Linear quadratic and model predictive control

Lecture notes for EL2700

Mikael Johansson

Contents

Part One: System Theory

1. Discrete-time linear systems

This first chapter summarizes some basic theory for discrete-time linear systems. Particular attention is given to understanding how the state-equations can be used to predict the future evolution of the system, and under what conditions we are able to find a control signal which transfers the state from any initial value to any future target. We also discuss more traditional concepts of stability, the influence of pole and zero locations on the transient response, and how to design linear controllers using pole placement. In parallel, we also explore if and how we can estimate the complete state vector from measurement of a few output signals. Finally, we will elaborate on how to use theory for discrete-time linear systems to control and underlying continuous-time system.

1.1 State-space equations, system response and stability

We consider discrete-time linear systems on the form

$$\begin{aligned} x_{t+1} &= Ax_t + Bu_t \\ y_t &= Cx_t + Du_t \end{aligned} \tag{1.1}$$

Here, $x_t \in \mathbb{R}^n$ is the *state vector* at time t, $u_t \in \mathbb{R}^m$ is the *control vector*, and $y_t \in \mathbb{R}^p$ is the *output vector*. The time index $t \in \mathbb{Z}$ is integer valued, so (1.1) describes how an input sequence $\{u_0, u_1, \dots, \}$ and the internal system dynamics affect the state evolution $\{x_1, x_2, \dots\}$, starting from an initial state x_0 at time $t = 0$. It is often convenient to use the short-hand notation (A, B, C, D) for a system whose state-space representation is given by (1.1).

The free system response

Let us first consider the response of the system when $u_t = 0$ for all t, *i.e.* the response of the autonomous system

$$x_{t+1} = Ax_t. \tag{1.2}$$

The solution is found by repeated application of the system equations from a given initial state x_0:

$$\begin{aligned} x_1 &= Ax_0 \\ x_2 &= Ax_1 = A^2x_0 \\ &\vdots \\ x_t &= A^t x_0 \end{aligned}$$

The free system response may converge to zero, diverge, or maintain a sustained oscillation. The following simple example illustrates this fact.

■ **Example 1.1** At time $t = 0$, you place w_0 dollars in a bank account with interest rate $r\%$. Your wealth w_t then evolves according to

$$w_{t+1} = aw_t$$

where $a = 1 + r/100$. The free response is

$$w_t = a^t w_0.$$

Clearly w_t stays constant if $a = 1$ (corresponding to zero interest rate), diverges if $|a| > 1$ (corresponding to a positive interest rate) and converges to zero if $|a| < 1$ (negative interest rate). Trajectories of the system for different values of a are shown in Figure 1.1.

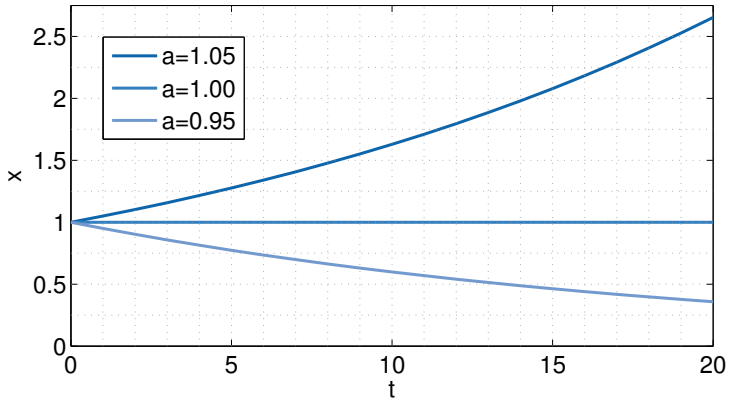

Figure 1.1:

■

Our main interest will be to ensure that $\|x(t)\| \to 0$ as $t \to \infty$, irrespectively of the initial value. This system property is referred to as asymptotic stability:

Definition 1.1.1 — Asymptotic stability. The autonomous discrete-time linear system (1.2) is *globally asymptotically stable* if its solution $\{x_t\}$ satisfies $\|x_t\| \to 0$ as $t \to \infty$ for all $x_0 \in \mathbb{R}^n$.

