Feedback Systems:
Notes on Linear Systems Theory

Richard M. Murray
Control and Dynamical Systems
California Institute of Technology

DRAFT – Fall 2019

October 19, 2019

These notes are a supplement for the second edition of *Feedback Systems* by Åström and Murray (referred to as FBS2e), focused on providing some additional mathematical background and theory for the study of linear systems.

Contents

1 Signals and Systems **5**

1.1 Linear Spaces and Mappings 5

1.2 Input/Output Dynamical Systems 8

1.3 Linear Systems and Transfer Functions 12

1.4 System Norms 13

1.5 Exercises 15

2 Linear Input/Output Systems **19**

2.1 Matrix Exponential 19

2.2 Convolution Equation 20

2.3 Linear System Subspaces 21

2.4 Input/output stability 24

2.5 Time-Varying Systems 25

2.6 Exercises 27

3 Reachability and Stabilization **31**

3.1 Concepts and Definitions 31

3.2 Reachability for Linear State Space Systems 32

3.3 System Norms 36

3.4 Stabilization via Linear Feedback 37

3.5 Exercises 37

4 Optimal Control **41**

4.1 Review: Optimization 41

4.2 Optimal Control of Systems 44

4.3 Examples 47

4.4 Linear Quadratic Regulators 49

4.5 Choosing LQR weights 52

4.6 Advanced Topics 54

4.7 Further Reading 56

5 State Estimation **61**

5.1 Concepts and Definitions 61

5.2 Observability for Linear State Space Systems 62

5.3 Combining Estimation and Control 63

6 Transfer Functions **65**

6.1 State Space Realizations of Transfer Functions 65

Chapter 1

Signals and Systems

The study of linear systems builds on the concept of linear maps over vector spaces, with inputs and outputs represented as function of time and linear systems represented as a linear map over functions. In this chapter we review the basic concepts of linear operators over (infinite-dimensional) vector spaces, define the notation of a linear system, and define metrics on signal spaces that can be used to determine norms for a linear system. We assume a basic background in linear algebra.

1.1 Linear Spaces and Mappings

We briefly review here the basic definitions for linear spaces, being careful to take a general view that will allow the underlying space to be a signal space (as opposed to a finite dimensional linear space).

Definition 1.1. A set V is a *linear space over* $\mathbb{R}$ if the following axioms hold:

1. Addition: For every $x, y \in V$ there is a unique element $x + y \in V$ where the addition operator $+$ satisfies:

   (a) Commutativity: $x + y = y + x$.

   (b) Associativity: $(x + y) + z = x + (y + z)$.

   (c) Additive identity element: there exists an element $0 \in V$ such that $x + 0 = x$ for all $x \in V$.

   (d) Additive inverse: For every $x \in V$ there exists a unique element $-x \in V$ such that $x + (-x) = 0$.

2. Scalar multiplication: For every $\alpha \in \mathbb{R}$ and $x \in V$ there exists a unique vector $\alpha x \in V$ and the scaling operator satisfies:

   (a) Associativity: $(\alpha\beta) = \alpha(\beta x)$.

   (b) Distributivity over addition in V: $\alpha(x + y) = \alpha x + \alpha y$.

   (c) Distributivity over addition in $\mathbb{R}$: $(\alpha + \beta)x = \alpha x + \beta x$.

   (d) Multiplicative identity: $1 \cdot x = x$ for all $x \in V$.

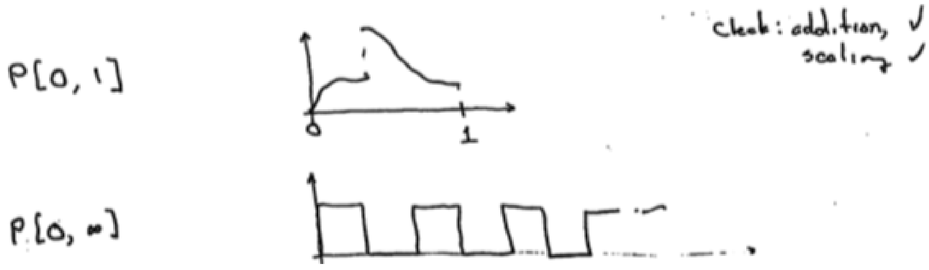
More generally, we can replace $\mathbb{R}$ with any *field* (such as complex number $\mathbb{C}$). The terms “vector space”, “linear space”, and “linear vector space” will be used interchangeably throughout the text.

A vector space V is said to have a basis $\mathcal{B} = \{v_1, v_2, \dots, v_n\}$ is any element $v \in V$ can be written as a linear combination of the basis vectors v_i and the elements of $\mathcal{B}$ are linearly independent. If such a basis exists for a finite n, then V is said to be finite-dimensional of dimension n. If no such basis exists for any finite n then the vector space is said to be infinite-dimensional.

Example 1.1 ($\mathbb{R}^n$)**.** The finite-dimensional vector space $V = \mathbb{R}^n$ consisting of elements $x = (x_1, \dots, x_n)$ is a vector space over the reals, with the addition and scaling operations defined as

$$\begin{aligned} x + y &= (x_1 + y_1, \dots, x_n + y_n) \\ \alpha x &= (\alpha x_1, \dots, \alpha_n) \end{aligned}$$

Example 1.2 ($\mathcal{P}[t_0, t_1]$)**.** The space of piecewise continuous mappings from a time interval $[t_0, t_1] \subset \mathbb{R}$ to $\mathbb{R}$ is defined as the set of functions $F : [t_0, t_1] \to \mathbb{R}$ that have a finite set of discontinuities on every bounded subinterval.



As an exercise, the reader should verify that the axioms of a linear space are satisfied.

Extensions and special cases include:

1. $\mathcal{P}^n[t_0, t_1]$: the space of piecewise continuous functions taking values in $\mathbb{R}^n$.

2. $\mathcal{C}^n[t_0, t_1]$: the space of continuous functions $F : [t_0, t_1] \to \mathbb{R}^n$.

All of these vector spaces are infinite dimensional.

Example 1.3 ($V_1 \times V_2$)**.** Given two linear spaces V_1 and V_2 of the same type, the Cartesian product $V_1 \times V_2$ is a linear space with addition and scaling defined component-wise. For example, $\mathbb{R}^n \times \mathbb{R}^m$ is the linear space $\mathbb{R}^{m+n}$ and the linear space $\mathcal{C}[t_0, t_1] \times \mathcal{C}[t_0, t_1]$ is a linear space $\mathcal{C}^2[t_0, t_1]$ with the operations

$$(f, g)(t) = (f(t), g(t)), \tag{S1.1}$$

$$(f_1, g_1) + (f_2, g_2) = (f_1 + g_1, f_2 + g_2), \tag{S1.2}$$

$$\alpha(f, g) = (\alpha f, \alpha g). \tag{S1.3}$$

Given a vector space V over the reals, we can define a *norm* on the vector space that associates with each element $x \in V$ a real number $\|x\| \in \mathbb{R}$.

Definition 1.2. A mapping $\|\cdot\| : V \to \mathbb{R}$ is a *norm* on V if it satisfies the following axioms:

1. $\|x\| \geq 0$ for all $x \in V$.

2. $\|x\| = 0$ if and only if $x = 0$.

3. $\|\alpha x\| = |\alpha| \, \|x\|$ for all $x \in V$ and $\alpha \in \mathbb{R}$.

4. $\|x + y\| \leq \|x\| + \|y\|$ for all $x, y \in V$ (called the *triangle inequality*).

These definitions are easy to verify for finite-dimensional vector spaces, but they hold even if a vector space is infinite-dimensional.

The following table describes some standard norms for finite-dimensional and infinite dimensional linear spaces.

Name	$V = \mathbb{R}^n$	$V = \{\mathbb{Z}_+ \to \mathbb{R}^n\}$	$V = \{(-\infty, \infty) \to \mathbb{R}\}$
1-norm, $\lVert \cdot \rVert_1$	$\sum_i \lvert x_i \rvert$	$\sum_k \lVert x[k] \rVert$	$\int_{-\infty}^{\infty} \lvert u(\tau) \rvert, d\tau$
2-norm, $\lVert \cdot \rVert_2$	$\sqrt{\sum_i \lvert x_i \rvert^2}$	$\left(\sum_k \lVert x[k] \rVert^2\right)^{1/2}$	$\left(\int_{-\infty}^{\infty} \lvert u(\tau) \rvert^2, d\tau\right)^{1/2}$
p-norm, $\lVert \cdot \rVert_p$	$\sqrt[p]{\sum_i \lvert x_i \rvert^p}$	$\left(\sum_k \lVert x[k] \rVert^2\right)^{1/p}$	$\left(\int_{-\infty}^{\infty} \lvert u(\tau) \rvert^p, d\tau\right)^{1/p}$
∞-norm, $\lVert \cdot \rVert_\infty$	$\max_i \lvert x_i \rvert$	$\max_k \lVert x[k] \rVert$	$\sup_t \lvert u(t) \rvert$

(The function sup is the supremum, where $\sup_t u(t)$ is the smallest number $\bar{u}$ such that $u(t) \leq \bar{u}$ for all t.)

A linear space equipped with a norm is called a *normed linear space.* A normed linear space is said to be *complete* if every Cauchy sequence in V converges to a point in V. (A sequence $\{x_i\}$ is a Cauchy sequence if for every $\epsilon > 0$ there exists an integer N such that $\|x_p - x_q\| < \epsilon$ for all $p, q > N$.) Not every normed linear space is complete. For example, the normed linear space $\mathcal{C}[0, \infty)$, consisting of continuous, real-valued functions is not complete since it is possible to construct a sequence of continuous functions that converge to a discontinuous function (for example a step function). The space $\mathcal{P}[0, \infty)$ consisting of piecewise continuous functions is complete. A complete normed linear space is called a *Banach space.*

Let V and W be linear spaces over $\mathbb{R}$ (or any common field). A mapping $A : V \to W$ is a linear map if

$$A(\alpha_1 v_1 + \alpha_2 v_2) = \alpha A v_1 + \alpha_2 V_2$$

for all $\alpha_1, \alpha_2 \in \mathbb{R}$ and $v_1, v_2 \in V$. Examples include:

1. Matrix multiplication on $\mathbb{R}^n$.

2. Integration operators on $\mathcal{P}[0, 1]$: $Av = \int_0^1 v(t) \, dt$.

3. Convolution operators: let $h \in \mathcal{P}[0, \infty)$ and define the linear operator C_h as

$$(C_h v)(t) = \int_0^t h(t - \tau) v(\tau) \, d\tau$$

This last item provides a hint of how we will define a linear system.

Definition 1.3. An *inner product* on a linear space V is a mapping $\langle \cdot \, , \cdot \rangle : V \times V \to \mathbb{R}$ with the following properties:

1. Bilinear: $\langle \alpha_1 v_1 + \alpha_2 v_2, w \rangle = \alpha_1 \langle v_1, w \rangle + \alpha_2 \langle v_2, w \rangle$ and the same for the second argument.

2. Symmetric: $\langle v, w \rangle = \langle w, v \rangle$

3. Positive definite: $\langle v, v \rangle > 0$ if $v \neq 0$.

A (complete) linear space with an inner product is called a *Hilbert* space. The inner produce also defines a norm given by $\|v\| = \langle v, v \rangle$. A property of the inner product is that $|\langle u, v \rangle| \leq \|u\|_2 \cdot \|v\|_2$ (the Cauchy-Schwartz inequality), which we leave as an exercise (hint: rewrite u as $u = z + (\langle u, v \rangle / \|v\|)v$ where z can be shown to be orthogonal to u).

Example 1.4 (2-norm)**.** Let $V = \mathcal{C}(-\infty, \infty)$. Then $\|\cdot\|_2$ can be verified to be a norm by checking each of the axioms:

1. $\|u\|_2 = \left(\int_{-\infty}^{\infty} |u(t)|^2 \, dt\right)^{1/2} > 0.$

2. If $u(t) = 0$ for all t then $\|u\|_2 = 0$ by definition. To see the converse, assume that $\|u\|_2 = 0$. Then by definition we must have
$$\int_{-\infty}^{\infty} |u(t)|^2 \, dt = 0$$
and therefore $\|u\|_2 = 0$ on any subset of $(-\infty, \infty)$. Since $\mathcal{C}(-\infty, \infty)$ consists of continuous functions, it follows that $u(t) = 0$ at all points t (if not, then there would be a subset of $(-\infty, \infty)$ on which $|u(t)| > 0$ and the integral would not be zero.

3. $\|\alpha u\|_2 = \left(\int_{-\infty}^{\infty} |\alpha u(t)|^2 \, dt\right)^{1/2} = \alpha \|u\|_2.$

4. To show the triangle inequality for the 2-norm, we make use of the Cauchy-Schwartz inequality by defining the inner product between two elements of V as
$$\langle u, v \rangle = \int_{-\infty}^{\infty} u(t) v(t) \, dt.$$
It can be shown that this satisfies the properties of an inner product. Using the fact that $\|u\|_2 = \langle u, u \rangle$ we can show that
$$\begin{aligned} \|u + v\|_2^2 &= \int_{-\infty}^{\infty} |u(t)|^2 + 2u(t)v(t) + |v(t)|^2 \, dt \\ &= \|u\|_2^2 + 2\langle u(t), v(t) \rangle \, dt + \|v\|_2^2 \\ &\leq \|u\|_2^2 + 2|\langle u(t), v(t) \rangle| \, dt + \|v\|_2^2 \\ &\leq \|u\|_2^2 + 2\|u\|_2 \cdot \|v\|_2 + \|v\|_2^2 = (\|u\|_2 + \|v\|_2)^2 \end{aligned}$$

1.2 Input/Output Dynamical Systems

We now proceed to define an input/output dynamical system, with an eventual focus on linear input/output dynamical systems. It is useful to distinguish between three different conceptual aspects of a “dynamical system:

- A *physical system* represents a physical (or biological or chemical) system that we are trying to analyze or design. An example of a physical system would be a vectored thrust aircraft or perhaps a laboratory experiment intended to test different control algorithms.

- A *system model* is an idealized version of the physical system. There may be many different system models for a given physical system, depending on what questions we are trying to answer. A model for a vectored thrust aircraft might be a simplified, planar version of the system (relevant for understanding basic tradeoffs), a nonlinear model that takes into account actuation and sensing characteristics (relevant for designing controllers that would be implemented on the physical system), or a complex model including bending modes, thermal properties and other details (relevant for doing model-based assessment of complex specifications involving those attributes).

- A *system representation* is a mathematical description of the system using one or more mathematical frameworks (e.g., ODEs, PDEs, automata, etc).

In the material that follows, we will use the word "system" to refer to the system representation, but keeping in mind that this is just a mathematical abstraction of a system model that is itself an approximation of the actual physical system.

Definition 1.4. Let $\mathcal{T}$ be a subset of $\mathbb{R}$ (usually $\mathcal{T} = [0, \infty)$ or $\mathcal{T} = \mathbb{Z}_+$). A *dynamical system* on $\mathcal{T}$ is a representation consisting of a tuple $\mathcal{D} = (\mathcal{U}, \Sigma, \mathcal{Y}, s, r)$ where

- the *input space* $\mathcal{U}$ is a set of functions mapping $\mathcal{T}$ to a set U representing the set of possible inputs to the system (typically $\mathcal{U} = \mathcal{P}^m[0, \infty)$);

- the *state space* Σ is a set representing the state of the system (usually $\mathbb{R}^n$, but can also be infinite dimensional, for example when time delays or partial differential equations are used);

- the *output space* $\mathcal{Y}$ is set of functions mapping $\mathcal{T}$ to a set Y representing the set of measured outputs of the system (typically $\mathcal{Y} = \mathcal{P}^p[0, \infty)$);

- the *state transition function* $s : \mathcal{T} \times \mathcal{T} \times \Sigma \times \mathcal{U} \to \Sigma$ is a function of the form $s(t_1, t_0, x_0, u(\,\cdot\,))$ that returns the state $x(t_1)$ of the system at time t_1 reached from state x_0 at time t_0 as a result of applying an input $u \in \mathcal{U}$;

- the *readout function* $r : \mathcal{T} \times \Sigma \times U \to Y$ is a function of the form $r(t, x, u)$ that returns the output $y(t) \in Y$ representing the value of the measured outputs of the system at time $t \in \mathcal{T}$ given that we are at state $x \in \Sigma$ and applying input $u \in U$.

Furthermore, the following axioms must be satisfied:

(A1) State transition axiom: for any $t_0, t_1 \in \mathcal{T}$ and $x_0 \in \Sigma$ with $t_1 \geq t_0$, if $u(\,\cdot\,), \tilde{u}(\,\cdot\,) \in \mathcal{U}$ and

$$u(t) = \tilde{u}(t) \quad \text{for all} \quad t \in [t_0, t_1] \cap \mathcal{T}$$

then

$$s(t_1, t_0, x_0, u(\,\cdot\,)) = s(t_1, t_0, x_0, \tilde{u}(\,\cdot\,)).$$

(A2) Semi-group axiom: For all $t_0 \leq t_1 \leq t_2 \in \mathcal{T}$, all $x_0 \in \Sigma$, and all $u(\,\cdot\,) \in \mathcal{U}$

$$s(t_2, t_1, s(t_1, t_0, x_0, u(\,\cdot\,)), u(\,\cdot\,) = s(t_2, t_0, x_0, u(\,\cdot\,)).$$

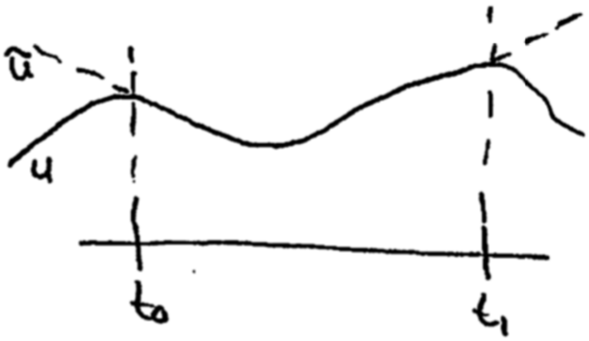


Figure S1.1: Illustration of the state transition axiom.

The definition of a dynamical system captures precisely the notion of a system that has an internal "state" $x \in \Sigma$ and that this state summarizes all information about the system at a given time. Axiom A1 states that inputs differ before reaching a state x_0 and after reaching a state x_1 but are otherwise the same will generate the same trajectory in state space, as illustrated in Figure S1.1. Axiom A2 has the interpretation that we can compute the state at time t_2 by first calculating the state at some intermediate time t_1. In both cases, these are formal statements that the state $x(t)$ summarizes all effects due to the input prior to time t.

Example 1.5 (Input/output differential equation representation)**.** A nonlinear input/output system can be represented as the differential equation

$$\frac{dx}{dt} = f(x,u), \qquad y = h(x,u), \tag{S1.4}$$

where x is a vector of state variables, u is a vector of control signals, and y is a vector of measurements. The term dx/dt represents the derivative of the vector x with respect to time, and f and h are (possibly nonlinear) mappings of their arguments to vectors of the appropriate dimension.

For mechanical systems, the state consists of the configuration variables $q \in \mathbb{R}^n$ and time derivatives of the configuration variables $\dot{q} \in \mathbb{R}^n$ (representing the generalized velocity of the system), so that $x = (q, \dot{q}) \in \mathbb{R}^{2n}$. Note that in the dynamical system formulation of mechanical systems we model the dynamics as first-order differential equations, rather than the more traditional second-order form (e.g., Lagrange's equations), but it can be shown that first order differential equations can capture the dynamics of higher-order differential equations by appropriate definition of the state and the maps f and h.

A model is called a *linear* state space model if the functions f and h are linear in x and u. A linear state space model can thus be represented by

$$\frac{dx}{dt} = A(t)x + B(t)u, \qquad y = C(t)x + D(t)u, \tag{S1.5}$$

where $x \in \mathbb{R}^n$, $u \in \mathbb{R}^m$, $y \in \mathbb{R}^p$ and $A(t)$, $B(t)$, $C(t)$, and $D(t)$ are constant matrices of appropriate dimension. The matrix A is called the *dynamics matrix*, the matrix B is called the *control matrix*, the matrix C is called the *sensor matrix*, and the matrix D is called the *direct term*. Frequently models will not have a direct term, indicating that the input signal u does not influence the output directly.

