Classical Control Theory

Jerzy Zabczyk*

Institute of Mathematics, Polish Academy of Sciences, Warsaw, Poland

Lectures given at the
Summer School on Mathematical Control Theory
Trieste, 3-28 September 2001

LNS028001

*zabczyk@impan.gov.pl

Abstract

The notes introduce basic concepts and results of the classical control theory. The following topics: controllability, observability, minimum energy control, stability and stabilizability as well as linear quadratic control problem and the associated Riccati equations are discussed in details.

Contents

0 Introduction

The aim of the lectures is to introduce and motivate basic concepts of the classical control theory. Due to the time limitation several of the important topics like, realisation, control with partial observation, systems on manifolds and infinite dimensional systems will be not covered in the notes. We follow basically our book [31]. For additional information we suggest the reader to consult other sources listed in references, in particular Sontag's book [25]. The reader should try to solve exercises and as a test of good understanding we recommend Exercises 1.7, 1.8, 2.2, 3.7.

A departure point of control theory is the differential equation

$$\dot{y} = f(y,u), \quad y(0) = x \in \mathbb{R}^n, \tag{0.1}$$

with the right-hand side depending on a parameter u from a set $U \subset \mathbb{R}^m$. The set U is called *the set of control parameters.* Differential equations depending on a parameter have been objects of the theory of differential equations for a long time. In particular an important question of continuous dependence of the solutions on parameters has been asked and answered under appropriate conditions. Problems studied in mathematical control theory are, however, of different nature, and a basic role in their formulation is played by the concept of *control.* One distinguishes controls of two types: *open* and *closed loop. An open loop control* can be basically an arbitrary function $u(\,\cdot\,) : [0,+\infty) \longrightarrow U$, for which the equation

$$\dot{y}(t) = f(y(t)),u(t)), \quad t \geq 0,\ y(0) = x, \tag{0.2}$$

has a well defined solution.

A closed loop control can be identified with a mapping $k : \mathbb{R}^n \longrightarrow U$, which may depend on $t \geq 0$, such that the equation

$$\dot{y}(t) = f(y(t),k(y(t))), \quad t \geq 0,\ y(0) = x, \tag{0.3}$$

has a well defined solution. The mapping $k(\,\cdot\,)$ is called *feedback.* Controls are called also *strategies* or *inputs*, and the corresponding solutions of (0.2) or (0.3) are *outputs* of the system.

One of the main aims of control theory is to find a strategy such that the corresponding output has desired properties. Depending on the properties involved one gets more specific questions.

Controllability. One says that a state $z \in \mathbb{R}^n$ is *reachable* from x in time T, if there exists an open loop control $u(\cdot)$ such that, for the output $y(\cdot)$, $y(0) = x$, $y(T) = z$. If an arbitrary state z is reachable from an arbitrary state x in a time T, then the system (0.1) is said to be *controllable.* In several situations one requires a weaker property of transferring an arbitrary state into a given one, in particular into the origin. A formulation of effective characterizations of controllable systems is an important task of control theory only partially solved.

Stabilizability. An equally important issue is that of stabilizability. Assume that for some $\bar{x} \in \mathbb{R}^n$ and $\bar{u} \in U$, $f(\bar{x}, \bar{u}) = 0$. A function $k : \mathbb{R}^n \longrightarrow U$, such that $k(\bar{x}) = \bar{u}$, is called a *stabilizing feedback* if $\bar{x}$ is a stable equilibrium for the system

$$\dot{y}(t) = f(y(t), k(y(t))), \quad t \geq 0, \ y(0) = x. \tag{0.4}$$

In the theory of differential equations there exist several methods to determine whether a given equilibrium state is a stable one. The question of whether, in the class of all equations of the form (0.4), there exists one for which $\bar{x}$ is a stable equilibrium is of a new qualitative type.

Observability. In many situations of practical interest one observes not the state $y(t)$ but its function $h(y(t))$, $t \geq 0$. It is therefore often necessary to investigate the pair of equations

$$\dot{y} = f(y, u), \quad y(0) = x, \tag{0.5}$$

$$w = h(y). \tag{0.6}$$

Relation (0.6) is called an *observation equation.* The system (0.5)–(0.6) is said to be *observable* if, knowing a control $u(\cdot)$ and an observation $w(\cdot)$, on a given interval $[0, T]$, one can determine uniquely the initial condition x.

Optimality. Besides the above problems of structural character, in control theory, with at least the same intensity, one asks optimality questions. In the so-called time-optimal problem one is looking for a control which not only transfers a state x onto z but does it in the minimal time T. In other situations the time $T > 0$ is fixed and one is looking for a control $u(\cdot)$ which minimizes the integral

$$\int_0^T g(y(t), u(t))\, dt + G(y(T)),$$

in which g and G are given functions.

We present now some examples to show that the models and problems discussed in control theory have an immediate real meaning.

Example 0.1 *Electrically heated oven.* Let us consider a simple model of an electrically heated oven, which consists of a jacket with a coil directly heating the jacket and of an interior part. Let T_0 denote the outside temperature. We make a simplifying assumption, that at an arbitrary moment $t \geq 0$, temperatures in the jacket and in the interior part are uniformly distributed and equal to $T_1(t), T_2(t)$. We assume also that the flow of heat through a surface is proportional to the area of the surface and to the difference of temperature between the separated media. Let $u(t)$ be the intensity of the heat input produced by the coil at moment $t \geq 0$. Let moreover a_1, a_2 denote the area of exterior and interior surfaces of the jacket, c_1, c_2 denote heat capacities of the jacket and the interior of the oven and r_1, r_2 denote radiation coefficients of the exterior and interior surfaces of the jacket. An increase of heat in the jacket is equal to the amount of heat produced by the coil reduced by the amount of heat which entered the interior and exterior of the oven. Therefore, for the interval $[t, t + \Delta t]$, we have the following balance:

$$c_1(T_1(t+\Delta t) - T_1(t)) \approx u(t)\Delta t - (T_1(t) - T_2(t))a_1 r_1 \Delta t - (T_1(t) - T_0)a_2 r_2 \Delta t.$$

Similarly, an increase of heat in the interior of the oven is equal to the amount of heat radiated by the jacket:

$$c_2(T_2(t + \Delta t) - T_2(t)) = (T_1(t) - T_2(t))a_1 r_2 \Delta t.$$

Dividing the obtained identities by Δt and taking the limit, as $\Delta t \downarrow 0$, we obtain

$$\begin{aligned} c_1 \frac{dT_1}{dt} &= u - (T_1 - T_2)a_1 r_1 - (T_1 - T_0)a_2 r_2, \\ c_2 \frac{dT_2}{dt} &= (T_1 - T_2)a_1 r_1. \end{aligned}$$

Let us remark that, according to the physical interpretation, $u(t) \geq 0$ for $t \geq 0$. Introducing new variables $x_1 = T_1 - T_0$ and $x_2 = T_2 - T_0$, we have

$$\frac{d}{dt}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -\dfrac{r_1 a_1 + r_2 a_2}{c_1} & \dfrac{r_1 a_1}{c_1} \\ \dfrac{r_1 a_1}{c_2} & -\dfrac{r_1 a_1}{c_2} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} c_1^{-1} \\ 0 \end{bmatrix} u.$$

It is natural to limit the considerations to the case when $x_1(0) \geq 0$ and $x_2(0) \geq 0$. It is physically obvious that if $u(t) \geq 0$ for $t \geq 0$, then also $x_1(t) \geq 0,\ x_2(t) \geq 0,\ t \geq 0$. One can prove this mathematically.

Let us assume that we want to obtain, in the interior part of the oven, a temperature T and keep it at this level infinitely long. Is this possible? Does the answer depend on initial temperatures $T_1 \geq T_0$, $T_2 \geq T_0$?

Example 0.2 *Soft landing.* Let us consider a spacecraft of total mass M moving vertically with the gas thruster directed toward the landing surface. Let h be the height of the spacecraft above the surface, u the thrust of its engine produced by the expulsion of gas from the jet. The gas is a product of the combustion of the fuel. The combustion decreases the total mass of the spacecraft, and the thrust u is proportional to the speed with which the mass decreases. Assuming that there is no atmosphere above the surface and that g is gravitational acceleration, one arrives at the following equations [13]:

$$M\ddot{h} = -gM + u, \tag{0.7}$$

$$\dot{M} = -ku, \tag{0.8}$$

with the initial conditions $M(0) = M_0$, $h(0) = h_0$, $\dot{h}(0) = h_1$; k a positive constant. One imposes additional constraints on the control parameter of the type $0 \leq u \leq \alpha$ and $M \geq m$, where m is the mass of the spacecraft without fuel. Let us fix $T > 0$. The soft landing problem consists of finding a control $u(\cdot)$ such that for the solutions $M(\cdot)$, $h(\cdot)$ of equation (0.7)

$$M(t) \geq m, \quad h(t) \geq 0, \quad t \in [0,T], \quad \text{and} \quad h(T) = \dot{h}(T) = 0.$$

The problem of the existence of such a control is equivalent to the controllability of the system (0.7)–(0.9).

A natural optimization question arises when the moment T is not fixed and one is minimizing the landing time. The latter problem can be formulated equivalently as the *minimum fuel problem.* In fact, let $v = \dot{h}$ denote the velocity of the spacecraft, and let $M(t) > 0$ for $t \in [0,T]$. Then

$$\frac{\dot{M}(t)}{M(t)} = -k\dot{v}(t) - gk, \quad t \in [0,T].$$

Therefore, after integration,

$$M(T) = e^{-v(T)k - gkT + v(0)k} M(0).$$

Thus a soft landing is taking place at a moment $T > 0$ ($v(T) = 0$) if and only if

$$M(T) = e^{-gkT} e^{v(0)k} M(0).$$

Consequently, the minimization of the landing time T is equivalent to the minimization of the amount of fuel $M(0) - M(T)$ needed for landing.

Example 0.3 *Optimal consumption.* The capital $y(t) \geq 0$ of an economy at any moment t is divided into two parts: $u(t)y(t)$ and $(1-u(t))y(t)$, where $u(t)$ is a number from the interval $[0,1]$. The first part goes for investments and contributes to the increase in capital according to the formula

$$\dot{y} = uy, \quad y(0) = x > 0.$$

The remaining part is for consumption evaluated by the *satisfaction*

$$J_T(x, u(\cdot)) = \int_0^T ((1 - u(t))y(t))^\alpha \, dt + ay^\alpha(T). \tag{0.9}$$

In definition (0.9), the number a is nonnegative and $\alpha \in (0,1)$. In the described situation one is trying to divide the capital to maximize the satisfaction.

Remark For more information about the electrically heated oven we refer to [2], [24]. The soft landing and optimal consumption models are extensively discussed in [14].

1 Controllability and Observability

1.1 Preliminaries

As we have already mentioned the basic object of classical control theory is a linear system described by a differential equation

$$\frac{dy}{dt} = Ay(t) + Bu(t), \quad y(0) = x \in \mathbb{R}^n, \tag{1.1}$$

and an observation relation

$$w(t) = Cy(t), \quad t \geq 0. \tag{1.2}$$

For completeness of the presentation we recall first basic concepts and notation related to linear differential equations.

Linear transformations $A : \mathbb{R}^n \longrightarrow \mathbb{R}^n$, $B : \mathbb{R}^m \longrightarrow \mathbb{R}^n$, $C : \mathbb{R}^m \longrightarrow \mathbb{R}^k$ in (1.1) and (1.2) will be identified with representing matrices and elements of $\mathbb{R}^n$, $\mathbb{R}^m$, $\mathbb{R}^k$ with one column matrices. The set of all matrices with n rows and m columns will be denoted by $\mathbf{M}(n,m)$ and the identity transformation as well as the identity matrix by I. The scalar product $\langle x,y\rangle$ and the norm $|x|$, of elements $x, y \in \mathbb{R}^n$ with coordinates $\xi_1,\ldots,\xi_n$ and $\eta_1,\ldots,\eta_n$, are defined by

$$\langle x,y\rangle = \sum_{j=1}^{n} \xi_j \eta_j, \quad |x| = \left(\sum_{j=1}^{n} \xi_j^2\right)^{1/2}.$$

The adjoint transformation of a linear transformation A as well as the transpose matrix of A are denoted by A^*. A matrix $A \in \mathbf{M}(n,n)$ is called *symmetric* if $A = A^*$. The set of all symmetric matrices is partially ordered by the relation $A_1 \geq A_2$ if $\langle A_1 x, x\rangle \geq \langle A_2 x, x\rangle$ for arbitrary $x \in \mathbb{R}^n$. If $A \geq 0$ then one says that matrix A is *nonnegative definite* and if, in addition, $\langle Ax, x\rangle > 0$ for $x \neq 0$ that A is *positive definite*. Treating $x \in \mathbb{R}^n$ as an element of $\mathbf{M}(n,1)$ we have $x^* \in \mathbf{M}(1,n)$. In particular we can write $\langle x,y\rangle = x^* y$ and $|x|^2 = x^* x$. The inverse transformation of A and the inverse matrix of A will be denoted by A^{-1}.

If $F(t) = [f_{ij}(t);\ i = 1,\ldots,n,\ j = 1,\ldots,m] \in \mathbf{M}(n,m)$, $t \in [0,T]$, then, by definition,

$$\int_0^T F(t)\,dt = \left[\int_0^T f_{ij}(t)\,dt;\ i = 1,\ldots,n,\ j = 1,\ldots,m\right], \tag{1.3}$$

under the condition that elements of $F(\,\cdot\,)$ are integrable.

Derivatives of the 1st and 2nd order of a function $y(t)$, $t \in \mathbb{R}$, are denoted by $\frac{dy}{dt}$, $\frac{d^2y}{dt^2}$ or by $\dot{y}$, $\ddot{y}$ and the nth order derivative, by $\frac{d^{(n)}y}{dt^{(n)}}$.

We will need some basic results on linear equations

$$\frac{dq}{dt} = A(t)q(t) + a(t), \quad q(t_0) = q_0 \in \mathbb{R}^n, \tag{1.4}$$

on a fixed interval $[0,T]$; $t_0 \in [0,T]$, where $A(t) \in \mathbf{M}(n,n)$, $A(t) = [a_{ij}(t);\ i = 1,\ldots,n,\ j = 1,\ldots,m]$, $a(t) \in \mathbb{R}^n$, $a(t) = (a_i(t);\ i = 1,\ldots,n)$, $t \in [0,T]$.

Theorem 1.1 *Assume that elements of the function $A(\,\cdot\,)$ are locally integrable. Then there exists exactly one function $S(t)$, $t \in [0,T]$ with values in $\mathbf{M}(n,n)$ and with absolutely continuous elements such that*

$$\frac{d}{dt}S(t) = A(t)S(t) \quad \textit{for almost all } t \in [0,T], \tag{1.5}$$

$$S(0) = I. \tag{1.6}$$

In addition, a matrix $S(t)$ is invertible for an arbitrary $t \in [0,T]$, and the unique solution of the equation (1.4) *is of the form*

$$q(t) = S(t)S^{-1}(t_0)q_0 + \int_{t_0}^{t} S(t)S^{-1}(s)a(s)\,ds, \quad t \in [0,T]. \tag{1.7}$$

Here is a sketch a proof of the theorem.