The qualitative properties in the scalar example carry over to the general vector case. This is obvious when A is diagonal, *i.e.* $A = \text{diag}(a_{11}, \ldots, a_{nn})$, since this implies that $[x_t]_i = a_{ii}^t [x_0]_i$. It is also easy to prove when A has n linearly independent eigenvectors v_i with associated eigenvalues λ_i. In this case

$$A = V^{-1}\Lambda V$$

where $V = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}$ and $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$. Hence

$$A^t = (V^{-1}\Lambda V)^t = V^{-1}\Lambda^t V$$

which vanishes as $t \to \infty$ if and only if $|\lambda_i| < 1$ for all $i = 1, \ldots, n$. The challenge comes from the fact that not all real square matrices are diagonalizable.

■ **Example 1.2** Consider the autonomous linear system (1.2) with

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

The matrix A has two eigenvalues at $\lambda = 1/2$ with corresponding (linearly dependent) eigenvectors

$$v_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad v_2 = -v_1 = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$$

and is therefore not diagonalizable. ■

Nevertheless, the following result holds even when A is not diagonalizable.

Theorem 1.1.1 — Asymptotic stability of linear systems. The discrete-time linear system $x_{t+1} = Ax_t$ with $A \in \mathbb{R}^{n \times n}$ is asymptotically stable if and only if all eigenvalues of A have magnitude less than one, *i.e.*

$$|\lambda_i(A)| < 1, \qquad \forall i = 1, \ldots, n. \tag{1.3}$$

Matrices whose A whose eigenvalues satisfy (1.3) are called *Schur*. Although it is sometimes possible to compute the eigenvalues of the matrix A analytically, *e.g.* as solutions to the characteristic equation $\det(\lambda I - A) = 0$, we will typically have to rely on numerical calculations to assess stability.

Geometrically, the condition (1.3) requires that all eigenvalues lie inside the unit circle in the complex plane. The unit circle is thus the *stability boundary* of discrete-time linear systems. Some representative pole locations and corresponding (free) system responses are shown in Figure 1.2.

In contrast to continuous-time linear system, where the free response can only tend to zero asymptotically, there are discrete-time linear systems whose state vector converges to zero in finite time. One such example is

$$x_{t+1} = \begin{bmatrix} 0 & 1/2 \\ 0 & 0 \end{bmatrix} x_t$$

It is easy to verify that $x_2 = 0$, no matter which initial value x_0 we choose. Matrices A such that $A^k = \mathbf{0}$ for some finite vaue of k are called *nilpotent*, and have all their eigenvalues at origin.

The driven response

Let us now consider the response of the system driven by the input u. We can proceed in the same way as we did for the free response, i.e. by simply iterating the system equations forward:

$$\begin{aligned} x_1 &= Ax_0 + Bu_0 \\ x_2 &= Ax_1 + Bu_1 = A(Ax_0 + Bu_0) + Bu_1 = A^2x_0 + ABu_0 + Bu_1 \\ &\vdots \\ x_t &= A^t x_0 + \sum_{k=0}^{t-1} A^k B u_{t-1-k} \end{aligned}$$

