilde{\mathbb{F}}_1)^{-1} R^{-1}\left(\mathbb{F}^* \widetilde{Q} + N\right)\Psi - \mathscr{R}^{-1}(\mathbb{F}^* Q + N)\Psi. \end{aligned} \tag{5.47}$$

Note that in the last equality of (5.47) we have used (5.37).

Altogether, we are able to determine $\widetilde{\Psi}_1$ and $\widetilde{\mathbb{F}}_1$ in terms of Ψ and $\mathbb{F}$ as in Proposition 5.6. If we could additionally compute $(I - \widetilde{\mathbb{F}}_1)^{-1}$ in terms of Ψ and $\mathbb{F}$, then we would be able to determine X_2 completely in terms of Ψ and $\mathbb{F}$ (see (5.42)).

5.4 Monotonicity and convergence of the Newton-Kleinman iterations

The principal objective of this work was to prove convergence of the Newton-Kleinman algorithm for externally stable regular linear systems with unbounded control and observation operators. As we have already mentioned in Section 5.3.1, our idea to represent X_k in terms of Ψ and $\mathbb{F}$ would result into intricate lines of proof. In this regard, one could try to use the Lyapunov equations (5.5) to show convergence of the Newton-Kleinman iterations, similarly to the finite-dimensional case (see, e.g., [38, Chapter 11]). However, the main obstacle of this approach is to characterize $D(A + BF_k)$ at each Newton's iteration in the case where the control operator B is unbounded (i.e., $B \in \mathcal{L}(\mathscr{U}, \mathscr{X}_{-1})$). Specifically, for all $k \in \mathbb{N}$, the relation between $D(A_k)$ and $D(A_{k+1})$ is not known to us at this very moment in time. Nevertheless, we could circumvent this hurdle by assuming that the control operator B is bounded, i.e., $B \in \mathcal{L}(\mathscr{U}, \mathscr{X})$. In this case we can show that (see Appendix A.6)

$$D(A) = D(A_k) = D(A^{\mathrm{opt}}).$$

In addition, we let $\Sigma = (\mathbb{T}, \Phi, \Psi, \mathbb{F})$ be strongly stable, i.e., we let $\mathbb{T}$ be strongly stable (cf. Definition 2.6.c). This assumption is needed so that we can apply Theorem 3.11 to our setting.

In [12] monotonicity and convergence of the Newton-Kleinman iterations was proven for exponentially stable well-posed linear systems with bounded control and observation operators. The authors used Lyapunov equations of type (5.5) to prove monotonicity and convergence. In this section we extend their approach to show convergence of Newton's iterations for strongly stable regular linear systems with bounded control and unbounded observation operators. In addition, as for the previous sections, we let the dual system be regular.

Proposition 5.7. *Let $\Sigma = (\mathbb{T}, \Phi, \Psi, \mathbb{F})$ be a strongly stable regular linear system such that its dual system is also regular. Let $(A, B, C, 0)$ be the generating operators of Σ such that $B \in \mathcal{L}(\mathscr{U}, \mathscr{X})$. Let J be the cost functional from (3.2) such that $R > 0$ and $\widetilde{Q} = Q - N^*R^{-1}N \geq 0$. For all $k \in \mathbb{N}$, let $F_k = -R^{-1}(B_\Lambda^* X_k + NC_\Lambda)$ and $A_k = A + BF_k$ be the operators generated by Algorithm 5.*
Then, F_k is an admissible feedback operator for Σ (cf. Definition 2.43) and A_k generates a strongly stable semigroup $\mathbb{T}^k$.

Proof. For almost all $t \geq 0$, let us consider Σ with feedback relation $u(t) = F_k x(t) + v(t)$, where $u, v \in L^2(0, \infty; \mathscr{U})$. This means that we have the closed-loop system

$$\begin{cases} \dot{x}(t) = Ax(t) + Bu(t), \\ y(t) = C_\Lambda x(t), \\ z(t) = F_k x(t), \\ u(t) = z(t) + v(t). \end{cases} \tag{5.48}$$

Let $\widetilde{\Psi}_k$ and $\widetilde{\mathbb{F}}_k$ denote respectively the output map and the input-output map corresponding to the observation operator F_k in (5.48). As a result, for all $t \geq 0$ we have

$$\begin{cases} x(t) = \mathbb{T}_t x_0 + \Phi_t u, \\ y = \Psi x_0 + \mathbb{F} u, \\ z = \widetilde{\Psi}_k x_0 + \widetilde{\mathbb{F}}_k u, \\ u = z + v = \widetilde{\Psi}_k x_0 + \widetilde{\mathbb{F}}_k u + v. \end{cases} \tag{5.49}$$

In what follows we will show that $I - \widetilde{\mathbb{F}}_k$ is boundedly invertible, which means that F_k is an admissible feedback operator for Σ (compare with Section 5.3.1). To this end, we show that the transfer function of $\widetilde{\mathbb{F}}_k$ is uniformly line-regular with feedthrough operator $\widetilde{D}_k = 0$ (cf. Definition 2.29.d).

Let $\widetilde{\mathbf{G}}_k$ denote the transfer function of $\widetilde{\mathbb{F}}_k$. Because the semigroup $\mathbb{T}$ is strongly stable, the growth bound of $\mathbb{T}$ (cf. Definition 2.2) satisfies $\omega_0(\mathbb{T}) = 0$ and therefore

$$\widetilde{\mathbf{G}}_k(s) = F_k \left(sI - A\right)^{-1} B, \qquad \text{for} \ \ \mathrm{Re}(s) > 0.$$

As shown in the proof of Theorem 5.3, F_k is an infinite-time admissible observation operator for the semigroup $\mathbb{T}$. As a result, it follows from Theorem 2.14 that for every $\alpha > 0$, there exists $K_\alpha \geq 0$ such that

$$\left\| F_k \left(sI - A\right)^{-1} \right\| \leq \frac{K_\alpha}{\sqrt{\mathrm{Re}(s) - \alpha}}, \qquad \forall\, s \in \mathbb{C}_\alpha.$$

Therefore, for all $s \in \mathbb{C}_\alpha$ we obtain

$$\left\| \widetilde{\mathbf{G}}_k(s) \right\| \leq \frac{K_\alpha \left\| B \right\|}{\sqrt{\mathrm{Re}(s) - \alpha}}, \qquad \forall\, s \in \mathbb{C}_\alpha. \tag{5.50}$$

It follows from (5.50) that for any $\varepsilon > 0$, we can choose $\mathrm{Re}(s)$ sufficiently large such that

$$\left\| \widetilde{\mathbf{G}}_k(s) \right\| \leq \varepsilon.$$

Therefore, we obtain

$$\lim_{\mathrm{Re}(s) \to \infty} \left\| \widetilde{\mathbf{G}}_1(s) \right\| = 0. \tag{5.51}$$

Condition (5.51) means that $\widetilde{\mathbf{G}}_k$ is uniformly line-regular with feedthrough operator $\widetilde{D}_k = 0$. Since $I - \widetilde{D}_k = I$ is boundedly invertible, it follows from Proposition 2.30 that $I - \widetilde{\mathbb{F}}_k$ is also boundedly invertible. This means that F_k is an admissible feedback operator for Σ.

Since $I - \widetilde{\mathbb{F}}_k$ is boundedly invertible, we get from the last equation in (5.49) that

$$u = (I - \widetilde{\mathbb{F}}_k)^{-1} \widetilde{\Psi}_k x_0 + (I - \widetilde{\mathbb{F}}_k)^{-1} v. \tag{5.52}$$

By substituting (5.52) in the other equations of (5.49), we obtain for all $t \geq 0$ that

$$\begin{cases} x(t) = \left(\mathbb{T}_t + \Phi_t (I - \widetilde{\mathbb{F}}_k)^{-1} \widetilde{\Psi}_k \right) x_0 + \Phi_t (I - \widetilde{\mathbb{F}}_k)^{-1} v, \\ y = \left(\Psi + \mathbb{F} (I - \widetilde{\mathbb{F}}_k)^{-1} \widetilde{\Psi}_k \right) x_0 + \mathbb{F} (I - \widetilde{\mathbb{F}}_k)^{-1} v, \\ z = \left(\widetilde{\Psi}_k + \widetilde{\mathbb{F}}_k (I - \widetilde{\mathbb{F}}_k)^{-1} \widetilde{\Psi}_k \right) x_0 + \widetilde{\mathbb{F}}_k (I - \widetilde{\mathbb{F}}_k)^{-1} v, \\ u = (I - \widetilde{\mathbb{F}}_k)^{-1} \widetilde{\Psi}_k x_0 + (I - \widetilde{\mathbb{F}}_k)^{-1} v. \end{cases} \tag{5.53}$$

In the closed-loop system (5.48), if we substitute $u(t) = F_k x(t) + v(t)$ into $\dot{x}(t) = Ax(t) + Bu(t)$, we obtain for almost all $t \geq 0$

$$\dot{x}(t) = Ax(t) + B\left(F_k x(t) + v(t)\right) = (A + BF_k)x(t) + Bv(t) = A_k x(t) + Bv(t).$$

Therefore, (5.48) can be reformed as

$$\begin{cases} \dot{x}(t) = A_k x(t) + Bv(t), \\ y(t) = C_\Lambda x(t), \\ z(t) = F_k x(t), \\ u(t) = z(t) + v(t). \end{cases} \tag{5.54}$$

By comparing (5.54) with (5.53), we conclude that the strongly continuous semigroup

$$\mathbb{T}^k = \mathbb{T} + \Phi(I - \widetilde{\mathbb{F}}_k)^{-1}\widetilde{\Psi}_k$$

is generated by $A_k = A + BF_k$. Since $\mathbb{T}$ is strongly stable, it follows from Lemma 2.46 that $\mathbb{T}^k$ is also strongly stable. □

We are now ready to provide the main theorems of this section. Theorem 5.8 shows that the operator sequence $(X_k)_{k\in\mathbb{N}} \in \mathcal{L}(\mathscr{X})$ generated by Algorithm 5 is monotonically non-increasing. Subsequently, Theorem 5.9 demonstrates convergence of this sequence to the Riccati operator (3.10) in appropriate norms.

Theorem 5.8. *Let $\Sigma = (\mathbb{T}, \Phi, \Psi, \mathbb{F})$ be a strongly stable regular linear system such that its dual system is also regular. Let $(A, B, C, 0)$ be the generating operators of Σ such that $B \in \mathcal{L}(\mathscr{U}, \mathscr{X})$. In addition, let J be the cost functional (3.2) such that $R > 0$ and $\widetilde{Q} = Q - N^* R^{-1} N \geq 0$. Let X_k, F_k, and A_k be the operators produced by Algorithm 5. Then, the self-adjoint operators $X_k \in \mathcal{L}(\mathscr{X})$ satisfy*

$$\langle x_0, X_k x_0\rangle_{\mathscr{X}} \geq \langle x_0, X_{k+1} x_0\rangle_{\mathscr{X}} \geq \langle x_0, X x_0\rangle_{\mathscr{X}},$$

for every $x_0 \in \mathscr{X}$ and all $k \in \mathbb{N}$, where $X \in \mathcal{L}(\mathscr{X})$ is the Riccati operator (3.10).

Proof. It follows from Theorem 5.3 that for all $k \in \mathbb{N}$, $A_k = A + BF_k$ generates a strongly continuous semigroup $\mathbb{T}^k$ and the feedback operator F_k is an infinite-time admissible observation operator for $\mathbb{T}^k$. Furthermore, by Remark 3.7.a, we know that F^{opt} is an infinite-time admissible observation operator for the semigroup $\mathbb{T}^{\mathrm{opt}}$ with generator A^{opt}.

Since B is bounded, i.e., $B \in \mathcal{L}(\mathscr{X}, \mathscr{U})$, it follows from Theorem A.6 (with $C = F_k$ and $C = F^{\mathrm{opt}}$) that $D(A) = D(A_k) = D(A^{\mathrm{opt}})$. Hence, the Lyapunov equation (5.5) can be written as

$$\begin{aligned} &\langle A_k x_0, X_{k+1} z_0\rangle_{\mathscr{X}} + \langle X_{k+1} x_0, A_k z_0\rangle_{\mathscr{X}} = \\ &\qquad - \left\langle C_\Lambda x_0, \widetilde{Q} C_\Lambda z_0\right\rangle_{\mathscr{Y}} - \left\langle R^{-1} B^* X_k x_0, B^* X_k z_0\right\rangle_{\mathscr{U}}, \end{aligned} \tag{5.55}$$

for all $x_0, z_0 \in D(A_k) = D(A) = D(A^{\mathrm{opt}})$. Since Σ is a strongly stable regular linear system, it follows from Proposition 5.7 that the semigroup $\mathbb{T}^k$ is strongly stable for all $k \in$

$\mathbb{N}$. Therefore, by part (5) of Theorem 2.21, we conclude that the infinite-time observability Gramian

$$X_{k+1} = \left(\Psi_k^{\text{aug}}\right)^* \Psi_k^{\text{aug}}$$

is the unique solution of (5.55). Because B is bounded, the feedback operator F_k generated by Algorithm 5 satisfies $F_k = -R^{-1}(B^* X_k + NC_\Lambda)$. By adding the expression

$$\langle B(F_{k+1} - F_k)x_0, X_{k+1}z_0\rangle_{\mathscr{X}} + \langle X_{k+1}x_0, B(F_{k+1} - F_k)z_0\rangle_{\mathscr{X}}$$

to both sides of (5.55) and rearranging the terms, we obtain

$$\begin{aligned}\langle A_{k+1}x_0, X_{k+1}z_0\rangle_{\mathscr{X}} + \langle X_{k+1}x_0, A_{k+1}z_0\rangle_{\mathscr{X}} = \\ -\left\langle C_\Lambda x_0, \widetilde{Q}C_\Lambda z_0\right\rangle_{\mathscr{U}} - \left\langle R^{-1}B^* X_{k+1}x_0, B^* X_{k+1}z_0\right\rangle_{\mathscr{U}} \\ -\left\langle R^{-1}B^*(X_{k+1} - X_k)x_0, B^*(X_{k+1} - X_k)z_0\right\rangle_{\mathscr{U}} .\end{aligned} \tag{5.56}$$

Furthermore, by increasing the index in (5.55), we get

$$\begin{aligned}\langle A_{k+1}x_0, X_{k+2}z_0\rangle_{\mathscr{X}} + \langle X_{k+2}x_0, A_{k+1}z_0\rangle_{\mathscr{X}} = \\ -\left\langle C_\Lambda x_0, \widetilde{Q}C_\Lambda z_0\right\rangle_{\mathscr{Y}} - \left\langle R^{-1}B^* X_{k+1}x_0, B^* X_{k+1}z_0\right\rangle_{\mathscr{U}} ,\end{aligned} \tag{5.57}$$

for all $z_0, x_0 \in D(A_{k+1}) = D(A)$. Next, we subtract (5.57) from (5.56) to obtain

$$\begin{aligned}\langle A_{k+1}x_0, (X_{k+1} - X_{k+2})z_0\rangle_{\mathscr{X}} + \langle (X_{k+1} - X_{k+2})x_0, A_{k+1}z_0\rangle_{\mathscr{X}} \\ = -\left\langle R^{-1}B^*(X_{k+1} - X_k)x_0, B^*(X_{k+1} - X_k)z_0\right\rangle_{\mathscr{U}} .\end{aligned} \tag{5.58}$$

Since A_{k+1} generates a strongly stable semigroup $\mathbb{T}^{k+1}$ (see Proposition 5.7), it follows from part (5) of Theorem 2.21 that $X_{k+1} - X_{k+2}$ is the unique solution of (5.58) which satisfies

$$\langle x_0, (X_{k+1} - X_{k+2})x_0\rangle \geq 0, \qquad \forall\ x_0 \in \mathscr{X} .$$

By Theorem 5.5, we know that for all $x_0 \in \mathscr{X}$ and every $k \in \mathbb{N}$ there holds

$$\langle x_0, X_k x_0\rangle_{\mathscr{X}} \geq \langle x_0, X x_0\rangle_{\mathscr{X}} .$$

Hence, we conclude that for all $k \in \mathbb{N}$ and every $x_0 \in \mathscr{X}$

$$\langle x_0, X_k x_0\rangle_{\mathscr{X}} \geq \langle x_0, X_{k+1} x_0\rangle_{\mathscr{X}} \geq \langle x_0, X x_0\rangle_{\mathscr{X}} .$$

□

Now we are ready to prove convergence of the Newton-Kleinman algorithm to the Riccati operator (3.10).

Theorem 5.9. *With the assumptions of Theorem 5.8, the sequence of self-adjoint operators $(X_k)_{k\in\mathbb{N}} \in \mathcal{L}(\mathscr{X})$ produced by Algorithm 5 converges to the Riccati operator (3.10) in the strong operator topology as $k \to \infty$, i.e.,*

$$\lim_{k\to\infty} X_k x = X x, \qquad \forall\ x \in \mathscr{X} .$$

In addition,

(i) if X_1 is compact, then

$$\lim_{k\to\infty} \|X_k - X\|_{\mathcal{L}(\mathscr{X})} = 0.$$

(ii) if X_1 is in the Schatten class $\mathcal{S}_p(\mathscr{X})$ for some $p \in [1,\infty)$, then

$$\lim_{k\to\infty} \|X_k - X\|_{\mathcal{S}_p(\mathscr{X})} = 0.$$

Proof. By Theorem 5.8, we know that for all $x_0 \in \mathscr{X}$ and every $k \in \mathbb{N}$ there holds

$$\langle x_0, X_k x_0\rangle_{\mathscr{X}} \geq \langle x_0, X_{k+1} x_0\rangle_{\mathscr{X}} \geq \langle x_0, X x_0\rangle_{\mathscr{X}}.$$

As a result, it follows from Theorem A.2 that X_k converges to some operator $X_\infty \in \mathcal{L}(\mathscr{X})$ which satisfies

$$\langle x_0, X_\infty x_0\rangle_{\mathscr{X}} \geq \langle x_0, X x_0\rangle_{\mathscr{X}}, \qquad \forall\, x_0 \in \mathscr{X}.$$

In what follows we show that $X_\infty = X$. In terms of (5.24), this means that

$$\lim_{k\to\infty} \Pi_k^* \mathscr{R} \Pi_k = 0.$$

Let us assume that $\lim_{k\to\infty} \Pi_k^* \mathscr{R} \Pi_k = \Delta_\infty \geq 0$. As a result, we have

$$X_\infty = X + \Delta_\infty. \tag{5.59}$$

Recall that for every $x_0, z_0 \in D(A_k) = D(A)$ and all $k \in \mathbb{N}$, the operator X_{k+1} satisfies the Lyapunov equation

$$\begin{aligned} &\langle A_k x_0, X_{k+1} z_0\rangle_{\mathscr{X}} + \langle X_{k+1} x_0, A_k z_0\rangle_{\mathscr{X}} = \\ &\qquad - \left\langle C_\Lambda x_0, \widetilde{Q} C_\Lambda z_0 \right\rangle_{\mathscr{Y}} - \left\langle R^{-1} B^* X_k x_0, B^* X_k z_0\right\rangle_{\mathscr{U}}. \end{aligned} \tag{5.60}$$

By adding the expression

$$\left\langle B(F^{\mathrm{opt}} - F_k) x_0, X_{k+1} z_0\right\rangle_{\mathscr{X}} + \left\langle X_{k+1} x_0, B(F^{\mathrm{opt}} - F_k) z_0\right\rangle_{\mathscr{X}}$$

to both sides of (5.60) and rearranging the terms, we obtain

$$\begin{aligned} &\left\langle A^{\mathrm{opt}} x_0, X_{k+1} z_0\right\rangle_{\mathscr{X}} + \left\langle X_{k+1} x_0, A^{\mathrm{opt}} z_0\right\rangle_{\mathscr{X}} = \\ &\qquad - \left\langle C_\Lambda x_0, \widetilde{Q} C_\Lambda z_0 \right\rangle_{\mathscr{U}} - \left\langle R^{-1} B^* X x_0, B^* X z_0\right\rangle_{\mathscr{U}} \\ &\qquad - \left\langle R^{-1} B^* (X_k - X_{k+1}) x_0, B^* (X_k - X_{k+1}) z_0\right\rangle_{\mathscr{U}} \\ &\qquad + \left\langle R^{-1} B^* (X_{k+1} - X) x_0, B^* (X_{k+1} - X) z_0\right\rangle_{\mathscr{U}}. \end{aligned} \tag{5.61}$$

Now we subtract the Riccati-like equation (5.4) from (5.61) to get

$$\begin{aligned} &\left\langle A^{\mathrm{opt}} x_0, (X_{k+1} - X) z_0\right\rangle + \left\langle (X_{k+1} - X) x_0, A^{\mathrm{opt}} z_0\right\rangle = \\ &\qquad - \left\langle R^{-1} B^* (X_k - X_{k+1}) x_0, B^* (X_k - X_{k+1}) z_0\right\rangle_{\mathscr{U}} \\ &\qquad + \left\langle R^{-1} B^* (X_{k+1} - X) x_0, B^* (X_{k+1} - X) z_0\right\rangle_{\mathscr{U}}. \end{aligned} \tag{5.62}$$

Passing the limit $k \to \infty$ in (5.62) results in

$$\left\langle A^{\mathrm{opt}} x_0, \Delta_\infty z_0 \right\rangle_{\mathscr{X}} + \left\langle \Delta_\infty x_0, A^{\mathrm{opt}} z_0 \right\rangle_{\mathscr{X}} = \left\langle B^* \Delta_\infty x_0, B^* \Delta_\infty z_0 \right\rangle_{\mathscr{U}} . \tag{5.63}$$

Let $\widetilde{\Psi}_\infty$ denote the output map corresponding to the pair $(A^{\mathrm{opt}}, B^*\Delta_\infty)$. This means that for every $x_0 \in \mathscr{X}$ and all $t \geq 0$ there holds

$$\left(\widetilde{\Psi}_\infty x_0\right)(t) = B^* \Delta_\infty \mathbb{T}_t^{\mathrm{opt}} x_0. \tag{5.64}$$

At this point, we follow the lines of the proof for Theorem 3.11 to show that $\Delta_\infty = 0$. If $x_0, z_0 \in D(A) = D(A^{\mathrm{opt}})$, then $t \mapsto \langle \mathbb{T}_t^{\mathrm{opt}} x_0, \Delta_\infty \mathbb{T}_t^{\mathrm{opt}} z_0 \rangle$ is continuously differentiable on $[0, \infty)$ and we obtain

$$\frac{d}{dt}\left\langle \mathbb{T}_t^{\mathrm{opt}} x_0, \Delta_\infty \mathbb{T}_t^{\mathrm{opt}} z_0 \right\rangle_{\mathscr{X}} = \left\langle A^{\mathrm{opt}} \mathbb{T}_t^{\mathrm{opt}} x_0, \Delta_\infty \mathbb{T}_t^{\mathrm{opt}} z_0 \right\rangle_{\mathscr{X}} + \left\langle \mathbb{T}_t^{\mathrm{opt}} x_0, \Delta_\infty A^{\mathrm{opt}} \mathbb{T}_t^{\mathrm{opt}} z_0 \right\rangle_{\mathscr{X}} .$$

Since Δ_∞ satisfies (5.63), there holds

$$\begin{aligned} \frac{d}{dt}\left\langle \mathbb{T}_t^{\mathrm{opt}} x_0, \Delta_\infty \mathbb{T}_t^{\mathrm{opt}} z_0 \right\rangle_{\mathscr{X}} &= \left\langle B^* \Delta_\infty \mathbb{T}_t^{\mathrm{opt}} x_0, B^* \Delta_\infty \mathbb{T}_t^{\mathrm{opt}} z_0 \right\rangle_{\mathscr{U}} \\ &= \left\langle \left(\widetilde{\Psi}_\infty x_0\right)(t), \left(\widetilde{\Psi}_\infty z_0\right)(t) \right\rangle_{\mathscr{U}} , \end{aligned} \tag{5.65}$$

where we have used (5.64). By integrating (5.65), we obtain for each $t > 0$

$$\left\langle \mathbb{T}_t^{\mathrm{opt}} x_0, \Delta_\infty \mathbb{T}_t^{\mathrm{opt}} z_0 \right\rangle_{\mathscr{X}} - \langle x_0, \Delta_\infty z_0 \rangle_{\mathscr{X}} = \int_0^t \left\langle \left(\widetilde{\Psi}_\infty x_0\right)(\tau), \left(\widetilde{\Psi}_\infty z_0\right)(\tau) \right\rangle_{\mathscr{U}} d\tau. \tag{5.66}$$

Since $\mathbb{T}$ is a strongly stable semigroup, it follows from Lemma 2.46 that $\mathbb{T}^{\mathrm{opt}}$ is also a strongly stable semigroup (cf. Remark 3.6.b). This means that

$$\lim_{t \to \infty} \mathbb{T}_t^{\mathrm{opt}} x_0 = 0.$$

By letting $t \to \infty$ in (5.66), we conclude that

$$-\langle x_0, \Delta_\infty z_0 \rangle = \left\langle \widetilde{\Psi}_\infty x_0, \widetilde{\Psi}_\infty z_0 \right\rangle_{L^2(0,\infty;\mathscr{U})} .$$

This being true for all $x_0, z_0 \in D(A^{\mathrm{opt}})$, we must have

$$\Delta_\infty = -\widetilde{\Psi}_\infty^* \widetilde{\Psi}_\infty \leq 0.$$

As a result, the only possibility for Δ_∞ as above, is to fulfill $\Delta_\infty = 0$. Hence, we obtain from (5.59) that

$$X_\infty = X,$$

which shows convergence of the Newton-Kleinman iterations to the Riccati operator (3.10).