Proof. Equation (1.4) is equivalent to the integral equation

$$q(t) = a_0 + \int_{t_0}^{t} A(s)q(s)\,ds + \int_{t_0}^{t} a(s)\,ds, \quad t \in [0,T].$$

The formula

$$\mathcal{L}y(t) = a_0 + \int_{t_0}^{t} a(s)\,ds + \int_{t_0}^{t} A(s)y(s)\,ds, \quad t \in [0,T],$$

defines a continuous transformation from the space of continuous functions $C[0,T;\,\mathbb{R}^n]$ into itself, such that for arbitrary $y(\cdot), \tilde{y}(\cdot) \in C[0,T;\,\mathbb{R}^n]$

$$\sup_{t\in[0,T]} |\mathcal{L}y(t) - \mathcal{L}\tilde{y}(t)| \le \left(\int_0^T |A(s)|\,ds\right) \sup_{t\in[0,T]} |y(t) - \tilde{y}(t)|.$$

If $\int_0^T |A(s)|\,ds < 1$, then by the contraction mapping principle the equation $q = \mathcal{L}q$ has exactly one solution in $C[0,T;\,\mathbb{R}^n]$ which is the solution of the integral equation. The case $\int_0^T |A(s)|\,ds \ge 1$ can be reduced to the previous one by considering the equation on appropriately shorter intervals. In particular we obtain the existence and uniqueness of a matrix valued function satisfying (1.5) and (1.6).

To prove the second part of the theorem let us denote by $\psi(t)$, $t \in [0,T]$, the matrix solution of

$$\frac{d}{dt}\psi(t) = -\psi(t)A(t), \quad \psi(0) = I,\ t \in [0,T].$$

Assume that, for some $t \in [0,T]$, $\det S(t) = 0$. Let $T_0 = \min\{t \in [0,T];\ \det S(t) = 0\}$. Then $T_0 > 0$, and for $t \in [0,T_0)$

$$0 = \frac{d}{dt}\left(S(t)S^{-1}(t)\right) = \left(\frac{d}{dt}S(t)\right)S^{-1}(t) + S(t)\frac{d}{dt}S^{-1}(t).$$

Thus

$$-A(t) = S(t)\frac{d}{dt}S^{-1}(t),$$

and consequently

$$\frac{d}{dt}S^{-1}(t) = -S^{-1}(t)A(t), \quad t \in [0, T_0),$$

so $S^{-1}(t) = \psi(t),\ t \in [0, T_0)$.

The function $\det \psi(t)$, $t \in [0, T]$, is continuous and

$$\det \psi(t) = \frac{1}{\det S(t)}, \quad t \in [0, T_0),$$

therefore there exists a finite $\lim_{t \uparrow T_0} \det \psi(t)$. This way $\det S(T_0) = \lim_{t \uparrow T_0} S(t) \neq 0$, a contradiction. The validity of (1.6) follows now by elementary calculation. □

The function $S(t)$, $t \in [0, T]$ will be called *the fundamental solution* of equation (1.4). It follows from the proof that the fundamental solution of the "adjoint" equation

$$\frac{dp}{dt} = -A^*(t)p(t), \quad t \in [0, T],$$

is $(S^*(t))^{-1}$, $t \in [0, T]$.

Exercise 1.1 Show that for $A \in \mathbf{M}(n, n)$ the series

$$\sum_{n=1}^{+\infty} \frac{A^n}{n!} t^n, \quad t \in \mathbb{R},$$

is uniformly convergent, with all derivatives, on an arbitrary finite interval.

The sum of the series from Exercise 1.1 is often denoted by $\exp(tA)$ or e^{tA}, $t \in \mathbb{R}$. We check easily that

$$e^{tA}e^{sA} = e^{(t+s)A}, \quad t, s \in \mathbb{R},$$

in particular

$$(e^{tA})^{-1} = e^{-tA}, \quad t \in \mathbb{R}.$$

Therefore the solution of (1.1) has the form

$$\begin{aligned} y(t) &= e^{tA}x + \textstyle\int_0^t e^{(t-s)A}Bu(s)\,ds) \\ &= S(t)x + \textstyle\int_0^t S(t-s)Bu(s)\,ds, \quad t \in [0,T], \end{aligned} \tag{1.8}$$

where $S(t) = \exp tA,\ t \geq 0$.

The majority of the concepts and results discussed for systems (1.1)–(1.2) can be extended to time dependent matrices $A(t) \in \mathbf{M}(n,n)$, $B(t) \in \mathbf{M}(n,n)$, $C(t) \in \mathbf{M}(k,n)$, $t \in [0,T]$, and therefore for systems

$$\frac{dy}{dt} = A(t)y(t) + B(t)u(t), \quad y(0) = x \in \mathbb{R}^n, \tag{1.9}$$

$$w(t) = C(t)y(t), \quad t \in [0,T]. \tag{1.10}$$

1.2 The controllability matrix

An arbitrary function $u(\cdot)$ defined on $[0,+\infty)$ locally integrable and with values in $\mathbb{R}^m$ will be called a *control, strategy* or *input* of the system (1.1)–(1.2). The corresponding solution of equation (1.1) will be denoted by $y^{x,u}(\cdot)$, to underline the dependence on the initial condition x and the input $u(\cdot)$. Relationship (1.2) can be written in the following way:

$$w(t) = Cy^{x,u}(t), \quad t \in [0,T].$$

The function $w(\cdot)$ is the *output* of the controlled system.

We will assume now that $C = I$ or equivalently that $w(t) = y^{x,u}(t)$, $t \geq 0$.

We say that a control u *transfers* a state a to a state b at the time $T > 0$ if

$$y^{a,u}(T) = b. \tag{1.11}$$

We then also say that the state a can be *steered* to b at time T or that the state b is *reachable* or *attainable* from a at time T.

The proposition below gives a formula for a control transferring a to b. In this formula the matrix Q_T, called the *controllability matrix* or *controllability Gramian*, appears:

$$Q_T = \int_0^T S(r)BB^*S^*(r)\,dr, \quad T > 0.$$

We check easily that Q_T is symmetric and nonnegative definite.

Proposition 1.1 *Assume that for some $T > 0$ the matrix Q_T is nonsingular. Then*

(i) *for arbitrary $a, b \in \mathbb{R}^n$ the control*

$$\hat{u}(s) = -B^* S^*(T-s) Q_T^{-1}(S(T)a - b), \quad s \in [0, T], \tag{1.12}$$

transfers a to b at time T;

(ii) *among all controls $u(\cdot)$ steering a to b at time T the control $\hat{u}$ minimizes the integral $\int_0^T |u(s)|^2\, ds$. Moreover,*

$$\int_0^T |\hat{u}(s)|^2\, ds = \langle Q_T^{-1}(S(T)a - b),\ S(T)a - b\rangle. \tag{1.13}$$

Proof. It follows from (1.12) that the control $\hat{u}$ is smooth or even analytic. From (1.8) and (1.12) we obtain that

$$\begin{aligned} y^{a,\hat{u}}(T) &= S(T)a - \left(\int_0^T S(T-s)BB^*S^*(T-s)\, ds\right)(Q_T^{-1}(S(T)a - b)) \\ &= S(T)a - Q_T(Q_T^{-1}(S(T)a - b)) = b. \end{aligned}$$

This shows (i).

To prove (ii) let us remark that the formula (1.13) is a consequence of the following simple calculations:

$$\begin{aligned} \int_0^T |\hat{u}(s)|^2\, ds &= \int_0^T |B^*S^*(T-s)Q_T^{-1}(S(T)a - b)|^2\, ds \\ &= \left\langle \int_0^T S(T-s)BB^*S^*(T-s)(Q_T^{-1}(S(T)a - b))\, ds, \right. \\ &\qquad\qquad\qquad\qquad \left. Q_T^{-1}(S(T)a - b) \right\rangle \\ &= \langle Q_T Q_T^{-1}(S(T)a - b), Q_T^{-1}(S(T)a - b)\rangle \\ &= \langle Q_T^{-1}(S(T)a - b), S(T)a - b\rangle. \end{aligned}$$

Now let $u(\cdot)$ be an arbitrary control transferring a to b at time T. We can assume that $u(\cdot)$ is square integrable on $[0, T]$. Then

$$\begin{aligned} \int_0^T \langle u(s), \hat{u}(s)\rangle\, ds &= -\int_0^T \langle u(s), B^*S^*(T-s)Q_T^{-1}(S(T)b - a)\rangle\, ds \\ &= -\left\langle \int_0^T S(T-s)Bu(s)\, ds, Q_T^{-1}(S(T)a - b) \right\rangle \\ &= \langle S(T)a - b, Q_T^{-1}(S(T)a - b)\rangle. \end{aligned}$$

Hence

$$\int_0^T \langle u(s), \hat{u}(s)\rangle\, ds = \int_0^T \langle \hat{u}(s), \hat{u}(s)\rangle\, ds.$$

From this we obtain that

$$\int_0^T |u(s)|^2\, ds = \int_0^T |\hat{u}(s)|^2\, ds + \int_0^T |u(s) - \hat{u}(s)|^2\, ds$$

and consequently the desired minimality property. □

Exercise 1.2 Write equation

$$\frac{d^2y}{dt^2} = u, \quad y(0) = \xi_1, \quad \frac{dy}{dt}(0) = \xi_2, \quad \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} \in \mathbb{R}^2,$$

as a first order system. Prove that for the new system, the matrix Q_T is nonsingular, $T > 0$. Find the control u transferring the state $\begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix}$ to $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ at time $T > 0$ and minimizing the functional $\int_0^T |u(s)|^2\, ds$. Determine the minimal value m of the functional. Consider $\xi_1 = 1$, $\xi_2 = 0$.

Answer. The required control is of the form

$$\hat{u}(s) = -\frac{12}{T^3}\left(\frac{\xi_1 T}{2} + \frac{\xi_2 T^2}{3} - \frac{sT\xi_2}{2} - s\xi_1\right), \quad s \in [0, T],$$

and the minimal value m of the functional is equal to

$$m = \frac{12}{T^3}\left((\xi_1)^2 + \xi_1\xi_2 T - \frac{2T^2}{3}(\xi_2)^2\right).$$

In particular, when $\xi_1 = 1$, $\xi_2 = 0$,

$$\hat{u}(s) = \frac{12}{T^3}\left(s - \frac{T}{2}\right), \quad s \in [0, T],\ m = \frac{12}{T^3}.$$

We say that a state b is *attainable* or *reachable* from $a \in \mathbb{R}^n$ if it is attainable or reachable at some time $T > 0$.

System (1.1) is called *controllable* if an arbitrary state $b \in \mathbb{R}^n$ is attainable from any state $a \in \mathbb{R}^n$ at some time $T > 0$. Instead of saying that system (1.1) is controllable we will frequently say that the *pair* (A, B) is *controllable.*

If for arbitrary $a, b \in \mathbb{R}^n$ the attainability takes place at a given time $T > 0$, we say that the system is *controllable at time* T. Proposition 1.1 gives a sufficient condition for the system (1.1) to be controllable. It turns out that this condition is also a necessary one.

The following result holds.

Proposition 1.2 *If an arbitrary state $b \in \mathbb{R}^n$ is attainable from* 0*, then the matrix Q_T is nonsingular for an arbitrary $T > 0$.*

Proof. Let, for a control u and $T > 0$,

$$\mathcal{L}_T u = \int_0^T S(r) B u(T-r)\, dr. \tag{1.14}$$

The formula (1.14) defines a linear operator from $U_T = L^1[0, T; \mathbb{R}^m]$ into $\mathbb{R}^n$. Let us remark that

$$\mathcal{L}_T u = y^{0,u}(T). \tag{1.15}$$

Let $E_T = \mathcal{L}_T(U_T)$, $T > 0$. It follows from (1.14) that the family of the linear spaces E_T is nondecreasing in $T > 0$. Since $\bigcup_{T>0} E_T = \mathbb{R}^n$, taking into account the dimensions of E_T, we have that $E_{\widetilde{T}} = \mathbb{R}^n$ for some $\widetilde{T}$. Let us remark that, for arbitrary $T > 0$, $v \in \mathbb{R}^n$ and $u \in U_T$,

$$\langle Q_T v, v\rangle = \left\langle \left(\int_0^T S(r) B B^* S^*(r)\, dr \right) v, v \right\rangle = \int_0^T |B^* S^*(r) v|^2\, dr, \tag{1.16}$$

$$\langle \mathcal{L}_T u, v \rangle = \int_0^T \langle u(r), B^* S^*(T-r) v\rangle\, dr. \tag{1.17}$$

From identities (1.16) and (1.17) we obtain $Q_T v = 0$ for some $v \in \mathbb{R}^n$ if the space E_T is orthogonal to v or if the function $B^* S^*(\,\cdot\,) v$ is identically equal to zero on $[0, T]$. It follows from the analyticity of this function that it is equal to zero everywhere. Therefore if $Q_T v = 0$ for some $T > 0$ then $Q_T v = 0$ for all $T > 0$ and in particular $Q_{\widetilde{T}} v = 0$. Since $E_{\widetilde{T}} = \mathbb{R}^n$ we have that $v = 0$, and the nonsingularity of Q_T follows. □

A sufficient condition for controllability is that the rank of B is equal to n. This follows from the next exercise.

Exercise 1.3 Assume rank $B = n$ and let B^+ be a matrix such that $BB^+ = I$. Check that the control

$$u(s) = \frac{1}{T} B^+ e^{(s-T)A}(b - e^{TA} a), \quad s \in [0, T],$$

transfers a to b at time $T \geq 0$.

1.3 Rank condition

We now formulate an algebraic condition equivalent to controllability. For matrices $A \in \mathbf{M}(n,n)$, $B \in \mathbf{M}(n,m)$ denote by $[A|B]$ the matrix $[B, AB, \ldots, A^{n-1}B] \in \mathbf{M}(n, nm)$ which consists of consecutively written columns of matrices $B, AB, \ldots, A^{n-1}B$.

Theorem 1.2 *The following conditions are equivalent.*
(i) *An arbitrary state $b \in \mathbb{R}^n$ is attainable from* 0.
(ii) *System* (1.1) *is controllable.*
(iii) *System* (1.1) *is controllable at a given time $T > 0$.*
(iv) *Matrix Q_T is nonsingular for some $T > 0$.*
(v) *Matrix Q_T is nonsingular for an arbitrary $T > 0$.*
(vi) $\operatorname{rank}[A|B] = n$.

Condition (vi) is called the *Kalman rank condition*, or the rank condition for short.

The proof will use the Cayley-Hamilton theorem. Let us recall that a *characteristic polynomial* $p(\cdot)$ of a matrix $A \in \mathbf{M}(n,n)$ is defined by

$$p(\lambda) = \det(\lambda I - A), \quad \lambda \in \mathbb{C}. \tag{1.18}$$

Let

$$p(\lambda) = \lambda^n + a_1\lambda^{n-1} + \ldots + a_n, \quad \lambda \in \mathbb{C}. \tag{1.19}$$

The Cayley-Hamilton theorem has the following formulation (see [1, 358–359]):

Theorem 1.3 *For arbitrary $A \in \mathbf{M}(n,n)$, with the characteristic polynomial* (1.19),

$$A^n + a_1A^{n-1} + \ldots + a_nI = 0.$$

Symbolically, $p(A) = 0$.