This definition of a dynamical system is not the most general one possible. In particular, we note that our definition is restricted to model systems that are *causal*: the current state depends only on the past inputs. Furthermore, we have ignored the important class of stochastic dynamical systems, in which the inputs, outputs, and states are described by probability distributions rather than deterministic values. Similarly, this class of systems does not capture other types of nondeterministic systems where a single state may lead to more than one possible output, a situation that is not uncommon in automata theory.

In addition to restricting ourselves to deterministic, causal dynamical systems, we will also often be interested in the case where the system is time-invariant as well. To define time invariance we define the *shift operator* $T_\tau : \mathcal{U} \to \mathcal{U}$ as $(T_\tau u)(t) = u(t+\tau)$. We further define the *input/output map* $\rho : \mathcal{T} \times \mathcal{T} \times \Sigma \times \mathcal{U} \to Y$ as

$$\rho(t, t_0, x_0, u(\,\cdot\,)) = r(t, s(t, t_0, x_0, u(\,\cdot\,)), u(t)),$$

which allows us to evaluate the output of the system at time t given the initial state $x(t_0) = x_0$ and the input applied to the system.

Definition 1.5. A dynamical systems is *time invariant* if

1. $\mathcal{U}$ is closed under translation:
$$u(\,\cdot\,) \in \mathcal{U} \quad \implies \quad T_\tau u(\,\cdot\,) \in \mathcal{U}.$$
2. The input/output map is *shift invariant*:
$$\rho(t_1, t_0, x_0, u(\,\cdot\,)) = \rho(t_1 + \tau, t_0 + \tau, x_0, T_\tau u(\,\cdot\,)).$$

It is straightforward to show that a linear state space model is time invariant if the matrices $A(t)$, $B(t)$, $C(t)$, and $D(t)$ do not depend on time, leading to the representation

$$\frac{dx}{dt} = Ax + Bu, \qquad y = Cx + Du. \tag{S1.6}$$

For our purposes, we will use a slightly more general description of a linear dynamical system, focusing on input/output properties.

Definition 1.6. An input/output dynamical system is a *linear input/output dynamical system* if

1. $\mathcal{U}$, Σ, and $\mathcal{Y}$ are linear spaces over $\mathbb{R}$ (or some other common field, such as $\mathbb{C}$);
2. for fixed $t, t_0 \in \mathcal{T}$ with $t \geq t_0$, $\rho : \mathcal{T} \times \mathcal{T} \times \Sigma \times \mathcal{U} \to Y$ is linear in $\Sigma \times \mathcal{U}$ onto Y:
$$\begin{gathered}\rho(t, t_0, x_0, u(\,\cdot\,)) = \rho(t, t_0, x_0, 0) + \rho(t, t_0, 0, u(\,\cdot\,))\\ \rho(t, t_0, \alpha x + \beta x', 0) = \alpha\rho(t, t_0, x, 0) + \beta\rho(t, t_0, x', 0)\\ \rho(t, t_0, 0, \alpha u(\,\cdot\,) + \beta u'(\,\cdot\,)) = \alpha\rho(t, t_0, 0, u(\,\cdot\,)) + \beta\rho(t, t_0, 0, u'(\,\cdot\,)).\end{gathered}$$

It follows from this definition that if $\mathcal{D}$ is a linear dynamical system representation then the output response can be divided into an initial condition (zero-input) response and a force (zero-initial state) response:

$$\rho(t, t_0, x_0, u(\,\cdot\,)) = \underbrace{\rho(t, t_0, x_0, 0)}_{\text{zero-input response}} + \underbrace{\rho(t, t_0, 0, u(\,\cdot\,))}_{\text{zero-state response}}.$$

Furthermore, the principle of superposition holds for the zero-state response:

$$\rho(t, t_0, x_0, \alpha u(\,\cdot\,) + \beta u'(\,\cdot\,)) = \alpha\rho(t, t_0, x_0, u(\,\cdot\,)) + \beta\rho(t, t_0, x_0, u(\,\cdot\,)).$$

These properties will be familiar to readers who have already encountered linear input/output systems in signal processing or control theory, though we do note here the subtlety that these definitions and properties hold in the time-varying case as well as for time-invariant systems.

For the remainder of the notes we will restrict ourselves to linear, time-invariant (LTI) representations. We will also generally concentrate on the zero-state response, corresponding to the (pure) input/output response.

1.3 Linear Systems and Transfer Functions

Let G be a linear, time-invariant, causal, finite-dimensional system. A different way of defining G is to define the zero-state response as a *convolution equation*:

$$y = G * u, \qquad y(t) = \int_{-\infty}^{\infty} G(t-\tau)u(\tau)\,d\tau.$$

In this formulation, the function $G : (-\infty, \infty) \to \mathbb{R}^m$ is called the *impulse* response of the system and can be regarding as the response of the system to a unit impulse $\delta(t)$ (see FBS2e for the definition of the impulse function). The term $G(t-\tau)$ then represents the response of the system at time t to an input and time τ and the convolution equation is constructed by considering the input to be the convolution of the impulse function $\delta(\,\cdot\,)$ with the input $u(\,\cdot\,)$ and applying the principle of superposition. We also note that if the system is causal then $G(t) = 0$ for all $t < 0$ (if this is not the case, then $y(t)$ and depending on $u(\tau)$ for $\tau < t$).

An alternative to representation of the input/output response as a convolution integral is to make use of the (one-sided) Laplace transform of the inputs, outputs, and impulse response. Letting $\hat{Y}(s)$ represent the Laplace transform of the signal $y(t)$w where $s \in \mathbb{C}$ is the Laplace variable, we have

$$\begin{aligned}
\hat{Y}(s) &= \int_0^{\infty} y(t)e^{-st}\,dt \\
&= \int_0^{\infty} \left(\int_0^{\infty} G(t-\tau)u(\tau)\,d\tau\right) e^{-st}\,dt \\
&= \int_0^{\infty}\int_0^{\infty} \left(G(t-\tau)u(\tau)\,e^{-s(t-\tau)}\,dt\right)\,d\tau \\
&= \underbrace{\left(\int_0^{\infty} G(t)e^{-st}\,dt\right)}_{\hat{G}(s)}\underbrace{\left(\int_0^{\infty} u(\tau)e^{-s\tau}\,d\tau\right)}_{\hat{U}(s)}.
\end{aligned}$$

The Laplace transform of $y(t)$ is thus given by the product of the Laplace transform of the impulse response $G(t)$ and the Laplace transform of the input $u(t)$. The function $\hat{G}(s)$ is called the *transfer function* between input u and output y and represents the zero-state, input/output response of the system. Notationally, we will often write $\hat{G}_{yu}$ to represent the transfer function from u to y so that we have

$$\hat{Y}(s) = \hat{G}_{yu}(s)\hat{U}(s).$$

For a system with m inputs and p outputs, a transfer function $\hat{G}(s)$ represents a mapping from $\mathbb{C}$ to $\mathbb{R}^{p\times m}$. Similar to our definition of norms for signal spaces, we can define norms for Laplace transforms. For the single-input, single-output (SISO) case we define

$$\|\hat{G}\|_2 = \left(\frac{1}{2\pi}\int_{-\infty}^{\infty} |\hat{G}(j\omega)|^2\, d\omega\right)^{1/2}, \qquad \|\hat{G}\|_\infty = \sup_\omega |\hat{G}(j\omega)|.$$

It is left as an exercise to show that these are actually norms that satisfy the properties in Definition 1.2. The 2-norm is a measure of the energy of the impulse response of the system by making use of *Parseval's theorem*:

$$\|\hat{G}\|_2 = \int_{-\infty}^{\infty} |G(t)|^2\, dt.$$

The ∞-norm can be though of in multiple ways: it is the peak value of the frequency response of the system represented by $\hat{G}$ or, equivalently, the distance in the complex plane to the farthest point on the Nyquist plot of $\hat{G}$ (see FBS2e for the definition of the Nyquist plot). It can be shown that the ∞-norm is *submultiplicative*:

$$\|\hat{G}\hat{H}\|_\infty \leq \|\hat{G}\|_\infty \|\hat{H}\|_\infty.$$

For a linear, time-invariant (LTI) state space model of the form

$$\frac{dx}{dt} = Ax + Bu, \qquad y = Cx + Du,$$

with $x \in \mathbb{R}^n$, $u \in \mathbb{R}$, and $y \in \mathbb{R}$, it can be shown that the transfer function has the form

$$\hat{G}(s) = C(sI - A)^{-1}B + D = \frac{n(s)}{d(s)}$$

where $n(s)$ and $d(s)$ are polynomials and $d(s)$ has highest order n. The *poles* of G are the roots of the denominator polynomial and the *zeros* of G are the roots of the numerator polynomial. We say that a transfer function $\hat{G}$ is *proper* if $\hat{G}(j\infty)$ is finite (in which case $\deg d \geq \deg n$), *strictly proper* if $\hat{G}(j\infty) = 0$ ($\deg d > \deg n$), and *biproper* if $\hat{G}$ and $\hat{G}^{-1}$ are both proper ($\deg d = \deg n$). The transfer function is said to be *stable* if it is analytic in the closed right half-plane (i.e., there are no right half-plane poles).

The following result is sometimes useful in proofs and derivations.

Theorem 1.1. *The 2-norm (respectively ∞-norm) of a rational transfer function $\hat{G}$ is finite if and only if $\hat{G}$ is strictly proper (respectively proper) and has no poles on the imaginary axis.*

1.4 System Norms

Given a norm for input signals and a norm for output signals, we can define the *induced norm* for an input/output system. Although this can be done for the general case of nonlinear input/output systems, we restrict ourselves here to the case of a linear input/output system. We furthermore assume that the input/output response is represented by the transfer function (hence we consider only the zero-state response).

Definition 1.7. The *induced a to b norm* for a linear system G is given by

$$\|G\|_{b,a} = \sup_{\|u\|_a \leq 1} \|y\|_b \qquad \text{where } y = G * u.$$

The induced a-norm to b-norm for a system is also called the *system gain.*

Theorem 1.2. *Assume that $\hat{G}$ is stable and strictly proper and that $\mathcal{U}, \mathcal{Y} = \mathcal{P}(-\infty, \infty)$. Then the following table summarizes the induced norms of G:*

	$\|u\|_2$	$\|u\|_\infty$
$\|y\|_2$	$\|\hat{G}\|_\infty$	∞
$\|y\|_\infty$	$\|\hat{G}\|_2$	$\|G\|_1$

Sketch of proofs.

2-norm to 2-norm. We first show that the 2-norm to 2-norm system gain is less than or equal to $\|\hat{G}\|_\infty$:

$$\begin{aligned}
\|y\|_2^2 &= \|\hat{Y}\|_2^2 \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} |\hat{G}(j\omega)|^2 \cdot |\hat{U}(j\omega)|^2 \, d\omega \\
&\leq \|\hat{G}\|_\infty^2 \cdot \frac{1}{2\pi} \int_{-\infty}^{\infty} |\hat{U}(j\omega)|^2 \, d\omega \\
&\leq \|\hat{G}\|_\infty^2 \cdot \|\hat{U}\|_2^2 = \|\hat{G}\|_\infty^2 \cdot \|u\|_2^2.
\end{aligned}$$

To establish equality it suffices to show that we can find an input that achieves the bound.

Let ω_0 be a frequency such that $\|\hat{G}(j\omega_0)\| = \|\hat{G}\|_\infty$ (this exists because $\hat{G}$ is stable and strictly proper). Define a signal u_ϵ such that

$$|\hat{U}_\epsilon(j\omega)| = \begin{cases} \sqrt{\pi/3} & \text{if } \omega_0 - \epsilon \leq \omega \leq \omega_0 + \epsilon \\ 0 & \text{otherwise} \end{cases}$$

and hence $\|u_\epsilon\|_2 = 1$. Then

$$\begin{aligned}
\|\hat{Y}_\epsilon\|_2^2 &= \frac{1}{2\pi} \int_{\omega_0-\epsilon}^{\omega_0+\epsilon} |\hat{G}(j\omega)|^2 \left(\frac{\pi}{\epsilon}\right) \, d\omega \\
&= \frac{1}{2\pi} \int_{\omega_0-\epsilon}^{\omega_0+\epsilon} |\hat{G}(j\omega_0)|^2 \left(\frac{\pi}{\epsilon}\right) \, d\omega + \delta_\epsilon \\
&= \|\hat{G}\|_\infty^2 \|u_\epsilon\|_2^2 + \delta_\epsilon,
\end{aligned}$$

where δ_ϵ represents the error that we obtain by evaluating $\hat{G}$ at $s = j\omega_0$ instead of $s = j\omega$ in the integral. By definition $\delta_\epsilon \to 0$ as $\epsilon \to 0$ (since $\hat{G}$ is continuous) and hence

$$\|\hat{Y}_0\|_2^2 = \|\hat{G}\|_\infty^2 \|\hat{U}_0\|^2$$

and so this input achieves the bound. (Note: to be more formal we need to rely on the fact that $\mathcal{U}$ and $\mathcal{Y}$ are Banach spaces.

∞-norm to 2-norm. Consider the bounded input $u(t) = 1$. This gives a constant output $y(t) = G(0)u(t)$. Assuming that the system has non-zero gain at $\omega = 0$ then $\|y\|_2 = \infty$. (If the gain is zero at zero frequency, a similar argument is possible using a sinusoid $u = \sin(\omega t)$.)

2-norm to ∞-norm. We make use of the following corollary of the Cauchy-Schwartz inequality:

$$\left(\int_{t_0}^{t_1} u(t)v(t)\,dt\right)^2 \le \left(\int_{t_0}^{t_1} |u(t)|^2\,dt\right)\left(\int_{t_0}^{t_1} |v(t)|^2\,dt\right).$$

The output satisfies

$$\begin{aligned}|y(t)|^2 &= \left(\int_{-\infty}^{\infty} G(t-\tau)u(\tau)\,d\tau\right)^2\\ &\le \left(\int_{-\infty}^{\infty} |G(t-\tau)|^2\,d\tau\right)\cdot\left(\int_{-\infty}^{\infty} |u(\tau)|^2\,d\tau\right)\\ &= \|G\|_2^2\,\|u\|_2^2 = \|\hat{G}\|_2^2\,\|u\|_2^2.\end{aligned}$$

Since this holds for all t, it follows that

$$\|y\|_\infty \le \|\hat{G}\|_2\|u\|_2.$$

To get equality, we can apply the signal $u(t) = G(-t)/\|G\|_2$. We have the $\|u\|_2 = 1$ and

$$|y(0)| = \int_{-\infty}^{\infty} G(-t)G(-t)/\|G\|_2\,dt = \|G\|_2.$$

So $\|y\|_\infty \ge |y(0)| = \|\hat{G}\|_2\|u\|_2$. Combining the two inequalities we have that $\|y\|_\infty = \|\hat{G}\|_2\|u\|_2$.

∞-norm to ∞-norm. See DFT [4].

□

1.5 Exercises

1.1 (DFT 2.1) Suppose that $u(t)$ is a continuous signal whose derivative $\dot{u}(t)$ is also continuous. Which of the following quantities qualifies as a norm for u:

(a) $\sup_t |\dot{u}(t)|$

(b) $|u(0)| + \sup_t |\dot{u}(t)|$

(c) $\max\{\sup_t |u(t)|, \sup_t |\dot{u}(t)|\}$

(d) $\sup_t |u(t)| + \sup_t |\dot{u}(t)|$

Make sure to give a thorough answer (not just yes or no).

1.2 (DFT 2.2) Consider the Venn diagram in Figure 2.1 of DFT. Show that the functions u1 to u9, defined below, are located in the diagram as shown in Figure 2.2. All the functions are zero for $t < 0$xs.

$$
\begin{aligned}
u_1(t) &= \begin{cases} 1/\sqrt{t}, & \text{if } t \le 1 \\ 0, & \text{if } t > 1 \end{cases} \\
u_2(t) &= \begin{cases} 1/t^{\frac{1}{4}}, & \text{if } t \le 1 \\ 0, & \text{if } t > 1 \end{cases} \\
u_3(t) &= 1 \\
u_4(t) &= 1/(1+t) \\
u_5(t) &= u_2 + u_4 \\
u_6(t) &= 0 \\
u_7(t) &= u_2(t) + 1
\end{aligned}
$$

1.3 (DFT 2.4) Let D be a pure time delay of τ seconds with transfer function

$$\widehat{D}(s) = e^{-s\tau}.$$

A norm $\|\cdot\|$ on transfer functions is *time-delay invariant* if for every bounded transfer function $\widehat{G}$ and every $\tau > 0$ we have

$$\|\widehat{D}\widehat{G}\| = \|\widehat{G}\|$$

Determine if the 2-norm and ∞-norm are time-delay invariant.

1.4 Consider a discrete time system having dynamics

$$x[k+1] = Ax[k] + Bu[k], \qquad y[k] = Cx[k],$$

where $x[k] \in \mathbb{R}^n$ is the state of the system at time $k \in \mathbb{Z}$, $u[k] \in \mathbb{R}$ is the (scalar) input for the system, $y[k] \in \mathbb{R}$ is the (scalar) output for the system and A, B, and C are constant matrices of the appropriate size. We use the notation $x[k] = x(kh)$ to represent the state of the system at discrete time k where $h \in \mathbb{R}$ is the sampling time (and similarly for $u[k]$ and $y[k]$).

Let $\mathcal{T} = [0, h, \dots, Nh]$ represent a discrete time range, with $N \in \mathbb{Z}$.

(a) Considered as a dynamical system over $\mathcal{T}$, what is the input space $\mathcal{U}$, output space $\mathcal{Y}$, and state space Σ corresponding to the dynamics above? Show that each of these spaces is a linear space by verifying the required properties (you may assume that $\mathbb{R}^p$ is a linear space for appropriate p).

(b) What is the state transition function $s(t_1, t_0, x_0, u(\,\cdot\,))$? Show that this function satisfies the state transition axiom and the semi-group axiom.

(c) What is the readout function $r(t, x, u)$? Show that the input/output system is a linear input/output dynamical system over $\mathcal{T}$.

(d) What is the zero-input response for the system? What is the zero-state response for the system?

1.5 (DFT 2.5) Compute the 1-norm of the impulse response corresponding to the transfer function

$$\frac{1}{\tau s+1} \qquad \tau > 0.$$

1.6 (DFT 2.6) For $\hat{G}$ stable and strictly proper, show that $\|G\|_1 < \infty$ and find an inequality relating $\|\hat{G}\|_\infty$ and $\|G\|_1$. (Remember that G represents the impulse response corresponding to the transfer function $\hat{G}$.)

1.7 (DFT 2.7) Derive the ∞-norm to ∞-norm system gain for a stable, proper plant $\widehat{G}$. (Hint: write $\widehat{G} = c + \widehat{G}_1$ where c is a constant and $\widehat{G}_1$ is strictly proper.)

1.8 (DFT 2.8) Let $\widehat{G}$ be the transfer function for a stable, proper plant (but not necessarily strictly proper).

(a) Show that the ∞-norm of the output y given an input $u(t) = \sin(\omega t)$ is $|\widehat{G}(jw)|$.

(b) Show that the 2-norm to 2-norm system gain for $\widehat{G}$ is $\|\widehat{G}\|_\infty$ (just as in the strictly proper case).

1.9 (DFT 2.10) Consider a system with transfer function

$$\widehat{G}(s) = \frac{s+2}{4s+1}$$

and input u and output y. Compute

$$\|G\|_1 = \sup_{\|u\|_\infty = 1} \|y\|_\infty$$

and find an input that achieves the supremum.

1.10 (DFT 2.12) For a linear system with input u and output y, prove that

$$\sup_{\|u\| \leq 1} \|y\| = \sup_{\|u\|=1} \|y\|$$

where $\|\cdot\|$ is any norm on signals.

1.11 Consider a second order mechanical system with transfer function

$$\widehat{G}(s) = \frac{1}{s^2 + 2\omega_n \zeta s + \omega_n^2}$$

(ω_n is the natural frequency of the system and ζ is the damping ratio). Setting $\omega_n = 1$, plot the ∞-norm as a function of the damping ratio $\zeta > 0$. (You may use a computer to to this, but if you do then make sure to turn in a copy of your code with your solutions.)