If additionally, we assume that X_1 (generated by the first Newton's iteration) is compact, then the non-increasing sequence $(X_k)_{k \in \mathbb{N}}$ is bounded from above by a compact operator and therefore converges in the uniform operator topology. If X_1 is in the Schatten class, then the non-increasing sequence $(X_k)_{k \in \mathbb{N}}$ is bounded from above by a Schatten class operator and therefore converges in the Schatten class topology. □

If $\Sigma = (\mathbb{T}, \Phi, \Psi, \mathbb{F})$ is exponentially stable (i.e., $\mathbb{T}$ is exponentially stable), then the semigroup $\mathbb{T}^{\mathrm{opt}}$ is also exponentially stable (cf. Remark 3.6). In this case one can show the quadratic rate of convergence for the Newton-Kleinman iterations [12, Theorem 6.4]. To this end, we need the following result from [18, p. 252] (see also [12, Theorem 5.5]):

Proposition 5.10. *Let $\mathbb{T}$ denote a strongly continuous semigroup on the Hilbert space $\mathscr{X}$ with generating operator $A : D(A) \subset \mathscr{X} \to \mathscr{X}$. If $\mathbb{T}$ is exponentially stable and $S = S^* \in \mathcal{L}(\mathscr{X})$, then*

$$\Gamma = \int_0^\infty \mathbb{T}_t^* S \mathbb{T}_t \, dt$$

is the unique solution of the Lyapunov equation

$$\langle Ax_0, \Gamma z_0\rangle_{\mathscr{X}} + \langle \Gamma x_0, Az_0\rangle_{\mathscr{X}} + \langle x_0, Sz_0\rangle_{\mathscr{X}} = 0, \qquad \forall\, x_0, z_0 \in D(A).$$

With the help of Proposition 5.10, we are now ready to show the quadratic rate of convergence for Algorithm 5:

Theorem 5.11. *(see [12, Theorem 6.4]) Let $\Sigma = (\mathbb{T}, \Phi, \Psi, \mathbb{F})$ be an exponentially stable regular linear system (cf. Definition 2.6.d) such that its dual system is also regular. Let $(A, B, C, 0)$ be the generating operators of Σ such that $B \in \mathcal{L}(\mathscr{U}, \mathscr{X})$. In addition, let J be the cost functional (3.2) such that $R > 0$ and $\widetilde{Q} = Q - N^*R^{-1}N \geq 0$. Furthermore, let X be the Riccati operator (3.10) and for all $k \in \mathbb{N}$, let X_k, F_k, and A_k be the operators produced by Algorithm 5. Then, the self-adjoint operators X_k satisfy*

$$\|X_{k+1} - X\| \leq c\,\|X_k - X\|^2, \qquad \forall\, k \in \mathbb{N},$$

where

$$c = \int_0^\infty \left\| \left(\mathbb{T}_t^{\mathrm{opt}}\right)^* \right\| \left\|BR^{-1}B^*\right\| \left\|\mathbb{T}_t^{\mathrm{opt}}\right\| \, dt \leq \frac{M^2}{2\omega} \left\|BR^{-1}B^*\right\|,$$

with the constants $M \geq 1$ and $\omega > 0$ given by (2.14).

Proof. We follow the lines of the proof of [12, Theorem 6.4]. As we showed in the proof of Theorem 5.9, the following Lyapunov equation holds for all $x_0, z_0 \in D(A) = D(A^{\mathrm{opt}})$ (see (5.62)):

$$\begin{aligned}\left\langle A^{\mathrm{opt}}x_0, (X_{k+1} - X)z_0\right\rangle_{\mathscr{X}} + \left\langle (X_{k+1} - X)x_0, A^{\mathrm{opt}}z_0\right\rangle_{\mathscr{X}} = \\ -\left\langle R^{-1}B^*(X_k - X_{k+1})x_0, B^*(X_k - X_{k+1})z_0\right\rangle_{\mathscr{U}} \\ +\left\langle R^{-1}B^*(X_{k+1} - X)x_0, B^*(X_{k+1} - X)z_0\right\rangle_{\mathscr{U}}.\end{aligned}$$

This means that $\Delta_{k+1} := X_{k+1} - X$ satisfies the Lyapunov equation

$$\left\langle A^{\mathrm{opt}}x_0, \Delta_{k+1}z_0\right\rangle_{\mathscr{X}} + \left\langle \Delta_{k+1}x_0, A^{\mathrm{opt}}z_0\right\rangle_{\mathscr{X}} + \langle x_0, Sz_0\rangle = 0, \tag{5.67}$$

with

$$S = (X_k - X_{k+1})BR^{-1}B^*(X_k - X_{k+1}) - (X_{k+1} - X)BR^{-1}B^*(X_{k+1} - X).$$

Because $\mathbb{T}$ is exponentially stable, we know from Remark 3.6 that the semigroup $\mathbb{T}^{\text{opt}}$ (generated by A^{opt}) is also exponentially stable. As a result, it follows from Proposition 5.10 that Δ_{k+1} is the unique solution of the Lyapunov equation (5.67) and is given by

$$\Delta_{k+1} = \int_0^\infty \left(\mathbb{T}_t^{\text{opt}}\right)^* S\, \mathbb{T}_t^{\text{opt}}\, dt. \tag{5.68}$$

By Theorem 5.5, we know that $X_{k+1} - X = \Pi_k^* \mathscr{R}^{-1} \Pi_k \geq 0$. In addition, it is clear that

$$(X_{k+1} - X) B R^{-1} B^* (X_{k+1} - X) \geq 0.$$

Hence, we observe from (5.68) that

$$0 \leq X_{k+1} - X \leq \int_0^\infty \left(\mathbb{T}_t^{\text{opt}}\right)^* \left[(X_k - X_{k+1}) B R^{-1} B^* (X_k - X_{k+1})\right] \mathbb{T}_t^{\text{opt}}\, dt.$$

Taking norms in the above inequality results in

$$\begin{aligned} \|X_{k+1} - X\| &\leq \|X_k - X_{k+1}\|^2 \int_0^\infty \left\|\left(\mathbb{T}_t^{\text{opt}}\right)^*\right\| \|BR^{-1}B^*\| \left\|\mathbb{T}_t^{\text{opt}}\right\| dt \\ &= c\, \|X_k - X_{k+1}\|^2, \end{aligned} \tag{5.69}$$

where

$$\begin{aligned} c &= \int_0^\infty \left\|\left(\mathbb{T}_t^{\text{opt}}\right)^*\right\| \|BR^{-1}B^*\| \left\|\mathbb{T}_t^{\text{opt}}\right\|\ dt \\ &= \|BR^{-1}B^*\| \int_0^\infty \left\|\left(\mathbb{T}_t^{\text{opt}}\right)^*\right\| \left\|\mathbb{T}_t^{\text{opt}}\right\|\ dt \leq \frac{M^2}{2\omega} \|BR^{-1}B^*\|. \end{aligned}$$

We know from Theorem 5.8 that for every $x_0 \in \mathscr{X}$ and all $k \in \mathbb{N}$ there holds

$$\langle x_0, X x_0 \rangle \leq \langle x_0, X_{k+1} x_0 \rangle \leq \langle x_0, X_k x_0 \rangle.$$

As a consequence, we have

$$0 \leq \langle x_0, (X_k - X_{k+1})\, x_0 \rangle \leq \langle x_0, (X_k - X)\, x_0 \rangle,$$

and therefore

$$\|X_k - X_{k+1}\| \leq \|X_k - X\|. \tag{5.70}$$

Now by using (5.69) and (5.70), we can conclude the quadratic rate of convergence

$$0 \leq \|X_{k+1} - X\| \leq c \|X_k - X\|^2.$$

□

5.5 The ADI method for approximation of the augmented output map

At Step 5 of Algorithm 5, one has to calculate the augmented output map Ψ_k^{aug} corresponding to the pair $(A_k, C_k^{\mathrm{aug}})$. For all $k \in \mathbb{N}$, we approximate Ψ_k^{aug} by applying the ADI method from [41]. Specifically, we adapt [41, Algorithm 1] to our setting in order to approximate the infinite-time observability Gramian

$$X_{k+1} = \left(\Psi_k^{\mathrm{aug}}\right)^* \Psi_k^{\mathrm{aug}}.$$

The adaptation of [41, Algorithm 1] to our setting is presented in Algorithm 6. For all $k \in \mathbb{N}$, this algorithm produces $S^k \in \mathcal{L}\left(\mathscr{X}, \mathscr{Y}^{i_k} \times \mathscr{U}^{i_k}\right)$ which approximates Ψ_k^{aug} (the numbers $i_k \in \mathbb{N}$ are to be determined a posteriori by a suitable stopping criteria). If the input and output spaces are finite-dimensional (i.e., $\mathscr{U} = \mathbb{C}^m$ and $\mathscr{Y} = \mathbb{C}^p$, for some $m, p \in \mathbb{N}$), then Algorithm 6 provides approximative solutions of finite rank.

Algorithm 6 The ADI method for approximation of the output map Ψ_k^{aug}.

Input: Operators A_k and C_k^{aug} from Step 4 of Algorithm 5, such that A_k generates a bounded semigroup $\mathbb{T}^k$. Shift parameters $\alpha_1, \ldots, \alpha_{i_k} \in \mathbb{C}$ with $\mathrm{Re}(\alpha_i) > 0$.
Output: $S^k := S_{i_k}^k = \iota_{\mathscr{Y}_{i_k}}^* \Psi_k^{\mathrm{aug}} \in \mathcal{L}\left(\mathscr{X}, \mathscr{Y}^{i_k} \times \mathscr{U}^{i_k}\right)$, such that $\widetilde{X}_{k+1}^{(i_k)} := \left(S^k\right)^* S^k \approx X_{k+1}$, where X_{k+1} is the infinite-time observability Gramian of the pair $(A_k, C_k^{\mathrm{aug}})$.

1: $V_1^k = (\alpha_1 I - A_k^*)^{-1} \left(C_k^{\mathrm{aug}}\right)^*$

2: $S_1^k = \sqrt{2\mathrm{Re}(\alpha_1)} \cdot \left(V_1^k\right)^*$

3: **for** $i = 1, 2, \ldots, i_k$ **do**

4: $\quad V_i^k = V_{i-1}^k - (\alpha_i + \overline{\alpha_{i-1}}) \cdot (\alpha_i I - A_k^*)^{-1} V_{i-1}^k$

5: $\quad S_i^k = \left[\left(S_{i-1}^k\right)^* \quad \sqrt{2\mathrm{Re}(\alpha_i)} \cdot V_i^k\right]$

6: **end for**

Remark 5.12. In Algorithm 6, we assumed that A_k generates a bounded semigroup $\mathbb{T}^k$. Therefore, the growth bound of $\mathbb{T}^k$ satisfies $\omega_0(\mathbb{T}^k) = 0$ and the shift parameters $\alpha_1, \ldots, \alpha_{i_k} \in \mathbb{C}$ must satisfy $\mathrm{Re}(\alpha_i) > 0$. One could extend Algorithm 6 to the case where the semigroup $\mathbb{T}^k$ is not necessarily bounded. In this case the shift parameters must satisfy $\mathrm{Re}(\alpha_i) > \omega_0(\mathbb{T}^k)$.

If the shift parameters $\alpha_1, \ldots, \alpha_{i_k} \in \mathbb{C}$ satisfy the non-Blaschke condition (4.79), then the convergence of Algorithm 6 follows from [41, Theorem 4.7]. In fact, the following convergence theorem holds true:

Theorem 5.13. *(see [41, Theorem 4.7]) Let $\mathscr{U}$, $\mathscr{X}$, $\mathscr{Y}$ be Hilbert spaces. For all $k \in \mathbb{N}$, let A_k and C_k^{aug} be the operators generated by Algorithm 5, such that $A_k : D(A_k) \subset \mathscr{X} \to \mathscr{X}$*

generates a bounded semigroup $\mathbb{T}^k$*. Let* $X_{k+1} \in \mathcal{L}(\mathscr{X})$ *be the infinite-time observability Gramian of the pair* $(A_k, C_k^{\mathrm{aug}})$*, which is given by*

$$X_{k+1} = \left(\Psi_k^{\mathrm{aug}}\right)^* \Psi_k^{\mathrm{aug}}.$$

Assume that the sequence of shifts parameters $(\alpha_i)_{i\in\mathbb{N}} \in \mathbb{C}$ *satisfies the non-Blaschke condition*

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

Then, Algorithm 6 is feasible and the operator sequence

$$\left(\widetilde{X}_{k+1}^{(i)}\right)_{i\in\mathbb{N}} = \left((S_i^k)^* S_i^k\right)_{i\in\mathbb{N}}$$

is monotonically non-decreasing and converges strongly to X_{k+1}*. This means that*

$$\lim_{i\to\infty} \widetilde{X}_{k+1}^{(i)} = X_{k+1}.$$

Moreover, the following holds true:

(i) If the Gramian X_{k+1} *is compact, then the operator sequence* $(\widetilde{X}_{k+1}^{(i)})_{i\in\mathbb{N}}$ *is also compact and there holds*

$$\lim_{i\to\infty} \left\|\widetilde{X}_{k+1}^{(i)} - X_{k+1}\right\|_{\mathcal{L}(\mathscr{X})} = 0.$$

(ii) If, for some $p \in [1,\infty)$*, the Gramian* X_{k+1} *is of* p*-th Schatten class, then the operator sequence* $(\widetilde{X}_{k+1}^{(i)})_{i\in\mathbb{N}}$ *is also of* p*-th Schatten class and there holds*

$$\lim_{i\to\infty} \left\|\widetilde{X}_{k+1}^{(i)} - X_{k+1}\right\|_{\mathcal{S}_p(\mathscr{X})} = 0.$$

An a posteriori stopping criteria for Algorithm 6

As we have already mentioned, for all $k \in \mathbb{N}$, the numbers $i_k \in \mathbb{N}$ within Algorithm 6 are to be determined a posteriori. In this part we give an efficient stopping criterion which is taken from [41, Remark 4.1.f].

For $i \in \mathbb{N}$, $i \geq 2$, let S_i^k and S_{i-1}^k be two successive approximations of the augmented output map Ψ_k^{aug}. Hence,

$$\widetilde{X}_{k+1}^{(i)} := \left(S_i^k\right)^* S_i^k \qquad \text{and} \qquad \widetilde{X}_{k+1}^{(i-1)} := \left(S_{i-1}^k\right)^* S_{i-1}^k$$

are two consecutive approximations of the infinite-time observability Gramian X_{k+1}. Now the recursively-defined operator

$$S_i^k = \left[\left(S_{i-1}^k\right)^* \quad \sqrt{2\mathrm{Re}(\alpha_i)} \cdot V_i^k\right]^*$$

allows us to calculate the norm of $\widetilde{X}_{k+1}^{(i)} - \widetilde{X}_{k+1}^{(i-1)}$ as follows:

$$\begin{aligned} \left\|\widetilde{X}_{k+1}^{(i)} - \widetilde{X}_{k+1}^{(i-1)}\right\|_{\mathcal{L}(\mathscr{X})} &= 2\mathrm{Re}(\alpha_i) \left\|V_i^k \left(V_i^k\right)^*\right\|_{\mathcal{L}(\mathscr{X})} \\ &= 2\mathrm{Re}(\alpha_i) \left\|\left(V_i^k\right)^* V_i^k\right\|_{\mathcal{L}(\mathscr{Y}\times\mathscr{U})}. \end{aligned} \tag{5.71}$$

As a result, a suitable stopping criterion for Algorithm 6 is to check whether the norm of $(V_i^k)^* V_i^k$ is below a certain absolute or relative threshold. If in addition, the input and output spaces are finite-dimensional (i.e., $\mathscr{U} = \mathbb{C}^m$ and $\mathscr{Y} = \mathbb{C}^p$, for some $m, p \in \mathbb{N}$), then $(V_i^k)^* V_i^k$ becomes an $(m+p) \times (m+p)$ matrix. We note that (5.71) holds true also in the Schatten norm $\mathcal{S}_p$ for $p \in [1, \infty)$ (see (6.32)).

Remark 5.14. (a) In [54, Section 4.6] an overview on possible stopping criteria for the ADI iteration to solve matrix Lyapunov equations is presented. These stopping criteria include the *relative change based criterion*, *residual based criterion*, and *stagnation based criterion*. The stopping criterion presented in this section is a relative change based criterion.

(b) Assume that the infinite-time observability Gramian

$$X_{k+1} = \left(\Psi_k^{\mathrm{aug}}\right)^* \Psi_k^{\mathrm{aug}}$$

is nuclear, i.e., $X_{k+1} \in \mathcal{S}_1(\mathscr{X})$. Let $\widetilde{X}_{k+1}^{(i)}$ be an approximation of X_{k+1} generated by Algorithm 6. Then, it follows from Theorem 5.13 that $\widetilde{X}_{k+1}^{(i)} \in \mathcal{S}_1(\mathscr{X})$ and we have

$$\left\| X_{k+1} - \widetilde{X}_{k+1}^{(i)} \right\|_{\mathcal{S}_1(\mathscr{X})} = \operatorname{trace}\left[X_{k+1}\right] - \operatorname{trace}\left[\widetilde{X}_{k+1}^{(i)}\right].$$

The trace of $\widetilde{X}_{k+1}^{(i)}$ can be determined by

$$\begin{aligned} \operatorname{trace}\left[\widetilde{X}_{k+1}^{(i)}\right] &= \operatorname{trace}\left[\sum_{j=1}^{i} 2\mathrm{Re}(\alpha_j) V_j^k \left(V_j^k\right)^*\right] \\ &= \sum_{j=1}^{i} 2\mathrm{Re}(\alpha_j) \operatorname{trace}\left[V_j^k \left(V_j^k\right)^*\right] \\ &= \sum_{j=1}^{i} 2\mathrm{Re}(\alpha_j) \operatorname{trace}\left[\left(V_j^k\right)^* V_j^k\right]. \end{aligned} \tag{5.72}$$

Chapter 6

Numerical examples

Only those who will risk going too far can possibly find out how far one can go.

—T. S. Eliot

Contents

We show the applicability of the algorithms developed in the previous chapters by applying them to a concrete example. The example that we consider throughout this chapter, is a two-dimensional heat equation with Robin boundary control and boundary integral observation. This example is taken from [41, Section 6], with the addition of a one-dimensional boundary integral observation. For ease of reference, we gather all the necessary results in Section 6.1 and refer the reader to [41, Section 6] for more details and proofs. We demonstrate the expected performance of the ADI method (cf. Chapter 4) and the Newton-Kleinman method (cf. Chapter 5) to solve the linear-quadratic optimal control problem, mainly in terms of monotonicity and convergence behavior. All the calculations were done either using the C++ library deal.II [3] or MATLAB 8.5 (R2015a) on a 64-bit server with 24 CPU cores of type Intel Xeon X5650 at 2.67 GHz and 48 GB main memory available. In particular, all the finite-element discretizations and mesh adaptations were done using the deal.II library.

6.1 Heat equation with Robin boundary control and boundary integral observation

Let $\Omega \subset \mathbb{R}^2$ be a bounded domain with piecewise C^2-boundary $\partial\Omega$. We consider the two-dimensional heat equation

$$\frac{\partial x}{\partial t}(\xi, t) = \Delta x(\xi, t), \quad (\xi, t) \in \Omega \times [0, \infty), \tag{6.1}$$

with initial condition $x(0, \xi) = x_0(\xi)$, $\xi \in \Omega$. Moreover, we consider the Robin boundary control

$$u(t) = \nu(\xi)^\top \nabla x(\xi, t) + ax(\xi, t), \quad (\xi, t) \in \partial\Omega \times [0, \infty), \quad a \in \mathbb{R}, \tag{6.2}$$

and one-dimensional boundary integral observation

$$y(t) = \int_{\partial\Omega} x(\xi, t)\ d\sigma_\xi, \tag{6.3}$$

where $d\sigma_\xi$ denotes the surface measure and $\nu(\xi)$ denotes the outward normal. Equations (6.1)–(6.3) can be formulated as an abstract dynamical system

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

To this end, we define the state space

$$x(t) := x(\cdot, t) \in L^2(\Omega) := \mathscr{X},$$

and let the input and output spaces both be one-dimensional, i.e.,

$$\mathscr{Y} = \mathscr{U} = \mathbb{C}.$$

For the operator A, we have

$$\begin{aligned} Ax &= \Delta x, \qquad \forall\ x \in D(A), \\ D(A) &= \left\{x \in H^1(\Omega) \mid \Delta x \in L^2(\Omega),\ \nu^T \nabla x + ax = 0 \text{ on } \partial\Omega\right\}. \end{aligned} \tag{6.4}$$

The spaces $\mathscr{X}_1$, $\mathscr{X}_{-1}$, $\mathscr{Z}_1$, and $\mathscr{Z}_{-1}$ are defined as in Section 2.2. In particular, we have

$$D(A^*)' = \mathscr{X}_{-1}, \qquad D(A^*) = \mathscr{Z}_1.$$

For the control operator B, we have

$$\langle Bu, x\rangle_{\mathscr{X}_{-1}, \mathscr{Z}_1} = u \cdot \int_{\partial\Omega} \overline{x(\xi)}\ d\sigma_\xi,$$

and it follows from the above construction of B that

$$B^* x = Cx = \int_{\partial\Omega} x(\xi)\ d\sigma_\xi, \quad \forall\ x \in D(A^*) = D(A). \tag{6.5}$$

Note that the system has a zero feedthrough operator (i.e., $D = 0$), because there is no direct relation between the input u and the output y.

As a result of the above formulation, the operators A, B, and C satisfy the following properties [41, Section 6]:

- A is self-adjoint and negative.
- A has a compact resolvent with $0 \in \rho(A)$.
- The input operator fulfills $B \in \mathcal{S}_2(\mathscr{U}, D((-A)^{1/2}))$, which is equivalent to

$$(-A)^{-1/2}B \in \mathcal{S}_2(\mathscr{U}, \mathscr{X}).$$

- $C = B^* \in \mathcal{L}(\mathscr{X}, \mathscr{Y})$.

Remark 6.1. By [41, Corollary 6.1], A generates an exponentially stable semigroup $\mathbb{T}$ on $\mathscr{X}$ and B is admissible for $\mathbb{T}$. As a result, the associated input map Φ satisfies Assumption (b) in Definition 2.4. Furthermore, as a consequence of the exponential stability of $\mathbb{T}$, we have that the operators Ψ and $\mathbb{F}$ are bounded, that is

$$\Psi \in \mathcal{L}(\mathscr{X}, L^2(0,\infty;\mathscr{Y})), \qquad \mathbb{F} \in \mathcal{L}(L^2(0,\infty;\mathscr{U}), L^2(0,\infty;\mathscr{Y})).$$

Altogether, we have that the quadruple $(A, B, C, 0)$ generates an exponentially stable regular well-posed linear system $\Sigma = (\mathbb{T}, \Phi, \Psi, \mathbb{F})$ [13].