Proof of Theorem 1.2. Equivalences (i)–(v) follow from the proofs of Propositions 1.1 and 1.2 and the identity

$$y^{a,u}(T) = \mathcal{L}_T u + S(T)a.$$

To show the equivalences to condition (vi) it is convenient to introduce a linear mapping l_n from the Cartesian product of n copies $\mathbb{R}^m$ into $\mathbb{R}^n$:

$$l_n(u_0, \ldots, u_{n-1}) = \sum_{j=0}^{n-1} A^j B u_j, \quad u_j \in \mathbb{R}^m, \ j = 0, \ldots, n-1.$$

We prove first the following lemma.

Lemma 1.1 *The transformation $\mathcal{L}_T$, $T > 0$, has the same image as l_n. In particular $\mathcal{L}_T$ is onto if and only if l_n is onto.*

Proof. For arbitrary $v \in \mathbb{R}^n$, $u \in L^1[0,T;\,\mathbb{R}^m]$, $u_j \in \mathbb{R}^m$, $j = 0,\ldots,n-1$:

$$\langle \mathcal{L}_T u, v\rangle = \int_0^T \langle u(s), B^*S^*(T-s)v\rangle\, ds,$$
$$\langle l_n(u_0,\ldots,u_{n-1}), v\rangle = \langle u_0, B^*v\rangle + \ldots + \langle u_{n-1}, B^*(A^*)^{n-1}v\rangle.$$

Suppose that $\langle l_n(u_0,\ldots,u_{n-1}), v\rangle = 0$ for arbitrary $u_0,\ldots,u_{n-1} \in \mathbb{R}^m$. Then $B^*v = 0,\ldots,B^*(A^*)^{n-1}v = 0$. From Theorem 1.3, applied to matrix A^*, it follows that for some constants $c_0,\ldots,c_{n-1}$

$$(A^*)^n = \sum_{k=0}^{n-1} c_k (A^*)^k.$$

Thus, by induction, for arbitrary $l = 0,1,\ldots$ there exist constants $c_{l,0},\ldots,c_{l,n-1}$ such that

$$(A^*)^{n+1} = \sum_{k=0}^{n-1} c_{l,k}(A^*)^k.$$

Therefore $B^*(A^*)^k v = 0$ for $k = 0,1,\ldots$. Taking into account that

$$B^*S^*(t)v = \sum_{k=0}^{+\infty} B^*(A^*)^k v \frac{t^k}{k!}, \quad t \geq 0,$$

we deduce that for arbitrary $T > 0$ and $t \in [0,T]$

$$B^*S^*(t)v = 0,$$

so $\langle \mathcal{L}_T u, v\rangle = 0$ for arbitrary $u \in L^1[0,T;\,\mathbb{R}^m]$.

Assume, conversely, that for arbitrary $u \in L^1[0,T;\,\mathbb{R}^n]$, $\langle \mathcal{L}_T u, v\rangle = 0$. Then $B^*S^*(t)v = 0$ for $t \in [0,T]$. Differentiating the identity

$$\sum_{k=0}^{+\infty} B^*(A^*)^k v \frac{t^k}{k!} = 0, \quad t \in [0,T],$$

$0,1,\ldots,(n-1)$ times and inserting each time $t = 0$, we obtain that $B^*(A^*)^k v = 0$ for $k = 0,1,\ldots,n-1$. And therefore

$$\langle l_n(u_0,\ldots,u_{n-1}), v\rangle = 0 \quad \text{for arbitrary } u_0,\ldots,u_{n-1} \in \mathbb{R}^m.$$

This implies the lemma. □

Assume that the system (1.1) is controllable. Then the transformation $\mathcal{L}_T$ is onto $\mathbb{R}^n$ for arbitrary $T > 0$ and, by the above lemma, the matrix $[A|B]$ has rank n. Conversely, if the rank of $[A|B]$ is n then the mapping l_n is onto $\mathbb{R}^n$ and also, therefore, the transformation $\mathcal{L}_T$ is onto $\mathbb{R}^n$ and the controllability of (1.1) follows. □

If the rank condition is satisfied then the control $\hat{u}(\cdot)$ given by (1.12) transfers a to b at time T. We now give a different, more explicit, formula for the transfer control involving the matrix $[A|B]$ instead of the controllability matrix Q_T.

Note that if $\mathrm{rank}[A|B] = n$ then there exists a matrix $K \in \mathbf{M}(mn, n)$ such that $[A|B]K = I \in \mathbf{M}(n, n)$ or equivalently there exist matrices K_1, $K_2, \dots, K_n \in \mathbf{M}(m, n)$ such that

$$BK_1 + ABK_2 + \ldots + A^{n-1}BK_n = I. \tag{1.20}$$

Let, in addition, φ be a function of class C^{n-1} from $[0, T]$ into R such that

$$\frac{d^j\varphi}{ds^j}(0) = \frac{d^j\varphi}{ds^j}(T) = 0, \quad j = 0, 1, \dots, n-1, \tag{1.21}$$

$$\int_0^T \varphi(s)ds = 1. \tag{1.22}$$

Proposition 1.3 *Assume that* $\mathrm{rank}[A|B] = n$ *and* (1.20)–(1.22) *hold. Then the control*

$$\tilde{u}(s) = K_1\psi(s) + K_2\frac{d\psi}{ds}(s) + \ldots + K_n\frac{d^{n-1}\psi}{ds^{n-1}}(s), \quad s \in [0, T]$$

where

$$\psi(s) = S(s - T)(b - S(T)a)\varphi(s), \quad s \in [0, T] \tag{1.23}$$

transfers a *to* b *at time* $T \geq 0$.

Proof. Taking into account (1.21) and integrating by parts $(j - 1)$ times, we have

$$\begin{aligned}\int_0^T S(T-s)BK_j\frac{d^{j-1}}{ds^{j-1}}\psi(s)\,ds &= \int_0^T e^{A(T-s)}BK_j\frac{d^{j-1}}{ds^{j-1}}\psi(s)\,ds\\ &= \int_0^T e^{A(T-s)}A^{j-1}BK_j\psi(s)\,ds\\ &= \int_0^T S(T-s)A^{j-1}BK_j\psi(s)\,ds,\\ &\quad j=1,2,\ldots,n.\end{aligned}$$

Consequently

$$\begin{aligned}\int_0^T S(T-s)B\tilde{u}(s)ds &= \int_0^T S(t-s)[A|B]K\psi(s)ds\\ &= \int_0^T S(T-s)\psi(s)ds.\end{aligned}$$

By the definition of ψ and by (1.22) we finally have

$$\begin{aligned}y^{a,\tilde{u}}(T) &= S(T)a + \int_0^T S(T-s)(S(s-T)(b-S(T)a))\varphi(s)ds\\ &= S(T)a + (b-S(T)a)\int_0^T \varphi(s)ds = b.\end{aligned}$$

□

Remark Note that Proposition 1.3 is a generalization of Exercise 1.3.

Exercise 1.4 Assuming that $U = \mathbb{R}$ prove that the system describing the electrically heated oven from Example 0.1 is controllable.

Exercise 1.5 Let L_0 be a linear subspace dense in $L^1[0,T;\mathbb{R}^m]$. If system (1.1) is controllable then for arbitrary $a,b\in\mathbb{R}^n$ there exists $u(\cdot)\in L_0$ transferring a to b at time T.

Hint. Use the fact that the image of the closure of a set under a linear continuous mapping is contained in the closure of the image of the set.

Exercise 1.6 If system (1.1) is controllable then for arbitrary $T>0$ and arbitrary $a,b\in\mathbb{R}^n$ there exists a control $u(\cdot)$ of class C^∞ transferring a to b at time T and such that

$$\frac{d^{(j)}u}{dt^{(j)}}(0) = \frac{d^{(j)}u}{dt^{(j)}}(T) = 0 \quad \text{for } j=0,1,\ldots.$$

Exercise 1.7 Assuming that the pair (A, B) is controllable, show that the system

$$\begin{aligned}\dot{y} &= Ay + Bv\\ \dot{v} &= u,\end{aligned}$$

with the state space $\mathbb{R}^{n+m}$ and the set of control parameters $\mathbb{R}^m$, is also controllable. Deduce that for arbitrary $a, b \in \mathbb{R}^n$, $u_0, u_1 \in \mathbb{R}^m$ and $T > 0$ there exists a control $u(\,\cdot\,)$ of class C^∞ transferring a to b at time T and such that $u(0) = u_0$, $u(T) = u_1$.

Hint. Use Exercise 1.6 and the Kalman rank condition.

Exercise 1.8 Suppose that $A \in \mathbf{M}(n, n)$, $B \in \mathbf{M}(n, m)$. Prove that the system

$$\frac{d^2y}{dt^2} = Ay + Bu, \quad y(0) \in \mathbb{R}^n, \ \frac{dy}{dt}(0) \in \mathbb{R}^n,$$

is controllable in $\mathbb{R}^{2n}$ if and only if the pair (A, B) is controllable.

Exercise 1.9 Consider system (1.9) on $[0, T]$ with integrable matrix-valued functions $A(t)$, $B(t)$, $t \in [0, T]$. Let $S(t)$, $t \in [0, T]$ be the fundamental solution of the equation $\dot{q} = Aq$. Assume that the matrix

$$Q_T = \int_0^T S(T)S^{-1}(s)B(s)B^*(s)(S^{-1}(s))^*S^*(T)\,ds$$

is positive definite. Show that the control

$$\hat{u}(s) = B^*(S^{-1}(s))^*S^*(T)Q_T^{-1}(b - S(T)a), \quad s \in [0, T],$$

transfers a to b at time T minimizing the functional $u \longrightarrow \int_0^T |u(s)|^2\,ds$.

1.4 A classification of control systems

Let $y(t)$, $t \geq 0$, be a solution of the equation (1.1) corresponding to a control $u(t)$, $t \geq 0$, and let $P \in \mathbf{M}(n, n)$ and $S \in \mathbf{M}(m, m)$ be nonsingular matrices. Define

$$\tilde{y}(t) = Py(t), \quad \tilde{u}(t) = Su(t), \quad t \geq 0.$$

Then

$$\begin{aligned}\frac{d}{dt}\tilde{y}(t) &= P\frac{d}{dt}y(t) = PAy(t) + PBu(t)\\ &= PAP^{-1}\tilde{y}(t) + PBS^{-1}\tilde{u}(t)\\ &= \widetilde{A}\tilde{y}(t) + \widetilde{B}\tilde{u}(t), \quad t \geq 0,\end{aligned}$$

where

$$\widetilde{A} = PAP^{-1}, \quad \widetilde{B} = PBS^{-1}. \tag{1.24}$$

The control systems described by (A, B) and $(\widetilde{A}, \widetilde{B})$ are called *equivalent* if there exist nonsingular matrices $P \in \mathbf{M}(n,n)$, $S \in \mathbf{M}(m,m)$, such that (1.24) holds. Let us remark that P^{-1} and S^{-1} can be regarded as transition matrices from old to new bases in $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively. The introduced concept is an equivalence relation. It is clear that a pair (A, B) is controllable if and only if $(\widetilde{A}, \widetilde{B})$ is controllable.

We now give a complete description of equivalent classes of the introduced relation in the case when $m = 1$.

Let us first consider a system

$$\frac{d^{(n)}}{dt^{(n)}}z + a_1\frac{d^{(n-1)}}{dt^{(n-1)}}z + \ldots + a_n z = u, \tag{1.25}$$

with initial conditions

$$z(0) = \xi_1, \quad \frac{dz}{dt}(0) = \xi_2, \quad \ldots, \quad \frac{d^{(n-1)}z}{dt^{(n-1)}}(0) = \xi_n. \tag{1.26}$$

Let $z(t)$, $\frac{dz}{dt}(t), \ldots, \frac{d^{(n-1)}z}{dt^{(n-1)}}(t)$, $t \geq 0$, be coordinates of a function $y(t)$, $t \geq 0$, and $\xi_1, \ldots, \xi_n$ coordinates of a vector x. Then

$$\dot{y} = \widetilde{A}y + \widetilde{B}u, \quad y(0) = x \in \mathbb{R}^n, \tag{1.27}$$

where matrices $\widetilde{A}$ and $\widetilde{B}$ are of the form

$$\widetilde{A} = \begin{bmatrix} 0 & 1 & \ldots & 0 & 0\\ 0 & 0 & \ldots & 0 & 0\\ \vdots & \vdots & \ddots & \vdots & \vdots\\ 0 & 0 & \ldots & 0 & 1\\ -a_n & -a_{n-1} & \ldots & -a_2 & -a_1\end{bmatrix}, \quad \widetilde{B} = \begin{bmatrix} 0\\ \vdots\\ 0\\ 1\end{bmatrix}. \tag{1.28}$$

We easily check that on the main diagonal of the matrix $[\widetilde{A}|\widetilde{B}]$ there are only ones and above the diagonal only zeros. Therefore $\mathrm{rank}[\widetilde{A}|\widetilde{B}] = n$ and, by Theorem 1.2, the pair $(\widetilde{A}, \widetilde{B})$ is controllable. Interpreting this result in terms of the initial system (1.21)–(1.22) we can say that for two arbitrary sequences of n numbers $\xi_1, \ldots, \xi_n$ and $\eta_1, \ldots, \eta_n$ and for an arbitrary positive number T there exists an analytic function $u(t)$, $t \in [0, T]$, such that for the corresponding solution $z(t)$, $t \in [0, T]$, of the equation (1.25)–(1.26)

$$z(T) = \eta_1, \quad \frac{dz}{dt}(T) = \eta_2, \quad \ldots, \quad \frac{d^{(n-1)}z}{dt^{(n-1)}}(T) = \eta_n.$$

Theorem 1.4 states that an arbitrary controllable system with the one dimensional space of control parameters is equivalent to a system of the form (1.25)–(1.26).

Theorem 1.4 *If $A \in \mathbf{M}(n, n)$, $b \in \mathbf{M}(n, 1)$ and the system*

$$\dot{y} = Ay + bu, \;\; y(0) = x \in \mathbb{R}^n \tag{1.29}$$

is controllable then it is equivalent to exactly one system of the form (1.28). *Moreover the numbers $a_1, \ldots, a_n$ in the representation* (1.24) *are identical to the coefficients of the characteristic polynomial of the matrix A*:

$$p(\lambda) = \det[\lambda I - A] = \lambda^n + a_1\lambda^{n-1} + \ldots + a_n, \quad \lambda \in \mathbb{C}. \tag{1.30}$$

Proof. By the Cayley-Hamilton theorem, $A^n + a_1 A^{n-1} + \ldots + a_n I = 0$. In particular

$$A^n b = -a_1 A^{n-1} b - \ldots - a_n b.$$

Since $\operatorname{rank}[A|b] = n$, therefore vectors $e_1 = A^{n-1}b, \ldots, e_n = b$ are linearly independent and form a basis in $\mathbb{R}^n$. Let $\xi_1(t), \ldots, \xi_n(t)$ be coordinates of the vector $y(t)$ in this basis, $t \geq 0$. Then

$$\frac{d\xi}{dt} = \begin{bmatrix} -a_1 & 1 & 0 & \ldots & 0 & 0 \\ -a_2 & 0 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ -a_{n-1} & 0 & 0 & \ldots & 0 & 1 \\ -a_n & 0 & 0 & \ldots & 0 & 0 \end{bmatrix} \xi + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} u. \tag{1.31}$$

Therefore an arbitrary controllable system (1.29) is equivalent to (1.31) and the numbers $a_1, \ldots, a_n$ are the coefficients of the characteristic polynomial of A. On the other hand, direct calculation of the determinant of $[\lambda I - \widetilde{A}]$ gives

$$\det(\lambda I - \widetilde{A}) = \lambda^n + a_1\lambda^{n-1} + \ldots + a_n = p(\lambda), \quad \lambda \in \mathbb{C}.$$

Therefore the pair $(\widetilde{A}, \widetilde{B})$ is equivalent to the system (1.31) and consequently also to the pair (A, b). □

Remark The problem of an exact description of the equivalence classes in the case of arbitrary m is much more complicated; see [27] and [29].