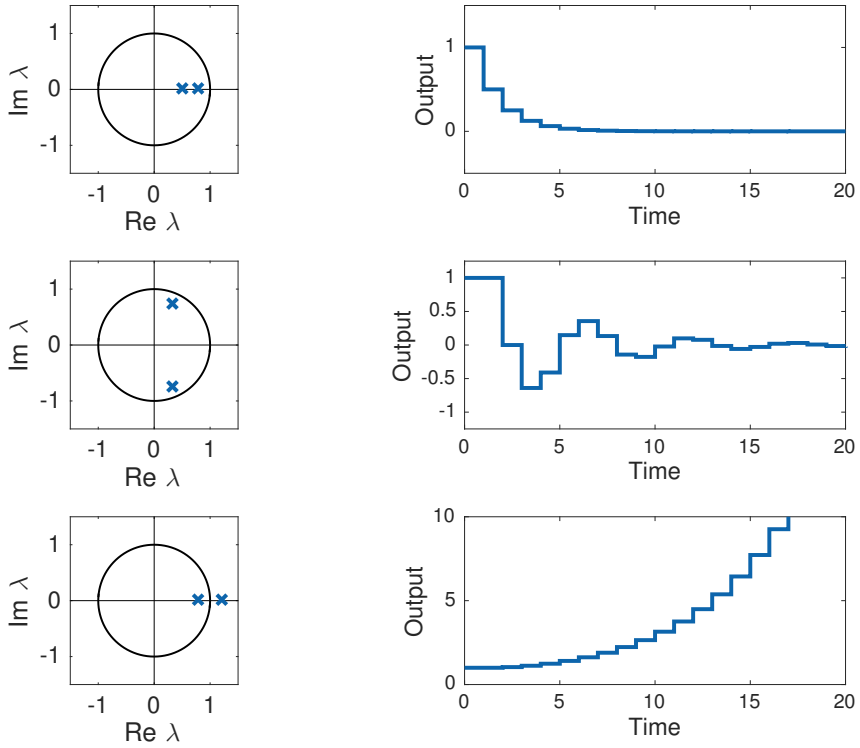

Figure 1.2: Eigenvalue locations (left) and associated unforced responses (right): eigenvalues on the positive real axis inside the unit circle give a well-damped response; eigenvalues close to the stability boundary results in an oscillatory response; a single eigenvalue outside the unit circle results in an unstable system.

We note that the response is divided into two terms: the *free (zero-input) response* $A^t x_0$ and the *driven (zero-state) response* $\sum_{k=0}^{t-1} A^k B u_{t-1-k}$. Since the output is just a linear combination of the state and the input, we have

$$y_t = Cx_t + Du_t = CA^t x_0 + \sum_{k=0}^{t-1} CA^k B u_{t-1-k} + Du_t \tag{1.4}$$

One important property of the driven response is that it is linear in x_0 and the input sequence $\{u_0, u_1, \dots\}$. It is convenient to represent this relation in vector form

$$\begin{aligned} x_t &= A^t x_0 + \begin{bmatrix} A^{t-1}B & A^{t-2}B & \dots & AB & B \end{bmatrix} \begin{bmatrix} u_0 \\ u_1 \\ \vdots \\ u_{t-1} \end{bmatrix} \\ &:= A^t x_0 + \mathcal{C}_t \mathcal{U}_t \end{aligned} \tag{1.5}$$

The matrix $\mathcal{C}_t$ is called the *controllability matrix* over hoizon t and will play an important role in our developments.

State transformations

It is often convenient to represent the state vector in another basis, *i.e.* to make a coordinate change

$$z = T^{-1} x$$

for some full rank matrix $T \in \mathbb{R}^{n\times n}$. We have already done so earlier in these notes to reveal the eigenstructure of the system matrix A. However, as we will see, coordinate changes are useful in many other contexts. In the new coordinates, the system state evolves according to

$$\begin{aligned} z_{t+1} &= T^{-1}x_{t+1} = T^{-1}(Ax_t + Bu_t) = T^{-1}ATz_t + T^{-1}Bu_t \\ y_t &= CTz_t + Du_t \end{aligned}$$

i.e., it can be represented by the discrete-time linear system $(T^{-1}AT, T^{-1}B, CT, D)$. Intuitively, a coordinate transformation of the state vector should not alter the input-output behaviour, which the following result asserts.

Theorem 1.1.2 Let $T \in \mathbb{R}^{n\times n}$ be a full rank matrix. The two discrete-time linear systems

$$\begin{aligned} x_{t+1} &= Ax_t + Bu_t \\ y_t &= Cx_t + Du_t \end{aligned}$$

and

$$\begin{aligned} z_{t+1} &= T^{-1}ATz_t + T^{-1}Bu_t \\ y_t &= CTz_t + Du_t \end{aligned}$$

realize the same input-output behavior, given that $z_0 = T^{-1}x_0$.

Proof. The input-output behavior for (A,B,C,D) is given by (1.4). For $(T^{-1}AT, T^{-1}B, CT, D)$, the expressions give

$$\begin{aligned} y_t &= CT(T^{-1}AT)^t z_0 + \sum_{k=0}^{t-1} CT(T^{-1}AT)^k T^{-1}Bu_{t-1-k} + Du_t = \\ &= CA^t x_0 + \sum_{k=0}^{t-1} CA^k Bu_{t-1-k} + Du_t \end{aligned}$$

which proves the equivalence. ■

1.2 Reachability and state transfer

Many control problems concern *state transfer*: given an initial state x_0, we would like to find a control sequence $\{u_0, u_1, \ldots, u_{t-1}\}$ that brings the system state to a target x_{tgt} at time t.