Remark 6.2. Since the output space is one-dimensional (i.e., $\mathscr{Y} = \mathbb{C}$), we have that weak regularity is equivalent to strong regularity (cf. Remark 2.40). Hence, Σ and its dual Σ^d are both regular (cf. Remark 5.2.b).

6.2 Regular linear-quadratic optimal control problem

We consider the regular optimal control problem for the heat equation (6.1) with the scalar input function u formed by the Robin boundary condition (6.2) and the scalar output function y formed by the integral of Dirichlet boundary values (6.3). We consider the cost functional (3.2) with $Q = R = 1$ and $N = 0$ (recall that $\mathscr{U} = \mathscr{Y} = \mathbb{C}$). As a result, the quadratic optimal control problem is to minimize the cost functional

$$J(x_0, u) = \int_0^\infty |y(t)|^2 + |u(t)|^2 \; dt, \tag{6.6}$$

subject to

$$\begin{aligned} x(t) &= \mathbb{T}_t x_0 + \Phi_t u, \qquad \forall\, t \in [0,\infty), \\ y &= \Psi x_0 + \mathbb{F} u. \end{aligned}$$

From Proposition 3.2 (with $Q = R = 1$ and $N = 0$) we obtain

$$\langle x_0, X x_0 \rangle = \min_{u \in L^2(0,\infty;\mathscr{U})} J(x_0, u),$$

where the Riccati operator X is given by

$$X = \Psi^*\Psi - \Psi^*\mathbb{F}(I + \mathbb{F}^*\mathbb{F})^{-1}\mathbb{F}^*\Psi. \tag{6.7}$$

By applying the identity $(I + \mathbb{F}^*\mathbb{F})^{-1}\mathbb{F}^* = \mathbb{F}^*(I + \mathbb{F}\mathbb{F}^*)^{-1}$ to (6.7), we can reformulate the Riccati operator as (see Proposition 3.4)

$$X = \Psi^*(I + \mathbb{F}\mathbb{F}^*)^{-1}\Psi.$$

Remark 6.3. Let (A, B, C) be the triple of operators defined in Section 6.1. Since $B^* = C$ is bounded (i.e., $B^* = C \in \mathcal{L}(\mathscr{X}, \mathscr{Y})$) and $R = 1$, it follows from [70, Proposition 10.5] that X satisfies the following Riccati equation: for all $x_0, z_0 \in D(A^{\mathrm{opt}})$,

$$\langle z_0, A^* X x_0 \rangle_{\mathscr{X}} + \langle A^* X z_0, x_0 \rangle_{\mathscr{X}} + \langle C z_0, C x_0 \rangle_{\mathscr{Y}} = \langle B^* X z_0, B^* X x_0 \rangle_{\mathscr{Y}} . \tag{6.8}$$

In the first part of Section 6.2.1 (labeled as "Part I"), we use the discretized version of (6.8) and compute its residual norm in order to have a measure of convergence for the matrix version of the Riccati-ADI method (see [35, Section 8]).

Remark 6.4. Let (A, B, C) be the triple of operators defined in Section 6.1. The operator A generates an exponentially stable analytic semigroup $\mathbb{T}$ on $\mathscr{X} = L^2(\Omega)$ (analyticity of $\mathbb{T}$ follows from [32, Chapter 3, Appendix 3A]; see also [16, Example 1]). As a result, the operators A, B, and C satisfy the assumptions of Theorem A.5. Therefore, the following holds true:

(i) The Riccati operator X from (6.7) is the unique solution of (6.8) for all $x_0, z_0 \in D(A)$.

(ii) X is nuclear, i.e., $X \in \mathcal{S}_1(\mathscr{X})$.

(iii) The feedback operator $F := B^* X$ is Hilbert-Schmidt, i.e., $F \in \mathcal{S}_2(\mathscr{X}, \mathscr{U})$.

6.2.1 The Riccati-ADI method

We apply the Riccati-ADI method (see Theorem 4.16) to find an approximation of the Riccati operator (6.7). To do so, we approximate the operators Ψ and $\mathbb{F}$ by Algorithm 1 to obtain the operators S_k and F_k. Then, the approximated Riccati operator is obtained by

$$X_k = S_k^* (I + F_k^* F_k)^{-1} S_k. \tag{6.9}$$

Throughout this section we consider the L-shaped domain $\Omega := (0, 1)^2 \backslash (0.5, 1)^2$ and choose the Robin boundary coefficient to be $a = 1$.

Part I: The choice of shift parameters

In the first part of this section, we show the importance of choosing the correct set of shift parameters. Specifically, we show that if the non-Blaschke condition (4.79) is not fulfilled, then the ADI method may converge to the wrong solution of the optimal control problem (this being true even in the finite-dimensional setting). To simplify the analysis in this part we perform all the iterations on a fixed grid. This is equivalent to using the matrix version of the Riccati-ADI algorithm as in [35]. Later on in "Part II" of this section, we will apply adaptive mesh refinements to construct an efficient approximation of the Riccati operator.

For the discretization of the PDE (6.1), we apply the finite element method with uniform square elements of maximal diameter h. On this mesh, we define the subspace $V_h \subset H^1(\Omega)$ using piecewise-linear basis functions. We refine the mesh 5 times globally and obtain 3072 (active) cells with 3201 degrees of freedom (for piecewise-linear basis functions). To illustrate our global refinement strategy, an example of the mesh for one and two times global refinement is shown in Figure 6.1. Note that the dimension of the discretized problem is chosen to be small enough, so that we are able to compare the Riccati-ADI method with a direct solver (e.g., the "care" routine in MATLAB).

(a) One time global refinement will result in 12 cells with 16 degrees of freedom for piecewise-linear basis functions.

(b) Two times global refinement will result in 48 cells with 65 degrees of freedom for piecewise-linear basis functions.

Figure 6.1: Uniform global refinement of the domain $\Omega = (0,1)^2 \backslash (0.5,1)^2$ with square elements.

By discretizing the PDE (6.1), we obtain a finite-dimensional dynamical system

$$\begin{aligned} E\dot{x}(t) &= Ax(t) + Bu(t), \qquad x(0) = x_0 \in \mathbb{C}^n, \\ y(t) &= Cx(t), \end{aligned} \tag{6.10}$$

with the state space dimension $n = 3201$. The matrix $E \in \mathbb{R}^{n\times n}$ is a symmetric positive definite mass matrix, $A \in \mathbb{R}^{n\times n}$ is a symmetric stiffness matrix, $B \in \mathbb{R}^{n\times 1}$ is the input matrix, and $C \in \mathbb{R}^{1\times n}$ the output matrix. The system (6.10) is asymptotically stable and the matrix A is negative definite. Furthermore, we have $B = C^*$ and a simple calculation shows that the system is passive.

Remark 6.5. For invertible $E \in \mathbb{C}^{n\times n}$, if we consider the cost functional (6.6) subject to the discretized system (6.10), then its unique minimum satisfies

$$\langle Ex_0, XEx_0\rangle = \min_{u\in L^2(0,\infty)} \int_0^\infty |u(t)|^2 + |y(t)|^2 \, dt,$$

where $X \in \mathbb{C}^{n\times n}$ is the unique positive semidefinite solution of the algebraic Riccati equation

$$A^*XE + E^*XA + C^*C - E^*XBB^*XE = 0. \tag{6.11}$$

If we modify Steps 1 and 7 in Algorithm 1 to

1: $\quad V_1 = (\alpha_1 E^* - A^*)^{-1} C^*,$

7: $\quad V_k = V_{k-1} - (\alpha_k + \overline{\alpha_{k-1}}) \cdot (\alpha_k E^* - A^*)^{-1} E^* V_{k-1},$

then for X_k as computed by the Riccati-ADI method we have

$$\langle Ex_0, X_k Ex_0\rangle = \min_{u\in L^2(0,\infty)} \int_0^\infty |u(t)|^2 + |(P_{\mathscr{Y}_k} y)(t)|^2 \, dt.$$

We note that if E is the positive definite mass matrix of a finite element discretization, then $\langle Ex_0, XEx_0\rangle$ equals $\langle x_0, XEx_0\rangle$, where the inner-product in the latter expression is the one induced by the underlying function space.

We find approximations of the Riccati operator corresponding to the cost functional (6.6) subject to the discretized system (6.10) once using the "care" routine of MATLAB and once using the Riccati-ADI method with the modification of Algorithm 1 as in Remark 6.5. We note that although "care" does work for the example considered, the computation takes about 2 hours. For comparison, Riccati-ADI requires just 20 seconds. We denote by X the approximate solution obtained from the "care" routine and use it as a reference for the comparisons with the approximations obtained by Algorithm 1 (denoted by X_k).

The choice of the shift parameters has a major effect on the convergence speed of the ADI algorithm. We illustrate that if the shift parameters do not satisfy the non-Blaschke condition (4.79), then the matrix X_k obtained by the Riccati-ADI method may converge to a positive semidefinite matrix which is not a solution to (6.11) (cf. Theorem 4.22). To this end, we choose the following two different sets of shift parameters in our example:

1. The first set of shift parameters is chosen using Penzl's heuristic procedure [44] on the matrix pencil $\lambda E - A$. The underlying Arnoldi process is initialized with a random vector in $\mathbb{R}^n$. We compute 32 Ritz values by the Arnoldi process to approximate the eigenvalues of the matrix pencil $\lambda E - A$. Out of these 32 Ritz values, 11 values are calculated using the inverse Arnoldi method (to increase the accuracy of approximation). By this choice, we generate a set of 10 shift parameters, which we re-use every 10 iterations. We sort these 10 shift parameters in an increasing order with respect to the values of their real parts in order to obtain a smooth convergence in Algorithm 1. This cyclic choice of shift parameters satisfies the non-Blaschke condition (4.79). We note that since the matrices E and A are self-adjoint, the computed shift parameters are positive real numbers.

2. As a second set of shift parameters, we choose the infinite sequence $\alpha_k = k^3$, $k = 1, 2, \ldots$, for which the non-Blaschke condition is not satisfied.

We perform the simulations using the above two sets of shift parameters. At each iteration, we observe the absolute residual norm of (6.11) using the approach proposed in [34, Section 3.3]. This means that we exploit the low-rank form of the approximate solution $X_k = S_k(I + F_k F_k^*)^{-1} S_k^*$ to calculate the residual norm. Figure 6.2 shows the absolute residual norm with respect to the iteration for $n = 3201$ degrees of freedom.

Considering Figure 6.2, we observe that by choosing the second set of shift parameters, $\alpha_k = k^3$, our sequence converges to a matrix which is not the solution of the corresponding algebraic Riccati equation. On the contrary, the first choice of shift parameters provides convergence to the desired solution. Specifically, with a tolerance of 10^{-13} on the absolute residual norm, the Riccati-ADI algorithm converges to the solution of (6.11) in fewer than 50 iterations for the state space dimension $n = 3201$. Accordingly, we use the first set of shift parameters to continue with further analyses in "Part II" of this section.

We finish our analyses in this part by observing the relative 2-norm difference

$$\frac{\|X_k - X\|_2}{\|X\|_2}$$

at every iteration to show the convergence behavior of the Riccati-ADI algorithm. Figure 6.3 shows the relative 2-norm difference of the approximations obtained by the Riccati-ADI method with respect to the approximate solution obtained by the "care" routine in

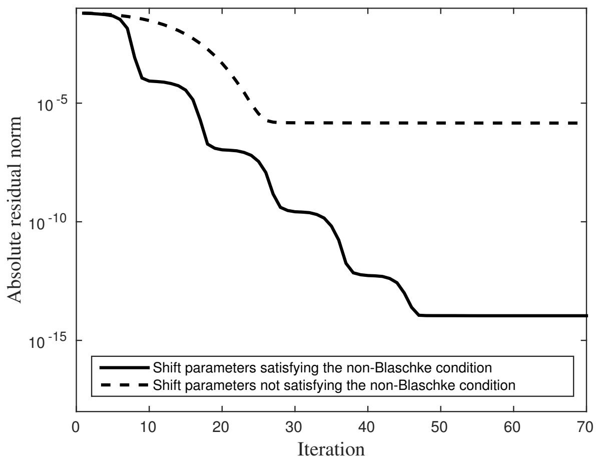


Figure 6.2: Comparison of two sets of shift parameters for Riccati-ADI: absolute residual norm of the Riccati equation (6.11) with respect to the iteration for $n = 3201$ degrees of freedom.

MATLAB. Note that since the matrices X_k and X are self-adjoint, their 2-norm difference equals the absolute value of the largest eigenvalue of $(X_k - X)$. This eigenvalue can be approximated efficiently using a power iteration without forming the product $X_k = S_k(I + F_k F_k^*)^{-1} S_k^*$ (see, e.g., [9]).

Part II: Monotonicity and convergence in the nuclear norm

In the second part of this section, we illustrate monotonicity and convergence of the Riccati-ADI method, which are proven in Theorems 4.21 and 4.22, respectively. Moreover, we demonstrate the efficiency of applying adaptive refinement techniques in Algorithm 1. In fact, we show that Riccati-ADI provides a good approximation of the Riccati operator using fewer degrees of freedom, when compared to the same method applied to a fixed grid.

To start the analysis, we refine the mesh 6 times globally (cf. Figure 6.1). As a result, we obtain 12288 (active) cells with 12545 degrees of freedom (for piecewise-linear basis functions). In all the analysis of this part, we use the first set of shift parameters from "Part I" (i.e., Penzl's heuristic procedure on the matrix pencil $\lambda E - A$).

The main computational effort of Algorithm 1 is the calculation of operators V_k, $k = 1, 2, \ldots,$ in Steps 1 and 7, namely

$$\begin{aligned} &1: \qquad V_1 = (\alpha_1 I - A^*)^{-1} C^*, \\ &7: \qquad V_k = V_{k-1} - (\alpha_k + \overline{\alpha_{k-1}}) \cdot (\alpha_k I - A^*)^{-1} V_{k-1}, \quad k = 2, 3, \ldots, \end{aligned}$$

Since the output space is one-dimensional (i.e., $\mathscr{Y} = \mathbb{C}$), we have $v_k := V_k \in L^2(\Omega)$. Besides, since in our example we have $A^* = A$ and $C^* = B$, Step 1 in Algorithm 1 can be written as

$$v_1 = (\alpha_1 I - A)^{-1} B,$$

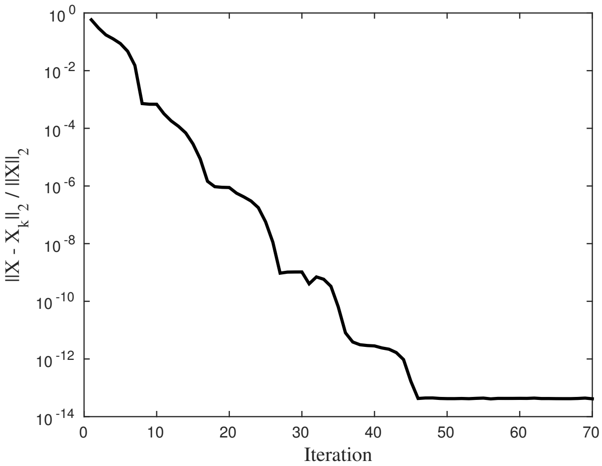


Figure 6.3: The relative 2-norm difference of the approximations obtained by the Riccati-ADI method with respect to the approximate solution obtained by the "care" routine in MATLAB.

and is therefore equivalent to the following boundary value Helmholtz problem [41]:
Find the function $v_1 \in L^2(\Omega)$ which satisfies

$$\begin{aligned} \alpha_1 \cdot v_1(\xi) - \Delta v_1(\xi) &= 0, \qquad && \xi \in \Omega, \\ \nu(\xi)^T \nabla v_1(\xi) + a v_1(\xi) &= 1, \qquad && \xi \in \partial\Omega. \end{aligned} \tag{6.12}$$

Similarly, expression $\hat{v}_k := (\alpha_k I - A^*)^{-1} v_{k-1}$ in Step 7 of Algorithm 1 is equivalent to solving an inhomogeneous boundary value Helmholtz problem:
Given $v_{k-1} \in L^2(\Omega)$, find the function $\hat{v}_k \in L^2(\Omega)$ which satisfies the differential equation

$$\begin{aligned} \alpha_k \cdot \hat{v}_k(\xi) - \Delta \hat{v}_k(\xi) &= v_{k-1}, \qquad && \xi \in \Omega, \\ \nu(\xi)^T \nabla \hat{v}_k(\xi) + a \hat{v}_k(\xi) &= 0, \qquad && \xi \in \partial\Omega. \end{aligned} \tag{6.13}$$

Eventually, we can determine v_k via

$$v_k = v_{k-1} - (\alpha_k + \overline{\alpha_{k-1}}) \hat{v}_k. \tag{6.14}$$

In the following, we put particular focus on calculating the solutions of (6.12) and (6.13). We solve these equations efficiently by applying the adaptive finite element method. Note that if we chose a fixed grid, then our algorithm would be equivalent to the approach of semi-discretization of the heat equation on a fixed grid with respect to space and then applying the matrix version of the Riccati-ADI method as shown in Part I of this section.

To ensure the reproducibility of the results presented in this section we list the important aspects regarding our implementation of the example using the C++ library deal.II:

(i) To discretize equations (6.12) and (6.13), we use uniform square elements and define the subspace $V_h \subset H^1(\Omega)$ using piecewise-linear basis functions (as in Part I of this section).

(ii) For numerical integration, we use the Gauss-Legendre quadrature rule of degree 2. This family is implemented in the class "QGauss" in the deal.II library.

(iii) To solve the discretized Helmholtz equations (6.12) and (6.13), we use a preconditioned conjugate gradient method for symmetric positive definite matrices. As a preconditioner, we use the symmetric successive overrelaxation (SSOR), see [52, Sections 9 & 10].

(iv) To approximate the discretization error of the PDEs (6.12) and (6.13), we use an a posteriori error estimator based on the discrete approximation of the gradient, as explained in Appendix A.1.1. This error estimator is computationally efficient, easy to implement, and works well for the example considered in this section. However, one may improve the approximation of the discretization error by applying a standard residual based L^2-error estimator (see, e.g., [1]). We note that more investigations are required to inspect the relation between the error estimates of the discretization and the convergence errors of the ADI algorithm. This is beyond the scope of this dissertation and is left as an open problem.

(v) For the successive mesh adaptations, we apply the *bulk criterion* from Appendix A.1.2, with top fraction (r) and bottom fraction (c) set to be

$$r = c = 0.01.$$

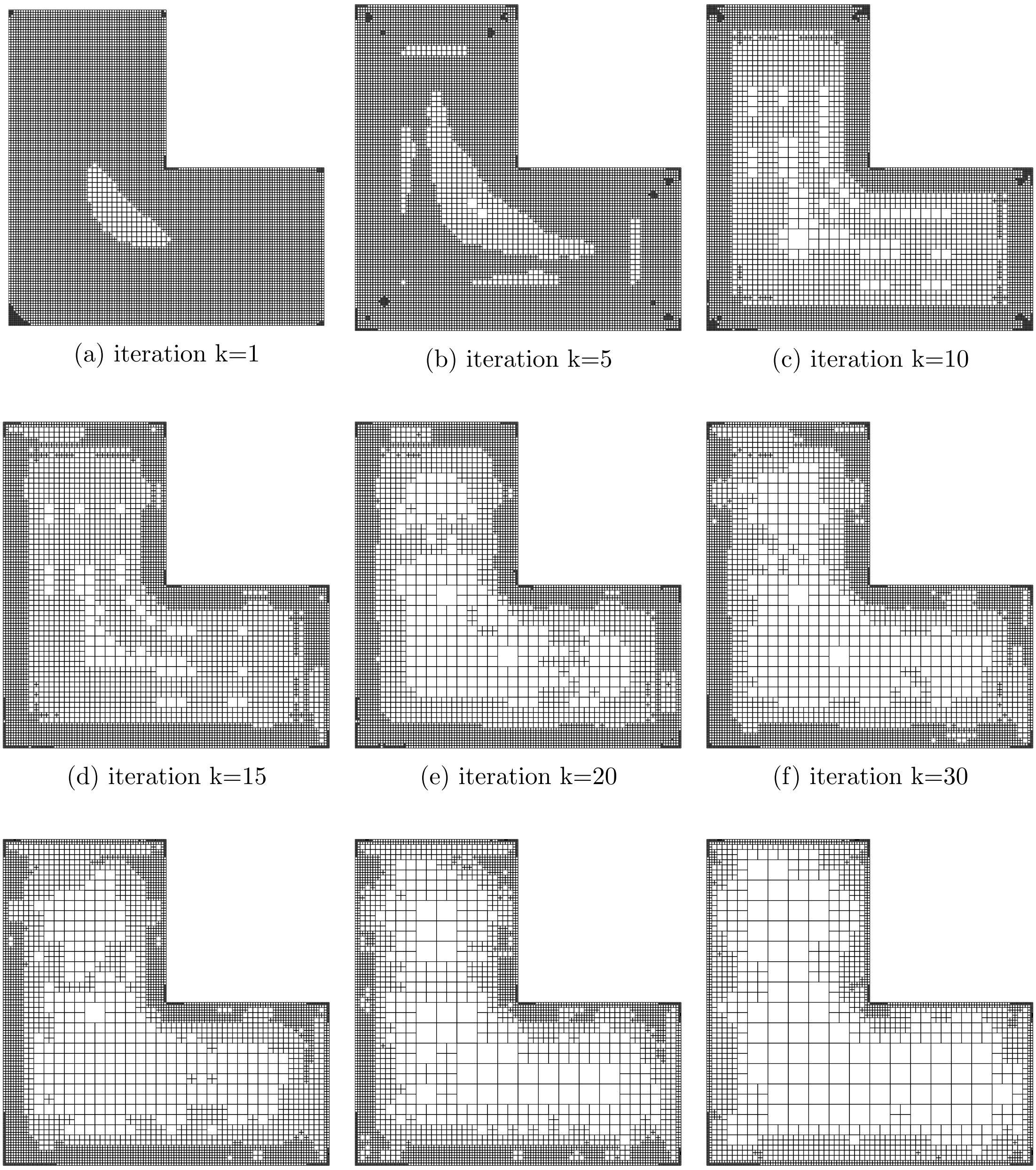
In order to capture the numerical error at the edges, the cells at the boundary are just allowed to be refined (i.e., we clear all the coarsening flags for the cells at the boundary).

Figure 6.4 shows a sequence of the adapted meshes produced by Algorithm 1. It is important to note that after each mesh adaptation one needs to interpolate the previous solution v_{k-1} on the new mesh, because v_{k-1} acts as the right-hand side of (6.13). This interpolation can be done by employing the class "SolutionTransfer" in the deal.II library. Figure 6.5 shows a sequence of solutions v_k on the adapted meshes produced by Algorithm 1.

Now we demonstrate monotonicity and convergence of the Riccati-ADI method. We recall from Remark 6.4 that the Riccati operator (6.7) corresponding to our example is nuclear, i.e., $X \in \mathcal{S}_1(\mathscr{X})$. By Theorem 4.22, we have that $X_k \in \mathcal{S}_1(\mathscr{X})$ for all $k \in \mathbb{N}$ and the sequence $(X_k)_{k\in\mathbb{N}}$ convergences to X in the nuclear norm (provided that the shift parameters satisfy the non-Blaschke condition (4.79)).