1.5 Kalman decomposition

Theorem 1.2 gives several characterizations of controllable systems. Here we deal with uncontrollable ones.

Theorem 1.5 *Assume that*

$$\text{rank}[A|B] = l < n.$$

There exists a nonsingular matrix $P \in \mathbf{M}(n,n)$ such that

$$PAP^{-1} = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}, \quad PB = \begin{bmatrix} B_1 \\ 0 \end{bmatrix},$$

where $A_{11} \in \mathbf{M}(l,l)$, $A_{22} \in \mathbf{M}(n-l, n-l)$, $B_1 \in \mathbf{M}(l,m)$. In addition the pair

$$(A_{11}, B_1)$$

is controllable.

The theorem states that there exists a basis in $\mathbb{R}^n$ such that system (1.1) written with respect to that basis has a representation

$$\begin{aligned} \dot{\xi}_1 &= A_{11}\xi_1 + A_{12}\xi_2 + B_1 u, \quad && \xi_1(0) \in \mathbb{R}^l, \\ \dot{\xi}_2 &= A_{22}\xi_2, && \xi_2(0) \in \mathbb{R}^{n-l}, \end{aligned}$$

in which (A_{11}, B_1) is a controllable pair. The first equation describes the so-called *controllable part* and the second the *completely uncontrollable* part of the system.

Proof. It follows from Lemma 1.1 that the subspace $E_0 = \mathcal{L}_T(L^1[0,T;\mathbb{R}^m])$ is identical with the image of the transformation l_n. Therefore it consists of all elements of the form $Bu_1 + ABu_1 + \ldots + A^{n-1}Bu_n$, $u_1, \ldots, u_n \in \mathbb{R}^m$ and is of dimension l. In addition it contains the image of B and by the Cayley-Hamilton theorem, it is invariant with respect to the transformation A. Let E_1 be any linear subspace of $\mathbb{R}^n$ complementing E_0 and let $e_1, \ldots, e_l$

and $e_{l+1}, \ldots, e_n$ be bases in E_0 and E_1 and P the transition matrix from the new to the old basis. Let $\widetilde{A} = PAP^{-1}$, $\widetilde{B} = PB$,

$$\widetilde{A}\begin{bmatrix}\xi_1\\ \xi_2\end{bmatrix} = \begin{bmatrix}A_{11}\xi_1 + A_{12}\xi_2\\ A_{21}\xi_1 + A_{22}\xi_2\end{bmatrix}, \quad \widetilde{B} = \begin{bmatrix}B_1 u\\ B_2 u\end{bmatrix},$$

$\xi_1 \in \mathbb{R}^l$, $\xi_2 \in \mathbb{R}^{n-l}$, $u \in \mathbb{R}^m$. Since the space E_0 is invariant with respect to A, therefore

$$\widetilde{A}\begin{bmatrix}\xi_1\\ 0\end{bmatrix} = \begin{bmatrix}A_{11}\xi_1\\ 0\end{bmatrix}, \quad \xi_1 \in \mathbb{R}^l.$$

Taking into account that $B(\mathbb{R}^m) \subset E_0$,

$$B_2 u = 0 \quad \text{for } u \in \mathbb{R}^m.$$

Consequently the elements of the matrices A_{22} and B_2 are zero. This finishes the proof of the first part of the theorem.

To prove the final part, let us remark that for the nonsingular matrix P

$$\operatorname{rank}[A|B] = \operatorname{rank}(P[A|B]) = \operatorname{rank}[\widetilde{A}|\widetilde{B}].$$

Since

$$[\widetilde{A}|\widetilde{B}] = \begin{bmatrix}B_1 & A_{11}B_1 & \ldots & A_{11}^{n-1}B_1\\ 0 & 0 & \ldots & 0\end{bmatrix},$$

so

$$l = \operatorname{rank}[\widetilde{A}|\widetilde{B}] = \operatorname{rank}[A_{11}|B_1].$$

Taking into account that $A_{11} \in \mathbf{M}(l,l)$, one gets the required property. □

Remark Note that the subspace E_0 consists of all points attainable from 0. It follows from the proof of Theorem 1.5 that E_0 is the smallest subspace of R^n invariant with respect to A and containing the image of B, and it is identical to the image of the transformation represented by $[A|B]$.

Exercise 1.10 Give a complete classification of controllable systems when $m = 1$ and the dimension of E_0 is $l < n$.

1.6 Observability

Assume that $B = 0$. Then system (1.1) is identical with the linear equation

$$\dot{z} = Az, \quad z(0) = x. \tag{1.32}$$

The observation relation (1.2) we leave unchanged:

$$w = Cz. \tag{1.33}$$

The solution to (1.32) will be denoted by $z^x(t)$, $t \geq 0$. Obviously

$$z^x(t) = S(t)x, \quad x \in \mathbb{R}^n.$$

The system (1.32)- (1.33), or the pair (A, C), is said to be *observable* if for arbitrary $x \in \mathbb{R}^n$, $x \neq 0$, there exists a $t > 0$ such that

$$w(t) = Cz^x(t) \neq 0.$$

If for a given $T > 0$ and for arbitrary $x \neq 0$ there exists $t \in [0, T]$ with the above property, then the system (1.32)- (1.33) or the pair (A, C) are said to be *observable at time* T. Let us introduce the so-called *observability matrix*:

$$R_T = \int_0^T S^*(r)C^*CS(r)\, dr.$$

The following theorem, dual to Theorem 1.2, holds.

Theorem 1.6 *The following conditions are equivalent.*

(i) *System (1.32)-(1.33) is observable.*
(ii) *System (1.32)-(1.33) is observable at a given time* $T > 0$.
(iii) *The matrix* R_T *is nonsingular for some* $T > 0$.
(iv) *The matrix* R_T *is nonsingular for arbitrary* $T > 0$.
(v) $\operatorname{rank}[A^*|C^*] = n$.

Proof. Analysis of the function $w(\cdot)$ implies the equivalence of (i) and (ii). Besides,

$$\begin{aligned}
\int_0^T |w(r)|^2\, dr &= \int_0^T |Cz^x(r)|^2\, dr \\
&= \int_0^T \langle S^*(r)C^*CS(r)x, x\rangle\, dr \\
&= \langle R_T x, x\rangle.
\end{aligned}$$

Therefore observability at time $T \geq 0$ is equivalent to $\langle R_T x, x\rangle \neq 0$ for arbitrary $x \neq 0$ and consequently to nonsingularity of the nonnegative, symmetric matrix R_T. The remaining equivalences are consequences of Theorem 1.2

and the observation that the controllability matrix corresponding to (A^*, C^*) is exactly R_T. □

Example 1.1. Let us consider the equation

$$\frac{d^{(n)}z}{dt^{(n)}} + a_1 \frac{d^{(n-1)}z}{dt^{(n-1)}} + \ldots + a_n z = 0, \tag{1.34}$$

and assume that

$$w(t) = z(t), \quad t \geq 0. \tag{1.35}$$

Matrices A and C corresponding to (1.34)-(1.35) are of the form

$$A = \begin{bmatrix} 0 & 1 & \ldots & 0 & 0 \\ 0 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 0 & 1 \\ -a_n & -a_{n-1} & \ldots & -a_2 & -a_1 \end{bmatrix}, \quad C = [1, 0, \ldots, 0].$$

We check directly that rank $[A^*|C^*] = n$ and thus the pair (A, C) is observable.

The next theorem is analogous to Theorem 1.5 and gives a decomposition of system (1.32)-(1.33) into observable and completely unobservable parts.

Theorem 1.7. *Assume that rank* $[A^*|C^*] = l < n$. *Then there exists a nonsingular matrix* $P \in \mathbf{M}(n, n)$ *such that*

$$PAP^{-1} = \begin{bmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{bmatrix}, \quad CP^{-1} = [C_1, 0],$$

where $A_{11} \in \mathbf{M}(l, l)$, $A_{22} \in \mathbf{M}(n - l, n - l)$ *and* $C_1 \in \mathbf{M}(k, l)$ *and the pair* (A_{11}, C_1) *is observable.*

Proof. The theorem follows directly from Theorem 1.5 and from the observation that a pair (A, C) is observable if and only if the pair (A^*, C^*) is controllable. □

Remark. It follows from the above theorem that there exists a basis in $\mathbb{R}^n$ such that the system (1.1)-(1.2) has representation

$$\begin{aligned} \dot{\xi}_1 &= A_{11}\xi_1 + B_1 u, \\ \dot{\xi}_2 &= A_{21}\xi_1 + A_{22}\xi_2 + B_2 u, \\ \eta &= C_1 \xi_1, \end{aligned}$$

and the pair $(A_{11}, C_1$ is observable.

Remark Basic concepts of the chapter are due to R. Kalman [16]. He is also the author of Theorems 1.2, 1.5 and 1.6. Exercise 1.3 as well as Proposition 1.3 are due to R. Triggiani [26].

2 Stability and stabilizability

2.1 Stable linear systems

In this chapter stable linear systems are characterized in terms of associated characteristic polynomials. A formulation of the Routh theorem on stable polynomials is given as well as a complete description of completely stabilizable systems.

Let $A \in \mathbf{M}(n,n)$ and consider linear systems

$$\dot{z} = Az, \quad z(0) = x \in \mathbb{R}^n. \tag{2.1}$$

Solutions of equation (2.1) will be denoted by $z^x(t)$, $t \geq 0$. In accordance with earlier notations we have that

$$z^x(t) = S(t)x = (\exp tA)x, \quad t \geq 0.$$

The system (2.1) is called *stable* if for arbitrary $x \in \mathbb{R}^n$

$$z^x(t) \longrightarrow 0, \quad \text{as } t \uparrow +\infty.$$

Instead of saying that (2.1) is stable we will often say that the matrix A is stable. Let us remark that the concept of stability does not depend on the choice of the basis in $\mathbb{R}^n$. Therefore if P is a nonsingular matrix and A is a stable one, then matrix PAP^{-1} is stable.

In what follows we will need the Jordan theorem [31] on canonical representation of matrices. Denote by $\mathbf{M}(n,m;\mathbb{C})$ the set of all matrices with n rows and m columns and with complex elements. Let us recall that a number $\lambda \in \mathbb{C}$ is called an *eigenvalue* of a matrix $A \in \mathbf{M}(n,n;\mathbb{C})$ if there exists a vector $a \in \mathbb{C}^n$, $a \neq 0$, such that $Aa = \lambda a$. The set of all eigenvalues of a matrix A will be denoted by $\sigma(A)$. Since $\lambda \in \sigma(A)$ if and only if the matrix $\lambda I - A$ is singular, therefore $\lambda \in \sigma(A)$ if and only if $p(\lambda) = 0$, where p is a *characteristic polynomial* of $A: p(\lambda) = \det[\lambda I - A]$, $\lambda \in \mathbb{C}$. The set $\sigma(A)$ consists of at most n elements and is nonempty.

Theorem 2.1 *For an arbitrary matrix $A \in \mathbf{M}(n,n;\mathbb{C})$ there exists a nonsingular matrix $P \in \mathbf{M}(n,n;\mathbb{C})$ such that*

$$PAP^{-1} = \begin{bmatrix} J_1 & 0 & \dots & 0 & 0 \\ 0 & J_2 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & J_{r-1} & 0 \\ 0 & 0 & \dots & 0 & J_r \end{bmatrix} = \widetilde{A}, \tag{2.2}$$

where $J_1, J_2, \dots, J_r$ are the so-called Jordan blocks

$$J_k = \begin{bmatrix} \lambda_k & \gamma_k & \dots & 0 & 0 \\ 0 & \lambda_k & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \lambda_k & \gamma_k \\ 0 & 0 & \dots & 0 & \lambda_k \end{bmatrix}, \quad \gamma_k \neq 0 \ \text{or} \ J_k = [\lambda_k], \ k = 1, \dots, r.$$

In the representation (2.2) at least one Jordan block corresponds to an eigenvalue $\lambda_k \in \sigma(A)$. Selecting matrix P properly one can obtain a representation with numbers $\gamma_k \neq 0$ given in advance.

For matrices with real elements the representation theorem has the following form:

Theorem 2.2 *For an arbitrary matrix $A \in \mathbf{M}(n,n)$ there exists a nonsingular matrix $P \in \mathbf{M}(n,n)$ such that (2.2) holds with "real" blocks I_k. Blocks I_k, $k = 1, \dots, r$, corresponding to real eigenvalues $\lambda_k = \alpha_k \in \mathbb{R}$ are of the form*

$$[\alpha_k] \quad \text{or} \quad \begin{bmatrix} \alpha_k & \gamma_k & \dots & 0 & 0 \\ 0 & \alpha_k & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \alpha_k & \gamma_k \\ 0 & 0 & \dots & 0 & \alpha_k \end{bmatrix}, \quad \gamma_k \neq 0, \ \gamma_k \in \mathbb{R},$$

and corresponding to complex eigenvalues $\lambda_k = \alpha_k + i\beta_k$, $\beta_k \neq 0$, $\alpha_k, \beta_k \in \mathbb{R}$,

$$\begin{bmatrix} K_k & L_k & \dots & 0 & 0 \\ 0 & K_k & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & K_k & L_k \\ 0 & 0 & \dots & 0 & K_k \end{bmatrix} \text{ where } K_k = \begin{bmatrix} \alpha_k & \beta_k \\ -\beta_k & \alpha_k \end{bmatrix},\ L_k = \begin{bmatrix} \gamma_k & 0 \\ 0 & \gamma_k \end{bmatrix},$$

compare [2].

We now prove the following theorem.

Theorem 2.3 *Assume that* $A \in \mathbf{M}(n,n)$. *The following conditions are equivalent:*

(i) $z^x(t) \longrightarrow 0$ *as* $t \uparrow +\infty$, *for arbitrary* $x \in \mathbb{R}^n$.
(ii) $z^x(t) \longrightarrow 0$ *exponentially as* $t \uparrow +\infty$, *for arbitrary* $x \in \mathbb{R}^n$.
(iii) $\omega(A) = \sup\{\mathrm{Re}\,\lambda;\ \lambda \in \sigma(A)\} < 0$.
(iv) $\int_0^{+\infty} |z^x(t)|^2\, dt < +\infty$ *for arbitrary* $x \in \mathbb{R}^n$.

For the proof we will need the following lemma.