For now, we will only consider the basic question of when a given discrete-time system admits a control sequence that performs the requested state transfer. Later, we will try to find control sequences that perform the state transfer in an optimal way (for example, using minimal energy or in minimal time) while respecting various constraints. But let us consider the basic question first: To reach $x_t = x_{\text{tgt}}$ from a given initial state x_0, (1.5) yields that the control sequence must satisfy

$$x_{\text{tgt}} - A^t x_0 = \mathcal{C}_t \mathcal{U}_t \tag{1.6}$$

This equation has a solution if and only if $x_{\text{tgt}} - A^t x_0$ lies in the range of $\mathcal{C}_t$. In particular, if $\mathcal{C}_t$ has full rank, then we can reach any x_{tgt} from any x_0. This observation can formalized as follows.

Definition 1.2.1 — Reachability. Consider the discrete-time linear system (1.1). We say that a state $x_{\text{tgt}} \in \mathbb{R}^n$ is reachable in k steps if there is an input sequence $\{u_0, u_1, \ldots, u_{k-1}\}$ that drives

the system from initial state $x_0 = 0$ to $x_k = x_{\text{tgt}}$. If there exists a finite value of k such that all $x_{\text{tgt}} \in \mathbb{R}^n$ are reachable in k steps, we say that the system (1.1) is reachable.

Theorem 1.2.1 — Reachability. The linear system (1.1) is reachable if and only if

$$\text{rank}\,(\mathcal{C}_n) = n$$

Proof. If the controllability matrix has rank n, then the linear system of equations

$$z_{\text{tgt}} = \mathcal{C}_t \mathcal{U}_t \tag{1.7}$$

admits a solution for every x_{tgt}. In other words, there exists a control sequence that drives the system from $x(0) = 0$ to $x(n) = x_{\text{tgt}}$ for every x_{tgt}, which implies that the system is reachable.

To show that reachability of a system implies reachability in n steps, recall the definition of the controllability matrix

$$\mathcal{C}_t = \begin{bmatrix} A^{t-1}B & A^{t-2}B & \cdots & AB & B \end{bmatrix} = \begin{bmatrix} A^{t-1}B & \mathcal{C}_{t-1} \end{bmatrix}$$

since we add new columns to the reachability matrix when we increase t, we have that the range of $\mathcal{C}_{t+1}$ is greater or equal to the range of $\mathcal{C}_t$. However, by the Cayley-Hamiton theorem,

$$A^n = \alpha_{n-1}A^{n-1} + \alpha_{n-2}A^{n-1} + \cdots + \alpha_1 A + \alpha_0 I$$

for some scalars $\alpha_0, \alpha_1, \ldots \alpha_{n-1}$. This means that

$$\text{range}(\mathcal{C}_{n+1}) = \text{range}(\begin{bmatrix} A^n B & \mathcal{C}_n \end{bmatrix})$$

since the new columns added to $\mathcal{C}_{n+1}$ are linear combinations of the columns already present in $\mathcal{C}_n$. Repeated use of the Cayley-Hamilton theorem reveals that $\text{range}(\mathcal{C}_t) = \text{range}(\mathcal{C}_n)$ for all $t \geq n$. ■

An important consequence of Theorem 1.2.1 is that if a discrete-time system is reachable, then there exists a control sequence which can drive the system state from the origin to any target state in at most n steps. However, the argument in the proof does not necessary hold if the control is constrained (for example, bounded in magnitude) and the control sequence obtained by solving (1.7) is not guaranteed to be optimal in any way. Finding optimal control policies under state and control constraints is one of the key objectives of this course.

In our reachability definition, we have assumed that $x_0 = 0$. However, it is clear that if the system is reachable then we can find a sequence $\mathcal{U}_n$ that transfer the system state from any initial state to any final state, since

$$x_{\text{tgt}} - A^n x_0 = \mathcal{C}_n \mathcal{U}_n$$

admits a solution for any left-hand side argument.

When a system is not controllable, there are states or linear combinations of states, which we cannot influence with the controls. The following example illustrates this fact.