Chapter 2

Linear Input/Output Systems

2.1 Matrix Exponential

Let $x(t) \in \mathbb{R}^n$ represent that state of a system whose dynamics satisfy the linear differential equation

$$\frac{d}{dt}x(t) = Ax(t), \qquad A \in \mathbb{R}^{n\times n},\ t \in [0,\infty).$$

The *initial value problem* is to find $x(t)$ given $x(0)$. The approach that we take is to show that there is a unique solution of the form $x(t) = e^{At}x(0)$ and then determine the properties of the solution (e.g., stability) as a function of the properties of the matrix A.

Definition 2.1. Let $S \in \mathbb{R}^{n\times n}$ be a square matrix. The *matrix exponential* of S is given by

$$e^S = I + S + \frac{1}{2}S^2 + \frac{1}{3!}S^2 + \cdots + \frac{1}{k!}S^k + \ldots$$

Proposition 2.1. *The series $\sum_{k=0}^{\infty}\frac{1}{k!}S^k$ converges for all $S \in \mathbb{R}^{n\times n}$.*

Proof. Simple case: Suppose S has a basis of eigenvectors $\{v_1,\ldots,v_n\}$. Then

$$\begin{aligned} e^S v_i &= (I + S + \cdots + \frac{1}{k!}S^k + \ldots)v_i \\ &= (1 + \lambda_i + \cdots + \frac{1}{k!}\lambda_i^k + \ldots)v_i \\ &= e^{\lambda_i}v_i, \end{aligned}$$

which implies that $e^S x$ is well defined and finite (since this is true for all basis elements.

General case: Let $\|S\| = a$. Then

$$\|\frac{1}{k!}S^k\| \leq \frac{1}{k!}\|S\|^k = \frac{a^k}{k!}.$$

Hence

$$\|e^S\| \leq \sum_{k=1}^{\infty}\frac{a^k}{k!} = e^a = e^{\|S\|}$$

and so $e^S x$ is well-defined and finite. □

Proposition 2.2. *If $P, T \in \mathbb{R}^n \to \mathbb{R}^n$ and $S = PTP^{-1}$ then*

$$e^S = e^{PTP^{-1}} = Pe^TP^{-1}.$$

Proof.

$$\begin{aligned} e^S &= \sum \frac{1}{k!} S^k = \sum \frac{1}{k!} (PTP^{-1})^k \\ &= \cdots \frac{1}{k!} (PTP^{-1}) \cdot (PTP^{-1}) \cdots (PTP^{-1}) \cdots \\ &= \sum \frac{1}{k!} PT^kP^{-1}) = P\Big(\sum \frac{T^k}{k!}\Big)P^{-1}) = Pe^TP^{-1}. \end{aligned}$$

□

Proposition 2.3. *If $S, T : \mathbb{R}^n \to \mathbb{R}^n$ commute ($ST = TS$) then $e^{S+T} = e^S e^T$.*

Proof. (basic idea) The first few terms of expansion for the matrix exponential are given by

$$\begin{aligned} (S+T)^0 &= I \\ (S+T)^1 &= S+T \\ (S+T)^2 &= (S+T)(S+T) = S^2 + ST + TS + T^2 \\ &= S^2 + 2ST + T^2 \quad \text{only if ST = TS!} \\ (S+T)^0 &= (S+T)(S+T)(S+T) \\ &= S^3 + S^2T + STS + ST^2 + TS^2 + TST + T^2S + T^3 \\ &= S^3 + 3S^2T + 3ST^2 + T^2 \quad \text{only if ST = TS!.} \end{aligned}$$

The general form becomes

$$(S+T)^k = \underbrace{\sum_{i=1}^{k} \binom{k}{i} S^iT^{k-i}}_{\text{binomial theorem}} = \sum_{i=1}^{k} \frac{k!}{i!(k-i)!} S^iT^{k-i}.$$

□

2.2 Convolution Equation

We now extend our results to include an input. Consider the non-autonomous differential equation

$$\dot{x} = Ax + b(t), \qquad x(0) = x_0. \tag{S2.1}$$

Theorem 2.4. *If $b(t)$ is a (piecewise) continuous signal, then there is a unique $x(t)$ satisfying equation* (S2.1) *given by*

$$x(t) = e^{At}x_0 + \int_0^t e^{T(t-\tau)} b(\tau)\, d\tau.$$

Proof. (existence only) Note that $x(0) = x_0$ and

$$\begin{aligned}
\frac{d}{dt}x(t) &= Ax(t) + \frac{d}{dt}\left(e^{At}\int_0^t e^{-A\tau}b(\tau)\,d\tau\right) \\
&= Ax(t) = Ae^{At}\left(\int_0^t e^{-A\tau}b(\tau)\,d\tau\right) + e^{At}\left(e^{-At}b(t)\right) \\
&= \left[Ax(t) + A\int_0^t e^{A(t-\tau)}b(\tau)\,d\tau\right] + b(t) \\
&= Ax(t) + b(t).
\end{aligned}$$

□

Note that the form of the solution is a combination of the initial condition response ($e^{At}x_0$) and the forced response ($\int_0^t \dots$). Linearity in the initial condition and the input follows from linearity of matrix multiplication and integration.

An alternative form of the solution can be obtained by defining the *fundamental matrix* $\Phi(t) = e^{At}$ as the solution of the matrix differential equation

$$\dot{\Phi} = A\Phi, \quad \Phi(0) = I.$$

Then the solution can be written as

$$x(t) = \Phi(t)x_0 + \underbrace{\int_0^t \Phi(t-\tau)b(\tau)\,d\tau}_{\textit{convolution} \text{ of } \Phi \text{ and } b(t)}.$$

Φ thus acts as a Green's function.

A common situation is that $b(t) = B \cdot a\sin(\omega t)$ where $B \in \mathbb{R}^n$ is a vector and $a\sin(\omega t)$ is a sinusoid with amplitude a and frequency ω. In addition, we wish to consider a specific combination of states $y = Cx$, where $C : \mathbb{R}^n \to \mathbb{R}$:

$$\begin{aligned}
\dot{x}(t) &= Ax + Bu(t) \qquad & u(t) &= a\sin(\omega t) \\
y(t) &= Cx & x(0) &= x_0.
\end{aligned} \tag{S2.2}$$

Theorem 2.5. *Let* $H(s) = C(sI - A)^{-1}B$ *and define* $M = |H(i\omega)|$, $\phi = \arg H(i\omega)$. *Then the sinusoidal response for the system in equation* (S2.2) *is given by*

$$y(t) = Ce^{At}x(0) + aM\sin(\omega t + \phi).$$

A proof can be found in FBS or worked out by using $\sin(\omega t) = \frac{1}{2}(e^{i\omega t} - e^{-i\omega t})$. The function $H(i\omega)$ gives the *frequency response* for the linear system. The function $H : \mathbb{C} \to \mathbb{C}$ is called the *transfer function* for the system.

2.3 Linear System Subspaces

To study the properties of a linear dynamical system, we study the properties of the eigenvalues and eigenvectors of the dynamics matrix $A \in \mathbb{R}^{n\times n}$. We will make use of the *Jordan canonical form* for a matrix. Recall that given any matrix $A \in \mathbb{R}^{n\times n}$ there exists a transformation $T \in \mathbb{C}^{n\times n}$ such that

$$J = TA^{-1}T = \begin{bmatrix} J_1 & & 0 \\ & \ddots & \\ 0 & & J_N \end{bmatrix}, \qquad J_k = \begin{bmatrix} \lambda_k & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_k \end{bmatrix} \in \mathbb{R}^{m_k \times m_k}.$$

This is the complex version of the Jordan form. There is also a real version with $T \in \mathbb{R}^{n\times n}$ in which case the Jordan blocks representing complex eigenvalues have the form

$$J_k = \left[\begin{array}{c|c|c|c} \begin{matrix} a_k & -b_k \\ b_k & a_k \end{matrix} & \begin{matrix} 0 & 1 \\ 1 & 0 \end{matrix} & 0 & 0 \\ \hline 0 & \ddots & \ddots & 0 \\ \hline 0 & 0 & \ddots & \begin{matrix} 0 & 1 \\ 1 & 0 \end{matrix} \\ \hline 0 & 0 & 0 & \begin{matrix} a_k & -b_k \\ b_k & a_k \end{matrix} \end{array}\right]$$

In both the real and complex cases the transformation matrices T consist of a set of generalized eigenvectors $w_{k_1}, \dots, w_{k_{m_k}}$ corresponding to the eigenvalue λ_k.

Returning now to the dynamics of a linear system, let $A \in \mathbb{R}^{n\times n}$ be a square matrix representing the dynamics matrix with eigenvalues $\lambda_j = a_j + ib_j$ and corresponding (generalized) eigenvectors $w_j = u_j + iv_j$ (with $v_j = 0$ if $b_j = 0$). Let $\mathcal{B}$ be a basis of $\mathbb{R}^n$ given by

$$\mathcal{B} = \{\underbrace{u_1, \dots, u_p,}_{\text{real } \lambda_j} \underbrace{u_{p+1}, v_{p+1}, \dots, u_{p+q}, v_{p+q}}_{\text{complex } \lambda_j}\}. \tag{S2.3}$$

Definition 2.2. Given $A \in \mathbb{R}^n$ and basis vector $\mathcal{B}$ as in equation (S2.3), define

1. *Stable subspace*: $E^s = \operatorname{span}\{u_j, v_j : a_j < 0\}$;

2. *Unstable subspace*: $E^u = \operatorname{span}\{u_j, v_j : a_j > 0\}$;

3. *Center subspace*: $E^c = \operatorname{span}\{u_j, v_j : a_j = 0\}$.

These three subspaces can be used to characterize the behavior an unforced linear system. Since $E^s \cap E^u = \{0\}$, $E^s \cap E^c = \{0\}$, and $E^c \cap E^u = \{0\}$, it follows that any vector x can be written as a unique decomposition

$$x = u + v + w, \qquad u \in E^s,\ v \in E^c,\ w \in E^u,$$

and thus $\mathbb{R}^n = E^s \oplus E^c \oplus E^u$ where $\oplus$ is the direct sum of two linear subspaces, defined as $S_1 \oplus S_2 = \{u + v : u \in S_1,\ v \in S_2\}$. If all eigenvalues of A have nonzero real part, so that $E^c = \{0\}$ then the linear system $\dot{x} = Ax$ is said to be *hyperbolic*.

Definition 2.3. A subspace $E \subset \mathbb{R}^n$ is *invariant* with represent to the matrix $A \in \mathbb{R}^{n\times n}$ if $AE \subset E$ and is *invariant with respect to the flow* $e^{At} : \mathbb{R}^n \to \mathbb{R}^n$ if $e^{At}E \subset E$ for all t.

Proposition 2.6. *Let E be the generalized eigenspace of A corresponding to an eigenvalue λ. Then $AE \subset E$.*

Proof. Let $\{v_1, \dots, v_k\}$ be a basis for the generalized eigenspace of A. Then for every $v \in E$ we can write

$$v = \sum \alpha_j v_j \qquad \Longrightarrow \qquad Av = \sum \alpha_j A v_j$$

where α_j is the eigenvalue for the generalized eigenspace. Since $v_1, \dots, v_k$ space the generalized eigenvectors, we know that $(\lambda I - A)^k_j v_j = 0$ for some minimal k_j associated with v_j. It follows that

$$(A - \lambda I)v_j \in \ker(A - \lambda I)^{k_j - 1} \subset E$$

and hence $Av_j = w + \alpha_j v_j$ where $w \in E$, which implies that $Av_j \in E$. □

Proposition 2.7. *The subspaces E^s, E^c, and E^u are all invariant under A and e^{At}.*

Proof. Invariance under A follows from Proposition 2.6. To show invariance of the flow note that

$$e^{At} = (I + At + \frac{1}{2}A^2t^2 + \dots)$$

so

$$e^{At}E \subset E \oplus AE \oplus A^2E \oplus \dots \subset E.$$

□

Theorem 2.8 (Stability of linear systems)**.** *The following statements are equivalent*

1. *$E^s = \mathbb{R}^n$ (i.e., all eigenvalues have negative real part);*
2. *For all $x_0 \in \mathbb{R}^n$, $\lim_{t\to\infty} e^{At}x_0 = 0$ (trajectories converge to the origin);*
3. *There exist constants $a, c, m, M > 0$ such that*

   $$me^{-at}\|x_0\| \leq \|e^{At}x_0\| \leq Me^{-ct}\|x_0\|$$

   (exponential rate of convergence).

Proof. To show the equivalence of (1) and (2) we assume without loss of generality that the matrix is transformed into (real) Jordan canonical form. It can be shown that each Jordan block J_k can be decomposed into a diagonal matrix $S_k = \lambda_k I$ and a nilpotent matrix N_k consisting of 1's on the superdiagonal. The properties of the decomposition $J_k = S_k + N_k$ are that S_k and N_k commute and $N_k^{m_k} = 0$ (so that N_k is *nilpotent*). From these two properties we have that

$$e^{J_k t} = e^{\lambda_k I t} e^{N_k t} = e^{\lambda_k t}(I + N_k + \frac{1}{2}N^2t^2 + \dots + \frac{1}{(m_k - 1)!}N^{m_k - 1}t^{m_k - 1}).$$

A similar decomposition is possible for complex eigenvalues, with the diagonal elements of $e^{S_k t}$ taking the form

$$e^{a_k t}\begin{bmatrix} \cos(b_k t) & -\sin(b_k t) \\ \sin(b_k t) & \cos(b_k t) \end{bmatrix}$$

and the matrix N_k being a block matrix with superdiagonal elements given by the 2×2 identity matrix.

For the real blocks we have $\lambda_k < 0$ and for the complex blocks we have the $a_k < 0$ and it follows that $e^{Jt} \to 0$ as $t \to \infty$ (making use of the fact that $e^{\lambda t} t^m \to 0$ for any $\lambda > 0$ and $m \geq 0$). It follows that (1) and (2) are thus equivalent.

To show (3) we need two additional facts, which we state without proof.

Lemma 2.9. *Let $T \in \mathbb{R}^{n \times n}$ be an invertible transformation and let $y = Tx$. Then there exists constants m and M such that*

$$m\|x\| \leq \|y\| \leq M\|x\|.$$

Lemma 2.10. *Let $A \in \mathbb{R}^{n \times n}$ and assume $\alpha < |\mathrm{Re}\,(\lambda)| < \beta$ for all eigenvalues λ. Then there exists a set of coordinates $y = Tx$ such that*

$$\alpha\|y\|^2 \leq y^T(TAT^{-1})y \leq \beta\|y\|^2.$$

Using these two lemmas we can account for the transformation in converting the system into Jordan canonical form. The only remaining element to prove is that a function of the form $h(t) = e^{\lambda_k t} t^m < \gamma e^{\lambda t}$ for some λ and $\gamma > 0$. This follows from the fact that a function of the form $e^{-\epsilon t} t^m l$ is continuous and zero at $t = 0$ and at $t = \infty$ and thus $e^{-\epsilon t} t^m l$ is bounded above and below. From this we can show (with a bit more work) that for any Jordan block J_k there exists $\gamma > 0$ and $\lambda_k < \lambda < 0$ such that $me^{-at} < \|e^{J_k t} x_0\| < Me^{-ct}$ where $a < \lambda_k < c < 0$. The full result follows by combining all of the various bounds. □

A number of other stability results can be derived along the same lines as the arguments above. For example, if $\mathbb{R}^n = E^u$ (all eigenvalues have positive real part) then all solutions to the initial value problem diverge, exponentially fast. If $\mathbb{R}^n = E^u \oplus E^s$ then we have a mixture of stable and unstable spaces. Any initial condition with a component in E^u diverges, but if $x_0 \in E^s$ then the solution converges to zero.

The unresolved case is when $E^c \neq \{0\}$. In this case, the solutions corresponding to this subspace will have the form

$$(I + Nt + \frac{1}{2}N^2t^2 + \cdots + \frac{1}{k!}N^kt^k$$

for real eigenvalues and

$$\begin{bmatrix} \cos(b_k t) & -\sin(b_k t) \\ \sin(b_k t) & \cos(b_k t) \end{bmatrix} (I + Nt + \frac{1}{2}N^2t^2 + \cdots + \frac{1}{k!}N^kt^k$$

for complex eigenvalues. Convergence in this subspace depends on N. If $N = 0$ then the solutions remain bounded but do not converge to the original (stable in the sense of Lyapunov). If $N \neq 0$ the solutions diverge, but closer than the exponential case. The case of a nonlinear system whose linearization as a non-trivial center subspace leads to a center "manifold" for the nonlinear system and stability depends on the nonlinear characteristics of the system.

2.4 Input/output stability

A system is called bounded input/bounded output (BIBO) stable if a bounded input gives a bounded output for all initial states. A system is called input to state stable (ISS) if $\|x(t)\| \leq \beta(\|x(0)\|) + \gamma(\|u\|)$ where β and γ are monotonically increasing functions that vanish at the origin.

2.5 Time-Varying Systems

Suppose that we have a time-varying ("non-autonomous"), nonhomogeneous linear system with dynamics of the form

$$\begin{aligned} \frac{dx}{dt} &= A(t)x + b(t), \\ y &= C(t)x + d(t), \end{aligned} \qquad x(0) = x_0. \tag{S2.4}$$

A matrix $\Phi(t,s) \in \mathbb{R}^{n\times n}$ is called the *fundamental matrix* for $\dot{x} = A(t)x$ if

1. $\frac{d}{dt}\Phi(t,s) = A(t)\Phi(t,s)$ for all s;

2. $\Phi(s,s) = I$;

3. $\det \Phi(t,s) \neq$ for all s, t.

Proposition 2.11. *If $\Phi(t,s)$ exists for $\dot{x} = A(t)x$ then the solution to equation* (S2.4) *is given by*

$$x(t) = \Phi(t,0)x_0 + \int_0^t \Phi(t,\tau)b(\tau)\,d\tau.$$

This solution generalizes the solution for linear time-invariant systems and we see that the structure of the solution—an initial condition response combined with a convolution integral—is preserved. If $A(t) = A$ is a constant, then $\Phi(t,s) = e^{A(t-s)}$ and we recover our previous solution. The matrix $\Phi(t,s)$ is also called the *state transition matrix* since $x(t) = \Phi(t,s)x(s)$ and $\Phi(t,\tau)\Phi(\tau,s) = \Phi(t,s)$ for all $t > \tau > s$. Solutions for $\Phi(t,s)$ exists for many different time-varying systems, including periodic systems, systems that are sufficiently smooth and bounded, etc.

Example 2.1. Let

$$A(t) = \begin{bmatrix} -1 & e^{at} \\ 0 & -1 \end{bmatrix}.$$

To find $\Phi(t,s)$ we have to solve the matrix differential equation

$$\begin{bmatrix} \dot{\Phi}_{11} & \dot{\Phi}_{12} \\ \dot{\phi}_{21} & \dot{\Phi}_{22} \end{bmatrix} = \begin{bmatrix} -1 & e^{at} \\ 0^-1 & \end{bmatrix} \begin{bmatrix} \Phi_{11} & \Phi_{12} \\ \phi_{21} & \Phi_{22} \end{bmatrix},$$

with $\Phi(s,s) = I$, which serves as an initial condition for the system. We can break the matrix equation into its individual elements. Beginning with the equations for the bottom row of the matrix, we have

$$\begin{aligned} \dot{\Phi}_{21} &= -\Phi_{21} &\implies& \quad \Phi_{21}(t,s) = e^{-(t-s)}\Phi_{21}(s,s) = 0, \\ \dot{\Phi}_{22} &= -\Phi_{22} &\implies& \quad \Phi_{22}(t,s) = e^{-(t-s)}\Phi_{22}(s,s) = e^{-(t-s)}. \end{aligned}$$

Now use Φ_{21} and Φ_{22} to solve for Φ_{11} and Φ_{12}:

$$\begin{aligned} \dot{\Phi}_{11} &= -\Phi_{11} \quad \implies \quad \Phi_{11}(t,s) = e^{-(t-s)} \\ \dot{\Phi}_{12} &= -\Phi_{12} + e^{at}e^{-(t-s)} = -\Phi_{12}(t,s) = e^{(a-1)t+s}. \end{aligned}$$

This last equation is of the form $\dot{x} = -x + b(t)$ and so we can solve it using the solution for a linear differential equation:

$$\begin{aligned}
\Phi_{12}(t,s) &= e^{-(t-s)}\Phi_{12}(s,s) + \int_s^t e^{-(t-\tau)} e^{(a-1)\tau+s}\, d\tau \\
&= e^{-(t-s)} \int_s^t e^{a\tau}\, d\tau = e^{-(t-s)} \left(\frac{1}{a} e^{a\tau} \right)\Big|_s^t \\
&= e^{-(t-s)} \left(\frac{1}{a} e^{at} - \frac{1}{a} e^{as} \right) \\
&= \frac{1}{a} e^{at-(t-s)} - \frac{1}{a} e^{as-(t-s)}.
\end{aligned}$$

Combining all of the elements, the fundamental matrix is thus given by

$$\Phi(t,s) = \begin{bmatrix} e^{-(t-s)} & \frac{1}{a}\left(e^{at-(t-s)} - e^{as-(t-s)}\right) \\ 0 & e^{-(t-s)} \end{bmatrix}.$$

The properties of the fundamental matrix can be verified by direct calculation (and are left to the reader).