Lemma 2.1 *Let* $\omega > \omega(A)$. *For arbitrary norm* $\|\cdot\|$ *on* $\mathbb{R}^n$ *there exist constants* M *such that*

$$\|z^x(t)\| \le M e^{\omega t}\|x\| \quad \text{for } t \ge 0 \text{ and } x \in \mathbb{R}^n.$$

Proof. Let us consider equation (2.1) with the matrix A in the Jordan form (2.2)

$$\dot{x} = \widetilde{A}w, \quad w(0) = x \in \mathbb{C}^n.$$

For $a = a_1 + ia_2$, where $a_1, a_2 \in \mathbb{R}^n$ set $\|a\| = \|a_1\| + \|a_2\|$. Let us decompose vector $w(t)$, $t \ge 0$ and the initial state x into sequences of vectors $w_1(t), \dots, w_r(t)$, $t > 0$ and $x_1, \dots, x_r$ according to the decomposition (2.2). Then

$$\dot{w}_k = J_k w_k, \quad w_k(0) = x_k,\ k = 1, \dots, r.$$

Let $j_1, \dots, j_r$ denote the dimensions of the matrices $J_1, \dots, J_r$, $j_1 + j_2 + \dots + j_r = n$.

If $j_k = 1$ then

$$w_k(t) = e^{\lambda_k t} x_k, \quad t \ge 0.$$

So $\|w_k(t)\| = e^{(\mathrm{Re}\,\lambda_k)t}\|x_k\|$, $t \ge 0$.

If $j_k > 1$, then

$$w_k(t) = e^{\lambda_k t} \sum_{l=0}^{j_k-1} \begin{bmatrix} 0 & \gamma_k & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & \gamma_k \\ 0 & 0 & \dots & 0 & 0 \end{bmatrix}^l x_k \frac{t^l}{l!}.$$

So

$$\|w_k(t)\| \le e^{(\mathrm{Re}\,\lambda_k)t} \|x_k\| \sum_{l=0}^{j_k-1} (M_k)^l \frac{t^l}{l!}, \quad t \ge 0,$$

where M_k is the norm of the transformation represented by

$$\begin{bmatrix} 0 & \gamma_k & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & \gamma_k \\ 0 & 0 & \dots & 0 & 0 \end{bmatrix}.$$

Setting $\omega_0 = \omega(A)$ we get

$$\sum_{k=1}^{r} \|w_k(t)\| \le e^{\omega_0 t} q(t) \sum_{k=1}^{r} \|x_k\|, \quad t \ge 0,$$

where q is a polynomial of order at most $\max(j_k - 1)$, $k = 1, \dots, r$. If $\omega > \omega_0$ and

$$M_0 = \sup \left\{ q(t) e^{(\omega_0 - \omega)t}, \quad t \ge 0 \right\},$$

then $M_0 < +\infty$ and

$$\sum_{k=1}^{r} \|w_k(t)\| \le M_0 e^{\omega t} \sum_{k=1}^{r} \|x_k\|, \quad t \ge 0.$$

Therefore for a new constant M_1

$$\|w(t)\| \le M_1 e^{\omega t} \|x\|, \quad t \ge 0.$$

Finally

$$\|z^x(t)\| = \|P w(t) P^{-1}\| \le M_1 e^{\omega t} \|P\| \, \|P^{-1}\| \, \|x\|, \quad t \ge 0,$$

and this is enough to define $M = M_1 \|P\| \, \|P^{-1}\|$. □

Proof of the theorem. Assume $\omega_0 \geq 0$. There exist $\lambda = \alpha + i\beta$, $\operatorname{Re}\lambda = \alpha \geq 0$ and a vector $a \neq 0$, $a = a_1 + ia_2$, $a_1, a_2 \in \mathbb{R}^n$ such that

$$A(a_1 + ia_2) = (\alpha + i\beta)(a_1 + ia_2).$$

The function

$$z(t) = z_1(t) + iz_2(t) = e^{(\alpha+i\beta)t}a, \quad t \geq 0,$$

as well as its real and imaginary parts, is a solution of (2.1). Since $a \neq 0$, either $a_1 \neq 0$ or $a_2 \neq 0$. Let us assume, for instance, that $a_1 \neq 0$ and $\beta \neq 0$. Then

$$z_1(t) = e^{\alpha t}(\cos\beta t)a_1 - (\sin\beta t)a_2, \quad t \geq 0.$$

Inserting $t = 2\pi k/\beta$, we have

$$|z_1(t)| = e^{\alpha t}|a_1|$$

and, taking $k \uparrow +\infty$, we obtain $z_1(t) \not\to 0$.

Now let $\omega_0 < 0$ and $\alpha \in (0, -\omega_0)$. Then by the lemma

$$|z^x(t)| \leq Me^{-\alpha t}|x| \quad \text{for } t \geq 0 \text{ and } x \in \mathbb{R}^n.$$

This implies (ii) and therefore also (i).

It remains to consider (iv). It is clear that it follows from (ii) and thus also from (iii). Let us assume that condition (iv) holds and $\omega_0 \geq 0$. Then $|z_1(t)| = e^{\alpha t}|a_1|$, $t \geq 0$, and therefore

$$\int_0^{+\infty} |z_1(t)|^2 \, dt = +\infty,$$

a contradiction. The proof is complete. □

Exercise 2.1 The matrix

$$A = \begin{bmatrix} 0 & 1 \\ -2 & -2 \end{bmatrix}$$

corresponds to the equation $\ddot{z} + 2\dot{z} + 2z = 0$. Calculate $\omega(A)$. For $\omega > \omega(A)$ find the smallest constant $M = M(\omega)$ such that

$$|S(t)| \leq Me^{\omega t}, \quad t \geq 0.$$

Hint. Prove that $|S(t)| = \varphi(t)e^{-t}$, where

$$\varphi(t) = \frac{1}{2}\left(2 + 5\sin^2 t + (20\sin^2 t + 25\sin^4 t)^{1/2}\right)^{1/2}, \quad t \geq 0.$$

2.2 Stable polynomials

Theorem 2.3 reduces the problem of determining whether a matrix A is stable to the question of finding out whether all roots of the characteristic polynomial of A have negative real parts. Polynomials with this property will be called *stable*. Because of its importance, several efforts have been made to find necessary and sufficient conditions for the stability of an arbitrary polynomial

$$p(\lambda) = \lambda^n + a_1\lambda^{n-1} + \ldots + a_n, \quad \lambda \in \mathbb{C}, \tag{2.3}$$

with real coefficients, in term of the coefficients $a_1, \ldots, a_n$. Since there is no general formula for roots of polynomials of order greater than 4, the existence of such conditions is not obvious. Therefore their formulation in the nineteenth century by Routh was a kind of a sensation. Before formulating and proving a version of the Routh theorem we will characterize stable polynomials of degree smaller than or equal to 4 using only the fundamental theorem of algebra. We deduce also a useful necessary condition for stability.

Theorem 2.4
(1) *Polynomials with real coefficients*:

(i) $\lambda + a$,
(ii) $\lambda^2 + a\lambda + b$,
(iii) $\lambda^3 + a\lambda^2 + b\lambda + c$,
(iv) $\lambda^4 + a\lambda^3 + b\lambda^2 + c\lambda + d$

are stable if and only if, respectively

(i)* $a > 0$,
(ii)* $a > 0,\ b > 0$,
(iii)* $a > 0,\ b > 0,\ c > 0$ *and* $ab > c$,
(iv)* $a > 0,\ b > 0,\ c > 0,\ d > 0$ *and* $abc > c^2 + a^2d$.

(2) *If polynomial* (2.3) *is stable then all its coefficients* $a_1, \ldots, a_n$ *are positive.*

Proof. (1) Equivalence (i)$\Longleftrightarrow$(i)* is obvious.

To prove (ii)$\Longleftrightarrow$(ii)* assume that the roots of the polynomial are of the form $\lambda_1 = -\alpha + i\beta$, $\lambda_2 = -\alpha - i\beta$, $\beta \neq 0$. Then $p(\lambda) = \lambda^2 + 2\alpha\lambda + \beta^2$, $\lambda \in \mathbb{C}$ and therefore the stability conditions are $a > 0$ and $b > 0$. If the roots λ_1, λ_2 of the polynomial p are real then $a = -(\lambda_1 + \lambda_2)$, $b = \lambda_1\lambda_2$. Therefore they are negative if only if $a > 0$, $b > 0$.

To show that (iii)$\Longleftrightarrow$(iii)* let us remark that the fundamental theorem of algebra implies the following decomposition of the polynomial, with real coefficients α, β, γ:

$$p(\lambda) = \lambda^3 + a\lambda^2 + b\lambda + c = (\lambda + \alpha)(\lambda^2 + \beta\lambda + \gamma), \quad \lambda \in \mathbb{C}.$$

It therefore follows from (i) and (ii) that the polynomial p is stable if only if $\alpha > 0$, $\beta > 0$ and $\gamma > 0$. Comparing the coefficients gives

$$a = \alpha + \beta, \quad b = \gamma + \alpha\beta, \quad c = \alpha\gamma,$$

and therefore $ab - c = \beta(\alpha^2 + \gamma + \alpha\beta) = \beta(\alpha^2 + b)$.

Assume that $a > 0$, $b > 0$, $c > 0$ and $ab - c > 0$. It follows from $b > 0$ and $ab - c > 0$ that $\beta > 0$. Since $c = \alpha\gamma$, α and γ are either positive or negative. They cannot, however, be negative because then $b = \gamma + \alpha\beta < 0$. Thus $\alpha > 0$ and $\gamma > 0$ and consequently $\alpha > 0$, $\beta > 0$, $\gamma > 0$. It is clear from the above formulae that the positivity of α, β, γ implies inequalities (iii)*. To prove (iv)$\Longleftrightarrow$(iv)* we again apply the fundamental theorem of algebra to obtain the representation

$$\lambda^4 + a\lambda^3 + b\lambda^2 + c\lambda + d = (\lambda^2 + \alpha\lambda + \beta)(\lambda^2 + \gamma\lambda + \delta)$$

and the stability condition $\alpha > 0$, $\beta > 0$, $\gamma > 0$, $\delta > 0$.

From the decomposition

$$a = \alpha + \gamma, \quad b = \alpha\gamma + \beta + \delta, \quad c = \alpha\delta + \beta\gamma, \quad d = \beta\delta,$$

we check directly that

$$abc - c^2 - a^2 d = \alpha\gamma\left((\beta - \delta)^2 + ac\right).$$

It is therefore clear that $\alpha > 0$, $\beta > 0$, $\gamma > 0$ and $\delta > 0$, and then (iv)* holds. Assume now that the inequalities (iv)* are true. Then $\alpha\gamma > 0$, and, since $a = \alpha + \gamma > 0$, therefore $\alpha > 0$ and $\delta > 0$. Since, in addition, $d = \beta\delta > 0$ and $c = \alpha\delta + \beta\gamma > 0$, so $\beta > 0$, $\delta > 0$. Finally $\alpha > 0$, $\beta > 0$, $\gamma > 0$, $\delta > 0$, and the polynomial p is stable.

(2) By the fundamental theorem of algebra, the polynomial p is a product of polynomials of degrees at most 2 which, by (1), have positive coefficients. This implies the result. □

Exercise 2.2 Find necessary and sufficient conditions for the polynomial

$$\lambda^2 + a\lambda + b$$

with complex coefficients a and b to have all roots with negative real parts.

Hint. Consider the polynomial $(\lambda^2 + a\lambda + b)(\lambda^2 + \bar{a}\lambda + \bar{b})$ and apply Theorem 2.4.

We now formulate a theorem which allows us to check, in a finite number of steps, that a given polynomial $p(\lambda) = \lambda^n + a_1\lambda^{n-1} + \ldots + a_n$, $\lambda \in \mathbb{C}$, with real coefficients is stable. As we already know, a stable polynomial has all coefficients positive, but this condition is not sufficient for stability if $n > 3$. Let U and V be polynomials with real coefficients given by

$$U(x) + iV(x) = p(ix), \quad x \in \mathbb{R}.$$

Let us remark that $\deg U = n$, $\deg V = n - 1$ if n is an even number and $\deg U = n - 1$, $\deg V = n$, if n is an odd number. Denote $f_1 = U$, $f_2 = V$ if $\deg U = n$, $\deg V = n - 1$ and $f_1 = V$, $f_2 = U$ if $\deg V = n$, $\deg U = n - 1$. Let $f_3, f_4, \ldots, f_m$ be polynomials obtained from f_1, f_2 by an application of the Euclid algorithm. Thus $\deg f_{k+1} < \deg f_k$, $k = 2, \ldots, m - 1$ and there exist polynomials $\kappa_1, \ldots, \kappa_m$ such that

$$f_{k-1} = \kappa_k f_k - f_{k+1}, \quad f_{m-1} = \kappa_m f_m.$$

Moreover the polynomial f_m is equal to the largest common divisor of f_1, f_2 multiplied by a constant.

The following theorem is due to F. J. Routh [23]. For the proof, see [31].

Theorem 2.5 *A polynomial p is stable if and only if $m = n + 1$ and the signs of the leading coefficients of the polynomials $f_1, \ldots, f_{n+1}$ alternate.*

Let us apply the above theorem to polynomials of degree 4,

$$p(\lambda) = \lambda^4 + a\lambda^3 + b\lambda^2 + c\lambda + d, \quad \lambda \in \mathbb{C}.$$

In this case

$$\begin{aligned} U(x) &= x^4 - bx^2 + d = f_1(x), \\ V(x) &= -ax^3 + cx = f_2(x), \quad x \in \mathbb{R}. \end{aligned}$$

Performing appropriate divisions we obtain

$$\begin{aligned} f_3(x) &= \left(b - \frac{c}{a}\right) x^2 - d, \\ f_4(x) &= -\left(c - ad\left(b - \frac{c}{a}\right)^{-1}\right) x, \\ f_5(x) &= d. \end{aligned}$$

The leading coefficients of the polynomials $f_1, f_2, \dots, f_5$ are

$$1,\ -a,\ \left(b - \frac{c}{a}\right),\ -\left(c - ad\left(b - \frac{c}{a}\right)^{-1}\right),\ d.$$

We obtain therefore the following necessary and sufficient conditions for the stability of the polynomial p:

$$a > 0,\ b - \frac{c}{a} > 0,\ c - ad\left(b - \frac{c}{a}\right) > 0,\ d > 0,$$

equivalent to those stated in Theorem 2.4.

We leave as an exercise the proof that the Routh theorem leads to an explicit stability algorithm. To formulate it we have to define the so-called *Routh array.*

For arbitrary sequences (α_k), (β_k), the *Routh sequence* (γ_k) is defined by

$$\gamma_k = -\frac{1}{\beta_1}\det\begin{bmatrix}\alpha_1 & \alpha_{k+1}\\ \beta_1 & \beta_{k+1}\end{bmatrix}, \quad k = 1, 2, \dots$$

If $a_1, \dots, a_n$ are coefficients of a polynomial p, we set additionally $a_k = 0$ for $k > n = \deg p$. The *Routh array* is a matrix with infinite rows obtained from the first two rows:

$$1, a_2, a_4, a_6, \dots,$$
$$a_1, a_3, a_5, a_7, \dots,$$

by consecutive calculations of the Routh sequences from the two preceding rows. The calculations stop if 0 appears in the first column. The Routh algorithm can be now stated as the theorem

Theorem 2.6 *A polynomial p of degree n is stable if and only if the $n+1$ first elements of the first columns of the Routh array are positive.*

Exercise 2.3 Show that, for an arbitrary polynomial $p(\lambda) = \lambda^n + a_1\lambda^{n-1} + \dots + a_n$, $\lambda \in \mathbb{C}$, with complex coefficients $a_1, \dots, a_n$, the polynomial $(\lambda^n + a_1\lambda^{n-1} + \dots + a_n)(\lambda^n + \bar{a}_1\lambda^{n-1} + \dots + \bar{a}_n)$ has real coefficients. Formulate necessary and sufficient conditions for the polynomial p to have all roots with negative real parts.