■ **Example 1.3** Consider the system

$$x_{t+1} = \begin{bmatrix} a_{11} & a_{12} \\ 0 & a_{22} \end{bmatrix} x_t + \begin{bmatrix} b_1 \\ 0 \end{bmatrix} u_t$$

Note that u_t cannot affect the second state, neither directly nor indirectly through the first state (since the second state does not depend on the first). Hence, we cannot expect to be able reach an arbitrary target in the second state. ■

It is not always this straight-forward to determine controllable and uncontrollable states by visual inspection. However, the proof of Theorem 1.2.1 reveals an interesting property of the set of reachable states, which we present in the next lemma.

Proposition 1.2.2 The set of reachable states is A-invariant, *i.e.* if x_{tgt} is reachable, then so is Ax_{tgt}.

Proof. First observe that $\text{range}(\mathcal{C}_n) \subseteq \text{range}(A\mathcal{C}_n)$. For this, note that

$$A\mathcal{C}_n = \begin{bmatrix} A^n B & A^{n-1}B & \dots & AB \end{bmatrix}$$

The last $n-1$ blocks are present in the controllability matrix, and by the Cayley-Hamilton theorem, we can write $A^n B$ as a linear combination of blocks in the controllability matrix. Thus $\text{range}(A\mathcal{C}_n) \subset \text{range}(\mathcal{C}_n)$. Now, if x_{tgt} is reachable, then there exists $\mathcal{U}_n$ such that $x_{\text{tgt}} = \mathcal{C}_n \mathcal{U}_n$. This implies that $Ax_{\text{tgt}} = A\mathcal{C}_n \mathcal{U}_n = \mathcal{C}_n \mathcal{V}_n$, hence Ax_{tgt} is also reachable. ■

The proposition allows us to construct a state transformation that reveals the structure into controllable and uncontrollable subsystems.

Theorem 1.2.3 Let $\text{rank}(\mathcal{C}_n) = n_r < n$. Then there exists a state transformation $z = T^{-1}x$ such that the system in the transformed coordinates has the form

$$\begin{aligned} z_{t+1} &= \begin{bmatrix} A_r & A_{12} \\ 0 & A_{\bar{r}} \end{bmatrix} z_t + \begin{bmatrix} B_r \\ 0 \end{bmatrix} u_t \\ y_t &= \begin{bmatrix} C_r & C_{\bar{r}} \end{bmatrix} z_t + Du_t \end{aligned} \tag{1.8}$$

where (A_r, B_r) is reachable.

Proof. Let $V \in \mathbb{R}^{n \times n_r}$ be a matrix whose columns span the range of $\mathcal{C}_n$, and let $W \in \mathbb{R}^{n \times n - n_r}$ be a matrix whose columns are independent of each other and those of V. Consider the state transformation $z = T^{-1}x$ with $T = \begin{bmatrix} V & W \end{bmatrix}$, i.e.

$$x = \sum_{i=1}^{n_r} z_i v_i + \sum_{j=n_r+1}^{n} z_j w_{j-n_r}.$$

Note that $z_j = 0$ for all $j = n_r + 1, \dots, n$ if x lies in the range of $\mathcal{C}_n$. In other words, the last $n - n_r$ components of $T^{-1}x$ are zero if $x \in \text{range}(\mathcal{C}_n)$.

Since the range of $\mathcal{C}_n$ is A-invariant, Av_i lies in the range of $\mathcal{C}_n$ and the last $n - n_r$ components of $T^{-1}Av_i$ must equal zero. We thus conclude that $\text{x}T^{-1}AT = T^{-1}A \begin{bmatrix} V & U \end{bmatrix}$ has the desired block structure. Similarly, B lies in the range of $\mathcal{C}_n$ so $T^{-1}B$ must also have the desired structure. ■

The following theorem, known as the Popov-Belevitch-Hautus test gives an alternative way to verify reachability. It is a convenient theoretical tool that adds some geometrical insight.

Theorem 1.2.4 — PBH test for reachability. The system (1.1) is *unreachable* if and only if there exists $\lambda \in \mathbb{C}$ and $w \in \mathbb{R}^n$ with $w \neq 0$ such that

$$w^T A = \lambda w^T \qquad w^T B = 0 \tag{1.9}$$

i.e. if one of the left eigenvectors of A is orthogonal to the columns of B.