We compute the nuclear norm of X_k at each iteration of the Riccati-ADI algorithm. This can be done efficiently by using the low-rank factors of $X_k = S_k^*(I + F_k F_k^*)^{-1} S_k$. More specifically, by computing the Cholesky factorization $I + F_k F_k^* = L_k L_k^*$ we obtain

$$\begin{aligned}
\|X_k\|_{\mathcal{S}_1(\mathscr{X})} &= \operatorname{trace}[X_k] = \sum_i \langle e_i, X_k e_i \rangle_{\mathscr{X}} \\
&= \sum_i \left\langle e_i, S_k^*(L_k L_k^*)^{-1} S_k e_i \right\rangle_{\mathscr{X}} = \sum_i \left\langle e_i, S_k^* L_k^{-*} L_k^{-1} S_k e_i \right\rangle_{\mathscr{X}} \\
&= \sum_i \left\langle L_k^{-1} S_k e_i, L_k^{-1} S_k e_i \right\rangle_{\mathscr{Y}^k} = \sum_i \left\| L_k^{-1} S_k e_i \right\|_{\mathscr{Y}^k}^2,
\end{aligned}$$

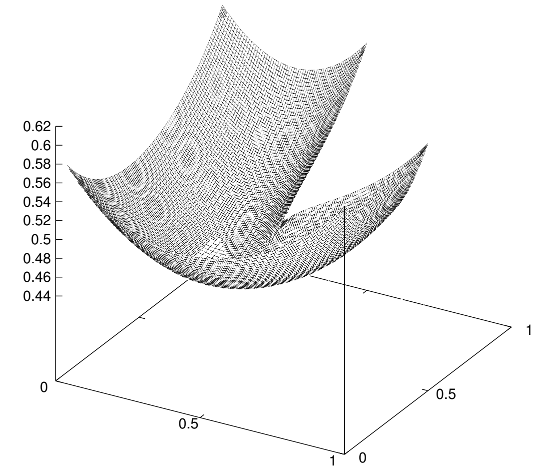


(a) iteration k=1 (b) iteration k=5 (c) iteration k=10

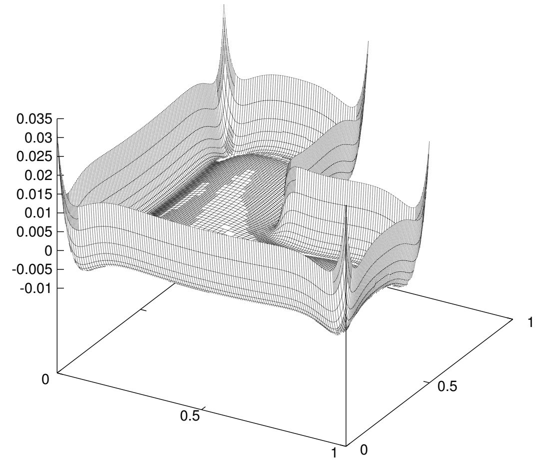
(d) iteration k=15 (e) iteration k=20 (f) iteration k=30

(g) iteration k=40 (h) iteration k=50 (i) iteration k=60

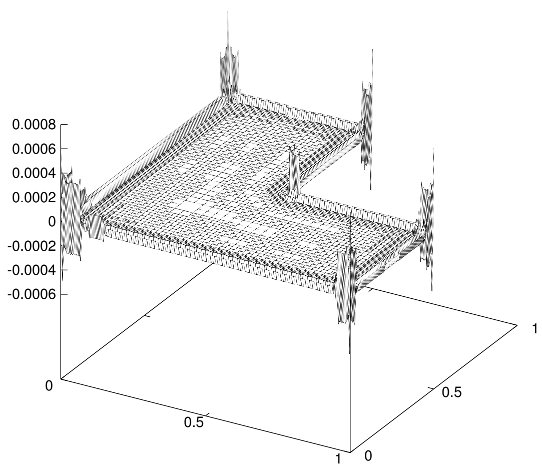
Figure 6.4: Sequence of adapted meshes produced by the ADI method using the a posteriori error estimator from Appendix A.1.1 and the bulk mesh adaptation strategy from Appendix A.1.2 with $r = c = 0.01$.



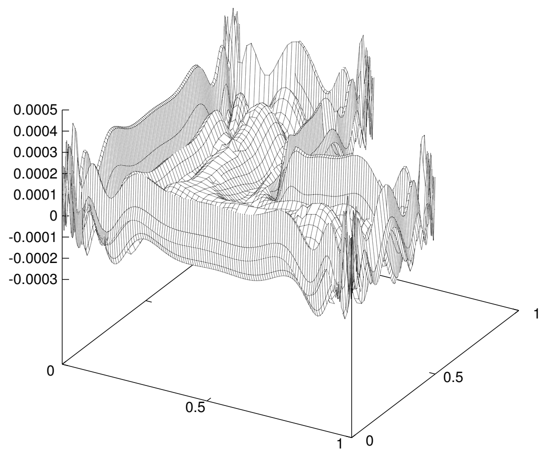
(a) iteration k=1



(b) iteration k=5

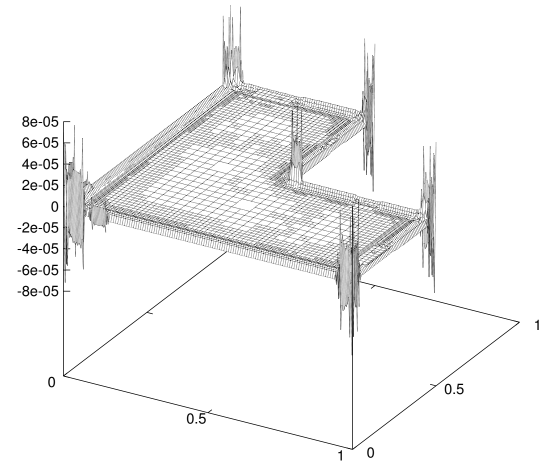


(c) iteration k=10

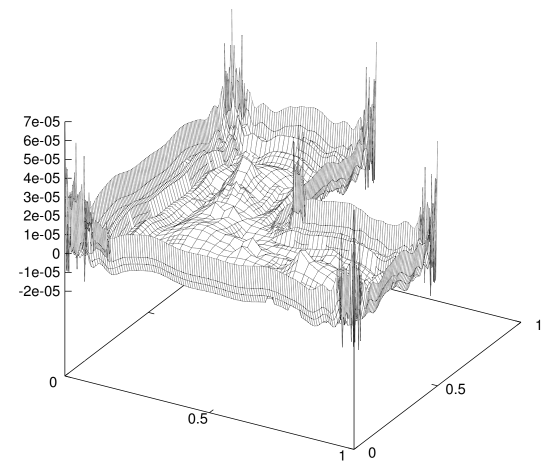


(d) iteration k=15

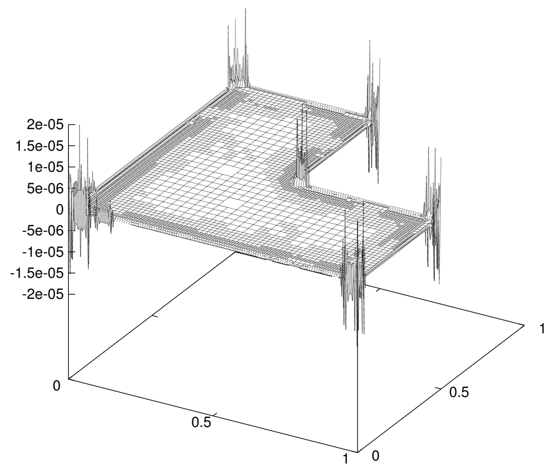
Figure 6.5: Sequence of solutions v_k ($k = 1, 5, 10, 15$) on the adapted meshes produced by Algorithm 1.



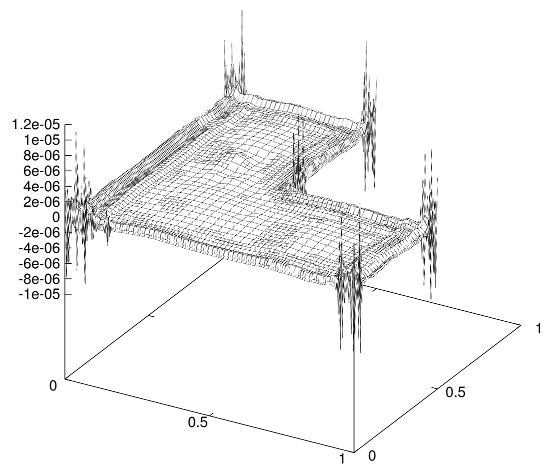
(a) iteration k=20



(b) iteration k=25



(c) iteration k=30



(d) iteration k=35

Figure 6.6: Sequence of solutions v_k ($k = 20, 25, 30, 35$) on the adapted meshes produced by Algorithm 1.

where the sequence (e_i) is an (arbitrary) orthonormal basis of $\mathscr{X} = L^2(\Omega)$. Figure 6.7 shows the sequence of nuclear norms for operators X_k. From this figure, we observe that

$$\|X_k\|_{\mathcal{S}_1(\mathscr{X})} \leq \|X_{k+1}\|_{\mathcal{S}_1(\mathscr{X})},$$

which is consistent with Theorem 4.21. In addition, it is important to note that our computations (see Table 6.1) provide an approximation for the nuclear norm of the Riccati operator (cf. Theorem 4.22):

$$\|X\|_{\mathcal{S}_1(\mathscr{X})} \approx 0.716352. \tag{6.15}$$

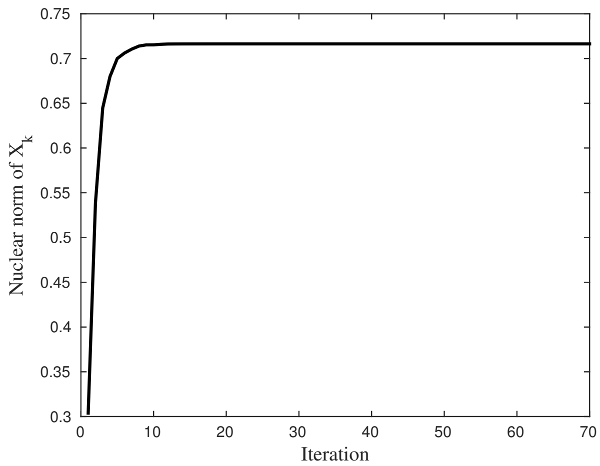


Figure 6.7: Monotonicity of the Riccati-ADI method: the sequence of nuclear norms for operators X_k.

A computationally efficient a posteriori relative stopping criterion for the Riccati-ADI algorithm in our example can be obtained by

$$C_{\text{adi}} := \frac{\|X_{k+1} - X_k\|_{\mathcal{S}_1(\mathscr{X})}}{\|X_{k+1}\|_{\mathcal{S}_1(\mathscr{X})}} = \frac{\operatorname{trace}[X_{k+1}] - \operatorname{trace}[X_k]}{\operatorname{trace}[X_{k+1}]}.$$

Figure 6.8 shows the values of C_{adi} at each iteration of the Riccati-ADI algorithm.

In the last part of this section we make a closer look at the iteration history of the Riccati-ADI algorithm for the fixed mesh and the successively adapted meshes. As shown in Table 6.1, including the possibility to successively adapt the mesh allows an almost identical approximation of the Riccati operator with severely fewer unknowns needed in the calculation. In this table we have used the abbreviation “DoF” for “degrees of freedom”. We note that the computation time of the Riccati-ADI method for this example using successively adapted meshes (including the computation of the error estimates and trace of the approximate solution at each iteration) is about 96 seconds for 70 iterations. For comparison, it takes about 210 seconds to perform 70 iterations of the Riccati-ADI method using a fixed mesh with 12545 degrees of freedom.

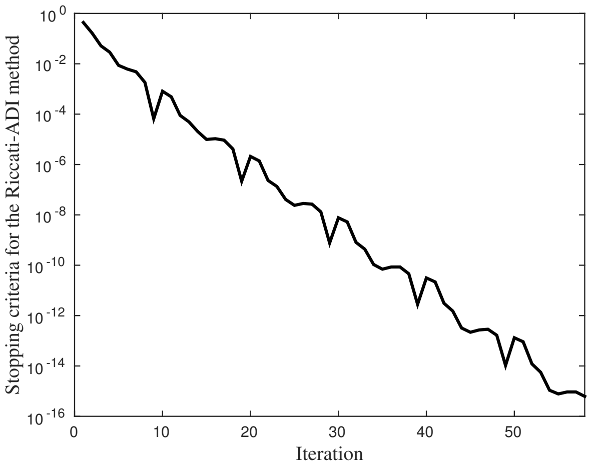


Figure 6.8: A posteriori relative stopping criterion (C_{adi}) for the Riccati-ADI algorithm.

Table 6.1: Iteration history for the fixed mesh (left) and successively adapted mesh (right).

Iteration	DoF	$\|X_k\|_{\mathcal{S}_1(\mathscr{X})}$	$\|v_k\|_{L^2(\Omega)}$	DoF	$\|X_k\|_{\mathcal{S}_1(\mathscr{X})}$	$\|v_k\|_{L^2(\Omega)}$
1	12545	0.303810	0.436148	12545	0.303812	0.436148
2	12545	0.538603	0.0515985	12455	0.538610	0.0515969
3	12545	0.644775	0.0238922	12270	0.644791	0.0238912
4	12545	0.679861	0.00899946	12148	0.679884	0.00900023
5	12545	0.699991	0.0030076	12221	0.699995	0.00300568
6	12545	0.706117	0.00122848	11922	0.706088	0.00122468
7	12545	0.710498	0.00060155	11356	0.710420	0.000597781
8	12545	0.713933	0.000249875	11048	0.713821	0.000248304
9	12545	0.715233	$7.66E-005$	10260	0.715206	$7.84E-005$
10	12545	0.715280	$7.73E-006$	8597	0.715373	$1.28E-005$
15	12545	0.716324	$8.05E-005$	7173	0.716518	$7.86E-005$
20	12545	0.716349	$4.44E-007$	6705	0.716607	$1.37E-006$
25	12545	0.716351	$3.57E-006$	6173	0.716648	$3.66E-006$
30	12545	0.716352	$2.64E-008$	6223	0.716671	$2.24E-007$
35	12545	0.716352	$1.81E-007$	5975	0.716692	$2.17E-007$
40	12545	0.716352	$1.60E-009$	5818	0.716718	$5.62E-008$
45	12545	0.716352	$9.75E-009$	5577	0.716743	$3.75E-008$
50	12545	0.716352	$9.77E-011$	5617	0.716971	$1.94E-008$
55	12545	0.716352	$5.48E-010$	4933	0.716994	$1.28E-008$
60	12545	0.716352	$6.02E-012$	4860	0.717653	$7.59E-009$
65	12545	0.716352	$3.16E-011$	3299	0.717673	$5.19E-009$
70	12545	0.716352	$3.72E-013$	3179	0.717970	$3.19E-009$

6.2.2 Newton-Kleinman algorithm

In this section we solve the regular optimal control problem corresponding to our example using the Newton-Kleinman algorithm (Algorithm 5). Although we proved monotonicity and convergence of the Newton-Kleinman iteration for bounded control operators (see Theorems 5.8 and 5.9), in this example we show the applicability of the algorithm even for the case where the control operator is unbounded. In fact, we will illustrate in this example that the sequence of operators produced by the Newton iteration is monotone and converges quadratically to the Riccati operator (6.7). As we showed in Chapter 5, at each step of the Newton-Kleinman iteration, one needs to solve a Lyapunov equation of the form (5.5). We solve this Lyapunov equation by applying the ADI method (see Algorithm 6). Hence, we refer to our implementation as the "Newton-ADI" method (see, e.g., [6]).

Throughout this section we use the first set of shift parameters from Section 6.2.1. These shift parameters are calculated a priori by employing the Penzl's heuristic procedure as explained in Part I of Section 6.2.1. In addition, we make use of the same finite element space, quadrature rule, preconditioner, and solver, as we did in Part II of Section 6.2.1. Hence, to avoid repetition, we mainly focus on those features which are exclusive to the Newton iteration. Nevertheless, to ensure consistency of our presentation, we recall some important implementation aspects of the ADI algorithm from the previous section. More details on the implementation of the Newton-ADI iteration in our example are included in Section A.3 of the appendix.

Now let us observe the first Newton iteration for our example: Let A and C be the operators given by (6.4) and (6.5), respectively. The first iteration is to solve the Lyapunov equation (5.5) with $k = 0$. By making the initialization $X_0 = 0$ we obtain

$$A^* X_1 + X_1 A + C^* C = 0, \tag{6.16}$$

for some $X_1 \in \mathcal{L}(\mathscr{X})$. Since A generates an exponentially stable semigroup $\mathbb{T}$, it follows from Theorem 2.21 that $X_1 = \Psi^* \Psi$ is the unique solution of (6.16). Furthermore, since operators A and C satisfy the assumptions of Theorem A.4, we conclude that the output map Ψ is Hilbert-Schmidt, i.e., $\Psi \in \mathcal{S}_2(\mathscr{X}, L^2(0, \infty; \mathscr{Y}))$. Therefore, there holds

$$X_1 = \Psi^* \Psi \in \mathcal{S}_1(\mathscr{X}).$$

Now it follows from Theorem 5.8 that the sequence of observability Gramians produced by Algorithm 5 satisfies

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

Since $X_1 \in \mathcal{S}_1(\mathscr{X})$, the min-max-Theorem of Courant-Fischer [47, Section 7.5] gives rise to

$$X_k \in \mathcal{S}_1(\mathscr{X}), \qquad \forall\, k \in \mathbb{N}, \tag{6.17}$$

and one can expect convergence of this sequence to the Riccati operator X in the nuclear norm, as shown in Theorem 5.9.

In order to approximate X_1, we apply the ADI method in Algorithm 7. This algorithm is an adaptation of [41, Algorithm 1] to our setting. Note that this is equivalent to choosing $B = 0$ in Algorithm 1. Since X_1 is nuclear, it follows from Theorem 4.22 (see

Algorithm 7 ADI iteration for Lyapunov equation (6.16).

Input: The generator A of a bounded strongly continuous semigroup $\mathbb{T}$, an infinite-time admissible observation operator $C \in \mathcal{L}(\mathscr{X}_1, \mathscr{Y})$, and shift parameters $\alpha_1, \alpha_2, \ldots, \alpha_{i_0} \in \mathbb{C}$, with $\mathrm{Re}(\alpha_i) > 0$.

Output: $S^0 := S_{i_0} \in \mathcal{L}(\mathscr{X}, \mathscr{Y}^{i_0})$, such that $\widetilde{X}_1^{(i_0)} := \left(S^0\right)^* S^0 \approx X_1$, where X_1 is the unique solution of the Lyapunov equation (6.16).

1: $V_1 = (\alpha_1 I - A^*)^{-1} C^*$

2: $S_1 = \sqrt{2\mathrm{Re}(\alpha_1)} \cdot V_1^*$

3: **for** $i = 2, 3, \ldots, i_0$ **do**

4: $\quad V_i = V_{i-1} - (\alpha_i + \overline{\alpha_{i-1}}) \cdot (\alpha_i I - A^*)^{-1} V_{i-1}$

5: $\quad S_i = [\, S_{i-1}^* \,, \sqrt{2\mathrm{Re}(\alpha_i)} \cdot V_i \,]^*$

6: **end for**

also Theorem 5.13) that the approximation of X_1 by Algorithm 7 converges in the nuclear norm.

As we have already stated in Section 6.2.2, since the output space is one-dimensional (i.e., $\mathscr{Y} = \mathbb{C}$), operators V_i calculated by Algorithm 7 satisfy $v_i := V_i \in L^2(\Omega)$. Since $A^* = A$ and $C^* = B$, Step 1 in Algorithm 7 can be written as

$$v_1 = (\alpha_1 I - A)^{-1} B,$$

and is therefore equivalent to solving the boundary value Helmholtz problem (6.12). Similarly, calculating $\hat{v}_i := (\alpha_i I - A^*)^{-1} v_{i-1}$ in Step 4 of Algorithm 7 is equivalent to solving the inhomogeneous boundary value Helmholtz problem (6.13). Eventually, we can determine v_i via (6.14).

Operator S^0 generated by Algorithm 7 approximates the output map Ψ corresponding to the pair (A, C) and we have

$$\widetilde{X}_1^{(i_0)} := \left(S^0\right)^* S^0 \approx X_1,$$

where X_1 is the unique solution of (6.16) and i_0 denotes the number of iterations performed by Algorithm 7. In our example, since $B^* = C \in \mathcal{L}(\mathscr{X}, \mathbb{C})$, we can define the *approximated feedback operator* at the first Newton iteration by

$$\widetilde{F}_1 := B^* \widetilde{X}_1^{(i_0)} \in \mathcal{L}(\mathscr{X}, \mathbb{C}).$$

For the rest of the iterations, we proceed as follows: For $k \in \mathbb{N}$, let $\widetilde{X}_k^{(i_{k-1})} := \left(S^{k-1}\right)^* S^{k-1}$ and $\widetilde{F}_k := B^* \widetilde{X}_k^{(i_{k-1})}$ be the operators generated by applying Algorithm 6 at the k-th Newton iteration. The $(k+1)$-st Newton iteration is to solve the following Lyapunov equation

$$A_k^* X_{k+1} + X_{k+1} A_k + \left(C_k^{\mathrm{aug}}\right)^* C_k^{\mathrm{aug}} = 0, \tag{6.18}$$

for $X_{k+1} \in \mathcal{L}(\mathscr{X})$, where

$$A_k = A - B\widetilde{F}_k, \qquad C_k^{\text{aug}} = \begin{bmatrix} C \\ \widetilde{F}_k \end{bmatrix}.$$

To solve (6.18), we apply Algorithm 6 to approximate the output map Ψ_k^{aug} generated by the pair (A_k, C_k^{aug}). From (6.17) we know that $X_{k+1} \in \mathcal{S}_1(\mathscr{X})$. Hence, it follows from Theorem 5.13 that the approximations of X_{k+1} by Algorithm 6 converge in the nuclear norm.

Proposition 6.6. *Let (A, B, C) be the triple of operators corresponding to equations (6.1) – (6.3) defined in Section 6.1. Let V_i^k be the operators generated by Algorithm 6 applied to the Lyapunov equation (6.18). Then, operators V_i^k have the structure*

$$V_i^k = \begin{bmatrix} v_i^k & v_i^{\widetilde{F}_k} \end{bmatrix}, \qquad i = 1, 2, \ldots, i_k,$$

where $v_i^k, v_i^{\widetilde{F}_k} \in L^2(\Omega)$ are determined recursively by

$$\begin{aligned} \begin{bmatrix} v_1^k & v_1^{\widetilde{F}_k} \end{bmatrix} &= \begin{bmatrix} v_1 - f_1^k \dfrac{b_1}{1 + c_1^k} & \quad f_1^k - f_1^k \dfrac{c_1^k}{1 + c_1^k} \end{bmatrix}, \\ \begin{bmatrix} v_i^k & v_i^{\widetilde{F}_k} \end{bmatrix} &= \begin{bmatrix} v_{i-1}^k & v_{i-1}^{\widetilde{F}_k} \end{bmatrix} - (\alpha_i + \overline{\alpha_{i-1}}) \cdot \begin{bmatrix} \hat{v}_i^k & \hat{v}_i^{\widetilde{F}_k} \end{bmatrix}, \qquad i = 2, 3, \ldots, i_k. \end{aligned} \tag{6.19}$$

The recursive formula (6.19) is evaluated as follows:

(i) The function $f_1^k \in L^2(\Omega)$ is the solution of the following boundary value Helmholtz problem:

$$\begin{aligned} \alpha_1 \cdot f_1^k(\xi) - \Delta f_1^k(\xi) &= \widetilde{F}_k^*, && \xi \in \Omega, \\ \nu(\xi)^T \nabla f_1^k(\xi) + \alpha f_1^k(\xi) &= 0, && \xi \in \partial\Omega, \end{aligned}$$

where

$$b_1 := B^* v_1 \in \mathbb{C}, \qquad c_1^k := B^* f_1^k \in \mathbb{C}.$$

(ii) The functions $\hat{v}_i^k$ and $\hat{v}_i^{\widetilde{F}_k}$ in (6.19) are determined recursively by

$$\begin{aligned} \hat{v}_i^k &= \tilde{v}_{i-1}^k - f_i^k \frac{\tilde{b}_{i-1}^k}{1 + c_i^k}, \\ \hat{v}_i^{\widetilde{F}_k} &= \tilde{v}_{i-1}^{\widetilde{F}_k} - f_i^k \frac{\tilde{b}_{i-1}^{\widetilde{F}_k}}{1 + c_i^k}, \end{aligned} \tag{6.20}$$

where

$$\tilde{b}_{i-1}^k := B^* \tilde{v}_{i-1}^k \in \mathbb{C}, \qquad \tilde{b}_{i-1}^{\widetilde{F}_k} := B^* \tilde{v}_{i-1}^{\widetilde{F}_k} \in \mathbb{C}, \qquad c_i^k := B^* f_i^k \in \mathbb{C}. \tag{6.21}$$

The functions $\tilde{v}_{i-1}^k$, $\tilde{v}_{i-1}^{\widetilde{F}_k}$, and f_i^k in (6.20) are the solutions of the following boundary value Helmholtz problem: Find $w \in L^2(\Omega)$, such that

$$\begin{aligned} \alpha_i \cdot w(\xi) - \Delta w(\xi) &= \breve{v}, && \xi \in \Omega, \\ \nu(\xi)^T \nabla w(\xi) + \alpha w(\xi) &= 0, && \xi \in \partial\Omega, \end{aligned} \tag{6.22}$$

with the corresponding right-hand side function $\breve{v} \in L^2(\Omega)$, $\breve{v} \in \{v_{i-1}^k, v_{i-1}^{\widetilde{F}_k}, \widetilde{F}_k^*\}$. *More specifically,*

- $w = \tilde{v}_{i-1}^k$ *corresponds to the right-hand side function* $\breve{v} = v_{i-1}^k$,
- $w = \tilde{v}_{i-1}^{\widetilde{F}_k}$ *corresponds to the right-hand side function* $\breve{v} = v_{i-1}^{\widetilde{F}_k}$,
- $w = f_i^k$ *corresponds to the right-hand side function* $\breve{v} = \widetilde{F}_k^*$.