2.3 Stabilizability and controllability

We say that the system

$$\dot{y} = Ay + Bu, \quad y(0) = x \in \mathbb{R}^n, \tag{2.4}$$

is *stabilizable* or that the pair (A, B) is *stabilizable* if there exists a matrix $K \in \mathbf{M}(m, n)$ such that the matrix $A + BK$ is stable. So if the pair (A, B) is stabilizable and a control $u(\cdot)$ is given in the *feedback* form

$$u(t) = Ky(t), \quad t \geq 0,$$

then all solutions of the equation

$$\dot{y}(t) = Ay(t) + BKy(t) = (A + BK)y(t), \quad y(0) = x, \ t \geq 0, \tag{2.5}$$

tend to zero as $t \uparrow +\infty$.

We say that system (2.4) is *completely stabilizable* if and only if for arbitrary $\omega > 0$ there exist a matrix K and a constant $M > 0$ such that for an arbitrary solution $y^x(t)$, $t \geq 0$, of (2.5)

$$|y^x(t)| \leq Me^{-\omega t}|x|, \quad t \geq 0. \tag{2.6}$$

By p_K we will denote the characteristic polynomial of the matrix $A + BK$. One of the most important results in the linear control theory is given by

Theorem 2.7 *The following conditions are equivalent:*

(i) *System* (2.4) *is completely stabilizable.*
(ii) *System* (2.4) *is controllable.*
(iii) *For arbitrary polynomial* $p(\lambda) = \lambda^n + \alpha_1\lambda^{n-1} + \ldots + \alpha_n$, $\lambda \in \mathbb{C}$, *with real coefficients, there exists a matrix* K *such that*

$$p(\lambda) = p_K(\lambda) \quad \textit{for } \lambda \in \mathbb{C}.$$

Proof. We start with the implication (ii)$\Longrightarrow$(iii) and prove it in three steps.

Step 1. The dimension of the space of control parameters $m = 1$. It follows from §1.4 that we can limit our considerations to systems of the form

$$\frac{d^{(n)}z}{dt^{(n)}}(t) + a_1\frac{d^{(n-1)}z}{dt^{(n-1)}}(t) + \ldots + a_n z(t) = u(t), \quad t \geq 0.$$

In this case, however, (iii) is obvious: It is enough to define the control u in the feedback form,

$$u(t) = (a_1 - \alpha_1)\frac{d^{(n-1)}z}{dt^{(n-1)}}(t) + \ldots + (a_n - \alpha_n)z(t), \quad t \geq 0,$$

and use the result (see §1.4) that the characteristic polynomial of the equation

$$\frac{d^{(n)}z}{dt^{(n)}} + \alpha_1\frac{d^{(n-1)}z}{dt^{(n-1)}} + \ldots + \alpha_n z = 0,$$

or, equivalently, of the matrix

$$\begin{bmatrix} 0 & 1 & \ldots & 0 & 0 \\ 0 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 0 & 1 \\ -\alpha_n & -\alpha_{n-1} & \ldots & -\alpha_2 & -\alpha_1 \end{bmatrix},$$

is exactly

$$p(\lambda) = \lambda^n + \alpha_1\lambda^{n-1} + \ldots + \alpha_n\lambda, \quad \lambda \in \mathbb{C}.$$

Step 2. The following lemma allows us to reduce the general case to $m = 1$. Note that in its formulation and proof its vectors from $\mathbb{R}^n$ are treated as one-column matrices.

Lemma 2.2 *If a pair (A, B) is controllable then there exist a matrix $L \in \mathbf{M}(m, n)$ and a vector $v \in \mathbb{R}^m$ such that the pair $(A+BL, Bv)$ is controllable.*

Proof of the lemma. It follows from the controllability of (A, B) that there exists $v \in \mathbb{R}^m$ such that $Bv \neq 0$. We show first that there exist vectors $u_1, \ldots, u_{n-1}$ in $\mathbb{R}^m$ such that the sequence $e_1, \ldots, e_n$ defined inductively

$$e_1 = Bv, \ e_{l+1} = Ae_l + Bu_l \quad \text{for } l = 1, 2, \ldots, n-1 \tag{2.7}$$

is a basis in $\mathbb{R}^n$. Assume that such a sequence does not exist. Then for some $k \geq 0$ vectors $e_1, \ldots, e_k$, corresponding to some $u_1, \ldots, u_k$ are linearly independent, and for arbitrary $u \in \mathbb{R}^m$ the vector $Ae_k + Bu$ belongs to the linear space E_0 spanned by $e_1, \ldots, e_k$. Taking $u = 0$ we obtain $Ae_k \in E_0$. Thus $Bu \in E_0$ for arbitrary $u \in \mathbb{R}^m$ and consequently $Ae_j \in E_0$ for $j = 1, \ldots, k$. This way we see that the space E_0 is invariant for A and contains the image of B. Controllability of (A, B) implies now that $E_0 = \mathbb{R}^n$, and compare the remark following Theorem 1.5. Consequently $k = n$ and the required sequences $e_1, \ldots, e_n$ and $u_1, \ldots, u_{n-1}$ exist. Let u_n be an arbitrary vector from $\mathbb{R}^m$.

We define the linear transformation L setting $Le_l = u_l$, for $l = 1, \ldots, n$. We have from (2.7)

$$\begin{aligned} e_{l+1} &= Ae_l + BLe_l = (A+BL)e_l \\ &= (A+BL)^l e_1 \\ &= (A+BL)^l Bv, \quad l = 0, 1, \ldots, n-1. \end{aligned}$$

Since

$$[A+BL|Bv] = [e_1, e_2, \ldots, e_n],$$

the pair $(A+BL, Bv)$ is controllable. □

Step 3. Let a polynomial p be given and let L and v be the matrix and vector constructed in Step 2. The system

$$\dot{y} = (A+BL)y + (Bv)u,$$

in which $u(\cdot)$ is a scalar control function, is controllable. It follows from Step 1 that there exists $k \in \mathbb{R}^n$ such that the characteristic polynomial of $(A+BL) + (Bv)k^* = A + B(L + vk^*)$ is identical with p.

The required feedback K can be defined as

$$K = L + vk^*.$$

We proceed to the proofs of the remaining implications. To show that (iii)$\Longrightarrow$(ii) assume that (A,B) is not controllable, that $\operatorname{rank}[A|B] = l < n$ and that K is a linear feedback. Let $P \in \mathbf{M}(n,n)$ be a nonsingular matrix from Theorem 1.5. Then

$$\begin{aligned} p_K(\lambda) &= \det[\lambda I - (A+BK)] \\ &= \det[\lambda I - (PAP^{-1} + PBKP^{-1})] \\ &= \det\begin{bmatrix} (\lambda I - (A_{11} + B_1K_1)) & -A_{12} \\ 0 & (\lambda I - A_{22}) \end{bmatrix} \\ &= \det[\lambda I - (A_{11} + B_1K_1)] \det[\lambda I - A_{22}], \quad \lambda \in \mathbb{C}, \end{aligned}$$

where $K_1 \in \mathbf{M}(m,n)$. Therefore for arbitrary $K \in \mathbf{M}(m,n)$ the polynomial p_K has a nonconstant divisor, equal to the characteristic polynomial of A_{22}, and therefore p_K cannot be arbitrary. This way the implication (iii)$\Longrightarrow$(ii) is true.

Assume now that condition (i) holds but that the system is not controllable. By the above argument we have for arbitrary $K \in \mathbf{M}(m,n)$ that $\sigma(A_{22}) \subset \sigma(A + BK)$. So if for some $M > 0$, $\omega > 0$ condition (2.6) holds then

$$\omega \leq -\sup\{\operatorname{Re}\lambda;\ \lambda \in \sigma(A_{22})\},$$

which contradicts complete stabilizability. Hence (i)$\Longrightarrow$(ii). Assume now that (ii) and therefore (iii) hold. Let $p(\lambda) = \lambda^n + a_1\lambda^{n-1} + \ldots + a_n$, $\lambda \in \mathbb{C}$ be a polynomial with all roots having real parts smaller than $-\omega$ (e.g., $p(\lambda) = (\lambda + \omega + \varepsilon)^n, \varepsilon > 0$). We have from (iii) that there exists a matrix K such that $p_K(\,\cdot\,) = p(\,\cdot\,)$. Consequently all eigenvalues of $A + BK$ have real parts smaller than $-\omega$. By Theorem 2.3, condition (i) holds. The proof of Theorem 2.7 is complete. □

Remark The proof of Theorem 2.7 is due to M. Wonham [28].

3 Linear quadratic problem

3.1 Introductory comments

This chapter starts from a derivation of the dynamic programming equations called Bellman's equations. They are used to solve the linear regulator problem on a finite time interval. A fundamental role is played here by the Riccati-type matrix differential equations. The stabilization problem is reduced to an analysis of an algebraic Riccati equation.

Our considerations will be devoted mainly to control systems

$$\dot{y} = f(y,u), \quad y(0) = x, \tag{3.1}$$

and to *criteria*, called also *cost functionals*,

$$J_T(x, u(\,\cdot\,)) = \int_0^T g(y(t), u(t))\,dt + G(y(T)), \tag{3.2}$$

when $T < +\infty$. If the control interval is $[0, +\infty]$, then the *cost functional*

$$J(x, u(\,\cdot\,)) = \int_0^{+\infty} g(y(t), u(t))\,dt. \tag{3.3}$$

Our aim will be to find a control $\hat{u}(\cdot)$ such that for all admissible controls $u(\cdot)$

$$J_T(x, \hat{u}(\cdot)) \leq J_T(x, u(\cdot)) \tag{3.4}$$

or

$$J(x, \hat{u}(\cdot)) \leq J(x, u(\cdot)). \tag{3.5}$$

There are basically two methods for finding controls minimizing cost functionals (3.2) or (3.3).

One of them *embeds* a given minimization problem into a parametrized family of similar problems. The embedding should be such that the minimal value, as a function of the parameter, satisfies an analytic relation. If the selected parameter is the initial state and the length of the control interval, then the minimal value of the cost functional is called the value function and the analytical relation, Bellman's equation. Knowing the solutions to the Bellman equation one can find the optimal strategy in the form of a closed loop control.

The other method leads to necessary conditions on the optimal, open-loop, strategy formulated in the form of the so-called maximum principle discovered by L. Pontryagin and his collaborators. They can be obtained (in the simplest case) by considering a parametrized family of controls and the corresponding values of the cost functional (3.2) and by an application of classical calculus.

3.2 Bellman's equation and the value function

Assume that the state space E of a control system is an open subset of $\mathbb{R}^n$ and let the set U of control parameters be included in $\mathbb{R}^m$. We assume that the functions f, g and G are continuous on $E \times U$ and E respectively and that g is nonnegative.

Theorem 3.1 *Assume that a real function $W(\cdot,\cdot)$, defined and continuous on $[0,T] \times E$, is of class C^1 on $(0,T) \times E$ and satisfies the equation*

$$\frac{\partial W}{\partial t}(t,x) = \inf_{u \in U}(g(x,u) + \langle W_x(t,x), f(x,u)\rangle), \quad (t,x) \in (0,T) \times E, \tag{3.6}$$

with the boundary condition

$$W(0,x) = G(x), \quad x \in E. \tag{3.7}$$

(i) *If $u(\cdot)$ is a control and $y(\cdot)$ the corresponding absolutely continuous, E-valued, solution of* (3.1), *then*

$$J_T(x, u(\cdot)) \geq W(T, x). \tag{3.8}$$

(ii) *Assume that for a certain function $\hat{v} : [0, T] \times E \longrightarrow U$*

$$g(x, \hat{v}(t, x)) + \langle W_x(t, x), f(x, \hat{v}(t, x))\rangle \\ \leq g(x, u) + \langle W_x(t, x), f(x, u)\rangle, \quad t \in (0, T),\ x \in E,\ u \in U, \tag{3.9}$$

and that $\hat{y}$ is an absolutely continuous, E-valued solution of the equation

$$\begin{aligned} \frac{d}{dt}\hat{y}(t) &= f(\hat{y}(t), \hat{v}(T - t, \hat{y}(t))), \quad t \in [0, T], \\ \hat{y}(0) &= x. \end{aligned} \tag{3.10}$$

Then, for the control $\hat{u}(t) = \hat{v}(T - t, \hat{y}(t))$, $t \in [0, T]$,

$$J_T(x, \hat{u}(\cdot)) = W(x, T).$$

Proof. (i) Let $w(t) = W(T - t, y(t))$, $t \in [0, T]$. Then $w(\cdot)$ is an absolutely continuous function on an arbitrary interval $[\alpha, \beta] \subset (0, T)$ and

$$\begin{aligned} \frac{dw}{dt}(t) &= -\frac{\partial W}{\partial t}(T - t, y(t)) + \left\langle W_x(T - t, y(t)), \frac{dy}{dt}(t)\right\rangle \\ &= -\frac{\partial W}{\partial t}(T - t, y(t)) + \langle W_x(T - t, y(t)), f(y(t), u(t))\rangle \end{aligned} \tag{3.11}$$

for almost all $t \in [0, T]$. Hence, from (3.6) and (3.7)

$$\begin{aligned} W(T - \beta, y(\beta)) - W(T - \alpha, y(\alpha)) &= w(\beta) - w(\alpha) = \int_\alpha^\beta \frac{dw}{dt}(t)\, dt \\ &= \int_\alpha^\beta \left[-\frac{\partial W}{\partial t}(T - t, y(t)) + \langle W_x(T - t, y(t)), f(y(t), u(t))\rangle\right] dt \\ &\geq -\int_\alpha^\beta g(y(t), u(t))\, dt. \end{aligned}$$

Letting α and β tend to 0 and T respectively we obtain

$$G(y(T)) - W(T, x) \geq -\int_0^T g(y(t), u(t))\, dt.$$

This proves (i).

(ii) In a similar way, taking into account (3.9), for the control $\hat{u}$ and the output $\hat{y}$,

$$G(\hat{y}(T)) - W(T,x) = \int_0^T \left[-\frac{\partial W}{\partial t}(T-t,\hat{y}(t)) + \langle W_x(T-t,\hat{y}(t)) \rangle \right] dt$$
$$= \int_0^T g(\hat{y}(t),\hat{u}(t))\, dt.$$

Therefore

$$G(\hat{y}(T)) + \int_0^T g(\hat{y}(s),\hat{u}(s))\, ds = W(T,x),$$

the required identity. □

Remark Equation (3.6) is called *Bellman's equation.* It follows from Theorem 3.1 that, under appropriate conditions, $W(T,x)$ is the minimal value of the functional $J_T(x,\cdot)$. Hence W is the *value function* for the problem of minimizing (3.2).