Proof. If $w \neq 0$ satisfies (1.9), then $w^T B = 0$, $w^T AB = \lambda w^T B$ and similarly $w^T A^t B = 0$ for all t. Hence, $w^T \mathcal{C}_n = 0$, i.e. the controllability matrix does not have full rank, and the system is

unreachable. Conversely, if the system is unreachable, then there is a state transformation $z = T^{-1}x$ which brings the system to the form (1.8). Now, let $w_{\bar{r}}$ be a left eigenvector of $A_{\bar{r}}$ with eigenvalue λ, i.e. $w_{\bar{r}}^T A_{\bar{r}} = \lambda w_{\bar{r}}^T$. Then $w^T = \begin{bmatrix} 0 & w_{\bar{r}}^T \end{bmatrix}$ satisfies (1.9) for the transformed system, and $w' = Tw$ for the original one. ■

It may appear strange that we call $\mathcal{C}_n$ the controllability matrix and not the reachability matrix. After all, we use it for determining the reachability properties of the underlying system. The reason for this naming convention is historical: for continuous-time linear systems, reachability coincides with the following concept of controllability:

Definition 1.2.2 — Controllability. Consider the discrete-time linear system (1.1). We say that the state $x_{\text{tgt}} \in \mathbb{R}^n$ is *controllable in k steps* if there exists an input sequence $\{u_0, u_1, \ldots, u_{k-1}\}$ that drives the system from initial state $x_0 = x_{\text{tgt}}$ to $x_k = 0$. If there exists a finite time k so that all $x_{\text{tgt}} \in \mathbb{R}^n$ are reachable in k steps, we say that the system (1.1) is controllable.

Controllability implies the existence of a k and a $\mathcal{U}_k \in \mathbb{R}^{m \times k}$ such that

$$-A^k x_{\text{tgt}} = \mathcal{C}_k \mathcal{U}_k$$

Clearly, any reachable system will also be controllable. However, since A may be nilpotent (there may be a finite t_0 such that $A^t = 0$ for all $t \geq t_0$), there are systems which are controllable but not reachable. One such system is

$$x_{t+1} = \begin{bmatrix} 0 & 1/2 \\ 0 & 0 \end{bmatrix} x_t + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u_t$$

The system is not reachable, since the controllability matrix has rank 1. No matter how you choose the input sequence, the second state will remain at zero for $t \geq 1$. However, the system is controllable since a zero input ensures that $x_t = 0$ for all $t \geq 2$, independently of the initial state. In fact, one can show that a system is controllable if and only if all eigenvalues of its uncontrollable subsystem are null. The concept of controllability can be weakened further by the notion of stabilizability:

Definition 1.2.3 The discrete-time linear system (1.1) is *stabilizable* if there exists an input sequence $\{u_k\}_{k=0}^{\infty}$ such that $\lim_{t \to \infty} x_t = 0$.

The fact that we no longer require the state to reach the origin in finite time means that we can allow for systems whose unstable modes are asymptotically stable. We summarize the observation in the following theorem, left without a proof.

Theorem 1.2.5 The system (1.1) is stabilizable if all eigenvalues λ_i of its unreachable subsystem satisfy $|\lambda_i| < 1$, *i.e.* if any $w \in \mathbb{R}^n$ with $w \neq 0$ satisfying

$$w^T A = \lambda_i w^T, \qquad w^T B = 0$$

has $|\lambda_i| < 1$.

1.3 Observability and state reconstruction

When we cannot observe the complete state vector x_t but only the output y_t, it is useful to be able to estimate the state from the input and output sequences. We will address this problem in more detail later in these notes, but here we consider the simpler problem of reconstructing the initial state from measured input and output sequences. At least conceptually, knowledge about the initial

state and the applied input sequence allows us to compute the current state exactly, provided that our model is a perfect description of the true system.