The solution for the unforced system ($b(t) = 0$) is given by

$$x(t) = \Phi(t,0)x(0) = \begin{bmatrix} e^{-t} & \frac{1}{a}\left(e^{(a-1)t} - e^{-t}\right) \\ 0 & e^{-t} \end{bmatrix} x(0).$$

We see that although the eigenvalues of $A(t)$ are both -1, if $a > 1$ then some solutions of the differential equation diverge. This is an example that illustrates that for a linear system $\dot{x} = A(t)x$ stability requires more than $\operatorname{Re}(\lambda_A) < 0$.

A common situation is one in which $A(t)$ is period with period T:

$$\frac{dx}{dt} = A(t)x(t), \qquad A(t+T) = A(t).$$

In this case, we can show that the fundamental matrix has the form

$$\Phi(t+T,s) = \Phi(T,0)\Phi(t,s).$$

This property allows us to compute the fundamental matrix just over the period $[0,T]$ (e.g., numerically) and use this to determine the fundamental matrix at any future time. Additional explotation of the structure of the problem is also possible, as the next theorem illustrates.

Theorem 2.12 (Floquet)**.** *Let $A(t)$ be piecewise continuous and T-periodic. Define $P(t) \in \mathbb{R}^{n\times n}$ as*

$$P(t) = \Phi(t,0)e^{-Bt}, \qquad B = \Phi(T,0).$$

Then

1. *$P(t+T) = P(t)$;*

2. *$P(0) = I$ and $\det P(t) \neq 0$;*

3. $\Phi(t,s) = P(t)e^{B(t-s)}P^{-1}(s)$*;*

4. If we set $z(t) = P^{-1}(t)x(t)$ *then* $\dot{z} = Bz$*.*

A consequence of this theorem is that in "rotating" ccoordinates z we can determine the stability properties of the system by examination of the matrix B. In particular, if $z(t) \to 0$ then $x(t) \to 0$. For a proof, see Callier and Desoer [3].

2.6 Exercises

2.1 (FBS2e 6.1) Show that if $y(t)$ is the output of a linear time-invariant system corresponding to input $u(t)$, then the output corresponding to an input $\dot{u}(t)$ is given by $\dot{y}(t)$. (Hint: Use the definition of the derivative: $\dot{z}(t) = \lim_{\epsilon \to 0}\big(z(t+\epsilon) - z(t)\big)/\epsilon$.)

2.2 (FBS2e 6.2) Show that a signal $u(t)$ can be decomposed in terms of the impulse function $\delta(t)$ as

$$u(t) = \int_0^t \delta(t-\tau)u(\tau)\,d\tau$$

and use this decomposition plus the principle of superposition to show that the response of a linear, time-invariant system to an input $u(t)$ (assuming a zero initial condition) can be written as a convolution equation

$$y(t) = \int_0^t h(t-\tau)u(\tau)\,d\tau,$$

where $h(t)$ is the impulse response of the system. (Hint: Use the definition of the Riemann integral.)

2.3 (FBS2e 6.4) Assume that $\zeta < 1$ and let $\omega_\mathrm{d} = \omega_0\sqrt{1-\zeta^2}$. Show that

$$\exp\begin{bmatrix} -\zeta\omega_0 & \omega_\mathrm{d} \\ -\omega_\mathrm{d} & -\zeta\omega_0 \end{bmatrix} t = e^{-\zeta\omega_0 t}\begin{bmatrix} \cos\omega_\mathrm{d}t & \sin\omega_\mathrm{d}t \\ -\sin\omega_\mathrm{d}t & \cos\omega_\mathrm{d}t \end{bmatrix}.$$

Also show that

$$\exp\left(\begin{bmatrix} -\omega_0 & \omega_0 \\ 0 & -\omega_0 \end{bmatrix} t\right) = e^{-\omega_0 t}\begin{bmatrix} 1 & \omega_0 t \\ 0 & 1 \end{bmatrix}.$$

Use the results of this problem and the convolution equation to compute the unit step response for a spring mass system

$$m\ddot{q} + c\dot{q} + kq = F$$

with initial condition $x(0)$.

2.4 (FBS2e 6.6) Consider a linear system with a Jordan form that is non-diagonal.

(a) Prove Proposition 6.3 in *Feedback Systems* by showing that if the system contains a real eigenvalue $\lambda = 0$ with a nontrivial Jordan block, then there exists an initial condition with a solution that grows in time.

(b) Extend this argument to the case of complex eigenvalues with $\operatorname{Re}\lambda = 0$ by using the block Jordan form

$$J_i = \begin{bmatrix} 0 & \omega & 1 & 0 \\ -\omega & 0 & 0 & 1 \\ 0 & 0 & 0 & \omega \\ 0 & 0 & -\omega & 0 \end{bmatrix}.$$

2.5 (FBS2e 6.8) Consider a linear discrete-time system of the form

$$x[k+1] = Ax[k] + Bu[k], \qquad y[k] \quad = Cx[k] + Du[k].$$

(a) Show that the general form of the output of a discrete-time linear system is given by the discrete-time convolution equation:

$$y[k] = CA^k x[0] + \sum_{j=0}^{k-1} CA^{k-j-1}Bu[j] + Du[k].$$

(b) Show that a discrete-time linear system is asymptotically stable if and only if all the eigenvalues of A have a magnitude strictly less than 1.

(c) Show that a discrete-time linear system is unstable if any of the eigenvalues of A have magnitude greater than 1.

(d) Derive conditions for stability of a discrete-time linear system having one or more eigenvalues with magnitude identically equal to 1. (Hint: use Jordan form.)

(e) Let $u[k] = \sin(\omega k)$ represent an oscillatory input with frequency $\omega < \pi$ (to avoid "aliasing"). Show that the steady-state component of the response has gain M and phase θ, where

$$Me^{i\theta} = C(e^{i\omega}I - A)^{-1}B + D.$$

(f) Show that if we have a nonlinear discrete-time system

$$\begin{aligned} x[k+1] &= f(x[k], u[k]), \qquad x[k] \in \mathbb{R}^n,\ u \in \mathbb{R}, \\ y[k] &= h(x[k], u[k]), \qquad y \in \mathbb{R}, \end{aligned}$$

then we can linearize the system around an equilibrium point $(x_{\mathrm{e}}, u_{\mathrm{e}})$ by defining the matrices A, B, C, and D as in equation (6.35).

2.6 Using the computation for the matrix exponential, show that equation (6.11) in *Feedback Systems* holds for the case of a 3×3 Jordan block. (Hint: Decompose the matrix into the form $S + N$, where S is a diagonal matrix.)

2.7 Consider a stable linear time-invariant system. Assume that the system is initially at rest and let the input be $u = \sin\omega t$, where ω is much larger than the magnitudes of the eigenvalues of the dynamics matrix. Show that the output is approximately given by

$$y(t) \approx |G(i\omega)| \sin\big(\omega t + \arg G(i\omega)\big) + \frac{1}{\omega}h(t),$$

where $G(s)$ is the frequency response of the system and $h(t)$ its impulse response.

2.8 Consider the system

$$\frac{dx}{dt} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} x + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u, \qquad y = \begin{bmatrix} 1 & 0 \end{bmatrix} x,$$

which is stable but not asymptotically stable. Show that if the system is driven by the bounded input $u = \cos t$ then the output is unbounded.

Chapter 3

Reachability and Stabilization

Preliminary reading The material in this chapter extends the material in Chapter 7 in FBS2e. Readers should be familiar with the material in Sections 7.1 and 7.2 in preparation for the more advanced concepts discussed here.

3.1 Concepts and Definitions

Consider an input/output dynamical system $\mathcal{D} = (\mathcal{U}, \Sigma, \mathcal{Y}, s, r)$ as defined in Section 1.2.

Definition 3.1 (Reachability)**.** A state x_f is *reachable from* x_0 *in time* T if there exists an input $u : [0, T] \to \mathbb{R}^m$ such that $x_\text{f} = s(T, t_0, x_0, u)$.

If x_f is reachable from x_0 in time T we will write

$$x_0 \underset{T}{\rightsquigarrow} x_\text{f}$$

or sometimes just $x_0 \rightsquigarrow x_\text{f}$ if there exists some T for which x_f is reachable from x_0 in time T. The set of all states that are reachable from x_0 in time less than or equal to T is written as

$$\mathcal{R}_{\leq T}(x_0) = \{x_\text{f} \in \mathbb{R}^n : x_0 \underset{\tau}{\rightsquigarrow} x_\text{f} \text{ for some } \tau \leq T\}.$$

Definition 3.2 (Reachable sytem)**.** An input/output dynamical system $\mathcal{D}$ is *reachable* if for every $x_0, x_\text{f} \in \mathbb{R}^n$ there exists $T > 0$ such that $x_0 \underset{T}{\rightsquigarrow} x_\text{f}$.

.

The notion of reachability captures the property that we can reach a any final point x_f starting from x_0 with some choice of input $u(\,\cdot\,)$. In many cases, it will be not be possible to reach *all* states x_f but it may be possible to reach an open neighborhood of such points.

Definition 3.3 (Small-time local controllability)**.** A system is *small-time locally controllable* (STLC) if for any $T > 0$ the set $\mathcal{R}_{\leq T}(x_0)$ contains a neighborhood of x_0.

The notions of reachability and (small-time local) controllability hold for arbitrary points in the state space, but we are often most interested in equilibrium points and our ability to stabilize a system via state feedback. To define this notion more precisely, we specialize to the case of a state space control systems whose dynamics can be written in the form

$$\frac{dx}{dt} = f(x, u), \qquad x(0) = x_0. \tag{S3.1}$$

Definition 3.4 (Stabilizability)**.** A control system with dynamics (S3.1) is *stabilizable* at an equilibrium x_e if there exists a control law $u = \alpha(x, x_\mathrm{e})$ such that

$$\frac{dx}{dt} = f(x, \alpha(x, x_\mathrm{e})) =: F(x)$$

is locally asymptotically stable at x_e.

The main distinction between reachability and stabilizability is that there may be regions of the state space that are not reachable via application of appropriate control inputs but the dynamics may be such that trajectories with initial conditions in those regions of the state space converge to the origin under the natural dynamics of the system. We will explore this concept more fully in the special case of linear time-invariant systems.

3.2 Reachability for Linear State Space Systems

Consider a linear, time-invariant system

$$\frac{dx}{dt} = Ax + Bu, \qquad x(0) = x_0, \tag{S3.2}$$

with $x \in \mathbb{R}^n$ and $u \in \mathbb{R}^m$, and having state transition function. In this case the state transition function is given by the convolution equation,

$$x(t) = s(t, 0, x_0, u(\,\cdot\,)) = e^{At}Bx_0 + \int_0^t e^{A(t-\tau)}Bu(\tau)\,d\tau.$$

It can be shown that if a linear system is small-time locally controllable at the origin then it is small-time locally controllable at any point x_f, and furthermore that small-time local controllability is equivalent to reachability between any two points (Exercise 3.9).

The problem of reachability for a linear time-invariant system is the same as the general case: we wish to find an input $u(\,\cdot\,)$ that can steer the system from an initial condition x_0 to a final condition x_f in a given time T. Because the system is linear, without loss of generality we can take $x_0 = 0$ (if not, replace the final position x_f with $x_\mathrm{f} - e^{AT}x_0$. In addition, since the state space dynamics depend only on the matrices A and B, we will often state that the pair (A, B) is reachable, stabilizable, etc.

The simplest (and most commonly) used test for reachability for a linear system is to check that the reachability matrix is full rank:

$$\operatorname{rank} \begin{bmatrix} B & AB & \cdots & A^{n-1}B \end{bmatrix} = n.$$

The rank test provides a simple method for checking for reachability, but has the disadvantage that doesn't provide any quantitative insight into how "hard" it might be to either reach a given state or to assign the eigenvalues of the closed loop systems.

A better method of characterizing the reachability properties of a linear system is to make use of the fact that the system defines a linear map between the input $u(\,\cdot\,) \in \mathcal{U}$ and the state $x(T) = x_f \in \Sigma$:

$$x(T) = \mathcal{L}_T u(\,\cdot\,) = \int_0^T e^{A(T-\tau)}Bu(\tau)\,d\tau. \tag{S3.3}$$

Recall that for a linear operator in finite dimensional spaces $L : \mathbb{R}^m \to \mathbb{R}^n$ with $m > n$ that the rank of the linear operator L is the same as the rank of the linear operator $LL^* : \mathbb{R}^n \to \mathbb{R}^n$ where $L^* : \mathbb{R}^n \to \mathbb{R}^m$ is the adjoint operator (given by L^T in the case of matrices). Furthermore, if L is surjective (onto) then the least squares inverse of L is given by

$$L^+ = L^*(LL^*)^{-1} \quad \text{and} \quad LL^+ = I \in \mathbb{R}^{n \times n}.$$

More generally, the adjoint operator can be defined on a linear map between Banach spaces by defining the dual of a Banach space V to be the space V^* of continuous linear functionals on V. Given a linear function $\omega \in V^*$ we write $\langle \omega, v \rangle := \omega(v)$ to represent the application of the function ω on the element v. In the case of finite dimensional vector spaces we can associate the set V^* with V and $\langle \sigma, v \rangle$ is of the form $w^\mathsf{T} v$ where $w \in \mathbb{R}^n$.

If we have a mapping between linear spaces V and W given by $L : V \to W$, the adjoint operator $L^* : W^* \to V^*$ is defined as the unique operator that satisfies

$$\langle L^* \sigma, v \rangle = \langle \sigma, Lv \rangle \quad \text{for all } v \in V \text{ and } \sigma \in W^*.$$

Note that the application of the linear function on the left occurs in the space V and on the right occurs in the space W.

For a signal space $\mathcal{U}$, a linear functional has the form of an integral

$$\langle \omega(\,\cdot\,), u(\,\cdot\,) \rangle = \int_0^\infty \omega(\tau) \cdot u(\tau)\, d\tau$$

and so we can associate each linear function in $\mathcal{U}^*$ with a function $\omega(t)$. Given a linear mapping $\mathcal{L}_T : \mathcal{U} \to \mathbb{R}^n$ of the form

$$\mathcal{L}_T(u(\,\cdot\,)) = \int_0^T h(T - \tau) u(\tau)\, d\tau$$

it can be shown that the adjoint operator $\mathcal{L}_T{}^* : \mathbb{R}^n \to \mathcal{U}^*$ is given by

$$\mathcal{L}_T^*(t) w^* = \begin{cases} \langle h(T-t), w \rangle & \text{if } t \leq T, \\ 0 & \text{otherwise} \end{cases}$$

where $w \in \mathbb{R}^n$.

To show that a system is reachable, we need to show that $\mathcal{L}_T : \mathcal{U} \to \mathbb{R}^n$ given by equation S3.3 is full rank. Using the analysis above, the adjoint operator $\mathcal{L}_T^* : \mathbb{R}^n \to \mathcal{U}^*$ is

$$(\mathcal{L}_T^* v)(t) = B^\mathsf{T} e^{A^\mathsf{T}(T-t)} v.$$

As in the finite dimensional case, the dimension of the range of the map $\mathcal{L}_T : \mathcal{U} \to \mathbb{R}^n$ is the same as the dimension of the range of the map $\mathcal{L}_T \mathcal{L}_T^* : \mathbb{R}^n \to \mathbb{R}^n$, which is given by

$$\mathcal{L}_T \mathcal{L}_T^* = \int_0^T e^{A(T-\tau)} B B^\mathsf{T} e^{A^\mathsf{T}(T-\tau)}\, d\tau.$$

This analysis leads to the following result on reachability for a linear system. △

Theorem 3.1 (Gramian test)**.** *A pair (A, B) is reachable in time T if and only if*

$$W_c(T) = \int_0^T e^{A(T-\tau)} B B^\mathsf{T} e^{A^\mathsf{T}(T-\tau)}\, d\tau = \int_0^T e^{A\tau} B B^\mathsf{T} e^{A^\mathsf{T}\tau}\, d\tau$$

is positive definite.

The matrix $W_\mathrm{c}(T)$ provides a means to compute an input that steers a linear system from the origin to a point $x_\mathrm{f} \in \mathbb{R}^n$. Given $T > 0$, define

$$u(t) = B^\mathsf{T} e^{A^\mathsf{T}(T-t)} W_\mathrm{c}^{-1}(T) x_\mathrm{f}.$$

It follows from the definition of W_c that $x_0 \underset{T}{\rightsquigarrow} x_f$. Furthermore, it is possible to show that if the system is reachable for some $T > 0$ then it is reachable for *all* $T > 0$. Note that this computation of $u(\,\cdot\,)$ corresponds to the computation of the least squares inverse in the finite dimensional case ($u = \mathcal{L}_T^*(\mathcal{L}_T\mathcal{L}_T)^{-1}x_\mathrm{f}$).

Lemma 3.2. *If $W_\mathrm{c}(T)$ is positive definite for some $T > 0$ then it is positive definite for all $T > 0$.*

Proof. We prove the statement by contradiction. Suppose that $W_\mathrm{c}(T)$ is positive definite for a specific $T > 0$ but that there exists $T' > 0$ such that rank $W_\mathrm{c}(T') = k < n$. Then there exists a vector $v \in \mathbb{R}^n$ such that $v^\mathsf{T} W_\mathrm{c}(T') = 0$ and furthermore

$$v^\mathsf{T} W_\mathrm{c}(T') v = v^\mathsf{T} \left(\int_0^{T'} e^{A\tau} B B^\mathsf{T} e^{A^\mathsf{T}\tau} v \, d\tau \right) = 0.$$

Since the integrand is a symmetric matrix, it follows that we must have

$$v^\mathsf{T} e^{A\tau} B B^\mathsf{T} e^{A^\mathsf{T}\tau} v = 0 \quad \text{for all } \tau \le T',$$

and hence

$$\begin{gathered} v^\mathsf{T} e^{A\tau} B = 0 \quad \Longrightarrow \quad v^\mathsf{T} B = 0 \quad (\text{evaluating at } t = 0) \\ \frac{d}{d\tau}(v^\mathsf{T} e^{A\tau} B) = v^\mathsf{T} A e^{A\tau} B = 0 \quad \Longrightarrow \quad v^\mathsf{T} A B = 0 \\ \vdots \\ v^\mathsf{T} A^{n-1} B = 0. \end{gathered}$$

Therefore $v^\mathsf{T} e^{A\tau} B = 0$ for *all* τ (including $\tau > T'$) and hence $v^\mathsf{T} W_\mathrm{c}(t) = 0$ for all $t > 0$, contradicting our original hypothesis. □

If the eigenvalues of A all have negative real part, it can be shown that $W_c(t)$ converges to a constant matrix as $t \to \infty$ and we write this matrix as $W_c = W_c(\infty)$. This matrix is called the *controllability Gramian.* (Note that FBS2e uses W_r to represent the reachability *matrix* $[B\ AB\ A^2B\ \ldots]$. This is different than the controllability *Gramian.*)

Theorem 3.3. *$AW_c + W_cA^\mathsf{T} = -BB^\mathsf{T}$.*

Proof.

$$\begin{aligned}
AW_\mathrm{c} + W_\mathrm{c}A^\mathsf{T} &= \int_0^\infty Ae^{A\tau}BB^\mathsf{T}e^{A^\mathsf{T}\tau}\,d\tau + \int_0^\infty e^{A\tau}BB^\mathsf{T}e^{A^\mathsf{T}\tau}A^\mathsf{T}\,d\tau \\
&= \int_0^\infty \frac{d}{dt}\left(e^{A\tau}BB^\mathsf{T}e^{A^\mathsf{T}\tau}A^\mathsf{T}\right)\,d\tau \\
&= \left(e^{At}BB^\mathsf{T}e^{A^\mathsf{T}t}\right)\Big|_{t=0}^{\infty} \\
&= 0 - BB^\mathsf{T} = -BB^\mathsf{T}.
\end{aligned}$$

□

Theorem 3.4. *A linear time-invariant control system S3.2 is reachable if and only if W_c is full rank and the subspace of points that are reachable from the origin is given by the image of W_c.*

Proof. Left as an exercise. Use the fact that the range of $W_\mathrm{c}(T)$ is independent of T. □

Reachability is best captured by the Gramian since it relates directly to the map between an input vector and final state, and its norm is related to the difficulty of moving from the origin to an arbitrary state. Furthermore, the eigenvectors of W_c and the corresponding eigenvalues provide a measure of how much control effort is required to move in different directions. There are, however, several other tests for reachability that can be used for linear systems.