Proof. The first Step in Algorithm 6 is to solve

$$V_1^k := (\alpha_1 I - A_k^*)^{-1} \left(C_k^{\text{aug}}\right)^* = \left[(\alpha_1 I - A_k^*)^{-1} C^* \quad (\alpha_1 I - A_k^*)^{-1} \widetilde{F}_k^*\right]. \tag{6.23}$$

We define

$$v_1^k := (\alpha_1 I - A_k^*)^{-1} C^*, \qquad v_1^{\widetilde{F}_k} := (\alpha_1 I - A_k^*)^{-1} \widetilde{F}_k^*.$$

Since the input and output spaces are one-dimensional (i.e., $\mathscr{U} = \mathscr{Y} = \mathbb{C}$), there holds $v_1^k, v_1^{\widetilde{F}_k} \in L^2(\Omega)$. For the resolvent operator $(\alpha_1 I - A_k^*)^{-1}$, we use the generalized Sherman-Morrison-Woodbury formula (Theorem A.1) to obtain

$$\begin{aligned}(\alpha_1 I - A_k^*)^{-1} &= (\alpha_1 I - A^* + \widetilde{F}_k^* B^*)^{-1} \\ &= \left(I - (\alpha_1 I - A^*)^{-1} \widetilde{F}_k^* \left(I + B^*(\alpha_1 I - A^*)^{-1} \widetilde{F}_k^*\right)^{-1} B^*\right)(\alpha_1 I - A^*)^{-1}.\end{aligned}$$

As a result, the functions v_1^k and $v_1^{\widetilde{F}_k}$ can be determined by

$$\begin{aligned} v_1^k &= \left(1 - (\alpha_1 I - A^*)^{-1} \widetilde{F}_k^* \left(1 + B^*(\alpha_1 I - A^*)^{-1} \widetilde{F}_k^*\right)^{-1} B^*\right)(\alpha_1 I - A^*)^{-1} C^*, \\ v_1^{\widetilde{F}_k} &= \left(1 - (\alpha_1 I - A^*)^{-1} \widetilde{F}_k^* \left(1 + B^*(\alpha_1 I - A^*)^{-1} \widetilde{F}_k^*\right)^{-1} B^*\right)(\alpha_1 I - A^*)^{-1} \widetilde{F}_k^*. \end{aligned} \tag{6.24}$$

Hence, in order to find v_1^k and $v_1^{\widetilde{F}_k}$ we require the following two ingredients:

- $v_1 = (\alpha_1 I - A^*)^{-1} C^* \in L^2(\Omega)$, which has already been computed at the first Newton iteration and does not need to be computed again.
- $f_1^k = (\alpha_1 I - A^*)^{-1} \widetilde{F}_k^* \in L^2(\Omega)$, which is the solution of the following boundary value Helmholtz problem:

$$\begin{aligned} \alpha_1 \cdot f_1^k - \Delta f_1^k(\xi) &= \widetilde{F}_k^*, \qquad && \xi \in \Omega, \\ \nu(\xi)^T \nabla f_1^k(\xi) + \alpha f_1^k(\xi) &= 0, && \xi \in \partial\Omega. \end{aligned} \tag{6.25}$$

Consequently, we have

$$\begin{aligned} v_1^k &= \left(1 - (\alpha_1 I - A^*)^{-1} \widetilde{F}_k^* \left(1 + B^*(\alpha_1 I - A^*)^{-1} \widetilde{F}_k^*\right)^{-1} B^*\right)(\alpha_1 I - A^*)^{-1} C^* \\ &= v_1 - f_1^k \left(1 + B^* f_1^k\right)^{-1} B^* v_1 \\ &= v_1 - f_1^k \frac{b_1}{1 + c_1^k}, \end{aligned}$$

with
$$b_1 := B^* v_1 \in \mathbb{C}, \qquad c_1^k := B^* f_1^k \in \mathbb{C}.$$
Similarly, we obtain
$$\begin{aligned}
v_1^{\widetilde{F}_k} &= \left(1 - (\alpha_1 I - A^*)^{-1} \widetilde{F}_k^* \left(1 + B^*(\alpha_1 I - A^*)^{-1} \widetilde{F}_k^*\right)^{-1} B^*\right)(\alpha_1 I - A^*)^{-1} \widetilde{F}_k^* \\
&= f_1^k - f_1^k \left(1 + B^* f_1^k\right)^{-1} B^* f_1^k \\
&= f_1^k - f_1^k \frac{c_1^k}{1 + c_1^k}.
\end{aligned}$$
From the above observations, we conclude that V_1^k in (6.23) can be determined by solving (6.25) for f_1^k and subsequently applying the formula
$$V_1^k = \begin{bmatrix} v_1^k & v_1^{\widetilde{F}_k} \end{bmatrix} = \begin{bmatrix} v_1 - f_1^k \dfrac{b_1}{1 + c_1^k} & f_1^k - f_1^k \dfrac{c_1^k}{1 + c_1^k} \end{bmatrix}. \tag{6.26}$$
For $i = 2, 3, 4, \ldots, i_k$, operators V_i^k are determined recursively by Step 4 of Algorithm 6, namely,
$$V_i^k = V_{i-1}^k - (\alpha_i + \overline{\alpha_{i-1}}) \cdot (\alpha_i I - A_k^*)^{-1} V_{i-1}^k. \tag{6.27}$$
In view of (6.26) it is reasonable to make the ansatz
$$V_i^k = \begin{bmatrix} v_i^k & v_i^{\widetilde{F}_k} \end{bmatrix}. \tag{6.28}$$
By substituting (6.28) in (6.27) we obtain
$$\begin{bmatrix} v_i^k & v_i^{\widetilde{F}_k} \end{bmatrix} = \begin{bmatrix} v_{i-1}^k & v_{i-1}^{\widetilde{F}_k} \end{bmatrix} - (\alpha_i + \overline{\alpha_{i-1}}) \cdot (\alpha_i I - A_k^*)^{-1} \begin{bmatrix} v_{i-1}^k & v_{i-1}^{\widetilde{F}_k} \end{bmatrix}.$$
We define
$$\hat{v}_i^k := (\alpha_i I - A_k^*)^{-1} v_{i-1}^k, \qquad \hat{v}_i^{\widetilde{F}_k} := (\alpha_i I - A_k^*)^{-1} v_{i-1}^{\widetilde{F}_k}.$$
Similarly to Step 1 of Algorithm 6, we apply the Sherman-Morrison-Woodbury formula (Theorem A.1) to the resolvent $(\alpha_i I - A_k^*)^{-1}$ and obtain
$$\begin{aligned}
\hat{v}_i^k &= \left(1 - (\alpha_i I - A^*)^{-1} \widetilde{F}_k^* \left(1 + B^*(\alpha_i I - A^*)^{-1} \widetilde{F}_k^*\right)^{-1} B^*\right)(\alpha_i I - A^*)^{-1} v_{i-1}^k, \\
\hat{v}_i^{\widetilde{F}_k} &= \left(1 - (\alpha_i I - A^*)^{-1} \widetilde{F}_k^* \left(1 + B^*(\alpha_i I - A^*)^{-1} \widetilde{F}_k^*\right)^{-1} B^*\right)(\alpha_i I - A^*)^{-1} v_{i-1}^{\widetilde{F}_k}.
\end{aligned} \tag{6.29}$$
As a result, we need the following three ingredients to determine $\hat{v}_i^k$ and $\hat{v}_i^{\widetilde{F}_k}$:
$$\begin{aligned}
\tilde{v}_{i-1}^k &:= (\alpha_i I - A^*)^{-1} v_{i-1}^k \in L^2(\Omega), \\
\tilde{v}_{i-1}^{\widetilde{F}_k} &:= (\alpha_i I - A^*)^{-1} v_{i-1}^{\widetilde{F}_k} \in L^2(\Omega), \\
f_i^k &:= (\alpha_i I - A^*)^{-1} \widetilde{F}_k^* \in L^2(\Omega).
\end{aligned}$$

The functions $\tilde{v}_{i-1}^{k}$, $\tilde{v}_{i-1}^{\widetilde{F}_k}$, and f_i^k are the solutions of the following boundary value Helmholtz problem: Find $w \in L^2(\Omega)$, such that

$$\begin{aligned} \alpha_i \cdot w(\xi) - \Delta w(\xi) &= \breve{v}, \qquad && \xi \in \Omega, \\ \nu(\xi)^T \nabla w(\xi) + \alpha w(\xi) &= 0, && \xi \in \partial\Omega, \end{aligned} \tag{6.30}$$

with the corresponding right-hand side function $\breve{v} \in L^2(\Omega)$, $\breve{v} \in \{v_{i-1}^k, v_{i-1}^{\widetilde{F}_k}, \widetilde{F}_k^*\}$. Altogether, we obtain

$$\begin{aligned} \hat{v}_i^k &= \tilde{v}_{i-1}^k - f_i^k \left(1 + B^* f_i^k\right)^{-1} B^* \tilde{v}_{i-1}^k \\ &= \tilde{v}_{i-1}^k - f_i^k \frac{\tilde{b}_{i-1}^k}{1 + c_i^k}, \end{aligned}$$

$$\begin{aligned} \hat{v}_i^{\widetilde{F}_k} &= \tilde{v}_{i-1}^{\widetilde{F}_k} - f_i^k \left(1 + B^* f_i^k\right)^{-1} B^* \tilde{v}_{i-1}^{\widetilde{F}_k} \\ &= \tilde{v}_{i-1}^{\widetilde{F}_k} - f_i^k \frac{\tilde{b}_{i-1}^{\widetilde{F}_k}}{1 + c_i^k}, \end{aligned}$$

where we have defined

$$\tilde{b}_{i-1}^k := B^* \tilde{v}_{i-1}^k \in \mathbb{C}, \qquad \tilde{b}_{i-1}^{\widetilde{F}_k} := B^* \tilde{v}_{i-1}^{\widetilde{F}_k} \in \mathbb{C}, \qquad c_i^k := B^* f_i^k \in \mathbb{C}.$$

As a result, in order to determine $V_i^k = [v_i^k \quad v_i^{\widetilde{F}_k}]$, one has to solve three boundary value Helmholtz problems (i.e., equations (6.30) with the right-hand sides $\breve{v} \in \{v_{i-1}^k, v_{i-1}^{\widetilde{F}_k}, \widetilde{F}_k^*\}$) to find the functions $\tilde{v}_{i-1}^k$, $\tilde{v}_{i-1}^{\widetilde{F}_k}$, f_i^k and then apply the recursive formulas (6.20) and (6.19). $\square$

Remark 6.7. By applying the Sherman-Morrison-Woodbury formula (Theorem A.1) in Steps 1 and 4 of Algorithm 6 (see, in particular, equations (6.24) and (6.29)), we can preserve the sparsity pattern of the finite element discretization of operator A in all Newton iterations. This reduces the computational costs of our algorithm significantly.

We solve the obtained sequence of boundary value Helmholtz equations by applying the adaptive finite element method. Similarly to Part II of Section 6.2.1, we start the iterations by refining the mesh 6 times globally (cf. Figure 6.1) to obtain 12545 degrees of freedom. To adaptively refine/coarsen the mesh, we use an error estimator based on the discrete approximation of $\nabla \tilde{v}_k$ (cf. Section A.1.1), where $\tilde{v}_k$ is the solution of (6.22) with the right-hand side function $\breve{v} = v_{i-1}^k$. Our mesh adaptation strategy is based on the bulk criterion (cf. Section A.1.2) with

$$r = 0.02, \qquad c = 0.01.$$

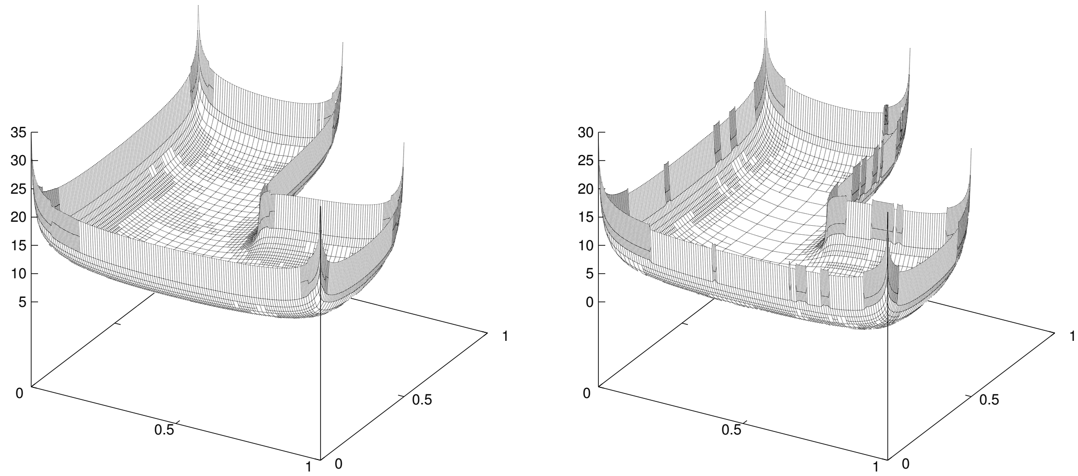
Algorithm 6 generates an approximation of the augmented output map Ψ_k^{aug}. If S^k denotes an approximation of Ψ_k^{aug}, then we have

$$\widetilde{X}_{k+1}^{(i_k)} := (S^k)^* S^k \approx X_{k+1},$$

where $X_{k+1} \in \mathcal{L}(\mathscr{X})$ is the solution of (6.18) and i_k denotes the number of iterations performed by Algorithm 6. Since $B^* = C \in \mathcal{L}(\mathscr{X}, \mathbb{C})$, we define the *approximated feedback operator* at the $(k+1)$-st Newton iteration by

$$\widetilde{F}_{k+1} := B^* \widetilde{X}_{k+1}^{(i_k)} \in \mathcal{L}(\mathscr{X}, \mathbb{C}).$$

More details on the approximation of the feedback operator $\widetilde{F}_k$ are presented in Section A.3. Figure 6.9 shows the approximated feedbacks $\widetilde{F}_{h,1}$ and $\widetilde{F}_{h,7}$ generated by the Newton-ADI method in our example.



(a) $\widetilde{F}_{h,1}$ (b) $\widetilde{F}_{h,7}$

Figure 6.9: Approximated feedbacks $\widetilde{F}_{h,1}$ and $\widetilde{F}_{h,7}$ generated by the Newton-ADI method.

We use the a posteriori stopping criterion (5.71) for Algorithm 6: For $i \in \mathbb{N}$, $i \geq 2$, let $\widetilde{X}_{k+1}^{(i)}$ be an approximation of X_{k+1}. Since $X_{k+1} \in \mathcal{S}_1(\mathscr{X})$, it follows from Theorem 5.13 that $\widetilde{X}_{k+1}^{(i)} \in \mathcal{S}_1(\mathscr{X})$. Hence, we can use the following a posteriori relative stopping criterion:

$$K_{\text{adi}} := \frac{\left\| \widetilde{X}_{k+1}^{(i)} - \widetilde{X}_{k+1}^{(i-1)} \right\|_{\mathcal{S}_1(\mathscr{X})}}{\left\| \widetilde{X}_{k+1}^{(i)} \right\|_{\mathcal{S}_1(\mathscr{X})}}. \tag{6.31}$$

To compute K_{adi}, we observe that

$$\begin{aligned} \left\| \widetilde{X}_{k+1}^{(i)} - \widetilde{X}_{k+1}^{(i-1)} \right\|_{\mathcal{S}_1(\mathscr{X})} &= 2\mathrm{Re}(\alpha_i) \left\| \left(V_i^k\right)^* V_i^k \right\|_{\mathcal{S}_1(\mathscr{Y}\times\mathscr{U})} \\ &= 2\mathrm{Re}(\alpha_i)\, \mathrm{trace}\left[\left(V_i^k\right)^* V_i^k \right]. \end{aligned} \tag{6.32}$$

Because of

$$\left(V_i^k\right)^* V_i^k = \begin{bmatrix} v_i^k & v_i^{F_k} \end{bmatrix}^* \begin{bmatrix} v_i^k & v_i^{F_k} \end{bmatrix} = \begin{bmatrix} \left(v_i^k\right)^* v_i^k & \left(v_i^k\right)^* v_i^{F_k} \\ \left(v_i^{F_k}\right)^* v_i^k & \left(v_i^{F_k}\right)^* v_i^{F_k} \end{bmatrix},$$

with $v_i^k, v_i^{F_k} \in L^2(\Omega)$, there holds

$$\left\| \widetilde{X}_{k+1}^{(i)} - \widetilde{X}_{k+1}^{(i-1)} \right\|_{\mathcal{S}_1(\mathscr{X})} = 2\mathrm{Re}(\alpha_i) \left[\left\| v_i^k \right\|_{L^2(\Omega)}^2 + \left\| v_i^{F_k} \right\|_{L^2(\Omega)}^2 \right].$$

Similarly, we have (see also (5.72))

$$\left\| \widetilde{X}_{k+1}^{(i)} \right\|_{\mathcal{S}_1(\mathscr{X})} = \mathrm{trace}\left[\widetilde{X}_{k+1}^{(i)} \right] = \sum_{j=1}^{i} 2\mathrm{Re}(\alpha_j) \left[\left\| v_j^k \right\|_{L^2(\Omega)}^2 + \left\| v_j^{F_k} \right\|_{L^2(\Omega)}^2 \right]. \tag{6.33}$$

At each Newton iteration we terminate Algorithm 6 whenever we reach $K_{\text{adi}} < 10^{-11}$. Figure 6.10 shows the sequence of adapted degrees of freedom and a posteriori relative stopping criteria (K_{adi} as in (6.31)) for the first three iterations of the Newton-ADI method.

To illustrate monotonicity of the Newton-Kleinman iteration (Theorem 5.8), at each Newton iteration we observe the nuclear norm of $\widetilde{X}_k := \widetilde{X}_k^{(i_{k-1})}$ (the approximation of X_k generated by Algorithm 6, where we drop the index i_{k-1} for simplicity). These values are depicted in Figure 6.11. From this figure we see that

$$\left\| \widetilde{X}_{k+1} \right\|_{\mathcal{S}_1(\mathscr{X})} \leq \left\| \widetilde{X}_k \right\|_{\mathcal{S}_1(\mathscr{X})}, \qquad \forall\, k \in \mathbb{N},$$

which is consistent with Theorem 5.8. Furthermore, we can determine an approximation for the nuclear norm of the Riccati operator (cf. Theorem 5.9):

$$\|X\|_{\mathcal{S}_1(\mathscr{X})} \approx \left\| \widetilde{X}_7 \right\|_{\mathcal{S}_1(\mathscr{X})} = 0.716532,$$

which is the same as the value obtained by the Riccati-ADI algorithm in the previous section (see (6.15)).

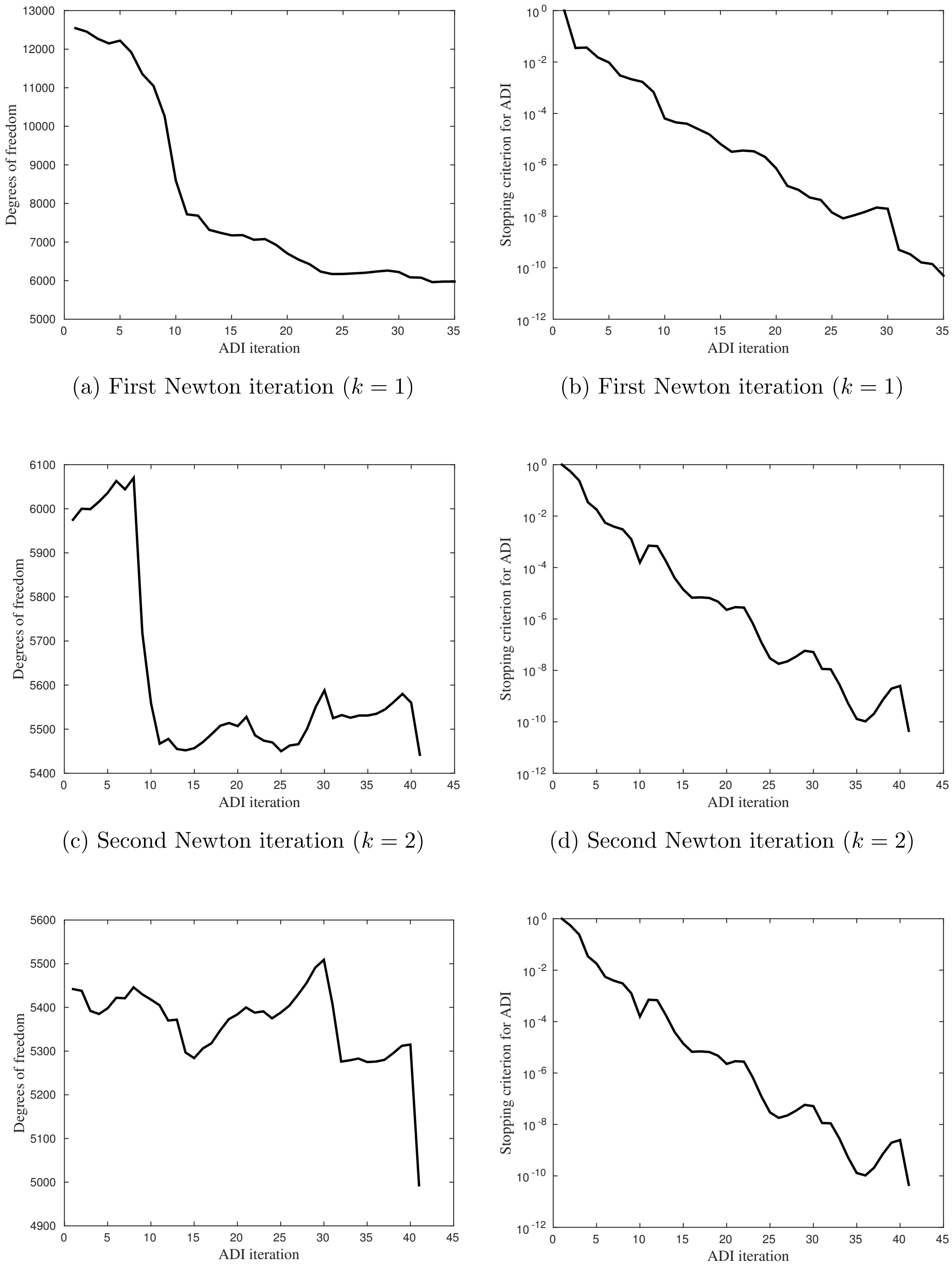
We close this section by providing an efficient a posteriori relative stopping criterion for the Newton iterations. We use the stopping criterion

$$\frac{\|X_{k+1} - X_k\|_{\mathcal{S}_1(\mathscr{X})}}{\|X_{k+1}\|_{\mathcal{S}_1(\mathscr{X})}} = \frac{\mathrm{trace}\,[X_{k+1}] - \mathrm{trace}\,[X_k]}{\mathrm{trace}\,[X_{k+1}]},$$

which can be approximated by operators $\widetilde{X}_k$:

$$K_{\text{nk}} := \frac{\mathrm{trace}\left[\widetilde{X}_{k+1}\right] - \mathrm{trace}\left[\widetilde{X}_k\right]}{\mathrm{trace}\left[\widetilde{X}_{k+1}\right]}, \tag{6.34}$$

where the trace of $\widetilde{X}_k$ can be computed efficiently as shown in (6.33). Table 6.2 shows the sequence of a posteriori relative stopping criteria K_{nk} for the Newton iterations.