Let $U(t,x)$ be the set of all control parameters $u \in U$ for which the infimum on the right-hand side of (3.6) is attained. The function $\hat{v}(\cdot,\cdot)$ from part (ii) of the theorem is a *selector* of the multivalued function $U(\cdot,\cdot)$ in the sense that

$$\hat{v}(t,x) \in U(t,x), \quad (t,x) \in [0,T] \times E.$$

Therefore, for the conditions of the theorem to be fulfilled, such a selector not only should exist, but the closed loop equation (3.10) should have a well defined, absolutely continuous, solution.

Remark A similar result holds for a more general cost functional

$$J_T(x,u(\,\cdot\,)) = \int_0^T e^{-\alpha t} g(y(t),u(t))\, dt + \bar{e}^{\alpha T} G(y(T)). \tag{3.12}$$

In this direction we propose to solve the following exercise.

Exercise 3.1 Taking into account a solution $W(\cdot,\cdot)$ of the equation

$$\frac{\partial W}{\partial t}(t,x) = \inf_{u \in U}(g(x,u) - \alpha W(t,x) + \langle W_x(t,x), f(x,u) \rangle),$$
$$W(0,x) = G(x), \quad x \in E,\ t \in (0,T),$$

and a selector $\hat{v}$ of the multivalued function

$$U(t,x) = \Big\{u \in U;\ g(x,u) + \langle W_x(t,x), f(x,u)\rangle = \inf_{u\in U}(g(x,u) + \langle W_x(t,x), f(x,u)\rangle)\Big\},$$

generalize Theorem 3.1 to the functional (3.12).

We will now describe an intuitive derivation of equation (3.6). Similar reasoning often helps to guess the proper form of the Bellman equation in situations different from the one covered by Theorem 3.1.

Let $W(t,x)$ be the minimal value of the functional $J_t(x,\,\cdot\,)$. For arbitrary $h > 0$ and arbitrary parameter $v \in U$ denote by $u^v(\,\cdot\,)$ a control which is constant and equal v on $[0,h)$ and is identical with the optimal strategy for the minimization problem on $[h, t+h]$. Let $z^{x,v}(t)$, $t \geq 0$, be the solution of the equation $\dot{z} = f(z,v)$, $z(0) = x$. Then

$$J_{t+h}(x, u^v(\,\cdot\,)) = \int_0^h g(z^{x,v}(s), v)\, ds + W(t, z^{x,v}(h))$$

and, approximately,

$$W(t+h,x) \approx \inf_{v\in U} J_{t+h}(x, u^v(\,\cdot\,)) \approx \inf_{v\in U} \int_0^h g(z^{x,v}(s), v)\, ds + W(t, z^{x,v}(h)).$$

Subtracting $W(t,x)$ we obtain that

$$\frac{1}{h}(W(t+h,x) - W(t,x)) \approx \inf_{u\in U}\left[\frac{1}{h}\int_0^h g(z^{x,v}(s), v)\, ds + \frac{1}{h}(W(t, z^{x,v}(h)) - W(t,x))\right].$$

Assuming that the function W is differentiable and taking the limits as $h \downarrow 0$ we arrive at (3.6). □

Exercise 3.2 Show that the solution of the Bellman equation corresponding to the optimal consumption model of Example 0.3, with $\alpha \in (0,1)$, is of the form

$$W(t,x) = p(t)x^{\alpha}, \quad t \geq 0,\ x \geq 0,$$

where the function $p(\cdot)$ is the unique solution of the following differential equation:

$$\dot{p} = \begin{cases} 1, & \text{for } p \le 1, \\ \alpha p + (1-\alpha)\left(\frac{1}{p}\right)^{\alpha/(1-\alpha)}, & \text{for } p \ge 1, \end{cases}$$
$$p(0) = a.$$

Find the optimal strategy.

Hint. First prove the following lemma.

Lemma 3.1 *Let $\psi_p(u) = \alpha u p + (1-u)^\alpha$, $p \ge 0$, $u \in [0,1]$. The maximal value $m(p)$ of the function $\psi_p(\cdot)$ is attained at*

$$u(p) = \begin{cases} 0, & \text{if } p > 1, \\ \left(\frac{1}{p}\right)^{1/(1-\alpha)}, & \text{if } p \in [0,1]. \end{cases}$$

Moreover

$$m(p) = \begin{cases} 1, & \text{if } p \ge 1, \\ \alpha p + (1-\alpha)\left(\frac{1}{p}\right)^{\alpha/(1-\alpha)}, & \text{if } p \in [0,1]. \end{cases}$$

3.3 The linear regulator problem and the Riccati equation

We now consider a special case of Problems (3.1) and (3.4) when the system equation is linear

$$\dot{y} = Ay + Bu, \quad y(0) = x \in \mathbb{R}^n, \tag{3.13}$$

$A \in \mathbf{M}(n,n)$, $B \in \mathbf{M}(n,m)$, the state space $E = \mathbb{R}^n$ and the set of control parameters $U = \mathbb{R}^m$. We assume that the cost functional is of the form

$$J_T = \int_0^T \left(\langle Qy(s), y(s)\rangle + \langle Ru(s), u(s)\rangle\right) ds + \langle P_0 y(T), y(T)\rangle, \tag{3.14}$$

where $Q \in \mathbf{M}(n,n)$, $R \in \mathbf{M}(m,m)$, $P_0 \in \mathbf{M}(n,n)$ are symmetric, nonnegative matrices and the matrix R is positive definite. The problem of minimizing (3.14) for a linear system (3.13) is called the *linear regulator problem* or the *linear-quadratic problem*.

The form of an optimal solution to (3.13) and (3.14) is strongly connected with the following *matrix Riccati equation*:

$$\dot{P} = Q + PA + A^*P - PBR^{-1}B^*P, \quad P(0) = P_0, \tag{3.15}$$

in which $P(s)$, $s \in [0,T]$, is the unknown function with values in $\mathbf{M}(n,n)$. The following theorem takes place.

Theorem 3.2 *Equation* (3.15) *has a unique global solution* $P(s)$, $s \geq 0$. *For arbitrary* $s \geq 0$ *the matrix* $P(s)$ *is symmetric and nonnegative definite. The minimal value of the functional* (3.14) *is equal to* $\langle P(T)x, x\rangle$ *and the optimal control is of the form*

$$\hat{u}(t) = -R^{-1}B^*P(T-t)\hat{y}(t), \quad t \in [0,T], \tag{3.16}$$

where

$$\dot{\hat{y}}(t) = (A - BR^{-1}B^*P(T-t))\hat{y}(t), \quad t \in [0,T], \quad \hat{y}(0) = x. \tag{3.17}$$

Proof. The proof will be given in several steps.

Step 1. For an arbitrary symmetric matrix P_0 equation (3.15) has exactly one local solution and the values of the solution are symmetric matrices.

Equation (3.15) is equivalent to a system of n^2 differential equations for elements $p_{ij}(\cdot)$, $i,j = 1,2,\ldots,n$ of the matrix $P(\cdot)$. The right-hand sides of these equations are polynomials of order 2, and therefore the system has a unique local solution being a smooth function of its argument. Let us remark that the same equation is also satisfied by $P^*(\cdot)$. This is because matrices Q, R and P_0 are symmetric. Since the solution is unique, $P(\cdot) = P^*(\cdot)$, and the values of $P(\cdot)$ are symmetric matrices.

Step 2. Let $P(s)$, $s \in [0,T_0)$, be a symmetric solution of (3.15) and let $T < T_0$. The function $W(s,x) = \langle P(s)x, x\rangle$, $s \in [0,T]$, $x \in \mathbb{R}^n$, is a solution of the Bellman equation (3.6)–(3.7) associated with the linear regular problem (3.13)–(3.14).

The condition (3.7) follows directly from the definitions. Moreover, for arbitrary $x \in \mathbb{R}^n$ and $t \in [0,T]$

$$\begin{aligned} &\inf_{u \in \mathbb{R}^n} \left(\langle Qx, x\rangle + \langle Ru, u\rangle + 2\langle P(t)x, Ax + Bu\rangle\right) \\ &= \langle Qx, x\rangle + \langle (A^*P(t) + P(t)A)x, x\rangle + \inf_{u \in \mathbb{R}^m} \left(\langle Ru, u\rangle + \langle u, 2B^*P(t)x\rangle\right). \end{aligned} \tag{3.18}$$

We need now the following lemma, the proof of which is left as an exercise.

Lemma 3.2 *If a matrix $R \in \mathbf{M}(m,m)$ is positive definite and $a \in \mathbb{R}^m$, then for arbitrary $u \in \mathbb{R}^m$*

$$\langle Ru, u\rangle + \langle a, u\rangle \geq -\frac{1}{4}\langle R^{-1}a, a\rangle.$$

Moreover, the equality holds if and only if

$$u = -\frac{1}{2}R^{-1}a.$$

It follows from the lemma that the expression (3.18) is equal to

$$\langle Q + A^*P(t) + P(t)A^* - P(t)BR^{-1}B^*P(A)x, x\rangle$$

and that the infimum in formula (3.18) is attained at exactly one point given by

$$-R^{-1}B^*P(t)x, \quad t \in [0,T].$$

Since $P(t)$, $t \in [0, T_0)$, satisfies the equation (3.15), the function W is a solution to the problem (3.6)–(3.7).

Step 3. The control $\hat{u}$ given by (3.16) on $[0,T]$, $T < T_0$, is optimal with respect to the functional $J_T(x,\cdot)$.

This fact is a direct consequence of Theorem 3.1.

Step 4. For arbitrary $t \in [0,T]$, $T < T_0$, the matrix $P(t)$ is nonnegative definite and

$$\langle P(t)x, x\rangle \leq \int_0^t \langle Q\tilde{y}^x(s), \tilde{y}^x(s)\rangle\, ds + \langle P_0\tilde{y}^x(t), \tilde{y}^x(t)\rangle, \tag{3.19}$$

where $\tilde{y}^x(\,\cdot\,)$ is the solution to the equation

$$\dot{\tilde{y}} = A\tilde{y}, \quad \tilde{y}(0) = x.$$

Applying Theorem 3.1 to the function $J_t(x,\cdot)$ we see that its minimal value is equal to $\langle P(t)x, x\rangle$. For arbitrary control $u(\,\cdot\,)$, $J_t(x,u) \geq 0$, the matrix $P(t)$ is nonnegative definite. In addition, estimate (3.19) holds because its right-hand side is the value of the functional $J_t(x,\cdot)$ for the control $u(s) = 0$, $s \in [0,t]$.

Step 5. For arbitrary $t \in [0, T_0)$ and $x \in \mathbb{R}^n$

$$0 \leq \langle P(t)x, x\rangle \leq \left\langle \left(\int_0^t S^*(r)QS(r)\, dr + S^*(t)P_0S(t)\right)x, x\right\rangle,$$

where $S(r) = e^{Ar}$, $r \geq 0$.

This result is an immediate consequence of the estimate (3.19).

Exercise 3.3 Show that if, for some symmetric matrices $P = (p_{ij}) \in \mathbf{M}(n,n)$ and $S = (s_{ij}) \in \mathbf{M}(n,n)$,

$$0 \le \langle Px, x\rangle \le \langle Sx, x\rangle, \quad x \in \mathbb{R}^n,$$

then

$$-\frac{1}{2}(s_{ii} + s_{jj}) \le p_{ij} \le s_{ij} + \frac{1}{2}(s_{ii} + s_{jj}), \quad i,j = 1,\dots,n.$$

It follows from Step 5 and Exercise 3.3 that solutions of (3.15) are bounded in $\mathbf{M}(n,n)$ and therefore an arbitrary maximal solution $P(\cdot)$ in $\mathbf{M}(n,n)$ exists for all $t \ge 0$.

The proof of the theorem is complete. □

Exercise 3.4 Solve the linear regulator problem with a more general cost functional

$$\int_0^T (\langle Q(y(t) - a), y(t) - a\rangle + \langle Ru(t), u(t)\rangle)\, dt + \langle P_0 y(T), y(T)\rangle,$$

where $a \in \mathbb{R}^n$ is a given vector.

Answer. Let $P(t)$, $q(t)$, $r(t)$, $t \ge 0$, be solutions of the following matrix, vector and scalar equations respectively,

$$\begin{aligned}
\dot{P} &= Q + A^*P + PA - PBR^{-1}B^*P, \quad P(0) = P_0,\\
\dot{q} &= A^*q - PBR^{-1}q - 2Qa, \quad q(0) = 0,\\
\dot{r} &= -\frac{1}{4}\langle R^{-1}q, q\rangle + \langle Qa, a\rangle, \quad r(0) = 0.
\end{aligned}$$

The minimal value of the functional is equal to

$$r(T) + \langle q(T), x\rangle + \langle P(T)x, x\rangle,$$

and the optimal, feedback strategy is of the form

$$u(t) = -\frac{1}{2}R^{-1}q(T-t) - R^{-1}B^*P(T-t)y(t), \quad t \in [0,T].$$

3.4 The linear regulator and stabilization

The obtained solution of the linear regulator problem suggests an important way to stabilize linear systems. It is related to the *algebraic Riccati equation*

$$Q + PA + A^*P - PBR^{-1}B^*P = 0, \quad P \ge 0, \tag{3.20}$$

in which the unknown is a nonnegative definite matrix P. If $\widetilde{P}$ is a solution to (3.20) and $\widetilde{P} \leq P$ for all the other solutions P, then $\widetilde{P}$ is called a *minimal solution* of (3.20). For arbitrary control $u(\cdot)$ defined on $[0, +\infty)$ we introduce the notation

$$J(x, u) = \int_0^{+\infty} (\langle Qy(s), y(s)\rangle + \langle Ru(s), u(s)\rangle) \, ds. \tag{3.21}$$

Theorem 3.3 *If there exists a nonnegative solution P of equation* (3.20) *then there also exists a unique minimal solution $\widetilde{P}$ of* (3.20)*, and the control $\tilde{u}$ given in the feedback form*

$$\tilde{u}(t) = -R^{-1}B^*\widetilde{P}y(t), \quad t \geq 0,$$

minimizes functional (3.21)*. Moreover the minimal value of the cost functional is equal to*

$$\langle \widetilde{P}x, x\rangle.$$

Proof. Let us first remark that if $P_1(t)$, $P_2(t)$, $t \geq 0$, are solutions of (3.15) and $P_1(0) \leq P_2(0)$ then $P_1(t) \leq P_2(t)$ for all $t \geq 0$. This is because the minimal value of the functional

$$J_t^1(x, u) = \int_0^t (\langle Qy(s), y(s)\rangle + \langle Ru(s), u(s)\rangle) \, ds + \langle P_1(0)y(t), y(t)\rangle$$

is not greater than the minimal value of the functional

$$J_t^2(x, u) = \int_0^t (\langle Qy(s), y(s)\rangle + \langle Ru(s), u(s)\rangle) \, ds + \langle P_2(0)y(t), y(t)\rangle,$$

and by Theorem 3.2 the minimal values are $\langle P_1(t)x, x\rangle$ and $\langle P_2(t)x, x\rangle$ respectively.