Theorem 3.5. *The following conditions are necessary and sufficient for reachability of a linear time-invariant system:*

- *Reachability matrix test:*

$$\mathit{rank}\begin{bmatrix}B & AB & \cdots & A^{n-1}B\end{bmatrix} = n.$$

- *Popov-Belman-Hautus (PBH) test:*

$$\mathit{rank}\left[\begin{array}{c|c} sI - A & B \end{array}\right] = n$$

for all $s \in \mathbb{C}$ (suffices to check for eigenvalues of A).

Proof. (Incomplete) PBH necessity: Suppose

$$\operatorname{rank}\left[\begin{array}{c|c} \lambda I - A & B \end{array}\right] < n.$$

Then there exists $v \neq 0$ such that

$$v^\mathsf{T}\left[\begin{array}{c|c} \lambda I - A & B \end{array}\right] = 0$$

and hence $x^\mathsf{T}A = \lambda x^\mathsf{T}$ and $x^\mathsf{T}B = 0$. It follows that $x^\mathsf{T}A^2 = \lambda^2 x^\mathsf{T}, \ldots, x^\mathsf{T}A^{n-1} = \lambda^{n-1}x^\mathsf{T}$ and thus

$$x^\mathsf{T}\begin{bmatrix}B & AB & \cdots & A^{n-1}B\end{bmatrix} = 0.$$

□

For both of these tests, we note that if the corresponding matrix is rank deficient, the left null space of that matrix gives directions in the state space that are unreachable (more accurately it consists of the directions in which the projected value of the state is constant along all trajectories of the system). The set of vectors orthogonal to this left null space defines a subspace V_r that represents the set of reachable states (exercise: prove this is a subspace).

Theorem 3.6. *Assume (A, B) is not reachable. Let rank$W_c = r < n$. Then there exists a transformation $T \in \mathbb{R}^{n\times n}$ such that*

$$TAT^{-1} = \begin{bmatrix} A_1 & A_2 \\ 0 & A_3 \end{bmatrix}, \quad TB = \begin{bmatrix} B_1 \\ 0 \end{bmatrix},$$

where $A_1 \in \mathbb{R}^{r\times r}$, $B_1 \in \mathbb{R}^{r\times m}$, and (A_1, B_1) is reachable.

Proof. (Sketch) Let V_r represent the null space of W_c and let $\mathcal{B}_{\bar{\mathrm{r}}} = \{w_1, \dots, w_{n-r}\}$ represent a basis for V_r. Complete this basis with a set of vectors $\{v_1, \dots, v_r\}$ such that $\{v_1, \dots, v_r, w_1, \dots, w_{n-r}\}$ is a basis for $\mathbb{R}^n$. Use these basis vectors as the columns of the transformation T. □

We note that the null space of W_c is uniquely defined, though the basis for that space is not unique. This subspace represents the set of linear functions on the state space whose values are constant and hence provides a characterization of the unreachable states of the system. The complement of that space is not a subspace, although if we look at the points that are reachable from the origin, this does form a subspace. We will return to this point in more detail when we discuss the Kalman decomposition in Chapter 5.

Finally, we note that a system that is reachable can be written in *reachable canonical form* (see FBS2e). This is primarily useful for proofs.

3.3 System Norms

Consider a stable state space system with no direct term and with system matrices A, B, and C. Let W_c be the controllability Gramian for the system and let $G(t)$ represent the impulse response function for the system and $\hat{G}(s)$ represent the corresponding transfer function (Laplace transform of the impulse response). Recall that the 2-norm to ∞-norm gain for a linear input/output system is given by $\|\hat{G}\|_2$.

Theorem 3.7. $\|\hat{G}\|_2 = \sqrt{CW_cC^\mathsf{T}}$.

Proof. The impulse response function given by

$$\begin{aligned} G(t) = Cx_\delta(t) &= C\int_0^t A^{A(t-\tau)}B\delta(\tau)\,d\tau \\ &= Ce^{At}B, \qquad t > 0. \end{aligned}$$

The system norm is given by

$$\begin{aligned} \|\hat{G}\|_2^2 &= \|G\|_2^2 \\ &= \int_0^\infty \big(Ce^{A\tau}B\big)\big(B^\mathsf{T}e^{A^\mathsf{T}\tau}C^\mathsf{T}\big)\,d\tau \\ &= C\left(\int_0^\infty e^{At}BB^\mathsf{T}e^{A^\mathsf{T}\tau}\,d\tau\right)C^\mathsf{T} \\ &= CW_\mathrm{c}C^\mathsf{T}. \end{aligned}$$

□

The more common norm in control system design is the 2-norm to 2-norm system gain, which is given by $\|\hat{G}\|_\infty$. To compute the ∞-norm of a transfer function, we define

$$H_\gamma = \begin{bmatrix} A & \frac{1}{\gamma} B B^\mathsf{T} \\ -C^\mathsf{T} C & -A^\mathsf{T} \end{bmatrix} \in \mathbb{R}^{2n \times 2n}.$$

The system gain can be determined in terms of H_γ as follows.

Theorem 3.8. *$\|\hat{G}\|_\infty < \gamma$ is an only if H_γ has no eigenvalues on the $j\omega$ axis.*

Proof. DGKF. □

To numerically compute the H_∞ norm, we can use the bisection method to determine γ to arbitrary accuracy.

3.4 Stabilization via Linear Feedback

We now consider the problem of stabilization, as defined in Definition 3.4. For a linear system, we will consider feedback laws of the form $u = -Kx$ (the negative sign is a convention associated with the use of “negative” feedback), so that

$$\frac{dx}{dt} = Ax + Bu = (A - BK)x.$$

One of the goals of introducing negative feedback is to stabilize an otherwise unstable system at the origin. In addition, state feedback can be used to “design the dynamics” of the close loop system by attempting to assign the eigenvalues of the closed loop system to specific values.

Theorem 7.3 states that if a (single-input) system is reachable then it is possible to assign the eigenvalues of the closed loop system to arbitrary values. This turns out to be true for the multi-input case as well and is proved in a similar manner (by using an appropriate normal form).

Using the decomposition theorem 3.6 it is easy to see that the question of stabilizability for a linear system comes down to the question of whether the dynamics in the unreachable space ($\dot{z} = A_3 z$) are stable, since these eigenvalues cannot be changed through the use of state feedback.

Although eigenvalue placement provides an easy method for designing the dynamics of the closed loop system, it is rarely used directly since it does not provide any guidelines for trading off the size of the inputs required to stabilize the dynamics versus the properties of the closed loop response. This is explored in a bit more detail in FBS2e Section 14.6 (Robust Pole Placement).

3.5 Exercises

3.1 (Sontag 3.1.2/3.1.3) Prove the following statements:

(a) If $(x, \sigma) \rightsquigarrow (z, \tau)$ and $(z, \tau) \rightsquigarrow (y, \mu)$, then $(x, \sigma) \rightsquigarrow (y, \mu)$.

(b) If $(x, \sigma) \rightsquigarrow (y, \mu)$ and if $\sigma < \tau < \mu$, then there exists a $z \in \mathcal{X}$ such that $(x, \sigma) \rightsquigarrow (z, \tau)$ and $(z, \tau) \rightsquigarrow (y, \mu)$.

(c) If $x \underset{T}{\rightsquigarrow} y$ for some $T > 0$ and if $0 < t < T$, then there is some $z \in \mathcal{X}$ such that $x \underset{t}{\rightsquigarrow} z$ and $z \underset{T-t}{\rightsquigarrow} y$.

(d) If $x \underset{t}{\rightsquigarrow} z$, $z \underset{s}{\rightsquigarrow} y$, and Σ is time-invariant, then $x \underset{t+s}{\rightsquigarrow} y$.

(e) If $x \rightsquigarrow z$, $z \rightsquigarrow y$, and Σ is time-invariant, then $x \rightsquigarrow y$.

(f) Given examples that show that properties (d) and (e) may be false if Σ is not time-invariant.

(g) Even for time-invariant systems, it is not necessarily true that $x \rightsquigarrow z$ implies that $z \rightsquigarrow x$ (so, "$\rightsquigarrow$" is not an equivalence relation).

3.2 (FBS2e 7.1) Consider the double integrator. Find a piecewise constant control strategy that drives the system from the origin to the state $x = (1, 1)$.

3.3 (FBS2e 7.2) Extend the argument in Section 7.1 in *Feedback Systems* to show that if a system is reachable from an initial state of zero, it is reachable from a nonzero initial state.

3.4 (FBS2e 7.3) Consider a system with the state x and z described by the equations

$$\frac{dx}{dt} = Ax + Bu, \qquad \frac{dz}{dt} = Az + Bu.$$

If $x(0) = z(0)$ it follows that $x(t) = z(t)$ for all t regardless of the input that is applied. Show that this violates the definition of reachability and further show that the reachability matrix W_r is not full rank. What is the rank of the reachability matrix?

3.5 (FBS2e 7.6) Show that the characteristic polynomial for a system in reachable canonical form is given by equation (7.7) and that

$$\frac{d^n z_k}{dt^n} + a_1 \frac{d^{n-1} z_k}{dt^{n-1}} + \cdots + a_{n-1} \frac{dz_k}{dt} + a_n z_k = \frac{d^{n-k} u}{dt^{n-k}},$$

where z_k is the kth state.

3.6 (FBS2e 7.7) Consider a system in reachable canonical form. Show that the inverse of the reachability matrix is given by

$$\tilde{W}_\mathrm{r}^{-1} = \begin{bmatrix} 1 & a_1 & a_2 & \cdots & a_{n-1} \\ & 1 & a_1 & \cdots & a_{n-2} \\ & & 1 & \ddots & \vdots \\ & 0 & & \ddots & a_1 \\ & & & & 1 \end{bmatrix}.$$

3.7 (FBS2e 7.10) Consider the system

$$\frac{dx}{dt} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} x + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u, \qquad y = \begin{bmatrix} 1 & 0 \end{bmatrix} x,$$

with the control law

$$u = -k_1 x_1 - k_2 x_2 + k_\mathrm{f} r.$$

Compute the rank of the reachability matrix for the system and show that eigenvalues of the system cannot be assigned to arbitrary values.

3.8 (FBS2e 7.11) Let $A \in \mathbb{R}^{n\times n}$ be a matrix with characteristic polynomial $\lambda(s) = \det(sI - A) = s^n + a_1 s^{n-1} + \cdots + a_{n-1}s + a_n$. Show that the matrix A satisfies

$$\lambda(A) = A^n + a_1 A^{n-1} + \cdots + a_{n-1}A + a_n I = 0,$$

where the zero on the right hand side represents a matrix of elements with all zeros. Use this result to show that A^n can be written in terms of lower order powers of A and hence any matrix polynomial in A can be rewritten using terms of order at most $n - 1$.

3.9 Show that for a linear time-invariant system, the following notions of controllability are equivalent:

(a) Reachability to the origin ($x_0 \rightsquigarrow 0$).

(b) Reachability from the origin ($0 \rightsquigarrow x_\mathrm{f}$).

(c) Small-time local controllability ($x_0 \rightsquigarrow B(x_0, \epsilon)$).

3.10 (Sontag 3.3.4) Assume that the pair (A, B) is not controllable with $\dim R(A, B) = \operatorname{rank} W_\mathrm{c} = r < n$. From Lemma 3.3.3, there exists an invertible matrix $T \in \mathbb{R}^{n\times n}$ such that the matrices $\tilde{A} := T^{-1}AT$ and $\tilde{B} := T^{-1}B$ have the block structure

$$\tilde{A} = \begin{bmatrix} A_1 & A_2 \\ 0 & A_3 \end{bmatrix}, \qquad \tilde{B} = \begin{bmatrix} B_1 \\ 0 \end{bmatrix},$$

where $A_1 \in \mathbb{R}^{r\times r}$ and $B_1 \in \mathbb{R}^{r\times m}$. Prove that (A_1, B_1) is itself a controllable pair.

3.11 (Sontag 3.3.6) Prove that if

$$A = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \lambda_{n-1} & 0 \\ 0 & 0 & \cdots & 0 & \lambda_n \end{bmatrix}, \qquad B = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_{n-1} \\ b_n \end{bmatrix}$$

then (A, B) is controllable if and only if $\lambda_i \neq \lambda_j$ for each $i \neq j$ and all $b_i \neq 0$.

3.12 (Sontag 3.3.14) Let (A, B) correspond to a time-invariant *discrete-time* linear system Σ. Recall that null-controllability means that every state can be controlled to zero. Prove that the following conditions are equivalent:

(a) Σ is null-controllable.

(b) The image of A^n is contained in the image of $R(A, B)$.

(c) In the decomposition in Sontag, Lemma 3.3.3, A_3 is nilpotent.

(d) $\operatorname{rank}[\lambda I - A, B] = n$ for all nonzero $\lambda \in \tilde{\mathbb{R}}$.

(e) $\operatorname{rank}[\lambda I - A, B] = n$ for all $\lambda \in \tilde{\mathbb{R}}$.

Chapter 4

Optimal Control

This set of notes expands on Chapter 6 of *Feedback Systems* by Åström and Murray (FBS2e), which introduces the concepts of reachability and state feedback. We also expand on topics in Section 7.5 of FBS2e in the area of feedforward compensation. Beginning with a review of optimization, we introduce the notion of Lagrange multipliers and provide a summary of the Pontryagin's maximum principle. Using these tools we derive the linear quadratic regulator for linear systems and describe its usage.

Prerequisites Readers should be familiar with modeling of input/output control systems using differential equations, linearization of a system around an equilibrium point and state space control of linear systems, including reachability and eigenvalue assignment. Some familiarity with optimization of nonlinear functions is also assumed.

4.1 Review: Optimization

*O*ptimization refers to the problem of choosing a set of parameters that maximize or minimize a given function. In control systems, we are often faced with having to choose a set of parameters for a control law so that the some performance condition is satisfied. In this chapter we will seek to *optimize* a given specification, choosing the parameters that maximize the performance (or minimize the cost). In this section we review the conditions for optimization of a static function, and then extend this to optimization of trajectories and control laws in the remainder of the chapter. More information on basic techniques in optimization can be found in [6] or the introductory chapter of [5].

Consider first the problem of finding the minimum of a smooth function $F : \mathbb{R}^n \to \mathbb{R}$. That is, we wish to find a point $x^* \in \mathbb{R}^n$ such that $F(x^*) \leq F(x)$ for all $x \in \mathbb{R}^n$. A necessary condition for x^* to be a minimum is that the gradient of the function be zero at x^*:

$$\frac{\partial F}{\partial x}(x^*) = 0.$$

The function $F(x)$ is often called a cost-function and x^* is the optimal-value for x. Figure S4.1 gives a graphical interpretation of the necessary condition for a minimum. Note that these are *not* sufficient conditions; the points x_1 and x_2 and x^* in the figure all satisfy the necessary condition but only one is the (global) minimum.

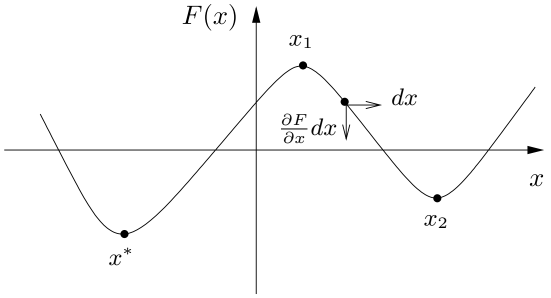
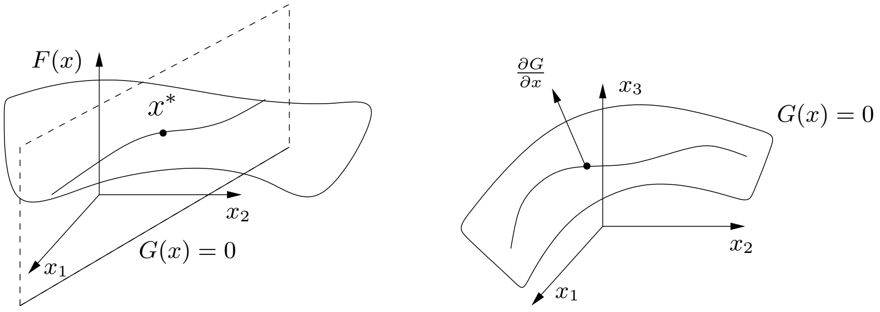


Figure S4.1: Optimization of functions. The minimum of a function occurs at a point where the gradient is zero.



(a) Constrained optimization

(b) Constraint normal vectors

Figure S4.2: Optimization with constraints. (a) We seek a point x^* that minimizes $F(x)$ while lying on the surface $G(x) = 0$ (a line in the $x_1 x_2$ plane). (b) We can parameterize the constrained directions by computing the gradient of the constraint G. Note that $x \in \mathbb{R}^2$ in (a), with the third dimension showing $F(x)$, while $x \in \mathbb{R}^3$ in (b).

The situation is more complicated if constraints are present. Let $G_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \dots, k$ be a set of smooth functions with $G_i(x) = 0$ representing the constraints. Suppose that we wish to find $x^* \in \mathbb{R}^n$ such that $G_i(x^*) = 0$ and $F(x^*) \leq F(x)$ for all $x \in \{x \in \mathbb{R}^n : G_i(x) = 0, i = 1, \dots, k\}$. This situation can be visualized as constraining the point to a surface (defined by the constraints) and searching for the minimum of the cost function along this surface, as illustrated in Figure S4.2a.

A necessary condition for being at a minimum is that there are no directions tangent to the constraints that also decrease the cost. Given a constraint function $G(x) = (G_1(x), \dots, G_k(x))$, $x \in \mathbb{R}^n$ we can represent the constraint as a $n - k$ dimensional surface in $\mathbb{R}^n$, as shown in Figure S4.2b. The tangent directions to the surface can be computed by considering small perturbations of the constraint that remain on the surface:

$$G_i(x + \delta x) \approx G_i(x) + \frac{\partial G_i}{\partial x}(x)\delta x = 0. \qquad \Longrightarrow \qquad \frac{\partial G_i}{\partial x}(x)\delta x = 0,$$

where $\delta x \in \mathbb{R}^n$ is a vanishingly small perturbation. It follows that the normal directions to the surface are spanned by $\partial G_i / \partial x$, since these are precisely the vectors that annihilate an admissible tangent vector δx.

Using this characterization of the tangent and normal vectors to the constraint, a necessary condition for optimization is that the gradient of F is spanned by vectors that are normal to the constraints, so that the only directions that increase the cost violate the constraints. We thus require that there exist scalars λ_i, $i = 1, \ldots, k$ such that

$$\frac{\partial F}{\partial x}(x^*) + \sum_{i=1}^{k} \lambda_i \frac{\partial G_i}{\partial x}(x^*) = 0.$$

If we let $G = \begin{bmatrix} G_1 & G_2 & \ldots & G_k \end{bmatrix}^T$, then we can write this condition as

$$\frac{\partial F}{\partial x} + \lambda^T \frac{\partial G}{\partial x} = 0 \tag{S4.1}$$

the term $\partial F/\partial x$ is the usual (gradient) optimality condition while the term $\partial G/\partial x$ is used to "cancel" the gradient in the directions normal to the constraint.

An alternative condition can be derived by modifying the cost function to incorporate the constraints. Defining $\widetilde{F} = F + \sum \lambda_i G_i$, the necessary condition becomes

$$\frac{\partial \widetilde{F}}{\partial x}(x^*) = 0.$$

The scalars λ_i are called *Lagrange multipliers*. Minimizing $\widetilde{F}$ is equivalent to the optimization given by

$$\min_x \left(F(x) + \lambda^T G(x) \right). \tag{S4.2}$$

The variables λ can be regarded as free variables, which implies that we need to choose x such that $G(x) = 0$ in order to insure the cost is minimized. Otherwise, we could choose λ to generate a large cost.