(a) First Newton iteration ($k = 1$)

(b) First Newton iteration ($k = 1$)

(c) Second Newton iteration ($k = 2$)

(d) Second Newton iteration ($k = 2$)

(e) Third Newton iteration ($k = 3$)

(f) Third Newton iteration ($k = 3$)

Figure 6.10: Sequence of adapted degrees of freedom and a posteriori relative stopping criterion (K_{adi}) for the first three iterations of the Newton-ADI method.

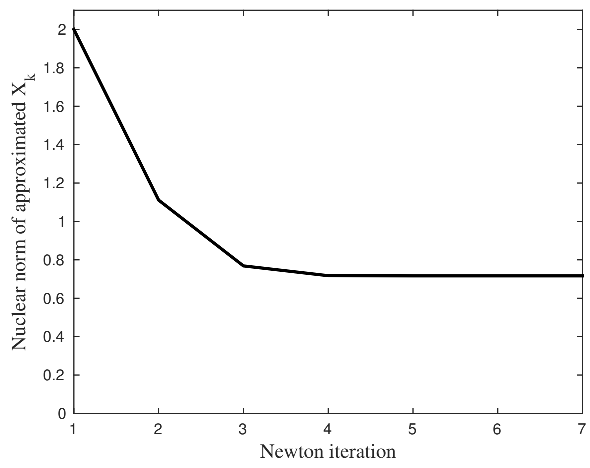


Figure 6.11: Nuclear norm of $\widetilde{X}_k$.

Looking at the values of K_{nk} in Table 6.2, one observes the quadratic rate of convergence for the Newton-Kleinman iterations, which we showed in Theorem 5.11 in case of bounded control and unbounded observation operators. Our numerical example shows that the Newton-Kleinman iteration may provide a quadratic rate of convergence, even in case of unbounded control operators. As already stated in Section 5.4, we were able to give a complete proof of convergence for the Newton-Kleinman iteration only in case of bounded control operators. Proving the convergence of Algorithm 5 in case of unbounded control operator is an open problem for our future research.

Table 6.2: Degrees of freedom (DoF) and the a posteriori relative stopping criterion (K_{nk} in (6.34)) at the end of each Newton iteration.

Iteration	DoF	K_{nk}
0	12545	–
1	5975	1
2	5442	0.739965
3	4993	0.307568
4	4993	0.0418406
5	4993	0.000733626
6	4993	2.39101×10^{-7}
7	4993	3.3239×10^{-14}

6.3 Singular linear-quadratic optimal control problem: positive real case

In this section we solve the positive real optimal control problem (cf. Section 3.6) for the heat equation (6.1) with the Robin boundary control (6.2) and the boundary integral observation (6.3) (see Section 6.1). In this case, the positive real optimal control problem is to maximize for $x_0 \in \mathscr{X}$

$$J(x_0, u) = -2\mathrm{Re} \int_0^\infty y(\tau)u(\tau)\, d\tau, \tag{6.35}$$

over all $u \in L^2(0, \infty; \mathscr{U})$ subject to (6.1)–(6.3). We apply Algorithm 4 to solve this problem and observe its expected performance in terms of monotonicity and convergence behavior.

Part I:

In the first part of this section, we focus on an appropriate choice of the shift parameters for Algorithm 4. To this end, we focus on the finite-dimensional positive real optimal control problem (see [37]) and present an effective strategy for choosing the shift parameters.

We discretize the PDE (6.1) by applying a finite element discretization with uniform square elements of maximal diameter h (cf. Figure 6.1). We define the subspace $V_h \subset H^1(\Omega)$ using piecewise-linear basis functions. As a result, we obtain the finite-dimensional dynamical system (6.10) with the state space dimension n. We note that the finite element discretization was done using the C++ library *deal.II* and the rest of the calculations in this part were done using MATLAB 8.5 (R2015a).

Remark 6.8. The finite-dimensional positive real optimal control problem is to maximize the cost function (6.35) subject to the linear system (6.10). The cost function (6.35) can be characterized by (cf. Section 3.6)

$$J(x_0, u) = \langle Ex_0, EXx_0\rangle - \|\Psi_\Xi x_0 + \mathbb{F}_\Xi u\|^2, \tag{6.36}$$

with

$$X = \Psi_\Xi^* \Psi_\Xi \in \mathbb{C}^{n\times n}.$$

The matrix $X \in \mathbb{C}^{n\times n}$ is the minimal solution of the positive real Lur'e equations

$$\begin{aligned} A^*XE + E^*XA &= -C_\Xi^* C_\Xi, \\ B^*XE - C &= -\overline{D_\Xi} C_\Xi, \\ 0 &= -\left|D_\Xi\right|^2. \end{aligned} \tag{6.37}$$

with $C_\Xi \in \mathbb{C}^{1\times n}$ and $D_\Xi \in \mathbb{C}$. It follows from (6.37) that $D_\Xi = 0$.

Remark 6.9. We consider the finite-dimensional version of Remark 3.13.a for the linear system (6.10) (recall that in our example we have $\mathscr{U} = \mathscr{Y} = \mathbb{C}$). It follows from (6.36) that the optimal cost

$$\sup_{u\in L^2(0,\infty;\mathbb{C})} J(x_0, u) = \langle Ex_0, EXx_0\rangle$$

holds true, if and only if,

$$\mathbb{F}_\Xi u^{\mathrm{opt}} + \Psi_\Xi x_0 = 0, \tag{6.38}$$

for some $u^{\mathrm{opt}} \in L^2(0,\infty;\mathbb{C})$. If $\mathbb{F}_\Xi \in \mathcal{L}(L^2(0,\infty;\mathbb{C}))$ is outer (i.e., $\mathbb{F}_\Xi$ has a dense range), then there exists a solution to (6.38). It follows from (6.10) and (6.38) that there exists some $x : [0,\infty) \to \mathbb{C}^n$ such that the differential-algebraic equation

$$\begin{bmatrix} E & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{x}(t) \\ \dot{u}(t) \end{bmatrix} = \begin{bmatrix} A & B \\ C_\Xi & D_\Xi \end{bmatrix} \begin{bmatrix} x(t) \\ u(t) \end{bmatrix}, \qquad x(0) = x_0 \in \mathscr{X},$$

is fulfilled for $C_\Xi \in \mathbb{C}^{1\times n}$ and $D_\Xi \in \mathbb{C}$. By a transformation of the matrix pencil

$$\begin{bmatrix} sE - A & -B \\ -C_\Xi & -D_\Xi \end{bmatrix}, \tag{6.39}$$

into the Kronecker form (see [21, Chap. XII, §7]), we can conclude that x and u can be expressed by sums of exponential functions of type

$$x(t) = \sum_{k=1}^{\ell} p_k(t) \mathrm{e}^{-\lambda_k t}, \qquad u(t) = \sum_{k=1}^{\ell} \widetilde{p}_k(t) \mathrm{e}^{-\lambda_k t}, \qquad \text{with} \qquad \ell \le n+1, \tag{6.40}$$

where $p_1,\dots,p_\ell$ and $\widetilde{p}_1,\dots,\widetilde{p}_\ell$ are vector-valued complex polynomials, and the pairwise distinct numbers $\lambda_1,\dots,\lambda_\ell$ are the generalized eigenvalues of the pencil (6.39). By using the three equations in (6.37), we obtain the deflating subspace relation

$$\begin{bmatrix} 0 & -sE + A & B \\ sE^* + A^* & 0 & -C^* \\ B^* & -C & 0 \end{bmatrix} \begin{bmatrix} XE & 0 \\ I & 0 \\ 0 & I \end{bmatrix} = \begin{bmatrix} -I & 0 \\ E^*X & C_\Xi^* \\ 0 & \overline{D_\Xi} \end{bmatrix} \begin{bmatrix} sE - A & -B \\ -C_\Xi & -D_\Xi \end{bmatrix}.$$

Hence, it follows from [48, Theorem 5.1] (see also [37, Remark 2]) that the generalized eigenvalues of the pencil (6.39) (denoted by $\lambda_1,\dots,\lambda_\ell$) are the negatives of the stable generalized eigenvalues of the *even matrix pencil*

$$s\mathcal{E} - \mathcal{A} = \begin{bmatrix} 0 & -sE + A & B \\ sE^* + A^* & 0 & -C^* \\ B^* & -C & 0 \end{bmatrix}. \tag{6.41}$$

We will make use of this fact to improve the numerical performance of our implementation by making a suitable choice of the shift parameters.

We refine the mesh 5 times globally (cf. Figure 6.1). As a result, we obtain 3072 (active) cells with 3201 degrees of freedom (for piecewise-linear basis functions). We find an approximate solution $X_k \in \mathbb{C}^{n\times n}$ of the positive real optimal control problem by applying the matrix version of Algorithm 4 (see [37]), where we use the modifications proposed in Remark 6.5 (see also [34, Remark 3.3]), which allow computations without explicit inversion of E. In addition, in Steps 6 and 7 of Algorithm 4, 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 Algorithm 1. In fact, we just need to access the last p rows of the matrix F_i in order to obtain the value of $Q_i(\overline{L_i} \otimes I_m)$ (cf. Remark 4.30).

The choice of shift parameters has a major effect on the convergence speed of the ADI algorithm. In our example, we propose the following strategy for choosing the shift parameters, which is motivated by Remark 6.9:

(i) We generate a set of 30 shift parameters by applying Penzl's heuristic procedure [44] on negatives of the stable eigenvalues of the even matrix pencil (6.41). In order to approximate the spectrum of this even matrix pencil, we calculate 450 Ritz values using the shift-and-invert Arnoldi process [53, Section 8.1.3] 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 the state-space dimension $n = 3201$ is about 58 seconds.

(ii) We add a large real shift parameter of order 10^{12} to the set of shift parameters from (i) and consider it to be the first parameter of the set. We use this large shift parameter just in the first iteration of Algorithm 4 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.

(iii) We sort the obtained 31 shift parameters from (i) and (ii) in an increasing order with respect to the values of their real part in order to obtain a smooth convergence for our algorithm. We perform 31 iterations of Algorithm 4 using these shift parameters. At each iteration, we observe the relative residual norm of the positive real Lur'e equations (6.37) using the approach proposed in [45, Section 6]. Subsequently, we extract a subset of 10 shift parameters which provide the highest reduction in the value of the residual norm. These parameters are then re-used every 10 iterations.

Figure 6.12 shows the relative residual norm with respect to the iteration for the state space dimension $n = 3201$ with the above choice of shift parameters. With a tolerance of 10^{-13} on the relative residual norm, our choice of shift parameters leads to convergence in 31 iterations. We note that the execution time of the ADI algorithm for this example (including the computation of the relative residual norm at each iteration) is about 188 seconds for 100 iterations.

Remark 6.10. In the example considered in this section the spectrum of the even matrix pencil (6.41) consists of only real values. If the spectrum was complex, containing eigenvalues with widely varying real and imaginary parts which dominate the behavior, then the selection of shift parameters would become a more delicate task. For example, see [37, Section 5], where the authors considered a convection-diffusion equation with the same boundary conditions as in (6.2) and (6.3). The authors illustrated that for convection dominated problems, an inaccurate choice of the shift parameters may result in a slow convergence of the ADI algorithm.

Part II:

In the second part of this section, we use the shift parameters obtained from Part I to show monotonicity and convergence of the ADI algorithm for the positive real optimal

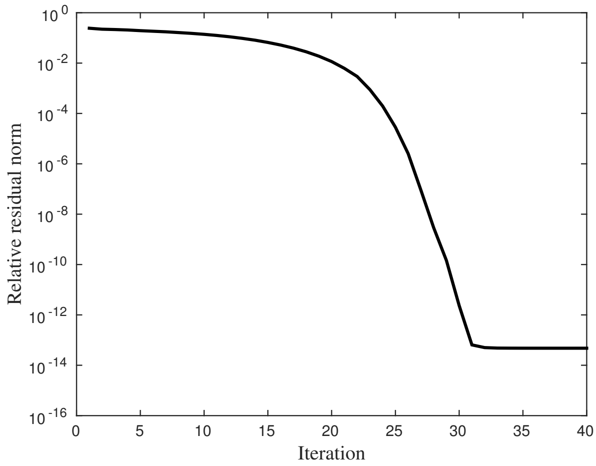


Figure 6.12: Relative residual norm of the positive real Lur'e equations (6.37) for the approximate solution X_k obtained by the matrix version of Algorithm 4 with the state space dimension $n = 3201$.

control problem. We note that Algorithm 4 requires the execution of Algorithm 1 in order to approximate the output map Ψ and the input-output map $\mathbb{F}$. As we have already discussed in Part II of Section 6.2.1, Steps 1 and 7 in Algorithm 1 require the solution of boundary value Helmholtz problems (6.12) and (6.13), respectively. These equations can be solved efficiently by applying the adaptive finite element method. We use the same setting as in Part II of Section 6.2.1 in terms of the finite element space and the adaptive solver. Hence, to avoid repetition, we just focus on the differing attributes of Algorithm 4.

In order to illustrate monotonicity of the ADI iteration (cf. Theorem 4.29), we observe the nuclear norm of $X_k = S_{\Xi,k}^* S_{\Xi,k}$ at each iteration of Algorithm 4. The recursive structure of $S_{\Xi,k}$ allows us to compute the nuclear norm of X_k efficiently. In fact, because the input and output spaces are one-dimensional (i.e., $\mathscr{Y} = \mathscr{U} = \mathbb{C}$), we have $v_k := V_k \in L^2(\Omega)$ for all $k \in \mathbb{N}$, and Steps 3, 9, and 10 of Algorithm 4 give rise to

$$S_{\Xi,k}^* = \begin{bmatrix} s_1 & s_2 & \dots & s_k \end{bmatrix},$$

where the functions $s_i \in L^2(\Omega)$, $i = 1, 2, \dots, k$, are determined recursively by

$$\begin{aligned} s_1 F_{\Xi,1} &= \sqrt{2\mathrm{Re}(\alpha_1)} v_1, \\ s_i F_{\Xi 22,i} &= \sqrt{2\mathrm{Re}(\alpha_i)} v_i - S_{\Xi,i-1}^* F_{\Xi 12,i}. \end{aligned}$$

As a result, the nuclear norm of $X_k = S_{\Xi,k}^* S_{\Xi,k}$ can be computed efficiently by

$$\|X_k\|_{\mathcal{S}_1(\mathscr{X})} = \operatorname{trace}\left[S_{\Xi,k}^* S_{\Xi,k}\right] = \sum_{j=1}^{k} \|s_j\|_{L^2(\Omega)}^2 . \tag{6.42}$$

Figure 6.13 shows the nuclear norm of X_k at each iteration of Algorithm 4. From this figure we observe that

$$\|X_k\|_{\mathcal{S}_1(\mathscr{X})} \leq \|X_{k+1}\|_{\mathcal{S}_1(\mathscr{X})}, \qquad \forall\ k \in \mathbb{N},$$

which is consistent with Theorem 4.29. Moreover, our calculations give an approximation for the nuclear norm of $X = \Psi_\Xi^* \Psi_\Xi \in \mathcal{L}(\mathscr{X})$ (cf. Theorem 4.29):

$$\|X\|_{\mathcal{S}_1(\mathscr{X})} \approx 2.226080.$$

We close this section by proposing the following computationally efficient a posteriori relative stopping criterion for Algorithm 4:

$$C_{\mathrm{pr}} := \frac{\|X_{k+1} - X_k\|_{\mathcal{S}_1(\mathscr{X})}}{\|X_{k+1}\|_{\mathcal{S}_1(\mathscr{X})}} = \frac{\operatorname{trace}[X_{k+1}] - \operatorname{trace}[X_k]}{\operatorname{trace}[X_{k+1}]} = \frac{\|s_k\|^2_{L^2(\Omega)}}{\sum_{j=1}^k \|s_j\|^2_{L^2(\Omega)}}.$$

Figure 6.14 shows the value of C_{pr} at each iteration of Algorithm 4.

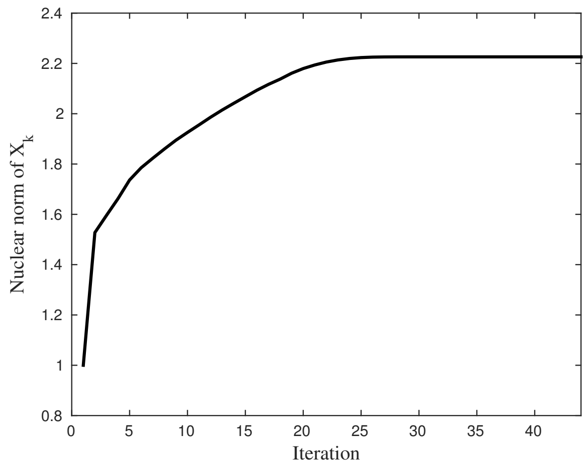


Figure 6.13: Nuclear norm of X_k.

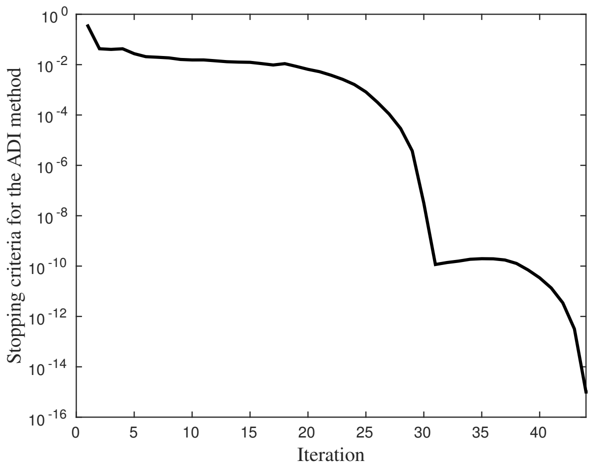


Figure 6.14: A posteriori relative stopping criterion (C_{pr}) for Algorithm 4.

Chapter 7

Summary and outlook

Change is the law of life. And those who look only to the past or present are certain to miss the future.

—John F. Kennedy

Contents

7.1 Summary and conclusions

The purpose of this PhD project was to develop two main algorithms for solving the linear-quadratic optimal control problem of externally stable well-posed linear systems, namely, the *ADI method* and the *Newton-Kleinman iteration*. These algorithms were the main contribution of this dissertation and were presented in Chapters 4 and 5. Furthermore, applicability and performance of these algorithms were illustrated by means of numerical examples arising from a two-dimensional heat equation with Robin boundary control and boundary integral observation (Chapter 6).

In the first part of Chapter 4 we proposed an extension of the ADI iteration to solve the regular linear-quadratic optimal control problem. This algorithm is called the *Riccati-ADI* method. We established a connection between this algorithm and the underlying linear-quadratic optimal control problem (Theorem 4.11). We assumed that $\Sigma = (\mathbb{T}, \Phi, \Psi, \mathbb{F})$ is a stable weakly regular linear system (cf. Definition 2.6.b), so that the output map Ψ and the input-output map $\mathbb{F}$ are bounded. This allowed us to apply an explicit formula for the solution of the linear-quadratic optimal control problem in terms of Ψ and $\mathbb{F}$ (Proposition 3.2). The link to the optimal control problem was established by considering a sequence of subspaces of $L^2(0, \infty)$. For these subspaces we chose the Takenaka–Malmquist basis (Definition 4.1), which allowed us to construct projections of the output map Ψ and the input-output map $\mathbb{F}$ (Algorithm 1). The sequence of subspaces is determined by the choice of shift parameters. We proved that the sequence of approximate solutions calculated by the Riccati-ADI algorithm is monotonically non-decreasing (Theorem 4.21). Furthermore, if the shift parameters satisfy the non-Blaschke condition (4.79), then the approximate solutions converge to the Riccati operator (3.10) in the strong operator topology. In addition, if the Riccati operator is of Schatten class (compact), then the convergence holds even in the Schatten norm (uniform operator topology). If the input and output spaces are finite-dimensional (i.e., $\mathscr{U} = \mathbb{C}^m$ and $\mathscr{Y} = \mathbb{C}^p$ for some $m, p \in \mathbb{N}$), then the Riccati-ADI algorithm provides approximate solutions in low-rank factored form.

In the second part of Chapter 4 we turned our focus to the singular linear-quadratic optimal control problem in the bounded real and positive real case. We showed that the ADI method can be applied to find approximate solutions of these singular optimal control problems (Algorithms 3 and 4). Following [15], we worked on the class of strongly stable weakly regular linear system (cf. Definition 2.6.c). Moreover, we assumed finite-dimensionality of the input and output spaces (which is justified in the actual applications). In this case, the ADI algorithms provide approximate solutions in low-rank factored form. This means that these algorithms produce sequences $(X_k)_{k\in\mathbb{N}} \in \mathcal{L}(\mathscr{X})$ of approximate solutions of the form $X_k = R_k^* R_k$ for some $R_k \in \mathcal{L}(\mathscr{X}, \mathbb{C}^{\ell_k})$, with small ℓ_k. In order to show convergence of the algorithms, we established a connection to the projected versions of the underlying singular linear-quadratic optimal control problem (Theorems 4.23 and 4.28). As in the regular case, we proved that the sequence of approximate solutions is monotonically non-decreasing. If the shift parameters are chosen appropriately, the sequence converges to the solution of the singular optimal control problem in the bounded real and positive real case.

Chapter 5 dealt with an extension of the Newton-Kleinman iteration [29] to infinite-dimensional spaces. We proposed an extension to solve the regular linear-quadratic optimal control problem subject to regular linear systems which have regular dual systems. The reason for this restriction is that weak regularity is not preserved under feedback (see for example [58, Remark 7.5.4]). Moreover, because the dual of a regular system is not necessarily regular (unless the output space is finite-dimensional), we needed the extra assumption of regularity also on the dual system. Algorithm 5 presented our extension of the Newton-Kleinman iteration. This algorithm constructs a sequence of infinite-time observability Gramians to approximate the Riccati operator (3.10). These Gramians have finite rank, if the input and output spaces are finite-dimensional. Inspired by [70, Section 8], feasibility of iterations was shown with the help of an interconnection of the system with its anticausal dual (Theorem 5.3). In addition, we established a direct connection between the Riccati operator (3.10) and the infinite-time observability Gramians generated by Algorithm 5. By assuming strong stability of the semigroup and boundedness of the control operator, we proved monotonicity and convergence of our algorithm to the Riccati operator (Theorems 5.8 and 5.9). If additionally, the first iteration of the Newton-Kleinman iteration is of Schatten class (compact), then the convergence holds even in the Schatten norm (uniform operator topology). Moreover, the quadratic rate of convergence of the Newton-Kleinman iterations was proven under the additional assumption of exponential stability of the semigroup. The results presented in Chapter 5 extend those of [12], which were developed for the class of exponentially stable well-posed linear systems with bounded control and observation operators. The presented numerical example in Section 6.2.2 suggests that it is even possible to apply Algorithm 5 in case of unbounded control operators. The proof of convergence in this case requires more investigations and is left as an open problem (see Section 5.3.1 for a potential idea).

The two algorithms given in this dissertation are developed for the class of well-posed linear systems. If these systems arise from abstract formulation of partial differential equations, then our algorithms allow us to find approximate solutions to the linear-quadratic optimal control problem using the approach “optimizing-then-discretizing”. In Chapter 6, applicability and performance of our algorithms were illustrated by means of numerical examples arising from a two-dimensional heat equation with Robin boundary control and boundary integral observation. Section 6.2 focused on the regular linear-quadratic optimal control problem. The Riccati operator (6.7) associated with our example satisfies the assumptions of Theorem A.5 and is therefore nuclear (Remark 6.4).

In Section 6.2.1, approximations of the Riccati operator (6.7) were obtained via the Riccati-ADI algorithm. In the first part of this section we focused on the choice of shift parameters. We showed that if the shift parameters do not satisfy the non-Blaschke condition (4.79), the Riccati-ADI method may converge to an operator which is not the correct approximation of the Riccati operator (6.7) (see Figure 6.2). In the second part of this section monotonicity and convergence of the Riccati-ADI algorithm were illustrated by observing the nuclear norm of the approximate solution at each iteration (Figure 6.7).