If, in particular, $P_1(0) = 0$ and $P_2(0) = P$ then $P_2(t) = P$ and therefore $P_1(t) \leq P$ for all $t \geq 0$. It also follows from Theorem 3.2 that the function $P_1(\cdot)$ is nondecreasing with respect to the natural order existing in the space of symmetric matrices. This easily implies that for arbitrary $i, j = 1, 2, \dots, n$ there exist finite limits $\tilde{p}_{ij} = \lim_{t \uparrow +\infty} \tilde{p}_{ij}(t)$, where $(\tilde{p}_{ij}(t)) = P_1(t)$, $t \geq 0$. Taking into account equation (3.15) we see that there exist finite limits

$$\lim_{t \uparrow +\infty} \frac{d}{dt}\tilde{p}_{ij}(t) = \gamma_{ij}, \quad i, j = 1, \dots, n.$$

These limits have to be equal to zero, for if $\gamma_{i,j} > 0$ or $\gamma_{i,j} < 0$ then $\lim_{t\uparrow+\infty} \tilde{p}_{ij}(t) = +\infty$. But $\lim_{t\uparrow+\infty} \tilde{p}_{ij}(t) = -\infty$, a contradiction. Hence the matrix $\widetilde{P} = (\tilde{p}_{ij})$ satisfies equation (3.20). It is clear that $\widetilde{P} \leq P$.

Now let $\tilde{y}(\cdot)$ be the output corresponding to the input $\tilde{u}(\cdot)$. By Theorem 3.2, for arbitrary $T \geq 0$ and $x \in \mathbb{R}^n$,

$$\langle \widetilde{P}x, x\rangle = \int_0^T (\langle Q\tilde{y}(t), \tilde{y}(t)\rangle + \langle R\tilde{u}(t), \tilde{u}(t)\rangle)\, dt + \langle \widetilde{P}\tilde{y}(T), \tilde{y}(T)\rangle, \tag{3.22}$$

and

$$\int_0^T (\langle Q\tilde{y}(t), \tilde{y}(t)\rangle + \langle R\tilde{u}(t), \tilde{u}(t)\rangle)\, dt \leq \langle \widetilde{P}x, x\rangle.$$

Letting T tend to $+\infty$ we obtain

$$J(x, \tilde{u}) \leq \langle \widetilde{P}x, x\rangle.$$

On the other hand, for arbitrary $T \geq 0$ and $x \in \mathbb{R}^m$,

$$\langle P_1(T)x, x\rangle \leq \int_0^T (\langle Q\tilde{y}(t), \tilde{y}(t)\rangle + \langle R\tilde{u}(t), \tilde{u}(t)\rangle)\, dt \leq J(x, \tilde{u}),$$

consequently, $\langle \widetilde{P}x, x\rangle \leq J(x, \tilde{u})$ and finally

$$J(x, \tilde{u}) = \langle \widetilde{P}x, x\rangle.$$

The proof is complete. □

Exercise 3.5 For the control system

$$\ddot{y} = u,$$

find the strategy which minimizes the functional

$$\int_0^{+\infty} (y^2 + u^2)\, dt$$

and the minimal value of this functional.

Answer. The solution of equation (3.20) in which $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$, $B = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, $Q = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, $R = [1]$, is matrix $P = \begin{bmatrix} \sqrt{2} & 1 \\ 1 & \sqrt{2} \end{bmatrix}$. The optimal strategy is of the form $u = -y - \sqrt{2}(\dot{y})$ and the minimal value of the functional is $\sqrt{2}(y(0))^2 + 2y(0)\dot{y}(0) + \sqrt{2}(\dot{y}(0))^2$.

For stabilizability the following result is essential. We need a new concept of *detectability*. A pair of matrices (A, C) is detectable if there exists a matrix L of proper dimension such that the matrix $A + LC$, is stable.

Theorem 3.4

(i) *If the pair (A, B) is stabilizable then equation* (3.20) *has at least one solution.*

(ii) *If $Q = C^*C$ and the pair (A, C) is detectable then equation* (3.20) *has at most one solution, and if P is the solution then the matrix $A - BR^{-1}B^*P$ is stable.*

Proof. (i) Let K be a matrix such that the matrix $A+BK$ is stable. Consider a feedback control $u(t) = Ky(t)$, $t \geq 0$. It follows from the stability of $A + BK$ that $y(t) \longrightarrow 0$, and therefore $u(t) \longrightarrow 0$ exponentially as $t \uparrow +\infty$. Thus for arbitrary $x \in \mathbb{R}^n$,

$$J(x, u(\cdot)) = \int_0^{+\infty} (\langle Qy(t), y(t)\rangle + \langle Ru(t), u(t)\rangle)\, dt < +\infty.$$

Since

$$\langle P_1(T)x, x\rangle \leq J(x, u(\cdot)) < +\infty, \quad T \geq 0,$$

for the solution $P_1(t)$, $t \geq 0$, of (3.15) with the initial condition $P_1(0) = 0$, there exists $\lim_{T\uparrow+\infty} P_1(T) = P$ which satisfies (3.20). (Compare the proof of the previous theorem.)

(ii) We prove first the following lemma.

Lemma 3.3

(i) *Assume that for some matrices $M \geq 0$ and K of appropriate dimensions,*

$$M(A - BK) + (A - BK)^*M + C^*C + K^*RK = 0. \tag{3.23}$$

If the pair (A, C) is detectable, then the matrix $A - BK$ is stable.

(ii) *If, in addition, P is a solution to* (3.20), *then $P \leq M$.*

Proof. (i) Let $S_1(t) = e^{(A-BK)t}$, $S_2(t) = e^{(A-LC)t}$, where L is a matrix such that $A - LC$ is stable and let $y(t) = S_1(t)x$, $t \geq 0$. Since

$$A - BK = (A - LC) + (LC - BK),$$

therefore

$$y(t) = S_2(t)x + \int_0^t S_2(t-s)(LC - BK)y(s)\,ds. \tag{3.24}$$

We show now that

$$\int_0^{+\infty} |Cy(s)|^2\,ds < +\infty \quad \text{and} \quad \int_0^{+\infty} |Ky(s)|^2\,ds < +\infty. \tag{3.25}$$

Let us remark that, for $t \geq 0$,

$$\dot{y}(t) = (A - BK)y(t) \quad \text{and} \quad \frac{d}{dt}\langle My(t), y(t)\rangle = 2\langle M\dot{y}(t), y(t)\rangle.$$

It therefore follows from (3.23) that

$$\frac{d}{dt}\langle My(t), y(t)\rangle + \langle Cy(t), Cy(t)\rangle + \langle RKy(t), Ky(t)\rangle = 0.$$

Hence, for $t \geq 0$,

$$\langle My(t), y(t)\rangle + \int_0^t |Cy(s)|^2\,ds + \int_0^t \langle RKy(s), Ky(s)\rangle\,ds = \langle Mx, x\rangle. \tag{3.26}$$

Since the matrix R is positive definite, (3.26) follows from (3.25). By (3.26),

$$|y(t)| \leq |S_2(t)x| + N\int_0^t |S_2(t-s)|(|Cy(s)| + |Ky(s)|)\,ds,$$

where $N = \max(|L|, |B|)$, $t \geq 0$. We need now the following classical result on convolutions of functions due to Young.

Lemma 3.4 *Assume that p, q, r are positive numbers such that $1/p + 1/q = 1 + 1/r$. If functions f, g belong respectively to L^p and L^q, then the convolution $f * g$ belongs to L^r and*

$$||f * g||_r \leq ||f||_p ||g||_q\,.$$

By Young's result and by (3.25),

$$\int_0^{+\infty} |y(s)|^2\,ds \leq N\int_0^{+\infty} |S_2(s)|\,ds \left(\int_0^{+\infty} (|Cy(s)| + |Ky(s)|)^2\,ds\right)^{1/2} + \left(\int_0^{+\infty} |S_2(s)|^2\,ds\right)^{1/2} |x| < +\infty.$$

It follows from Theorem 2.3(iv) that $y(t) \to 0$ as $t \to \infty$. This proves the required result.

Let us also remark that

$$M = \int_0^{+\infty} S_1^*(s)(C^*C + K^*RK)S_1(s)\, ds. \tag{3.27}$$

(ii) Define $K_0 = R^{-1}B^*P$ then $RK_0 = -B^*P$, $PB = -K_0^*R$.

Consequently,

$$P(A - BK) + (A - BK)^*P + K^*RK = -C^*C + (K - K_0)^*R(K - K_0)$$

and

$$M(A - BK) + (A - BK)^*M + K^*RK = -C^*C.$$

Hence if $V = M - P$ then

$$V(A - BK) + (A - BK)^*V + (K - K_0)^*R(K - K_0) = 0.$$

Since the matrix $A - BK$ is stable the above equation has only one solution given by the formula,

$$V = \int_0^{+\infty} S_1^*(s)(K - K_0)^*R(K - K_0)S_1(s)\, ds \geq 0,$$

and therefore $M \geq P$. The proof of the lemma is complete. □

To prove part (ii) of Theorem 3.4 assume that matrices $P \geq 0$, $P_1 \geq 0$ are solutions of (3.20). Define $K = R^{-1}B^*P$. Then

$$\begin{aligned} &P(A - BK) + (A - BK)^*P + C^*C + K^*RK \\ &\qquad = PA + A^*P + C^*C - PBR^{-1}B^*P = 0. \end{aligned} \tag{3.28}$$

Therefore, by Lemma 3.3(ii), $P_1 \leq P$. In the same way $P_1 \geq P$. Hence $P_1 = P$. Identity (3.28) and Lemma 3.3(i) imply the stability of $A - BK$.□

Let us recall that a pair (A, C) is observable if and only if the pair (A^*, C^*) is controllable. As a corollary from Theorem 3.4 we obtain

Theorem 3.5 *If the pair (A, B) is controllable, $Q = C^*C$ and the pair (A, C) is observable, then equation* (3.20) *has exactly one solution, and if P is this unique solution, then the matrix $A - BR^{-1}B^*P$ is stable.*

Theorem 3.5 indicates an effective way of stabilizing linear system (3.13). Controllability and observability tests in the form of the corresponding rank conditions are effective, and equation (3.20) can be solved numerically using methods similar to those for solving polynomial equations. The uniqueness of the solution of (3.20) is essential for numerical algorithms.

The following examples show that equation (3.20) does not always have a solution and that in some cases it may have many solutions.

Example 3.1 If, in (3.20), $B = 0$, then we arrive at the Lyapunov equation

$$PA + A^*P = Q, \quad P \geq 0. \tag{3.29}$$

If Q is positive definite, then equation (3.29) has at most one solution, and if, in addition, matrix A is not stable, then it does not have any solutions.

Exercise 3.6 If Q is a singular matrix then equation (3.20) may have many solutions. For if P is a solution to (3.20) and

$$\widetilde{A} = \begin{bmatrix} 0 & 0 \\ 0 & A \end{bmatrix}, \; \widetilde{Q} = \begin{bmatrix} 0 & 0 \\ 0 & Q \end{bmatrix}, \quad \widetilde{A} \in \mathbf{M}(k,k), \; k > n,$$

then, for an arbitrary nonnegative matrix $R \in \mathbf{M}(k-n, k-n)$, matrix

$$\widetilde{P} = \begin{bmatrix} R & 0 \\ 0 & P \end{bmatrix}$$

satisfies the equation

$$\widetilde{P}\widetilde{A} + \widetilde{A}^*\widetilde{P} = \widetilde{Q}.$$

Exercise 3.7 Solve the linear regulator problem on finite and infinite intervals when the control system is given by the equation;

$$\dot{y} = Ay + a + Bu, \quad y(0) = x \in \mathbb{R}^n, \tag{3.30}$$

where $a \in \mathbb{R}^n$ is a given vector.

Remark Dynamic programming ideas are presented in the monograph by R. Bellmann [3]. The results of the linear regulator problem are classic. Theorem 3.4 is due to W.M. Wonham [29]. In the proof of Lemma 3.3(i) we follow [30], see also [31].

References

[1] A. V. Balakrishnan, “Semigroup theory and control theory”, in *Proc. IFIP Congress*, Tokyo 1965.

[2] S. Barnett, *Introduction to Mathematical Control Theory* (Clarendon Press, Oxford, 1975).

[3] R. Bellman, *Dynamic Programming* (Princeton University Press, 1977).

[4] A. Bensoussan, G. Da Prato, M. Delfour and S. K. Mitter, *Representation and Control of Infinite Dimensional Systems* Birkhäuser, Vol. 1(1992), Vol. 2(1993).

[5] R. W. Brockett, *Finite Dimensional Linear Systems* (Wiley, New York, 1970).

[6] L. Cesari, *Optimization Theory and Applications* (Springer-Verlag, New York, 1963).

[7] F. H. Clarke, *Optimization and Nonsmooth Analysis* (Wiley Interscience, New York, 1983).

[8] E. A. Coddington and N. Levinson, *Theory of Ordinary Differential Equations* (McGraw-Hill, New York, 1955).

[9] R. F. Curtain and A. J. Pritchard, *Infinite Dimensional Linear Systems Theory* (Lecture Notes in Control and Information Sciences, Springer-Verlag, New York, 1978).

[10] G. Da Prato, J. Math. Pures et Appl. **52**, 353 (1973).

[11] N. Dunford and J. Schwartz, *Linear operators*, Part I (Interscience Publishers, New York, London, 1958).

[12] N. Dunford and J. Schwartz, *Linear operators*, Part II (Interscience Publishers, New York, London, 1963).

[13] H. O. Fattorini, SIAM J. Control, **4**, 686 (1966).

[14] W. H. Fleming and R. W. Rishel, *Deterministic and Stochastic Optimal Control* (Springer-Verlag, Berlin, Heidelberg, New York, 1975).

[15] F. R. Gantmacher, *Applications of the Theory of Matrices* (Interscience Publishers Inc., New York, 1959).

[16] R. E. Kalman, "On the general theory of control systems", in: *Automatic and Remote Control*, Proc. First Int. Congress of IFAC, Moscow, 1960, vol. 1 (Izdat. AN SSSR, Moskva, 1961) pp. 481–492.

[17] E. B. Lee and L. Markus, *Foundations of Optimal Control Theory* (Wiley, New York, 1967).

[18] G. Leitman, *An introduction to optimal control* (Mc Graw-Hill, New York, 1966).

[19] J. C. Maxwell, *On governers* (Proc. Royal Society, 1868).

[20] R. Pallu de la Barriere, *Cours d'automatique théorique* (Dunod, Paris, 1966).

[21] L. S. Pontryagin, *Ordinary Differential Equation* (Addison-Wesley, Reading, Mass., 1962).

[22] S. Rolewicz, *Functional Analysis and Control Theory* (Polish Scientific Publishers, Warszawa, and D. Reidel Publishing Company, Dordrecht, Boston, Lancaster, Tokyo, 1987).

[23] E. J. Routh, *Treatise on the Stability of a Given State of Motion* (Macmillan and Co., London, 1877).

[24] T. Schanbacher, SIAM J. Control and Optimization **27**, 457 (1989).

[25] E.D. Sontag, *Mathematical Control Theory*, Springer Verlag, 1990.

[26] R. Triggiani, *Constructive steering control functions for linear systems and abstract rank condition* (to appear).

[27] W. A. Wolvich, *Linear Multivariable Systems* (Springer-Verlag, New York, 1974).

[28] W. M. Wonham, IEEE Trans. Automat. Control **AC-12**, 660 (1967).

[29] W. M. Wonham, *Linear Multivariable Control: A Geometric Approach* (Springer-Verlag, New York, 1979).

[30] J. Zabczyk, Appl. Math. Optimization **3**, 383 (1976).

[31] J. Zabczyk, *Mathematical Control Theory: An Introduction*, Birkhauser, 1996.