Example 4.1 (Two free variables with a constraint)**.** Consider the cost function given by

$$F(x) = F_0 + (x_1 - a)^2 + (x_2 - b)^2,$$

which has an unconstrained minimum at $x = (a, b)$. Suppose that we add a constraint $G(x) = 0$ given by

$$G(x) = x_1 - x_2.$$

With this constraint, we seek to optimize F subject to $x_1 = x_2$. Although in this case we could do this by simple substitution, we instead carry out the more general procedure using Lagrange multipliers.

The augmented cost function is given by

$$\tilde{F}(x) = F_0 + (x_1 - a)^2 + (x_2 - b)^2 + \lambda(x_1 - x_2),$$

where λ is the Lagrange multiplier for the constraint. Taking the derivative of $\tilde{F}$, we have

$$\frac{\partial \tilde{F}}{\partial x} = \begin{bmatrix} 2x_1 - 2a + \lambda & 2x_2 - 2b - \lambda \end{bmatrix}.$$

Setting each of these equations equal to zero, we have that at the minimum

$$x_1^* = a - \lambda/2, \qquad x_2^* = b + \lambda/2.$$

The remaining equation that we need is the constraint, which requires that $x_1^* = x_2^*$. Using these three equations, we see that $\lambda^* = a - b$ and we have

$$x_1^* = \frac{a+b}{2}, \qquad x_2^* = \frac{a+b}{2}.$$

To verify the geometric view described above, note that the gradients of F and G are given by

$$\frac{\partial F}{\partial x} = \begin{bmatrix} 2x_1 - 2a & 2x_2 - 2b \end{bmatrix}, \qquad \frac{\partial G}{\partial x} = \begin{bmatrix} 1 & -1 \end{bmatrix}.$$

At the optimal value of the (constrained) optimization, we have

$$\frac{\partial F}{\partial x} = \begin{bmatrix} b - a & a - b \end{bmatrix}, \qquad \frac{\partial G}{\partial x} = \begin{bmatrix} 1 & -1 \end{bmatrix}.$$

Although the derivative of F is not zero, it is pointed in a direction that is normal to the constraint, and hence we cannot decrease the cost while staying on the constraint surface.

We have focused on finding the minimum of a function. We can switch back and forth between maximum and minimum by simply negating the cost function:

$$\max_x F(x) = \min_x \bigl(-F(x)\bigr)$$

We see that the conditions that we have derived are independent of the sign of F since they only depend on the gradient begin zero in approximate directions. Thus finding x^* that satisfies the conditions corresponds to finding an *extremum* for the function.

Very good software is available for numerically solving optimization problems of this sort. The NPSOL and SNOPT libraries are available in FORTRAN (and C). In MATLAB, the `fmin` function can be used to solve a constrained optimization problem.

4.2 Optimal Control of Systems

Consider now the *optimal control problem*:

$$\min_{u(\,\cdot\,)} \int_0^T L(x,u)\,dt + V\bigl(x(T)\bigr)$$

subject to the constraint

$$\dot{x} = f(x,u), \qquad x \in \mathbb{R}^n,\ u \in \mathbb{R}^m.$$

Abstractly, this is a constrained optimization problem where we seek a *feasible trajectory* $(x(t), u(t))$ that minimizes the cost function

$$J(x,u) = \int_0^T L(x,u)\,dt + V\bigl(x(T)\bigr).$$

More formally, this problem is equivalent to the "standard" problem of minimizing a cost function $J(x,u)$ where $(x,u) \in L_2[0,T]$ (the set of square integrable functions) and $h(z) = \dot{x}(t) - f(x(t),u(t)) = 0$ models the dynamics. The term $L(x,u)$ is referred to as the integral cost and $V(x(T))$ is the final (or terminal) cost.

There are many variations and special cases of the optimal control problem. We mention a few here:

Infinite horizon optimal control If we let $T = \infty$ and set $V = 0$, then we seek to optimize a cost function over all time. This is called the *infinite horizon* optimal control problem, versus the *finite horizon* problem with $T < \infty$. Note that if an infinite horizon problem has a solution with finite cost, then the integral cost term $L(x, u)$ must approach zero as $t \to \infty$.

Linear quadratic (LQ) optimal control If the dynamical system is linear and the cost function is quadratic, we obtain the *linear quadratic* optimal control problem:

$$\dot{x} = Ax + Bu, \qquad J = \int_0^T \left(x^T Q x + u^T R u\right) \, dt + x^T(T) P_1 x(T).$$

In this formulation, $Q \geq 0$ penalizes state error, $R > 0$ penalizes the input and $P_1 > 0$ penalizes terminal state. This problem can be modified to track a desired trajectory (x_d, u_d) by rewriting the cost function in terms of $(x - x_d)$ and $(u - u_d)$.

Terminal constraints It is often convenient to ask that the final value of the trajectory, denoted x_f, be specified. We can do this by requiring that $x(T) = x_f$ or by using a more general form of constraint:

$$\psi_i(x(T)) = 0, \qquad i = 1, \dots, q.$$

The fully constrained case is obtained by setting $q = n$ and defining $\psi_i(x(T)) = x_i(T) - x_{i,f}$. For a control problem with a full set of terminal constraints, $V(x(T))$ can be omitted (since its value is fixed).

Time optimal control If we constrain the terminal condition to $x(T) = x_f$, let the terminal time T be free (so that we can optimize over it) and choose $L(x, u) = 1$, we can find the *time-optimal* trajectory between an initial and final condition. This problem is usually only well-posed if we additionally constrain the inputs u to be bounded.

A very general set of conditions are available for the optimal control problem that captures most of these special cases in a unifying framework. Consider a nonlinear system

$$\begin{aligned} &\dot{x} = f(x, u), \quad x = \mathbb{R}^n, \\ &x(0) \text{ given}, \quad\ u \in \Omega \subset \mathbb{R}^m, \end{aligned}$$

where $f(x, u) = (f_1(x, u), \dots f_n(x, u)) : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$. We wish to minimize a cost function J with terminal constraints:

$$J = \int_0^T L(x, u) \, dt + V(x(T)), \qquad \psi(x(T)) = 0.$$

The function $\psi : \mathbb{R}^n \to \mathbb{R}^q$ gives a set of q terminal constraints. Analogous to the case of optimizing a function subject to constraints, we construct the *Hamiltonian*:

$$H = L + \lambda^T f = L + \sum \lambda_i f_i.$$

The variables λ are functions of time and are often referred to as the *costate variables*. A set of necessary conditions for a solution to be optimal was derived by Pontryagin [7].

Theorem 4.1 (Maximum Principle)**.** *If (x^*, u^*) is optimal, then there exists $\lambda^*(t) \in \mathbb{R}^n$ and $\nu^* \in \mathbb{R}^q$ such that*

$$\dot{x}_i = \frac{\partial H}{\partial \lambda_i} \qquad -\dot{\lambda}_i = \frac{\partial H}{\partial x_i} \qquad \begin{array}{l} x(0) \textit{ given, } \psi(x(T)) = 0 \\ \lambda(T) = \dfrac{\partial V}{\partial x}(x(T)) + \nu^T \dfrac{\partial \psi}{\partial x} \end{array}$$

and

$$H(x^*(t), u^*(t), \lambda^*(t)) \leq H(x^*(t), u, \lambda^*(t)) \quad \textit{for all} \quad u \in \Omega$$

The form of the optimal solution is given by the solution of a differential equation with boundary conditions. If $u = \arg\min H(x, u, \lambda)$ exists, we can use this to choose the control law u and solve for the resulting feasible trajectory that minimizes the cost. The boundary conditions are given by the n initial states $x(0)$, the q terminal constraints on the state $\psi(x(T)) = 0$ and the $n - q$ final values for the Lagrange multipliers

$$\lambda(T) = \frac{\partial V}{\partial x}(x(T)) + \nu^T \frac{\partial \psi}{\partial x}.$$

In this last equation, ν is a free variable and so there are n equations in $n + q$ free variables, leaving $n - q$ constraints on $\lambda(T)$. In total, we thus have $2n$ boundary values.

The maximum principle is a very general (and elegant) theorem. It allows the dynamics to be nonlinear and the input to be constrained to lie in a set Ω, allowing the possibility of bounded inputs. If $\Omega = \mathbb{R}^m$ (unconstrained input) and H is differentiable, then a necessary condition for the optimal input is

$$\frac{\partial H}{\partial u} = 0.$$

We note that even though we are *minimizing* the cost, this is still usually called the maximum principle (an artifact of history).

Sketch of proof. We follow the proof given by Lewis and Syrmos [5], omitting some of the details required for a fully rigorous proof. We use the method of Lagrange multipliers, augmenting our cost function by the dynamical constraints and the terminal constraints:

$$\begin{aligned} \tilde{J}(x(\,\cdot\,), u(\,\cdot\,), \lambda(\,\cdot\,), \nu) &= J(x, u) + \int_0^T -\lambda^T(t)\big(\dot{x}(t) - f(x, u)\big)\, dt + \nu^T \psi(x(T)) \\ &= \int_0^T \big(L(x, u) - \lambda^T(t)\big(\dot{x}(t) - f(x, u)\big)\, dt \\ &\qquad + V(x(T)) + \nu^T \psi(x(T)). \end{aligned}$$

Note that λ is a function of time, with each $\lambda(t)$ corresponding to the instantaneous constraint imposed by the dynamics. The integral over the interval $[0, T]$ plays the role of the sum of the finite constraints in the regular optimization.

Making use of the definition of the Hamiltonian, the augmented cost becomes

$$\tilde{J}(x(\,\cdot\,), u(\,\cdot\,), \lambda(\,\cdot\,), \nu) = \int_0^T \big(H(x, u) - \lambda^T(t)\dot{x}\big)\, dt + V(x(T)) + \nu^T \psi(x(T)).$$

We can now "linearize" the cost function around the optimal solution $x(t) = x^*(t) + \delta x(t)$, $u(t) = u^*(t) + \delta u(t)$, $\lambda(t) = \lambda^*(t) + \delta\lambda(t)$ and $\nu = \nu^* + \delta\nu$. Taking T as fixed for simplicity (see [5] for the more general case), the incremental cost can be written as

$$\begin{aligned}
\delta\tilde{J} &= \tilde{J}(x^* + \delta x, u^* + \delta u, \lambda^* + \delta\lambda, \nu^* + \delta\nu) - \tilde{J}(x^*, u^*, \lambda^*, \nu^*) \\
&\approx \int_0^T \left(\frac{\partial H}{\partial x}\delta x + \frac{\partial H}{\partial u}\delta u - \lambda^T \delta\dot{x} + \Bigl(\frac{\partial H}{\partial \lambda} - \dot{x}^T\Bigr)\delta\lambda \right) dt \\
&\qquad + \frac{\partial V}{\partial x}\delta x(T) + \nu^T \frac{\partial \psi}{\partial x}\delta x(T) + \delta\nu^T \psi\bigl(x(T), u(T)\bigr),
\end{aligned}$$

where we have omitted the time argument inside the integral and all derivatives are evaluated along the optimal solution.

We can eliminate the dependence on $\delta\dot{x}$ using integration by parts:

$$-\int_0^T \lambda^T \delta\dot{x}\, dt = -\lambda^T(T)\delta x(T) + \lambda^T(0)\delta x(0) + \int_0^T \dot{\lambda}^T \delta x\, dt.$$

Since we are requiring $x(0) = x_0$, the $\delta x(0)$ term vanishes and substituting this into $\delta\tilde{J}$ yields

$$\begin{aligned}
\delta\tilde{J} \approx \int_0^T &\left[\Bigl(\frac{\partial H}{\partial x} + \dot{\lambda}^T\Bigr)\delta x + \frac{\partial H}{\partial u}\delta u + \Bigl(\frac{\partial H}{\partial \lambda} - \dot{x}^T\Bigr)\delta\lambda \right] dt \\
&+ \Bigl(\frac{\partial V}{\partial x} + \nu^T \frac{\partial \psi}{\partial x} - \lambda^T(T)\Bigr)\delta x(T) + \delta\nu^T \psi\bigl(x(T), u(T)\bigr).
\end{aligned}$$

To be optimal, we require $\delta\tilde{J} = 0$ for all δx, δu, $\delta\lambda$ and $\delta\nu$, and we obtain the (local) conditions in the theorem. □

4.3 Examples

To illustrate the use of the maximum principle, we consider a number of analytical examples. Additional examples are given in the exercises.

Example 4.2 (Scalar linear system)**.** Consider the optimal control problem for the system

$$\dot{x} = ax + bu, \tag{S4.3}$$

where $x = \mathbb{R}$ is a scalar state, $u \in \mathbb{R}$ is the input, the initial state $x(t_0)$ is given, and $a, b \in \mathbb{R}$ are positive constants. We wish to find a trajectory $(x(t), u(t))$ that minimizes the cost function

$$J = \tfrac{1}{2}\int_{t_0}^{t_f} u^2(t)\, dt + \tfrac{1}{2}cx^2(t_f),$$

where the terminal time t_f is given and $c > 0$ is a constant. This cost function balances the final value of the state with the input required to get to that state.

To solve the problem, we define the various elements used in the maximum principle. Our integral and terminal costs are given by

$$L = \tfrac{1}{2}u^2(t) \qquad V = \tfrac{1}{2}cx^2(t_f).$$

We write the Hamiltonian of this system and derive the following expressions for the costate λ:

$$H = L + \lambda f = \tfrac{1}{2}u^2 + \lambda(ax + bu)$$
$$\dot{\lambda} = -\frac{\partial H}{\partial x} = -a\lambda, \qquad \lambda(t_f) = \frac{\partial V}{\partial x} = cx(t_f).$$

This is a final value problem for a linear differential equation in λ and the solution can be shown to be

$$\lambda(t) = cx(t_f)e^{a(t_f - t)}.$$

The optimal control is given by

$$\frac{\partial H}{\partial u} = u + b\lambda = 0 \quad \Rightarrow \quad u^*(t) = -b\lambda(t) = -bcx(t_f)e^{a(t_f-t)}.$$

Substituting this control into the dynamics given by equation (S4.3) yields a first-order ODE in x:

$$\dot{x} = ax - b^2cx(t_f)e^{a(t_f-t)}.$$

This can be solved explicitly as

$$x^*(t) = x(t_o)e^{a(t-t_o)} + \frac{b^2c}{2a}x^*(t_f)\left[e^{a(t_f-t)} - e^{a(t+t_f-2t_o)}\right].$$

Setting $t = t_f$ and solving for $x(t_f)$ gives

$$x^*(t_f) = \frac{2a\, e^{a(t_f-t_o)}x(t_o)}{2a - b^2c\left(1 - e^{2a(t_f-t_o)}\right)}$$

and finally we can write

$$u^*(t) = -\frac{2abc\, e^{a(2t_f-t_o-t)}x(t_o)}{2a - b^2c\left(1 - e^{2a(t_f-t_o)}\right)} \tag{S4.4}$$
$$x^*(t) = x(t_o)e^{a(t-t_o)} + \frac{b^2c\, e^{a(t_f-t_o)}x(t_o)}{2a - b^2c\left(1 - e^{2a(t_f-t_o)}\right)}\left[e^{a(t_f-t)} - e^{a(t+t_f-2t_o)}\right]. \tag{S4.5}$$

We can use the form of this expression to explore how our cost function affects the optimal trajectory. For example, we can ask what happens to the terminal state $x^*(t_f)$ and $c \to \infty$. Setting $t = t_f$ in equation (S4.5) and taking the limit we find that

$$\lim_{c\to\infty} x^*(t_f) = 0.$$

Example 4.3 (Bang-bang control)**.** The time-optimal control program for a linear system has a particularly simple solution. Consider a linear system with bounded input

$$\dot{x} = Ax + Bu, \qquad |u| \le 1,$$

and suppose we wish to minimize the time required to move from an initial state x_0 to a final state x_f. Without loss of generality we can take $x_f = 0$. We choose the cost functions and terminal constraints to satisfy

$$J = \int_0^T 1\, dt, \qquad \psi(x(T)) = x(T).$$

To find the optimal control, we form the Hamiltonian

$$H = 1 + \lambda^T(Ax + Bu) = 1 + (\lambda^T A)x + (\lambda^T B)u.$$

Now apply the conditions in the maximum principle:

$$\begin{aligned}
\dot{x} &= \frac{\partial H}{\partial \lambda} = Ax + Bu \\
-\dot{\lambda} &= \frac{\partial H}{\partial x} = A^T \lambda \\
u &= \text{arg min } H = -\text{sgn}(\lambda^T B)
\end{aligned}$$

The optimal solution always satisfies this equation (since the maximum principle gives a necessary condition) with $x(0) = x_0$ and $x(T) = 0$. It follows that the input is always either $+1$ or -1, depending on $\lambda^T B$. This type of control is called "bang-bang" control since the input is always on one of its limits. If $\lambda^T(t)B = 0$ then the control is not well defined, but if this is only true for a specific time instant (e.g., $\lambda^T(t)B$ crosses zero) then the analysis still holds.

4.4 Linear Quadratic Regulators

In addition to its use for computing optimal, feasible trajectories for a system, we can also use optimal control theory to design a feedback law $u = \alpha(x)$ that stabilizes a given equilibrium point. Roughly speaking, we do this by continuously re-solving the optimal control problem from our current state $x(t)$ and applying the resulting input $u(t)$. Of course, this approach is impractical unless we can solve explicitly for the optimal control and somehow rewrite the optimal control as a function of the current state in a simple way. In this section we explore exactly this approach for the linear quadratic optimal control problem.

We begin with the the finite horizon, linear quadratic regulator (LQR) problem, given by

$$\begin{aligned}
\dot{x} &= Ax + Bu, \qquad x \in \mathbb{R}^n, u \in \mathbb{R}^n, x_0 \text{ given}, \\
\tilde{J} &= \frac{1}{2}\int_0^T \left(x^T Q_x x + u^T Q_u u\right)\, dt + \frac{1}{2} x^T(T) P_1 x(T),
\end{aligned}$$

where $Q_x \geq 0$, $Q_u > 0$, $P_1 \geq 0$ are symmetric, positive (semi-) definite matrices. Note the factor of $\frac{1}{2}$ is usually left out, but we included it here to simplify the derivation. (The optimal control will be unchanged if we multiply the entire cost function by 2.)

To find the optimal control, we apply the maximum principle. We begin by computing the Hamiltonian H:

$$H = \frac{1}{2} x^T Q_x x + \frac{1}{2} u^T Q_u u + \lambda^T (Ax + Bu).$$

Applying the results of Theorem 4.1, we obtain the necessary conditions

$$\begin{aligned}
\dot{x} &= \left(\frac{\partial H}{\partial \lambda}\right)^T && = Ax + Bu && x(0) = x_0 \\
-\dot{\lambda} &= \left(\frac{\partial H}{\partial x}\right)^T && = Q_x x + A^T \lambda && \lambda(T) = P_1 x(T) \\
0 &= \frac{\partial H}{\partial u} && = Q_u u + \lambda^T B.
\end{aligned} \tag{S4.6}$$

The last condition can be solved to obtain the optimal controller

$$u = -Q_u^{-1} B^T \lambda,$$

which can be substituted into the dynamic equation (S4.6) To solve for the optimal control we must solve a *two point boundary value problem* using the initial condition $x(0)$ and the final condition $\lambda(T)$. Unfortunately, it is very hard to solve such problems in general.

Given the linear nature of the dynamics, we attempt to find a solution by setting $\lambda(t) = P(t)x(t)$ where $P(t) \in \mathbb{R}^{n \times n}$. Substituting this into the necessary condition, we obtain

$$\begin{aligned} \dot{\lambda} &= \dot{P}x + P\dot{x} = \dot{P}x + P(Ax - BQ_u^{-1}B^T P)x, \\ \implies \quad & -\dot{P}x - PAx + PBQ_u^{-1}BPx = Q_x x + A^T P x. \end{aligned}$$

This equation is satisfied if we can find $P(t)$ such that

$$-\dot{P} = PA + A^T P - PBQ_u^{-1}B^T P + Q_x, \qquad P(T) = P_1. \tag{S4.7}$$

This is a *matrix differential equation* that defines the elements of $P(t)$ from a final value $P(T)$. Solving it is conceptually no different than solving the initial value problem for vector-valued ordinary differential equations, except that we must solve for the individual elements of the matrix $P(t)$ backwards in time. Equation (S4.7) is called the *Riccati ODE.*

An important property of the solution to the optimal control problem when written in this form is that $P(t)$ can be solved without knowing either $x(t)$ or $u(t)$. This allows the two point boundary value problem to be separated into first solving a final-value problem and then solving a time-varying initial value problem. More specifically, given $P(t)$ satisfying equation (S4.7), we can apply the optimal input

$$u(t) = -Q_u^{-1} B^T P(t) x.$$

and then solve the original dynamics of the system forward in time from the initial condition $x(0) = x_0$. Note that this is a (time-varying) *feedback* control that describes how to move from *any* state to the origin in time T.