Section 6.2.2 dealt with the approximation of the Riccati operator (6.7) via the Newton-Kleinman iteration. At each Newton iteration, we solved the respective Lyapunov equation by adapting the ADI method from [41] to our setting (Algorithm 6). Hence, we referred to our method as the “Newton-ADI” iteration. Monotonicity and (quadratic rate of) convergence for the Newton iterations were illustrated by observing the nuclear norms of the approximate solutions (Figure 6.11 and Table 6.2).

Section 6.3 demonstrated an application of Algorithm 4 to solve the singular linear-quadratic optimal control problem in the positive real case. In the first part of this section we proposed an effective strategy for choosing the shift parameters. The shift parameters were obtained by applying the Penzl's heuristic procedure [44] on negatives of the stable eigenvalues of the even matrix pencil (6.41). The second part of Section 6.3 illustrated monotonicity and convergence of the approximate solutions produced by Algorithm 4. This was obtained by observing the nuclear norm of the approximate solution at each iteration (Figure 6.13)

All the algorithms applied in Chapter 6 required numerical solutions of a sequence of Helmholtz equations, which were solved efficiently by employing an adaptive finite-element solver. Section 6.2.1 provided the most important aspects regarding our implementation of the ADI algorithm. The efficiency of applying an adaptive finite-element solver was shown by comparing the results with the case of a fixed mesh (Table 6.1).

7.2 Future research perspectives

The work presented in this dissertation opens the path to a wide range of future research perspectives. Without attempting to give a complete overview, we mention several possibilities for extending the work in this thesis.

- The choice of shift parameters is essential for the speed of convergence in the Riccati-ADI algorithm. In [34, Section 3.2] it is stated that a choice based on the stable eigenvalues of the Hamiltonian

$$\mathcal{H} = \begin{bmatrix} A & -BB^* \\ -C^*C & -A^* \end{bmatrix}$$

  is effective. However, the efficient numerical computation of dominant stable eigenvalues of a Hamiltonian matrix seems not to have been explored so far. The Riccati-ADI method would be an application for this research area.

- In our numerical examples we calculated the shift parameters a priori (before starting the ADI iteration). In case of adaptively refined meshes, it might be reasonable to make a posteriori choices of shift parameters. The self-generating shift parameters as proposed by [7] is worthwhile to investigate for our examples.

- As we have shown in Section 6.3, an effective choice of shift parameters for Algorithm 4 can be made according to the generalized eigenvalues of the even matrix pencil (6.41). Nevertheless, selection of (sub-)optimal shift parameters in this case remains an open problem.

- In Chapter 6 we worked with suitable approximations of boundary value Helmholtz equations arising from ADI iterations. As a result, our approach could be referred to as *inexact ADI iteration* [41]. Providing an error analysis for inexact ADI iteration (as in [41, Section 5]) would be an interesting topic for further research.

- A distinct advantage of our algorithms is the possibility of using adaptive refinement techniques. To ensure convergence to correct solutions, one needs to choose accurate error estimators as well as appropriate mesh adaptation strategies. In fact, making inaccurate choices may slow down or destroy convergence of the resulting approximations. These aspects require more investigations in our implementations and are left as open problems.

- Monotonicity and convergence of the Newton-Kleinman iteration were proven in Section 5.4 by assuming boundedness of the control operator and strong stability of the semigroup. Convergence of Algorithm 5 in case of unbounded control operators is an open problem for our future research. It seems that Section 5.3.1 is an appropriate way of looking at this problem, because it is independent of the Lyapunov equation (5.5) and one does not require the characterization of $D(A_k)$.

- In this work we considered exclusively the class of well-posed linear systems. However, there are important input-state-output systems that do not belong to the class of well-posed linear systems. A particular example of such a system is the heat equation with Dirichlet control and Neumann observation [43]. It would be interesting to investigate the extension of our algorithms to non-well-posed linear systems.

Appendix A

Appendix

Life is found in the dance between your deepest desire and your greatest fear.

—Tony Robbins

Contents

A.1 Error estimator and mesh adaptation strategy

In this section we present an error indicator and a mesh adaptation strategy that are used in numerical simulations of Chapter 6. The material presented in this section are mainly from [3–5] and we refer to these references for more details. We start by defining the notation used in this section.

Assume that $\mathcal{T}_h$ is a decomposition of the d-dimensional domain $\Omega \subset \mathbb{R}^d$ into cells K (e.g., triangles or quadrilaterals in $\mathbb{R}^2$) of maximal diameter h. Let $P(K)$ denote a suitable space of ploynomial-like functions defined on the cell $K \in \mathcal{T}_h$. For a Hilbert space V, we define the finite element subspaces $V_h \subset V$ by

$$V_h = \left\{ v \in V : \ v_{|K} \in P(K), \quad K \in \mathcal{T}_h \right\}.$$

For a function $u \in V$, we denote its approximation on the finite-dimensional subspace $V_h \subset V$ by $u_h \in V_h$.

A.1.1 Error estimator based on a discrete approximation of the gradient

We present an error indicator that is based on the discrete approximation of the gradient. This error indicator is taken from [4, Tutorial programs, Step 9].

For any cell $K \in \mathcal{T}_h$, we denote its adjacent cell by K'. We connect the centers of K and K' by the vector $\mathbf{y}_{KK'}$, e.g., as depicted in Figure A.1.

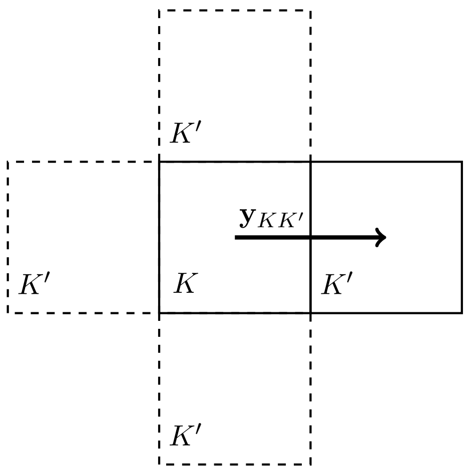


Figure A.1: Connecting the centers of two adjacent cells K and K' by the vector $\mathbf{y}_{KK'}$.

The directional derivative of a function $u \in H^1(\Omega)$ on the cell K can be approximated by

$$\frac{\mathbf{y}_{KK'}^T}{|\mathbf{y}_{KK'}|} \nabla u \approx \frac{u(K') - u(K)}{|\mathbf{y}_{KK'}|}, \tag{A.1}$$

where $u(K)$ and $u\,(K')$ denote the evaluation of the function $u \in H^1(\Omega)$ at the center of the cells K and K'. Multiplying (A.1) by $\frac{\mathbf{y}_{KK'}}{|\mathbf{y}_{KK'}|}$ and summing over all neighboring cells

K' of K result in

$$\left(\sum_{K'} \frac{\mathbf{y}_{KK'} \mathbf{y}_{KK'}^T}{|\mathbf{y}_{KK'}|^2} \right) \nabla u \approx \sum_{K'} \frac{\mathbf{y}_{KK'} \left[u(K') - u(K) \right]}{|\mathbf{y}_{KK'}|^2}.$$

If the cell K has neighbors in all directions, then the vectors $\mathbf{y}_{KK'}$ span the whole space $V_h \subset H^1(\Omega)$. As a result, the matrix

$$Y := \sum_{K'} \frac{\mathbf{y}_{KK'} \mathbf{y}_{KK'}^T}{|\mathbf{y}_{KK'}|^2}$$

is invertible and we have

$$\nabla u \approx Y^{-1} \sum_{K'} \frac{\mathbf{y}_{KK'} \left[u(K') - u(K) \right]}{|\mathbf{y}_{KK'}|^2}. \tag{A.2}$$

The approximation of the right-hand side in (A.2) on the cell K is denoted by $\nabla_h u(K)$ and the following error indicator will be used as refinement criterion:

$$\eta_K = h^{1+\frac{d}{2}} \left| \nabla_h u_h(K) \right|.$$

A.1.2 Bulk criterion for the mesh adaptation

We review a mesh adaptation strategy from [3] and [5]. This mesh adaptation is called the *bulk criterion*, which controls the total reduction of the error estimates on a given mesh $\mathcal{T}_h$.

Let $u_h \in V_h$ be an approximation of the function $u \in V$ and define the approximation error by $e := u - u_h$. Let η_K denote the local error indicator on the cell K (e.g., constructed as in Appendix A.1.1) such that the following error estimate holds

$$\|e\|_{L^2} \leq \eta := \sum_{K \in \mathcal{T}_h} \eta_K.$$

Let TOL denote a suitable tolerance for the mesh adaptation, that is

$$\eta \leq \mathrm{TOL}. \tag{A.3}$$

If condition A.3 is satisfied on the current mesh $\mathcal{T}_h$, then $u_h \in V_h$ is accepted as an approximation to $u \in V$ and we are allowed to coarsen the mesh. Otherwise, we refine the mesh so that condition A.3 is satisfied. Now the bulk criterion for the mesh adaptation reads as follows:

(i) We order the cells on the current mesh $\mathcal{T}_h$ according to their respective local error indicator η_K:

$$\eta_{K_1} \geq \eta_{K_2} \geq \cdots \geq \eta_{K_n}, \tag{A.4}$$

where $n \in \mathbb{N}$ is the number of cells on the current mesh.

(ii) Using the ordering (A.4), we define

- r: top fraction of cells to be refined
- c: bottom fraction of cells to be coarsened

(iii) We refine the smallest subset $\mathcal{R} \subset \mathcal{T}_h$, such that

$$r \cdot \eta \leq \sum_{K \in \mathcal{R} \subset \mathcal{T}_h} \eta_K,$$

and coarsen the biggest subset $\mathcal{C} \subset \mathcal{T}_h$, such that

$$c \cdot \eta \geq \sum_{K \in \mathcal{C} \subset \mathcal{T}_h} \eta_K.$$

A.2 Woodbury formula

In finite dimensions the Sherman-Morrison-Woodbury formula (or just Woodbury formula) reads as follows: Assume that the matrices $A \in \mathbb{C}^{n \times n}$ and $C \in \mathbb{C}^{k \times k}$ are both invertible. Let $U \in \mathbb{C}^{n \times k}$ and $V \in \mathbb{C}^{k \times n}$. Then $A + UCV$ is invertible, if and only if, $C^{-1} + VA^{-1}U$ is invertible. In this case there holds

$$(A + UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1}. \tag{A.5}$$

Formula (A.5) says that the inverse of a rank-k correction of an invertible matrix A, can be computed by the rank-k correction of A^{-1}. In (A.5) if we replace C by $-C^{-1}$, then we obtain an expression for the inverse of the *Schur complement* (see, e.g., [14]). We refer to [25] for a survey on the Woodbury formula and its applications.

The following Theorem gives the generalization of the Sherman-Morrison-Woodbury formula to Hilbert spaces. This formula is used in the proofs of Proposition 3.4 and Proposition 6.6.

Theorem A.1. *[19, Theorem 1.1] Let $\mathscr{X}$ and $\mathscr{Y}$ be Hilbert spaces. Let $A \in \mathcal{L}(\mathscr{X})$ and $G \in \mathcal{L}(\mathscr{Y})$ be both invertible and let $Y, Z \in \mathcal{L}(\mathscr{Y}, \mathscr{X})$. Then $A + YGZ^*$ is invertible, if and only if, $G^{-1} + Z^*A^{-1}Y$ is invertible. In this case there holds*

$$(A + YGZ^*)^{-1} = A^{-1} - A^{-1}Y(G^{-1} + Z^*A^{-1}Y)^{-1}Z^*A^{-1}.$$

A.3 Approximation of the feedback operator in the Newton-ADI method

In this section we give more details regarding the implementation of the Newton-ADI method for the example presented in Section 6.2.2.

Recall the first iteration of the Newton-ADI method in Section 6.2.2, in which an operator S^0 is generated such that

$$\widetilde{X}_1^{(i_0)} := \left(S^0\right)^* S^0 \approx X_1,$$

where X_1 is the unique solution of (6.16). In our example, since $B^* = C \in \mathcal{L}(\mathscr{X}, \mathbb{C})$, the approximate feedback operator $\widetilde{F}_1$ satisfies

$$\widetilde{F}_1 := B^* \widetilde{X}_1^{(i_0)} \in \mathcal{L}(\mathscr{X}, \mathbb{C}),$$

and can be calculated by

$$\begin{aligned}
\widetilde{F}_1 &= B^* \widetilde{X}_1^{(i_0)} \\
&= B^* (S^0)^* S^0 \\
&= B^* \begin{bmatrix} v_1 & v_2 & \cdots & v_{i_0} \end{bmatrix} \begin{bmatrix} 2\mathrm{Re}(\alpha_1) & & & \\ & 2\mathrm{Re}(\alpha_2) & & \\ & & \ddots & \\ & & & 2\mathrm{Re}(\alpha_{i_0}) \end{bmatrix} \begin{bmatrix} (v_1)^* \\ (v_2)^* \\ \vdots \\ (v_{i_0})^* \end{bmatrix} \\
&= \begin{bmatrix} B^* v_1 & B^* v_2 & \cdots & B^* v_{i_0} \end{bmatrix} \begin{bmatrix} 2\mathrm{Re}(\alpha_1) & & & \\ & 2\mathrm{Re}(\alpha_2) & & \\ & & \ddots & \\ & & & 2\mathrm{Re}(\alpha_{i_0}) \end{bmatrix} \begin{bmatrix} (v_1)^* \\ (v_2)^* \\ \vdots \\ (v_{i_0})^* \end{bmatrix}.
\end{aligned}$$

Since $B^* = C$ is a boundary integral operator, we have

$$B^* v_i = \int_\Gamma v_i(\xi)\, d\xi \ \in \mathbb{C}, \qquad i = 1, 2, \ldots, i_0.$$

Let $v_{h,i}$ be an approximation of v_i obtained by the finite element discretization of the Helmholtz equations (6.12) and (6.13). Then,

$$v_{h,i} = \sum_{j=1}^{n} x_{i,j} \varphi_j, \qquad x_{i,j} \in \mathbb{C}, \qquad \varphi_j \in V_h \subset H^1(\Omega), \qquad j = 1, 2, \ldots, n, \tag{A.6}$$

where $n \in \mathbb{N}$ is the number of degrees of freedom (DoF) corresponding to the finite element discretization. Consequently, we can calculate $B^* v_{h,i}$ by

$$b_i^0 := B^* v_{h,i} = \int_\Gamma v_{h,i}(\xi)\, d\xi = \sum_{j=1}^{n} x_{i,j} \int_\Gamma \varphi_j(\xi)\, d\xi \in \mathbb{C}. \tag{A.7}$$

Let $\widetilde{F}_{h,1}$ denote the approximation of $\widetilde{F}_1$. Then, we observe that

$$\begin{aligned}
\widetilde{F}_{h,1} &= \begin{bmatrix} b_1^0 & b_2^0 & \cdots & b_{i_0}^0 \end{bmatrix} \begin{bmatrix} 2\mathrm{Re}(\alpha_1) & & & \\ & 2\mathrm{Re}(\alpha_2) & & \\ & & \ddots & \\ & & & 2\mathrm{Re}(\alpha_{i_0}) \end{bmatrix} \begin{bmatrix} v_{h,1}^* \\ v_{h,2}^* \\ \vdots \\ v_{h,i_0}^* \end{bmatrix} \\
&= 2 \sum_{i=1}^{i_0} \mathrm{Re}(\alpha_i) b_i^0 \cdot v_{h,i}^*.
\end{aligned} \tag{A.8}$$

We note that if the functions $v_{h,i}$, $i = 1, 2, \ldots, i_0$, are computed using adaptive mesh refinements, then one has to interpolate all these functions on a unified mesh before computing the feedback $\widetilde{F}_1$ in (A.8).

We can generalize the above calculations to all iterations of the Newton-ADI method: Let S^k denote the approximation of the augmented output map Ψ_k^{aug}, generated by Algorithm 6. Then,

$$\widetilde{X}_{k+1}^{(i_k)} := (S^k)^* S^k \approx X_{k+1},$$

where $X_{k+1} \in \mathcal{L}(\mathscr{X})$ is the unique solution of (6.18). Since $B^* = C \in \mathcal{L}(\mathscr{X}, \mathbb{C})$, the approximated feedback operator generated by the $(k+1)$-Newton iteration satisfies

$$\widetilde{F}_{k+1} := B^* \widetilde{X}_{k+1}^{(i_k)} \in \mathcal{L}(\mathscr{X}, \mathbb{C}).$$

The feedback $\widetilde{F}_{k+1}$ can be calculated by

$$\begin{aligned}
\widetilde{F}_{k+1} &= B^* \widetilde{X}_{k+1}^{(i_k)} \\
&= B^* (S^k)^* S^k \\
&= B^* \begin{bmatrix} V_1^k & V_2^k & \cdots & V_{i_k}^k \end{bmatrix} \begin{bmatrix} 2\mathrm{Re}(\alpha_1) & & & \\ & 2\mathrm{Re}(\alpha_2) & & \\ & & \ddots & \\ & & & 2\mathrm{Re}(\alpha_{i_k}) \end{bmatrix} \begin{bmatrix} (V_1^k)^* \\ (V_2^k)^* \\ \vdots \\ (V_{i_k}^k)^* \end{bmatrix}.
\end{aligned}$$

From Proposition 6.6 we know that V_i^k have the structure $V_i^k = [v_i^k \quad v_i^{\widetilde{F}_k}]$. Since $B^* = C$ is a boundary integral operator, we have

$$B^* V_i^k = \left[\int_\Gamma v_i^k(\xi)\, d\xi \quad \int_\Gamma v_i^{\widetilde{F}_k}(\xi)\, d\xi \right] \in \mathbb{C}^{1\times 2}, \qquad i = 1, 2, \ldots, i_k.$$

Let $v_{h,i}^k$ and $v_{h,i}^{\widetilde{F}_k}$ be respectively the approximations of v_i^k and $v_i^{\widetilde{F}_k}$ (see (A.6)) obtained by the finite element discretization of equations (6.22) and a further application of the recursive formulas (6.19) and (6.20) (see Proposition 6.6). Then,

$$v_{h,i} = \sum_{j=1}^{n} y_{i,j} \varphi_j, \qquad v_{h,i}^{\widetilde{F}_k} = \sum_{j=1}^{n} z_{i,j} \varphi_j,$$

with $y_{i,j}, z_{i,j} \in \mathbb{C}$, $\varphi_j \in V_h \subset H^1(\Omega)$, for $j = 1, 2, \ldots, n$, where $n \in \mathbb{N}$ is the number of degrees of freedom (DoF) corresponding to the finite element discretization. We define

$$V_{h,i}^k := \left[v_{h,i}^k \quad v_{h,i}^{\widetilde{F}_k} \right].$$

Consequently, we obtain

$$\begin{aligned}
\mathbf{B}_i^k &:= B^* V_{h,i}^k \\
&= \left[B^* v_{h,i}^k \quad B^* v_{h,i}^{\widetilde{F}_k} \right] \\
&= \left[\int_\Gamma v_{h,i}^k(\xi)\, d\xi \quad \int_\Gamma v_{h,i}^{\widetilde{F}_k}(\xi)\, d\xi \right] \\
&= \sum_{j=1}^{n} \left[y_{i,j} \int_\Gamma \varphi_j(\xi)\, d\xi \quad z_{i,j} \int_\Gamma \varphi_j(\xi)\, d\xi \right] \in \mathbb{C}^{1\times 2}.
\end{aligned}$$

Eventually, if $\widetilde{F}_{h,k+1}$ denotes an approximation of $\widetilde{F}_{k+1}$, then we observe that

$$\widetilde{F}_{h,k+1} = \begin{bmatrix}\mathbf{B}_1^k & \mathbf{B}_2^k & \cdots & \mathbf{B}_{i_k}^k\end{bmatrix} \begin{bmatrix} 2\mathrm{Re}(\alpha_1) & & & \\ & 2\mathrm{Re}(\alpha_2) & & \\ & & \ddots & \\ & & & 2\mathrm{Re}(\alpha_{i_k}) \end{bmatrix} \begin{bmatrix} \left(V_{h,1}^k\right)^* \\ \left(V_{h,2}^k\right)^* \\ \vdots \\ \left(V_{h,i_k}^k\right)^* \end{bmatrix}$$

$$= 2\sum_{i=1}^{i_k} \mathrm{Re}(\alpha_i)\mathbf{B}_i^k \cdot \left(V_{h,i}^k\right)^*.$$

A.4 Convergence of operator sequences

The results presented in this section are taken directly from [41, Appendix A]. Our purpose is to simplify the referencing.

Theorem A.2. *(see [50, p. 263]) Let $\mathscr{X}$ be a Hilbert space and $(X_k)_{k\in\mathbb{N}} \in \mathcal{L}(\mathscr{X})$ be a sequence of self-adjoint operators such that $X_k \geq X_{k+1}$, for all $k \in \mathbb{N}$. Moreover, assume that there exists some $X \in \mathcal{L}(\mathscr{X})$ such that $X_k \geq X$, for all $k \in \mathbb{N}$.*
Then, there exists some self-adjoint operator $\widetilde{X} \in \mathcal{L}(\mathscr{X})$ such that $X_k \geq \widetilde{X} \geq X$, for all $k \in \mathbb{N}$, and the sequence $(X_k)_{k\in\mathbb{N}}$ converges to $\widetilde{X}$ in the strong operator topology, i.e.,

$$\lim_{k\to\infty} X_k x = \widetilde{X}x, \qquad \forall\, x \in \mathscr{X}.$$

Theorem A.3. *(see [22, Theorem III.6.3]) Let $\mathscr{X}_1$ and $\mathscr{X}_2$ be Hilbert spaces, and let $\Pi_k \in \mathcal{L}(\mathscr{X}_1)$ be a sequence of self-adjoint operators converging in the strong operator topology to $\Pi \in \mathcal{L}(\mathscr{X}_1)$.*

(a) If $T \in \mathcal{K}(\mathscr{X}_1, \mathscr{X}_2)$, then

$$\lim_{k\to\infty} \|\Pi_k T - \Pi T\|_{\mathcal{L}(\mathscr{X}_1,\mathscr{X}_2)} = 0.$$

(b) If $T \in \mathcal{S}_p(\mathscr{X}_1, \mathscr{X}_2)$ with $p \in [1,\infty)$, then

$$\lim_{k\to\infty} \|\Pi_k T - \Pi T\|_{\mathcal{S}_p(\mathscr{X}_1,\mathscr{X}_2)} = 0.$$

A.5 Nuclearity of the Riccati operator

In this section we gather two important theories that are needed in the analysis of Chapter 6. In particular, Theorem A.5 will help us to show that the Riccati operator (6.7) is nuclear (i.e., $X \in \mathcal{S}_1(\mathscr{X})$). The results presented here are from [16] and we refer to this paper as well as the references therein for more details.

Theorem A.4. *[16, Theorem 3.2] Suppose that A is the generator of an exponentially stable semigroup $\mathbb{T}$ on the Hilbert space $\mathscr{X}$, and that $C \in \mathcal{L}(\mathscr{X}, \mathbb{C}^p)$. Then the output map $\Psi : \mathscr{X} \to L^2(0,\infty;\mathbb{C}^p)$ defined by*

$$(\Psi x_0)(t) = C\mathbb{T}_t x_0, \qquad \forall\, t \geq 0, \quad \forall\, x_0 \in \mathscr{X},$$

is Hilbert–Schmidt, i.e., $\Psi \in \mathcal{S}_2(\mathscr{X}, L^2(0,\infty;\mathbb{C}^p))$.

Theorem A.5. *[16, Theorem 4.1 & 4.6] Suppose that:*

(i) A is the generator of a strongly continuous analytic semigroup $\mathbb{T}$ on a Hilbert space $\mathscr{X}$. Let $\omega_0(\mathbb{T})$ be the growth bound of $\mathbb{T}$ (cf. Definition 2.2).

(ii) There exists some $\beta \in (-1,0]$ such that $(\omega I - A)^\beta B \in \mathcal{L}(\mathscr{U}, \mathscr{X})$, for all $\omega > \omega_0(\mathbb{T})$. This means that $B \in \mathcal{L}(\mathscr{U}, \mathscr{X}_\beta)$.