An important variation of this problem is the case when we choose $T = \infty$ and eliminate the terminal cost (set $P_1 = 0$). This gives us the cost function

$$J = \int_0^\infty (x^T Q_x x + u^T Q_u u)\, dt. \tag{S4.8}$$

Since we do not have a terminal cost, there is no constraint on the final value of λ or, equivalently, $P(t)$. We can thus seek to find a constant P satisfying equation (S4.7). In other words, we seek to find P such that

$$PA + A^T P - PBQ_u^{-1}B^T P + Q_x = 0. \tag{S4.9}$$

This equation is called the *algebraic Riccati equation.* Given a solution, we can choose our input as

$$u = -Q_u^{-1} B^T P x.$$

This represents a *constant gain* $K = Q_u^{-1}B^T P$ where P is the solution of the algebraic Riccati equation.

The implications of this result are interesting and important. First, we notice that if $Q_x > 0$ and the control law corresponds to a finite minimum of the cost, then we must have that $\lim_{t \to \infty} x(t) = 0$, otherwise the cost will be unbounded. This means that the optimal control for moving from any state x to the origin can be achieved by applying a feedback $u = -Kx$ for K chosen as described as above and letting the system evolve in closed loop. More amazingly, the gain matrix K can be written in terms of the solution to a (matrix) quadratic equation (S4.9). This quadratic equation can be solved numerically: in MATLAB the command `K = lqr(A, B, Qx, Qu)` provides the optimal feedback compensator.

In deriving the optimal quadratic regulator, we have glossed over a number of important details. It is clear from the form of the solution that we must have $Q_u > 0$ since its inverse appears in the solution. We would typically also have $Q_x > 0$ so that the integral cost is only zero when $x = 0$, but in some instances we might only care about certain states, which would imply that $Q_x \geq 0$. For this case, if we let $Q_x = H^T H$ (always possible), our cost function becomes

$$J = \int_0^\infty x^T H^T H x + u^T Q_u u \, dt = \int_0^\infty \|Hx\|^2 + u^T Q_u u \, dt.$$

A technical condition for the optimal solution to exist is that the pair (A, H) be *detectable* (implied by observability). This makes sense intuitively by considering $y = Hx$ as an output. If y is not observable then there may be non-zero initial conditions that produce no output and so the cost would be zero. This would lead to an ill-conditioned problem and hence we will require that $Q_x \geq 0$ satisfy an appropriate observability condition.

We summarize the main results as a theorem.

Theorem 4.2. *Consider a linear control system with quadratic cost:*

$$\dot{x} = Ax + Bu, \qquad J = \int_0^\infty x^T Q_x x + u^T Q_u u \, dt.$$

Assume that the system defined by (A, B) is reachable, $Q_x = Q_x^T \geq 0$ and $Q_u = Q_u^T > 0$. Further assume that $Q_x = H^T H$ and that the linear system with dynamics matrix A and output matrix H is observable. Then the optimal controller satisfies

$$u = -Q_u^{-1} B^T P x, \qquad PA + A^T P - PBQ_u^{-1} B^T P = -Q_x,$$

and the minimum cost from initial condition $x(0)$ is given by $J^* = x^T(0) P x(0)$.

The basic form of the solution follows from the necessary conditions, with the theorem asserting that a constant solution exists for $T = \infty$ when the additional conditions are satisfied. The full proof can be found in standard texts on optimal control, such as Lewis and Syrmos [5] or Athans and Falb [1]. A simplified version, in which we first assume the optimal control is linear, is left as an exercise.

Example 4.4 (Optimal control of a double integrator)**.** Consider a double integrator system

$$\frac{dx}{dt} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} x + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u$$

with quadratic cost given by

$$Q_x = \begin{bmatrix} q^2 & 0 \\ 0 & 0 \end{bmatrix}, \qquad Q_u = 1.$$

The optimal control is given by the solution of matrix Riccati equation (S4.9). Let P be a symmetric positive definite matrix of the form

$$P = \begin{bmatrix} a & b \\ b & c \end{bmatrix}.$$

Then the Riccati equation becomes

$$\begin{bmatrix} -b^2 + q^2 & a - bc \\ a - bc & 2b - c^2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix},$$

which has solution

$$P = \begin{bmatrix} \sqrt{2q^3} & q \\ q & \sqrt{2q} \end{bmatrix}.$$

The controller is given by

$$K = Q_u^{-1} B^T P = [q \quad \sqrt{2q}].$$

The feedback law minimizing the given cost function is then $u = -Kx$.

To better understand the structure of the optimal solution, we examine the eigenstructure of the closed loop system. The closed-loop dynamics matrix is given by

$$A_{cl} = A - BK = \begin{bmatrix} 0 & 1 \\ -q & -\sqrt{2q} \end{bmatrix}.$$

The characteristic polynomial of this matrix is

$$\lambda^2 + \sqrt{2q}\lambda + q.$$

Comparing this to $\lambda^2 + 2\zeta\omega_0\lambda + \omega_0^2$, we see that

$$\omega_0 = \sqrt{q}, \qquad \zeta = \frac{1}{\sqrt{2}}.$$

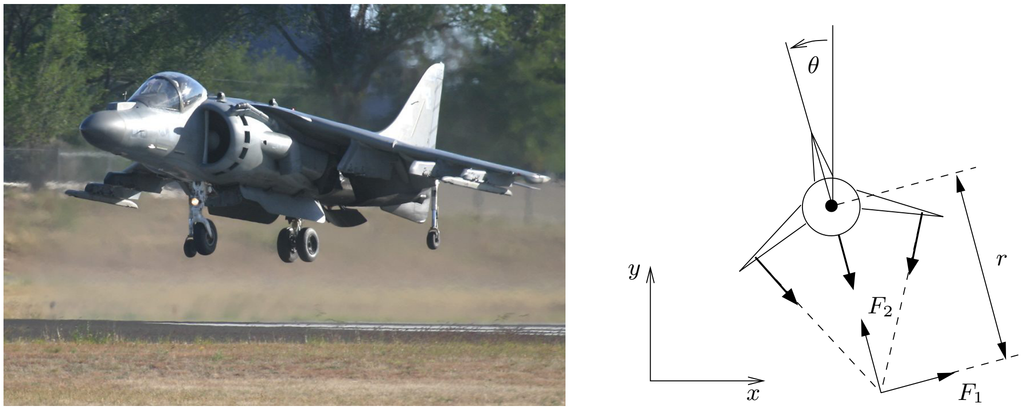
Thus the optimal controller gives a closed loop system with damping ratio $\zeta = 0.707$, giving a good tradeoff between rise time and overshoot (see FBS2e).

4.5 Choosing LQR weights

One of the key questions in LQR design is how to choose the weights Q_x and Q_u. To choose specific values for the cost function weights Q_x and Q_u, we must use our knowledge of the system we are trying to control. A particularly simple choice is to use diagonal weights

$$Q_x = \begin{bmatrix} q_1 & & 0 \\ & \ddots & \\ 0 & & q_n \end{bmatrix}, \qquad Q_u = \begin{bmatrix} \rho_1 & & 0 \\ & \ddots & \\ 0 & & \rho_n \end{bmatrix}.$$

For this choice of Q_x and Q_u, the individual diagonal elements describe how much each state and input (squared) should contribute to the overall cost. Hence, we can take states that should remain small and attach higher weight values to them. Similarly, we can penalize an input versus the states and other inputs through choice of the corresponding input weight ρ_j.



(a) Harrier "jump jet" (b) Simplified model

Figure S4.3: Vectored thrust aircraft. The Harrier AV-8B military aircraft (a) redirects its engine thrust downward so that it can "hover" above the ground. Some air from the engine is diverted to the wing tips to be used for maneuvering. As shown in (b), the net thrust on the aircraft can be decomposed into a horizontal force F_1 and a vertical force F_2 acting at a distance r from the center of mass.

Choosing the individual weights for the (diagonal) elements of the Q_x and Q_u matrix can be done by deciding on a weighting of the errors from the individual terms. Bryson and Ho [2] have suggested the following method for choosing the matrices Q_x and Q_u in equation (S4.8): (1) choose q_i and ρ_j as the inverse of the square of the maximum value for the corresponding x_i or u_j; (2) modify the elements to obtain a compromise among response time, damping and control effort. This second step can be performed by trial and error.

It is also possible to choose the weights such that only a given subset of variable are considered in the cost function. Let $z = Hx$ be the output we want to keep small and verify that (A, H) is observable. Then we can use a cost function of the form

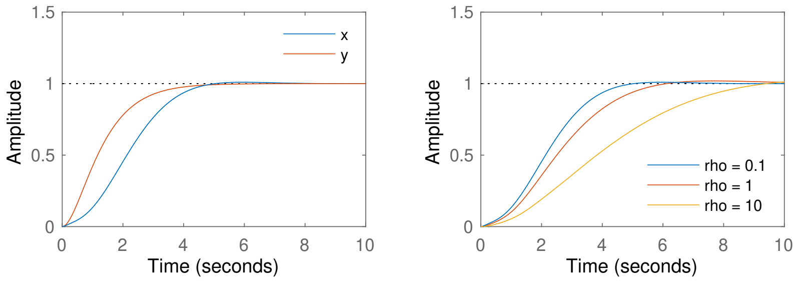
$$Q_x = H^T H \qquad Q_u = \rho I.$$

The constant ρ allows us to trade off $\|z\|^2$ versus $\rho\|u\|^2$.

We illustrate the various choices through an example application.

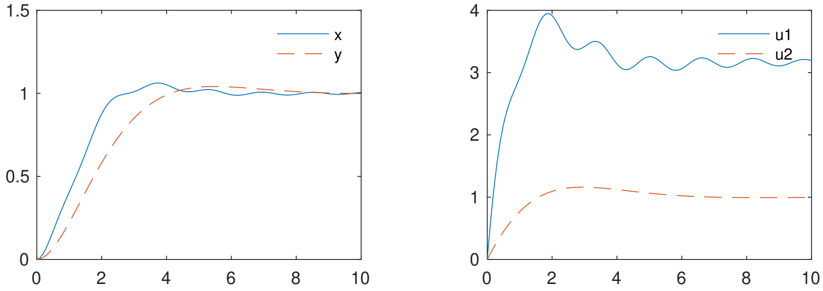
Example 4.5 (Thrust vectored aircraft)**.** Consider the thrust vectored aircraft example introduced in FBS2e, Example 2.9. The system is shown in Figure S4.3, reproduced from FBS2e. The linear quadratic regulator problem was illustrated in Example 6.8, where the weights were chosen as $Q_x = I$ and $Q_u = \rho I$. Figure S4.4 reproduces the step response for this case.

A more physically motivated weighted can be computing by specifying the comparable errors in each of the states and adjusting the weights accordingly. Suppose, for example that we consider a 1 cm error in x, a 10 cm error in y and a 5° error in θ to be equivalently bad. In addition, we wish to penalize the forces in the sidewards direction (F_1) since these results in a loss in efficiency.



(a) Step response in x and y (b) Effect of control weight ρ

Figure S4.4: Step response for a vectored thrust aircraft. The plot in (a) shows the x and y positions of the aircraft when it is commanded to move 1 m in each direction. In (b) the x motion is shown for control weights $\rho = 1$, 10^2, 10^4. A higher weight of the input term in the cost function causes a more sluggish response.



(a) Step response in x and y (b) Inputs for the step response

Figure S4.5: Step response for a vector thrust aircraft using physically motivated LQR weights (a). The rise time for x is much faster than in Figure S4.4a, but there is a small oscillation and the inputs required are quite large (b).

This can be accounted for in the LQR weights be choosing

$$Q_x = \begin{bmatrix} 100 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2\pi/9 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \qquad Q_u = \begin{bmatrix} 10 & 0 \\ 0 & 1 \end{bmatrix}.$$

The results of this choice of weights are shown in Figure S4.5.

4.6 Advanced Topics

In this section we briefly touch on some related topics in optimal control, with reference to more detailed treatments where appropriate.

Dynamic programming

General quadratic cost functions

$$\dot{x} = Ax + Bu \qquad J = \int_0^\infty \overbrace{\Big(x^T Q_x x + u^T Q_u u + x^T S u\Big)}^{L(x,u)}\, dt,$$

where the S term is almost always left out.

Singular extremals The necessary conditions in the maximum principle enforce the constraints through the of the Lagrange multipliers $\lambda(t)$. In some instances, we can get an extremal curve that has one or more of the λ's identically equal to zero. This corresponds to a situation in which the constraint is satisfied strictly through the minimization of the cost function and does not need to be explicitly enforced. We illustrate this case through an example.

Example 4.6 (Nonholonomic integrator)**.** Consider the minimum time optimization problem for the nonholonomic integrator introduced in Example **??** with input constraints $|u_i| \leq 1$. The Hamiltonian for the system is given by

$$H = 1 + \lambda_1 u_1 + \lambda_2 u_2 + \lambda_3 x_2 u_1$$

and the resulting equations for the Lagrange multipliers are

$$\dot{\lambda}_1 = 0, \qquad \dot{\lambda}_2 = \lambda_3 x_2, \qquad \dot{\lambda}_3 = 0. \tag{S4.10}$$

It follows from these equations that λ_1 and λ_3 are constant. To find the input u corresponding to the extremal curves, we see from the Hamiltonian that

$$u_1 = -\text{sgn}(\lambda_1 + \lambda_3 x_2 u_1), \qquad u_2 = -\text{sgn}\lambda_2.$$

These equations are well-defined as long as the arguments of $\text{sgn}(\,\cdot\,)$ are non-zero and we get switching of the inputs when the arguments pass through 0.

An example of an abnormal extremal is the optimal trajectory between $x_0 = (0,0,0)$ to $x_f = (\rho, 0, 0)$ where $\rho > 0$. The minimum time trajectory is clearly given by moving on a straight line with $u_1 = 1$ and $u_2 = 0$. This extremal satisfies the necessary conditions but with $\lambda_2 \equiv 0$, so that the "constraint" that $\dot{x}_2 = u_2$ is not strictly enforced through the Lagrange multipliers.

Abnormal extremals

$$\dot{x}_1 = u_1, \qquad \dot{x}_2 = u_2, \qquad \dot{x}_3 = x_2 u_1.$$

Minimum length curve:

$$J = \int_0^T u_1^2 + u_2^2\, dt$$

Show that x_0 to x_f gives multiple solutions.

4.7 Further Reading

There are a number of excellent books on optimal control. One of the first (and best) is the book by Pontryagin et al. [7]. During the 1960s and 1970s a number of additional books were written that provided many examples and served as standard textbooks in optimal control classes. Athans and Falb [1] and Bryson and Ho [2] are two such texts. A very elegant treatment of optimal control from the point of view of optimization over general linear spaces is given by Luenberger [6]. Finally, a modern engineering textbook that contains a very compact and concise derivation of the key results in optimal control is the book by Lewis and Syrmos [5].

Exercises

4.1 (a) Let $G_1, G_2, \dots, G_k$ be a set of row vectors on a $\mathbb{R}^n$. Let F be another row vector on $\mathbb{R}^n$ such that for every $x \in \mathbb{R}^n$ satisfying $G_i x = 0$, $i = 1, \dots, k$, we have $Fx = 0$. Show that there are constants $\lambda_1, \lambda_2, \dots, \lambda_k$ such that

$$F = \sum_{i=1}^{k} \lambda_i G_i.$$

(b) Let $x^* \in \mathbb{R}^n$ be an the extremal point (maximum or minimum) of a function f subject to the constraints $g_i(x) = 0$, $i = 1, \dots, k$. Assuming that the gradients $\partial g_i(x^*)/\partial x$ are linearly independent, show that there are k scalers λ_i, $i = 1, \dots, k$ such that the function

$$\tilde{f}(x) = f(x) + \sum_{i=1}^{k} \lambda_i g_i(x)$$

has an extremal point at x^*.

4.2 Consider the following control system

$$\begin{aligned} \dot{q} &= u \\ \dot{Y} &= q u^T - u q^T \end{aligned}$$

where $u \in \mathbb{R}^m$ and $Y \in \mathbb{R}^{m \times m}$ is a skew symmetric matrix, $Y^T = Y$.

(a) For the fixed end point problem, derive the form of the optimal controller minimizing the following integral

$$\frac{1}{2} \int_0^1 u^T u \, dt.$$

(b) For the boundary conditions $q(0) = q(1) = 0$, $Y(0) = 0$ and

$$Y(1) = \begin{bmatrix} 0 & -y_3 & y_2 \\ y_3 & 0 & -y_1 \\ -y_2 & y_1 & 0 \end{bmatrix}$$

for some $y \in \mathbb{R}^3$, give an explicit formula for the optimal inputs u.

(c) (Optional) Find the input u to steer the system from $(0, 0)$ to $(0, \tilde{Y}) \in \mathbb{R}^m \times \mathbb{R}^{m\times m}$ where $\tilde{Y}^T = -\tilde{Y}$.

(Hint: if you get stuck, there is a paper by Brockett on this problem.)

4.3 In this problem, you will use the maximum principle to show that the shortest path between two points is a straight line. We model the problem by constructing a control system

$$\dot{x} = u,$$

where $x \in \mathbb{R}^2$ is the position in the plane and $u \in \mathbb{R}^2$ is the velocity vector along the curve. Suppose we wish to find a curve of minimal length connecting $x(0) = x_0$ and $x(1) = x_f$. To minimize the length, we minimize the integral of the velocity along the curve,

$$J = \int_0^1 \|\dot{x}\|\,dt = \int_0^1 \sqrt{\dot{x}^T\dot{x}}\,dt,$$

subject to to the initial and final state constraints. Use the maximum principle to show that the minimal length path is indeed a straight line at maximum velocity. (Hint: try minimizing using the integral cost $\dot{x}^T\dot{x}$ first and then show this also optimizes the optimal control problem with integral cost $\|\dot{x}\|$.)

4.4 Consider the optimal control problem for the system

$$\dot{x} = -ax + bu,$$

where $x = \mathbb{R}$ is a scalar state, $u \in \mathbb{R}$ is the input, the initial state $x(t_0)$ is given, and $a, b \in \mathbb{R}$ are positive constants. (Note that this system is not quite the same as the one in Example 4.2.) The cost function is given by

$$J = \tfrac{1}{2}\int_{t_0}^{t_f} u^2(t)\,dt + \tfrac{1}{2}cx^2(t_f),$$

where the terminal time t_f is given and c is a constant.

(a) Solve explicitly for the optimal control $u^*(t)$ and the corresponding state $x^*(t)$ in terms of t_0, t_f, $x(t_0)$ and t and describe what happens to the terminal state $x^*(t_f)$ as $c \to \infty$.

(b) Show that the system is differentially flat with appropriate choice of output(s) and compute the state and input as a function of the flat output(s).

(c) Using the polynomial basis $\{t^k,\ k = 0, \dots, M-1\}$ with an appropriate choice of M, solve for the (non-optimal) trajectory between $x(t_0)$ and $x(t_f)$. Your answer should specify the explicit input $u_d(t)$ and state $x_d(t)$ in terms of t_0, t_f, $x(t_0)$, $x(t_f)$ and t.

(d) Let $a = 1$ and $c = 1$. Use your solution to the optimal control problem and the flatness-based trajectory generation to find a trajectory between $x(0) = 0$ and $x(1) = 1$. Plot the state and input trajectories for each solution and compare the costs of the two approaches.

(e) (Optional) Suppose that we choose more than the minimal number of basis functions for the differentially flat output. Show how to use the additional degrees of freedom to minimize the cost of the flat trajectory and demonstrate that you can obtain a cost that is closer to the optimal.

4.5 Repeat Exercise 4.4 using the system

$$\dot{x} = -ax^3 + bu.$$

For part (a) you need only write the conditions for the optimal cost.

4.6 Consider the problem of moving a two-wheeled mobile robot (e.g., a Segway) from one position and orientation to another. The dynamics for the system is given by the nonlinear differential equation

$$\dot{x} = \cos\theta\, v, \qquad \dot{y} \ = \sin\theta\, v, \qquad \dot{\theta} = \omega,$$

where (x, y) is the position of the rear wheels, θ is the angle of the robot with respect to the x axis, v is the forward velocity of the robot and ω is spinning rate. We wish to choose an input (v, ω) that minimizes the time that it takes to move between two configurations (x_0, y_0, θ_0) and (x_f, y_f, θ_f), subject to input constraints $|v| \leq L$ and $|\omega| \leq M$.

Use the maximum principle to show that any optimal trajectory consists of segments in which the robot is traveling at maximum velocity in either the forward or reverse direction, and going either straight, hard left ($\omega = -M$) or hard right ($\omega = +M$).

Note: one of the cases is a bit tricky and cannot be completely proven with the tools we have learned so far. However, you should be able to show the other cases and verify that the tricky case is possible.

4.7 Consider a linear system with input u and output y and suppose we wish to minimize the quadratic cost function

$$J = \int_0^\infty \left(y^T y + \rho u^T u\right)\, dt.$$

Show that if the corresponding linear system is observable, then the closed loop system obtained by using the optimal feedback $u = -Kx$ is guaranteed to be stable.

4.8 Consider the system transfer function

$$H(s) = \frac{s+b}{s(s+a)}, \qquad a, b > 0$$

with state space representation

$$\dot{x} = \begin{bmatrix} 0 & 1 \\ 0 & -a \end{bmatrix} x + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u,$$
$$y = \begin{bmatrix} b & 1 \end{bmatrix} x$$

and performance criterion

$$V = \int_0^\infty (x_1^2 + u^2) dt.$$

(a) Let

$$P = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix},$$

with $p_{12} = p_{21}$ and $P > 0$ (positive definite). Write the steady state Riccati equation as a system of four explicit equations in terms of the elements of P and the constants a and b.

(b) Find the gains for the optimal controller assuming the full state is available for feedback.

4.9 Consider the optimal control problem for the system

$$\dot{x} = ax + bu \qquad J = \tfrac{1}{2}\int_{t_0}^{t_f} u^2(t)\,dt + \tfrac{1}{2}cx^2(t_f),$$

where $x \in \mathbb{R}$ is a scalar state, $u \in \mathbb{R}$ is the input, the initial state $x(t_0)$ is given, and $a, b \in \mathbb{R}$ are positive constants. We take the terminal time t_f as given and let $c > 0$ be a constant that balances the final value of the state with the input required to get to that position. The optimal trajectory is derived in Example 4.2.

Now consider the infinite horizon cost

$$J = \tfrac{1}{2}\int_{t_0}^{\infty} u^2(t)\,dt$$

with $x(t)$ at $t = \infty$ constrained to be zero.

(a) Solve for $u^*(t) = -bPx^*(t)$ where P is the positive solution corresponding to the algebraic Riccati equation. Note that this gives an explicit feedback law ($u = -bPx$).

(b) Plot the state solution of the finite time optimal controller for the following parameter values

$$\begin{aligned} a &= 2, \qquad b = 0.5, \qquad x(t_0) = 4, \\ c &= 0.1,\ 10, \qquad t_f = 0.5,\ 1,\ 10. \end{aligned}$$

(This should give you a total of 6 curves.) Compare these to the infinite time optimal control solution. Which finite time solution is closest to the infinite time solution? Why?

4.10 Consider the lateral control problem for an autonomous ground vehicle from Example **??**. We assume that we are given a reference trajectory $r = (x_d, y_d)$ corresponding to the desired trajectory of the vehicle. For simplicity, we will assume that we wish to follow a straight line in the x direction at a constant velocity $v_d > 0$ and hence we focus on the y and θ dynamics:

$$\dot{y} = \sin\theta\, v_d, \qquad \dot{\theta} \;= \frac{1}{l}\tan\phi\, v_d.$$

We let $v_d = 10$ m/s and $l = 2$ m.

(a) Design an LQR controller that stabilizes the position y to $y_d = 0$. Plot the step and frequency response for your controller and determine the overshoot, rise time, bandwidth and phase margin for your design. (Hint: for the frequency domain specifications, break the loop just before the process dynamics and use the resulting SISO loop transfer function.)

(b) Suppose now that $y_d(t)$ is not identically zero, but is instead given by $y_d(t) = r(t)$. Modify your control law so that you track $r(t)$ and demonstrate the performance of your controller on a "slalom course" given by a sinusoidal trajectory with magnitude 1 meter and frequency 1 Hz.

Chapter 5

State Estimation

Preliminary reading The material in this chapter extends the material Chapter 8 of FBS2e. Readers should be familiar with the material in Sections 8.1–8.3 in preparation for the more advanced topics discussed here.

5.1 Concepts and Definitions

Let $\mathcal{D} = (\mathcal{U}, \sigma, \mathcal{Y}, s, r)$ be an input/output dynamical system on time range $\mathcal{T}$ with input/output map

$$\rho(t, t_0, x_0, u(\,\cdot\,)) = r(t, s(t, t_0, x_0, u(\,\cdot\,)), u(t)).$$

Let $x_0, z_0 \in \Sigma$ be different initial conditions and let $t_1, t_2 \in \mathcal{T}$ be two different times with $t_1 < t_2$. We write $\mathcal{U}_{[t_1,t_2]}$ represent input signals that are restricted to the time range $[t_1, t_2] \in \mathcal{T}$.

Definition 5.1 (Distinguishability)**.** A control $u \in \mathcal{U}_{[t_1,t_2]}$ *distinguishes between* $x(t_0) = x_0$ *and* $x(t_0) = z_0$ if

$$\rho(t_2, t_0, x_0, u(\,\cdot\,)) \neq \rho(t_2, t_0, z_0, u(\,\cdot\,).$$

The initial states (x_0, t_0) and (z_0, t_0) are *distinguishable on* $[t_1, t_2]$ if there exists $t \in [t_1, t_2]$ and input $u \in \mathcal{U}_{[t_1,t]}$ that distinguishes between x_0 and z_0 on $[t_0, t]$.

Definition 5.2 (Observability)**.** The system $\mathcal{D}$ is *observable on* $[t_1, t_2]$ if every pair of initial states $x_0, z_0 \in \Sigma$ are distinguishable. The system is *observable* if for every $t_1 < t_2$, $\mathcal{D}$ is observable on $[t_1, t_2]$.

If a system is observable then for any initial condition x_0 there exists an input $u(\,\cdot\,)$ that can be used to uniquely determine the initial condition by measurement of the output over an (arbitrary) interval of time. For many systems (including linear systems), the initial state can be determined for any input that is applied to the system.

For an observable system, it is possible to create a *state estimator* (also called observer) that measures the input $u(t)$ and output $y(t)$ of the system and provides an estimate $\hat{x}(t)$ of the current state. For a nonlinear system of the form

$$\frac{dx}{dt} = f(x, u), \qquad y = h(x)$$

a common form for an estimator is to construct a copy of the system dynamics and update the estimated state based on the error between the predicted output and the measured output:

$$\frac{d\hat{x}}{dt} = f(\hat{x}, u) + \alpha(y - h(\hat{x})).$$

We see that the estimator requires the current input u in order to update the estimated state to match the dynamics of the underlying model.

5.2 Observability for Linear State Space Systems

For linear systems we have that

$$\rho(t, t_0, x_0, u(\,\cdot\,)) = Ce^{At}x_0 + \int_{t_0}^{t} Ce^{A(t-\tau)Bu(\tau)}\, d\tau.$$

It follows that whether two initial states are distinguishable is independent of the input $u(\,\cdot\,)$ since the effect of the input is the same for any initial condition. Hence observability depends only on the pair (A, C), and we say that the system is observable if the pair (A, C) is observable.

In FBS2e, a simple characterization of observability is given by considering the output $y(t) = Cx(t)$ and its derivatives, leading to the observability rank condition. A more insightful analysis is obtained by considering the linear operator $\mathcal{M}_T : \mathbb{R}^n \to \mathcal{Y}$ given by

$$\big(\mathcal{M}_T(x_0)\big)(t) = Ce^{At}x_0.$$

The question of observability is equivalent to whether the map $\mathcal{M}_T$ is an injection (one-to-one) so that given any output $y(t)$ in the range of $\mathcal{M}_T$ there exists a unique x_0 such that $y(t) = \mathcal{M}_T x_0$.

To characterize the injectivity of $\mathcal{M}_T$ we compute the adjoint operator $\mathcal{M}_T^* : \mathcal{Y}^* \to \mathbb{R}^n$, which can be shown to be

$$\mathcal{M}_T^*(y(\,\cdot\,)) = \int_{t_0}^{T} y^\mathsf{T}(\tau) C e^{A\tau}\, d\tau.$$

The map $\mathcal{M}_T$ is injective if its rank is equal to n and the rank of $\mathcal{M}_T$ is equal to the rank of the operator $\mathcal{M}_T^*\mathcal{M}_T$, given by

$$W_\text{o}(T) = \int_{t_0}^{T} e^{A^\mathsf{T}\tau} C^T C e^{A\tau}\, d\tau.$$

W_o is an $n \times n$ square, symmetric matrix. In the case that $W_\text{o}(T)$ is not full rank, the null space of $W_\text{o}(T)$ gives the subspace of initial conditions whose values cannot be distinguished through measurement of the output. As in the case of reachability, it can be shown that the rank of $W_\text{o}(T)$ is independent of T.

Given the input $u(\,\cdot\,)$ and the output $y(\,\cdot\,)$ on an interval $[t_0, T]$ the values of the initial state can be computed using W_o. Assume first that $u(\,\cdot\,) = 0$ so that the system dynamics are given by $y(t) = Ce^{At}x_0$. The value for x_0 is given by

$$x_0 = (\mathcal{M}_T^*\mathcal{M}_T)^{-1}\mathcal{M}_T^* y(\,\cdot\,) = \big(W_0(T)\big)^{-1} \int_{t_0}^{T} y^\mathsf{T}(\tau) C e^{A\tau}\, d\tau.$$

In the case that the model is not correct and so the output trajectory is not in the range of $\mathcal{M}_T$, this estimate represents the best estimate (in a least squares sense) of the initial state. If the system has a nonzero input, then this estimate should be applied to the function

$$\tilde{y}(t) = y(t) - \int_{t_0}^{t} e^{A(t-\tau)} Bu(\tau)\, d\tau$$

which removes the contribution of the input from the measured state $y(t)$. △

If a linear system is stable, we can compute the *observability Gramian* $W_\text{o} \in \mathbb{R}^{n\times n}$:

$$W_\text{o} = \lim_{T\to\infty} W_\text{o}(T) = \int_{t_0}^{\infty} e^{A^\mathsf{T}\tau} C^T C e^{A\tau}\, d\tau.$$

A (stable) linear system is observable if and only if the observabilty Gramian has rank n. The observability Gramian can be computed using linear algebra:

Theorem 5.1. $A^\mathsf{T} W_o + W_o A = -C^\mathsf{T} C.$

As in the case of reachability, there are a number of equivalent conditions for observability of a linear system.

Theorem 5.2 (Observability conditions)**.** *A linear system with dynamics matrix A and output matrix C is observable if an only if the following equivalent conditions hold:*

1. *$W_o(T)$ has rank n for any $T > 0$.*
2. *W_o has rank n (requires A stable).*
3. *Popov-Bellman-Hautus (PBH) test:*

$$\mathit{rank} \begin{bmatrix} C \\ sI - A \end{bmatrix} = n$$

for all $s \in \mathbb{C}$ (suffices to check for eigenvalue of A).

4. *Observability rank test:*

$$\mathit{rank} \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix} = n.$$

5.3 Combining Estimation and Control

As discussed in FBS2e, a state estimator can be combined with a state feedback controller to design the dynamics of a reachable and observable linear system. In the case that a system is not reachable and/or not observable, the ability to stabilize the system to an equilibrium point will depend on whether the unreachable modes have unstable eigenvalues and the ability to detect the system state around an equilibrium will depend on whether the unobservable modes have stable eigenvalues. The *Kalman decomposition*, described briefly in Section 8.3 can be used to understand the structure of the system. We expand on this description here by describing some of the properties of the subspaces in the Kalman decomposition.

Consider a linear system in standard form

$$\frac{dx}{dt}x = Ax + Bu, \qquad y = Cx.$$

Assume that the system is neither reachable nor observable. As we saw in Section 3.2, if a system is not reachable then there exists a subspace E_{r} of reachable states. It can be shown that this subspace is A invariant (easy proof: look at the structure of A in Theorem 3.6) and $E_{\mathrm{r}} = \operatorname{range} W_{\mathrm{r}}$. Similarly, we see from the analysis above that there exists a subspace $E_{\overline{\mathrm{o}}}$ of states that are not observable, consisting of states in the null space of $W_{\mathrm{o}}(T)$. This subspace is also A invariant (Exercise **??**).

Another characerization of E_{r} and $E_{\overline{\mathrm{o}}}$ is that E_{r} is the smallest A-invariant subspace containing B which is equivalent to range Wc. Similarly, $E_{\overline{\mathrm{o}}}$ is the smallest A-invariant subspace that annihilates C. We can now decompose the state space as $\mathbb{R}^n = E_{\mathrm{ro}} \oplus E_{\mathrm{r}\overline{\mathrm{o}}} \oplus E_{\overline{\mathrm{r}}\mathrm{o}} \oplus E_{\overline{\mathrm{r}}\overline{\mathrm{o}}}$, where $A \oplus B = \{x + y : x \in A, y \in B\}$. Define $E_{\mathrm{r}\overline{\mathrm{o}}} = E_{\mathrm{r}} \cap E_{\overline{\mathrm{o}}}$. This space is uniquely defined and is A invariant.

Using these subspaces, it is possible to construct a transformation that extends the decomposition in Theorem 3.6 to account for both reachability and observability.

Theorem 5.3 (Kalman decomposition)**.** *Let (A, B, C) represent an input/output linear system with n states and define $r = rank W_c \leq n$ and $q = nullity W_o \leq n$. Then there exists a transformation $T \in \mathbb{R}^{n \times n}$ such that the dynamics can be transformed into the block matrix form*

$$\begin{aligned}
\frac{dx}{dt} &= \begin{bmatrix} A_{ro} & 0 & * & 0 \\ * & A_{r\overline{o}} & * & * \\ 0 & 0 & A_{\overline{r}o} & 0 \\ 0 & 0 & * & A_{\overline{r}\overline{o}} \end{bmatrix} + \begin{bmatrix} B_{ro} \\ B_{r\overline{o}} \\ 0 \\ 0 \end{bmatrix} \\
y &= \begin{bmatrix} C_{ro} & 0 & C_{\overline{r}o} & 0 \end{bmatrix},
\end{aligned}$$

where $*$ represents non-zero elements. Furthermore, in this basis:

$$E_{ro} = \operatorname{span} \begin{bmatrix} I \\ 0 \\ 0 \\ 0 \end{bmatrix}, \qquad E_{r\overline{o}} = \operatorname{span} \begin{bmatrix} 0 \\ I \\ 0 \\ 0 \end{bmatrix}, \qquad E_{\overline{r}o} = \operatorname{span} \begin{bmatrix} 0 \\ 0 \\ I \\ 0 \end{bmatrix}, \qquad E_{\overline{r}\overline{o}} = \operatorname{span} \begin{bmatrix} 0 \\ 0 \\ 0 \\ I \end{bmatrix}$$

Proof. (Sketch) Choose E_{ro} such that $E_{\mathrm{r}} = E_{\mathrm{ro}} \oplus E_{\mathrm{r}\overline{\mathrm{o}}}$. This space is not uniquely defined and is not A invariant, but can be constructed by completing the basis for E_{r}. Choose $E_{\overline{\mathrm{r}}\overline{\mathrm{o}}}$ such that $E_{\overline{\mathrm{o}}} = E_{\mathrm{r}\overline{\mathrm{o}}} \oplus E_{\overline{\mathrm{r}}\overline{\mathrm{o}}}$. This space is also not uniquely defined nor is it A invariant, but can be constructed by completing the basis for $E_{\overline{\mathrm{o}}}$. Finally, choose $E_{\overline{\mathrm{r}}\mathrm{o}}$ such that $\mathbb{R}^n = E_{\mathrm{ro}} \oplus E_{\mathrm{r}\overline{\mathrm{o}}} \oplus E_{\overline{\mathrm{r}}\mathrm{o}} \oplus E_{\overline{\mathrm{r}}\overline{\mathrm{o}}}$. $E_{\overline{\mathrm{r}}\mathrm{o}}$ is not uniquely defined and is not A invariant, but can be constructed by completing the basis for $\mathbb{R}^n$. The transformation T is constructed by using the basis elements for each subspace and rewriting the dynamics in terms of this basis. □

Chapter 6

Transfer Functions

6.1 State Space Realizations of Transfer Functions

Theorem 6.1. *Given any proper transfer function $\hat{G}(s)$, there exists a state space representation (A, B, C, D) such that $G(s) = C(sI - A)^{-1}B + D$. If $\hat{G}(s)$ is strictly proper, then $D = 0$.*

The representation for a transfer function is not unique. In particular, given any invertible map $T : \mathbb{R}^n \to \mathbb{R}^n$ define a new system

$$\begin{aligned} A' &= T^{-1}AT \\ B' &= T^{-1}B \\ C' &= CT \\ D' &= D \end{aligned} \qquad \Longrightarrow \qquad \begin{aligned} G'(s) &= CT\big(sI - (T^{-1}AT)\big)^{-1}T^{-1}B \\ &= CT\big(T^{-1}(sI - A)^{-1}T\big)T^{-1}B \\ &= C(sI - A)^{-1}B = G(s) \end{aligned}$$

A realization (A, B, C, D) is said to be *minimal* if there exists no other realization with a state space of smaller dimension. It can be shown that the minimum number of states in a realization is equal to the number of polels of $\hat{G}(s)$.

Bibliography

[1] M. Athans and P. L. Falb. *Optimal Control: An Introduction to the Theory and Its Applications.* Dover, 2006. Originally published in 1963.

[2] A. E. Bryson, Jr. and Y.-C. Ho. *Applied Optimal Control: Optimization, Estimation, and Control.* Wiley, New York, 1975.

[3] F. M. Callier and C. A. Desoer. *Linear System Theory.* Springer-Verlag, 1991.

[4] J. C. Doyle, B. A. Francis, and A. R. Tannenbaum. *Feedback Control Theory.* Macmillan Publishing Company, 1992.

[5] F. L. Lewis and V. L. Syrmos. *Optimal Control.* Wiley, second edition, 1995.

[6] D. G. Luenberger. *Optimization by Vector Space Methods.* Wiley, New York, 1997.

[7] L. S. Pontryagin, V. G. Boltyanskii, R. V. Gamkrelidze, and E. F. Mishchenko. *The Mathematical Theory of Optimal Processes.* Wiley-Interscience, 1962. (translated from Russian).

Index

algebraic Riccati equation, 50
asymptotic stability
    discrete-time systems, 28

bang-bang control, 49

center subspace, 22
characteristic polynomial, 39
    reachable canonical form, 38
control matrix, 10
control signal, 10
convolution equation, 27
    discrete-time, 28
cost function, 41
costate variables, 45

direct term, 10
dynamics matrix, 10

eigenvalues
    for discrete-time systems, 28
extremum, 44

feasible trajectory, 44
final cost, 44
finite horizon, 45
flow invariant, 22
frequency response, 21

Hamiltonian, 45
Harrier AV-8B aircraft, 53

infinite horizon, 45
integral cost, 44
invariant subspace, 22

Lagrange multipliers, 43
linear quadratic, 45
linear systems, 10

matrix differential equation, 50
measured signals, 10
mechanical systems, 10

nonlinear systems, 10
    linear approximation, 28

optimal control problem, 44
optimal value, 41
optimization, 41

Parseval's theorem, 13

Riccati ODE, 50

sensor matrix, 10
stable subspace, 22
state, of a dynamical system, 10
steady-state response
    discrete-time systems, 28
superposition, 12

terminal cost, 44
transfer function, 21
two point boundary value problem, 50

unstable subspace, 22