(iii) $C \in \mathcal{L}(\mathscr{X}, \mathscr{Y})$.

(iv) (Exponential detectability) There exists $L \in \mathcal{L}(\mathscr{Y}, \mathscr{X})$ such that $A + LC$ generates an exponentially stable analytic semigroup on $\mathscr{X}$.

(v) (Finite cost condition) For each $x_0 \in \mathscr{X}$, there exists $u \in L^2(0,\infty;\mathscr{U})$ such that the mild solution x to $\frac{d}{dt}x(t) = Ax(t) + Bu(t)$, $x(0) = x_0$, satisfies $Cx(\cdot) \in L^2(0,\infty;\mathscr{Y})$.

Then, there exists a self-adjoint, nonnegative $X \in \mathcal{L}(\mathscr{X})$ such that:

1. *X is the unique self-adjoint nonnegative solution of the following algebraic Riccati equation*

$$\langle Ax_0, Xz_0\rangle_{\mathscr{X}} + \langle Xx_0, Az_0\rangle_{\mathscr{X}} + \langle Cx_0, Cz_0\rangle_{\mathscr{Y}} = \langle B^*Xx_0, B^*Xz_0\rangle_{\mathscr{U}},$$

for all $x_0, z_0 \in D((\omega I - A)^\epsilon)$ and any $\epsilon > 0$.

2. *$B^*X \in \mathcal{L}(\mathscr{X}, \mathscr{U})$.*

If in addition to the assumptions (i)–(v), the input and the output spaces are finite-dimensional (i.e., $\mathscr{U} = \mathbb{C}^m$ and $\mathscr{Y} = \mathbb{C}^p$ for some $m, p \in \mathbb{N}$), then

(1) $X \in \mathcal{S}_1(\mathscr{X}_\gamma, (\mathscr{X}_\gamma)')$ for all $\gamma > -\frac{1}{2}$.

(2) If $\beta > -\frac{1}{2}$, then $B^*X \in \mathcal{S}_2(\mathscr{X}_\gamma, \mathscr{U})$ for all $\gamma > -\frac{1}{2}$.

A.6 Perturbation of semigroup generator by unbounded observation operator

Theorem A.6. *[62, Theorem 5.4.2] Let $\mathbb{T}$ be a strongly continuous semigroup on $\mathscr{X}$ with generating operator A. If $C : D(A) \subset \mathscr{X} \to \mathscr{Y}$ is an admissible observation operator for $\mathbb{T}$ and $B \in \mathcal{L}(\mathscr{Y}, \mathscr{X})$, then the operator $A + BC : D(A) \subset \mathscr{X} \to \mathscr{X}$ is the generator of a strongly continuous semigroup $\mathbb{T}^{cl}$ on $\mathscr{X}$. This semigroup satisfies the integral equation*

$$\mathbb{T}^{cl}_t x_0 = \mathbb{T}_t x_0 + \int_0^t \mathbb{T}_{t-\sigma} BC \mathbb{T}^{cl}_\sigma x_0 \, d\sigma, \qquad \forall\ x_0 \in D(A) = D(A+BC), \quad \forall\ t \geq 0.$$

Moreover, for any Hilbert space $\mathscr{Y}_1$, the space of all admissible observation operators for $\mathbb{T}$ that map into $\mathscr{Y}_1$ is equal to the corresponding space for $\mathbb{T}^{cl}$.

Bibliography

[1] M. Ainsworth and J. Oden, *A Posteriori Error Estimation in Finite Element Analysis*, Pure and Applied Mathematics (New York), Wiley-Interscience [John Wiley & Sons], New York, 2000.

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

[3] W. Bangerth, D. Davydov, T. Heister, L. Heltai, G. Kanschat, M. Kronbichler, M. Maier, B. Turcksin, and D. Wells, *The `deal.II` library, version 8.4*, Journal of Numerical Mathematics, 24 (2016).

[4] W. Bangerth, R. Hartmann, and G. Kanschat, `deal.II` *Differential Equations Analysis Library, Technical Reference.* http://www.dealii.org.

[5] W. Bangerth and R. Rannacher, *Adaptive Finite Element Methods for Differential Equations*, Birkhäuser, 2013.

[6] P. Benner, M. Heinkenschloss, J. Saak, and H. K. Weichelt, *An inexact low-rank Newton–ADI method for large-scale algebraic Riccati equations*, Applied Numerical Mathematics, 108 (2016), pp. 125–142.

[7] P. Benner, P. Kürschner, and J. Saak, *Self-generating and efficient shift parameters in ADI methods for large Lyapunov and Sylvester equations*, Electronic Transactions on Numerical Analysis, 43 (2014), pp. 142–162.

[8] P. Benner, J.-R. Li, and T. Penzl, *Numerical solution of large-scale Lyapunov equations, Riccati equations, and linear-quadratic optimal control problems*, Numerical Linear Algebra with Applications, 15 (2008), pp. 755–777.

[9] P. Benner and J. Saak, *A Galerkin-Newton-ADI method for solving the large-scale algebraic Riccati equations*, Technical Report SPP1253-090, Deutsche Forschungsgemeinschaft, Priority Program 1253, Bonn, Germany, (2010).

[10] D. A. Bini, B. Iannazzo, and B. Meini, *Numerical Solution of Algebraic Riccati Equations*, vol. 9, SIAM, 2012.

[11] J. R. Bunch and L. Kaufman, *Some stable methods for calculating inertia and solving symmetric linear systems*, Mathematics of Computation, (1977), pp. 163–179.

[12] J. Burns, E. Sachs, and L. Zietsman, *Mesh independence of Kleinman-Newton iterations for Riccati equations in Hilbert space*, SIAM J. Control Optim., 47 (2008), pp. 2663–2692.

[13] C. Byrnes, D. Gilliam, V. Shubov, and G. Weiss, *Regular linear systems governed by a boundary controlled heat equation*, J. Dyn. Control Syst., 8 (2002), pp. 341–370.

[14] R. W. Cottle, *Manifestations of the Schur complement*, Linear Algebra and its Applications, 8 (1974), pp. 189–211.

[15] R. Curtain, *Linear operator inequalities for strongly stable weakly regular linear systems*, Math. Control Signals Systems, 14 (2001), pp. 299–337.

[16] R. Curtain, K. Mikkola, and A. Sasane, *The Hilbert–Schmidt property of feedback operators*, Journal of Mathematical Analysis and Applications, 329 (2007), pp. 1145–1160.

[17] R. Curtain and G. Weiss, *Well-posedness of triples of operators (in the sense of linear systems theory)*, in Control and Estimation of Distributed Parameter Systems, W. Schappacher, F. Kappel, and K. Kunisch, eds., Progress in Systems and Control Theory, Birkhäuser-Verlag, Basel, 1989, pp. 41–59.

[18] R. Curtain and H. Zwart, *An Introduction to Infinite Dimensional Linear Systems Theory*, Springer, Berlin, Heidelberg, New York, 1995.

[19] C. Y. Deng, *A generalization of the Sherman-Morrison-Woodbury formula*, Applied Mathematics Letters, 24 (2011), pp. 1561–1564.

[20] A. Frazho and W. Bhosri, *An Operator Perspective on Signals and Systems*, vol. 204, Springer Science & Business Media, 2009.

[21] F. R. Gantmakher, *The Theory of Matrices*, vol. 131, American Mathematical Soc., 1998.

[22] I. Gohberg and M. Kreĭn, *Introduction to the Theory of Linear Nonselfadjoint Operators*, vol. 18 of Trans. Math. Monogr., AMS, Providence, RI, 1969.

[23] G. H. Golub and C. F. Van Loan, *Matrix Computations*, vol. 3, JHU Press, 2012.

[24] C. Guiver and M. R. Opmeer, *Bounded real and positive real balanced truncation for infinite-dimensional systems*, Mathematical Control and Related Fields, 3 (2013), pp. 83–119.

[25] W. W. Hager, *Updating the inverse of a matrix*, SIAM review, 31 (1989), pp. 221–239.

[26] S. Hansen and G. Weiss, *New results on the operator Carleson measure criterion*, IMA J. Math Control Inform., 14 (1997), pp. 3–32.

[27] B. Jacob and J. R. Partington, *Admissibility of control and observation operators for semigroups: a survey*, in Current Trends in Operator Theory and its Applications, Springer, 2004, pp. 199–221.

[28] B. Jacob, J. R. Partington, and S. Pott, *Conditions for admissibility of observation operators and boundedness of Hankel operators*, Integral Equations and Operator Theory, 47 (2003), pp. 315–338.

[29] D. Kleinman, *On an iterative technique for Riccati equation computations*, IEEE Trans. Automatic Control, 13 (1968), pp. 114–115.

[30] C. S. Kubrusly and P. C. Vieira, *Convergence and decomposition for tensor products of Hilbert space operators*, Oper. Matrices, 2 (2008), pp. 407–416.

[31] P. Lancaster and L. Rodman, *Algebraic Riccati Equations*, Clarendon Press, Oxford, 1995.

[32] I. Lasiecka and R. Triggiani, *Control Theory for Partial Differential Equations: Continuous and Approximation Theories, Part 1*, Cambridge University Press, 2000.

[33] A. Laub, *A Schur method for solving algebraic Riccati equations*, IEEE Transactions on Automatic Control, 24 (1979), pp. 913–921.

[34] Y. Lin and V. Simoncini, *A new subspace iteration method for the algebraic Riccati equation*, Numerical Linear Algebra with Applications, 22 (2015), pp. 26–47.

[35] A. Massoudi, M. Opmeer, and T. Reis, *Analysis of an iteration method for the algebraic Riccati equation*, SIAM J. Matrix Anal. Appl., 37 (2016), pp. 624 – 648.

[36] A. Massoudi, M. R. Opmeer, and T. Reis, *The ADI method for the algebraic Riccati equation*, Hamburger Beiträge zur Angewandten Mathematik, preprint, (2014).

[37] ——, *The ADI method for bounded real and positive real Lur'e equations*, Numerische Mathematik, (2016), pp. 1–28.

[38] V. Mehrmann, *The Autonomous Linear Quadratic Control Problem, Theory and Numerical Solution*, no. 163 in Lecture Notes in Control and Information Sciences, Springer-Verlag, Heidelberg, Germany, 1991.

[39] K. Mikkola et al., *Infinite-Dimensional Linear Systems, Optimal Control and Algebraic Riccati Equations*, Helsinki University of Technology, 2002.

[40] P. Opdenacker and E. Jonckheere, *A contraction mapping preserving balanced reduction scheme and its infinity norm error bounds*, IEEE Trans. Circuits Syst. I, 35 (1988), pp. 184–189.

[41] M. Opmeer, T. Reis, and W. Wollner, *Finite-rank ADI iteration for operator Lyapunov equations*, SIAM J. CONTROL OPTIM, 51 (2013), pp. 4084 – 4117.

[42] M. R. Opmeer, *The algebraic Riccati equation for infinite-dimensional systems*, in Proceedings of the 21st International Symposium on Mathematical Theory of Networks and Systems, 2014, pp. 1610–1614.

[43] M. R. Opmeer and R. F. Curtain, *Non-well-posed infinite-dimensional linear systems*, in Proceedings of the 45th IEEE Conference on Decision and Control, IEEE, 2006, pp. 4694–4699.

[44] T. Penzl, *A cyclic low-rank Smith method for large sparse Lyapunov equations*, SIAM J. Sci. Comput., 21 (1999/00), pp. 1401–1418 (electronic).

[45] F. Poloni and T. Reis, *A deflation approach for large-scale Lur'e equations*, SIAM Journal on Matrix Analysis and Applications, 33 (2012), pp. 1339–1368.

[46] M. Reed and B. Simon, *Methods of Modern Mathematical Physics I: Functional Analysis*, Academic Press, New York, London, 1972.

[47] ——, *Methods of Modern Mathematical Physics IV: Analysis of Operators*, Academic Press, New York, London, 1977.

[48] T. Reis, *Lur'e equations and even matrix pencils*, Linear Algebra and its Applications, 434 (2011), pp. 152–173.

[49] T. Reis and T. Stykel, *Positive real and bounded real balancing for model reduction of descriptor systems*, International Journal of Control, 83 (2010), pp. 74–88.

[50] F. Riesz and B. Sz.-Nagy, *Functional Analysis*, Blackie & Son Limited, London, Glasgow, 1956.

[51] M. Rosenblum and J. Rovnyak, *Hardy Classes and Operator Theory*, Courier Corporation, 1997.

[52] Y. Saad, *Iterative Methods for Sparse Linear Systems*, SIAM, 2003.

[53] ——, *Numerical Methods for Large Eigenvalue Problems: Revised Edition*, vol. 66, SIAM, 2011.

[54] J. Saak, *Efficient Numerical Solution of Large Scale Algebraic Matrix Equations in PDE Control and Model Order Reduction*, doctoral dissertation, Technische Universität Chemnitz, 2009.

[55] D. Salamon, *Infinite-dimensional linear systems with unbounded control and observation: a functional analytic approach*, Trans. Amer. Math. Soc., 300 (1987), pp. 383–431.

[56] D. Salamon, *Realization theory in Hilbert space*, Mathematical Systems Theory, 21 (1988), pp. 147–164.

[57] O. Staffans, *Quadratic optimal control of stable well-posed linear systems*, Trans. Amer. Math. Soc., 349 (1997), pp. 3679–3715.

[58] ——, *Well-Posed Linear Systems*, vol. 103 of Encyclopedia of Mathematics and its Applications, Cambridge University Press, Cambridge, 2005.

[59] O. Staffans and G. Weiss, *Transfer functions of regular linear systems. part II. the system operator and the Lax-Phillips semigroup*, Trans. Amer. Math. Soc., 354 (2002), pp. 3229–3262.

[60] O. J. Staffans, *On the discrete and continuous time infinite-dimensional algebraic Riccati equations*, Systems & control letters, 29 (1996), pp. 131–138.

[61] O. J. Staffans and G. Weiss, *Transfer functions of regular linear systems part III: inversions and duality*, Integral Equations and Operator Theory, 49 (2004), pp. 517–558.

[62] M. Tucsnak and G. Weiss, *Observation and Control for Operator Semigroups*, Birkhäuser Advanced Texts Basler Lehrbücher, Birkhäuser, Basel, 2009.

[63] ——, *Well-posed systems–the LTI case and beyond*, Automatica, 50 (2014), pp. 1757–1779.

[64] G. Weiss, *Admissibility of unbounded control operators*, SIAM J. Control Optim., 27 (1989), pp. 527–545.

[65] ——, *Admissible observation operators for linear semigroups*, Israel J. Math., 65 (1989), pp. 17–43.

[66] ——, *The representation of regular linear systems on Hilbert spaces*, in Control and Estimation of Distributed Parameter Systems, W. Schappacher, F. Kappel, and K. Kunisch, eds., Progress in Systems and Control Theory, Birkhäuser-Verlag, Basel, 1989, pp. 401–416.

[67] ——, *Regular linear systems with feedback*, Math. Control Signals Systems, 7 (1994), pp. 23–57.

[68] ——, *Transfer functions of regular linear systems. part I: characterization of regularity*, Trans. Amer. Math. Soc., 342 (1994), pp. 827–854.

[69] G. Weiss and H. Zwart, *An example in linear quadratic optimal control*, Systems & Control Letters, 33 (1998), pp. 339–349.

[70] M. Weiss and G. Weiss, *Optimal control of stable weakly regular linear systems*, Math. Control Signals Systems, 10 (1997), pp. 287–330.

[71] J. Willems, *Dissipative dynamical systems. II: Linear systems with quadratic supply rates*, Arch. Ration. Mech. Anal., 45 (1972), pp. 352–393.

Index

Abstract

In this dissertation we develop two algorithms for solving the linear-quadratic optimal control problem of externally stable well-posed linear systems.

The first algorithm is an extension of the *Alternating Direction Implicit (ADI) iteration* in order to solve the regular linear-quadratic optimal control problem. The algorithm is based on approximating the output map and the input-output map of the well-posed linear system using projections on appropriate subspaces. These projections are determined by the so-called "shift parameters". We prove that the approximation obtained by this algorithm expresses the optimal cost for a projected optimal control problem. Furthermore, we show that the sequence of approximate solutions obtained by this algorithm is monotonically non-decreasing. Under mild assumptions on the shift parameters, we prove convergence of this sequence to the optimal cost operator (Riccati operator).
Later on, we turn our focus to the singular linear-quadratic optimal control problem in the *bounded real* and *positive real* case. We show that the ADI iteration can be applied to find approximate solutions of these singular optimal control problems. By assuming finite-dimensionality of the input and output spaces (which is justified in the actual applications), our method provides approximate solutions in low-rank factored form. In order to show convergence, we establish a connection to the projected singular linear-quadratic optimal control problem. As in the regular case, we show that the sequence of approximate solutions is monotonically non-decreasing. If the shift parameters are chosen appropriately, the sequence converges to the optimal cost of the singular optimal control problem in the bounded real and positive real case.

The second algorithm is an extension of the *Newton-Kleinman iteration* to the infinite-dimensional spaces. We propose an extension for solving the regular linear-quadratic optimal control problem subject to regular linear systems which have regular dual systems. We construct a sequence of infinite-time observability Gramians to approximate the Riccati operator. These Gramians have finite rank, if the input and output spaces are finite-dimensional. The feasibility of iterations is shown with the help of an interconnection of the system with its anticausal dual. In addition, we establish a direct connection between the Riccati operator and the sequence of infinite-time observability Gramians. In order to prove monotonicity and convergence of our algorithm, we further assume strong stability of the semigroup and boundedness of the control operator. Moreover, the quadratic rate of convergence of the Newton-Kleinman iterations is proven under the additional assumption

of exponentially stability on the semigroup. The presented numerical example suggests that it is even possible to apply our method in case of unbounded control operators. However, the proof of convergence in this case requires more investigations and is left as an open problem.

The two algorithms given in this dissertation are developed for the class of well-posed linear systems. If these systems arise from abstract formulation of partial differential equations, then our algorithms allow numerical solutions of the optimal control problems using the approach "optimizing-then-discretizing". We verify the applicability of our algorithms by applying them to a two-dimensional heat equation with Robin boundary control and boundary integral observation. We present three numerical examples. The first two deal with the regular linear-quadratic optimal control problem. The Riccati operator associated with these examples is nuclear. We find approximations of the Riccati operator once by applying the ADI method and once by employing the Newton-Kleinman iteration.
The last example demonstrates the applicability of the ADI method for solving the singular linear-quadratic optimal control problem in the positive real case. A correct choice of shift parameters is crucial for convergence of the ADI method. We propose an effective strategy for choosing the shift parameters based on the stable eigenvalues of the even matrix pencil. In all the examples, monotonicity and convergence of the approximate solutions are illustrated by observing the nuclear norm of the approximations at each iteration.

Kurzfassung

In der vorliegenden Arbeit entwickeln wir zwei Algorithmen für die linear-quadratische optimale Steuerung von *extern-stabilen wohl-definierten linearen Systemen (externally stable well-posed linear systems).*

Der erste Algorithmus ist eine Erweiterung der *Alternating Direction Implicit (ADI) Iteration* für die Lösung des regulären linear-quadratischen optimalen Steuerungsproblems. Dieser Algorithmus basiert auf den Approximationen des *Ausgangsoperators* sowie des *Eingang-Ausgang Operators* anhand der Projektionen in geeignete Unterräume. Diese Projektionen werden durch die sogenannten "Shift-Parameter" definiert. Wir zeigen, dass die Approximation, die durch diesen Algorithmus erzeugt wird, dem optimalen Kostenoperator eines projizierten optimalen Steuerungsproblems entspricht. Außerdem zeigen wir, dass die Folge der approximierten Lösungen monoton wachsend ist. Unter milden Voraussetzungen an die Shift-Parameter beweisen wir die Konvergenz dieser Folge gegen den optimalen Kostenoperator (Riccati Operator).
Desweiteren betrachten wir das singuläre optimale Steuerungsproblem im "beschränkten reellen" (*bounded real*) und "positiven reellen" (*positive real*) Fall. Wir zeigen die Anwendbarkeit der ADI Iteration, um approximierte Lösungen dieser Probleme zu finden. Falls der Eingangs- und Ausgangsraum endlichdimensional sind, dann liefert unser Algorithmus approximierte Lösungen in faktorisierter Form mit endlichem Rang. Um die Konvergenz zu zeigen, erstellen wir eine Verbindung mit dem projizierten singulären optimalen Steuerungsproblem. Wir beweisen, dass die Folge der approximierten Lösungen monoton wachsend ist. Mit "geeigneten" Shift-Parametern konvergiert diese Folge gegen optimalen Kostenoperator im beschränkten reellen und positiven reellen Fall.

Der zweite Algorithmus ist eine Erweiterung der *Newton-Kleinman Iteration* auf unendlichdimensionalen Räumen. Wir schlagen eine Erweiterung dieser Iteration vor, um das reguläre optimale Steuerungsproblem gemäß regulären linearen Systemen mit regulären Dualsystemen zu lösen. Eine Folge der Gramschen-Steuerbarkeitsoperatoren wird konstruiert, um den Riccati Operator zu approximieren. Diese Gramschen-Steuerbarkeitsoperatoren haben endlichen Rang, wenn der Eingangs- und Ausgangsraum endlichdimensional sind. Die Durchführbarkeit der Iterationen wird anhand einer Verbindung zwischen dem System und seinem antikausalen Dualsystem gezeigt. Darüber hinaus stellen wir eine direkte Beziehung zwischen dem Riccati Operator und den Gramschen-Steuerbarkeitsoperatoren her. Um die Monotonie und Konvergenz unseres Algorithmus zu be-

weisen, nehmen wir zusätzlich an, dass die Halbgruppe stark stabil und der Kontrolloperator beschränkt ist. Anhand eines numerischen Beispieles zeigen wir die Anwendbarkeit unserer Methode sogar im Fall der unbeschränkten Kontrolloperatoren. In diesem Fall benötigt der Konvergenzbeweis weitere Recherche und bleibt ein offenes Problem.

Beide Algorithmen wurden für die Klasse der wohl-definierten linearen Systemen entwickelt. Wenn diese Systeme ihren Ursprung in abstrakter Form der partiellen Differentialgleichungen haben, dann ermöglichen unsere Algorithmen numerische Lösungen der optimalen Steuerungsprobleme anhand der sogenannten "Optimierung-dann-Diskretisierung" Methode. Wir überprüfen die Realisierbarkeit unserer Algorithmen, indem wir diese auf die zweidimensionale Wärmeleitungsgleichung mit *Robin Randsteuerung* und *Randintegral Beobachtung* anwenden. Im letzten Teil der Arbeit präsentieren wir drei numerische Beispiele. Die ersten beiden handeln von dem Fall eines regulären optimalen Steuerungsproblems. Der Riccati Operator in diesen Beispielen ist nuklear. Wir finden Approximationen des Riccati Operators einerseits mittels der ADI Methode und anderseits durch die Anwendung der Newton-Kleinman Iteration.
Das letzte Beispiel demonstriert die Anwendbarkeit der ADI Methode bei der Lösung des optimalen Steuerungsproblems im positiven reellen Fall. Die geeignete Wahl der Shift-Parameter ist entscheidend für die Konvergenz der ADI Methode. Weiter schlagen wir eine effektive Strategie vor, die auf stabilen Eigenwerten des geraden Matrixbüschels basiert. In allen Drei Beispielen werden Monotonie und Konvergenz der approximierten Lösungen durch Beobachtung der nuklearen Normen illustriert.

List of Publications

Refereed articles in journals:

[1] A. Massoudi, M. Opmeer, and T. Reis, *Analysis of an iteration method for the algebraic Riccati equation*, SIAM J. Matrix Anal. Appl., 37 (2016), pp. 624 – 648.

[2] A. Massoudi, M. Opmeer, and T. Reis, *The ADI method for bounded real and positive real Lur'e equations*, Numerische Mathematik, (2016), pp. 1 – 28.

Conference proceedings:

[1] A. Massoudi and T. Reis, *The Newton-Kleinman method for the optimal control of stable regular linear systems*, PAMM 16.1 (2016): 815 – 816.

Eidesstattliche Versicherung

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Hamburg, Dezember 31, 2016

Arash Massoudi