To this end, we consider the contribution of the output signal due to the initial state (that is, the measured output minus the output caused by previous inputs): It is convenient to re-write (1.4) as

$$\varepsilon_t = y_t - \left(\sum_{k=0}^{t-1} CA^k B u_{t-1-k} + Du_t \right)$$

and observe that

$$\varepsilon_t = CA^t x_0.$$

To be able to reconstruct the initial state, we must thus ensure that

$$\begin{bmatrix} \varepsilon_0 \\ \varepsilon_1 \\ \vdots \\ \varepsilon_{t-1} \end{bmatrix} = \begin{bmatrix} C \\ CA \\ \cdots \\ CA^{t-1} \end{bmatrix} x_0 := \mathcal{O}_t x_0$$

admits a unique solution x_0. Here, we have introduced $\mathcal{O}_t$, the *observability matrix* (over horizon t).

Definition 1.3.1 — Observability. A state $x_{\text{init}} \neq 0$ is said to be *unobservable* over k steps if, with $x_0 = x_{\text{init}}$ and $u_t = 0$ for $t = 0, \ldots, k-1$, we have $y_t = 0$ for $t = 0, \ldots, k-1$. The system (1.1) is called *unobservable* if it has some unobservable state, and called *observable* otherwise.

By this definition, an unobservable state is one which cannot be distinguished from $x_0 = 0$ based on output measurements. The development of observability criteria parallels the one on controllability. Specifically, we can derive the following results.

Theorem 1.3.1 — Observability. The linear system (1.1) is observable if and only if

$$\operatorname{rank}(\mathcal{O}_n) = n$$

Proposition 1.3.2 The set of unobservable states are A-invariant. In other words, if x_{init} is unobservable, then so is Ax_{init}.

Theorem 1.3.3 Let $\operatorname{rank}(\mathcal{O}_n) = n_o < n$. Then there exists a state transformation $z = T^{-1}x$ such that the system in the transformed coordinates has the form

$$\begin{aligned} z_{t+1} &= \begin{bmatrix} A_{\bar{o}} & A_{12} \\ 0 & A_o \end{bmatrix} z_t + \begin{bmatrix} B_{\bar{o}} \\ B_o \end{bmatrix} u_t \\ y_t &= \begin{bmatrix} 0 & C_o \end{bmatrix} z_t + Du_t \end{aligned}$$

where (A_o, B_o, C_o, D_o) is observable.

Theorem 1.3.4 The linear system (1.1) is unobservable if and only if there exists $v \in \mathbb{R}^n$ with $v \neq 0$ and $\lambda \in \mathbf{C}$ such that

$$Av = \lambda v \qquad Cv = 0$$

i.e., if one if the right eigenvectors of A is orthogonal to the rows of C.

Analogously to the discussion about reachability, controllability and stabilizability, we can weaken the notion of observability to systems which are reconstructable and detectable.

Definition 1.3.2 The discrete-time linear system (1.1) is *reconstructable in k steps* if, for every initial condition x_0, it holds that x_k is uniquely determined by $\{u_0, u_1, \ldots, u_{k-1}\}$ and $\{y_0, y_1, \ldots, y_{k-1}\}$. If the system is reconstructable in n steps, we say that the system is *reconstructable*.

A system is reconstructable if and only if the eigenvalues of its unobservable system matrix are null. The concept of detectability allows for systems where the unobservable modes are asymptotically stable:

Definition 1.3.3 The discrete-time linear system (1.1) is *detectable* if it is asymptotically reconstructable in k steps as $k \to \infty$.

Theorem 1.3.5 The system (1.1) is detectable if all eigenvalues λ_i of its unobservable subsystem satisfy $|\lambda_i| < 1$, *i.e.* any solution $v \in \mathbb{R}^n$ with $v \neq 0$ to

$$Av = \lambda_i v, \qquad Cv = 0$$

has $|\lambda_i| < 1$.

1.4 State feedback and observers

In the basic control course, we used state feedback control laws on the form

$$u_t = -Lx_t + l_r r_t \tag{1.10}$$

to assign desired eigenvalues to the closed-loop system matrix. The same approach works without changes for discrete-time linear systems. The control law (1.10) gives the closed-loop system

$$\begin{aligned} x_{t+1} &= (A - BL)x_t + Bl_r r_t \\ y_t &= (C - DL)x_t + Dl_r r_t \end{aligned}$$

Thus, the free response of the closed-loop system is characterized by the matrix $(A - BL)$, whose eigenvalues we can shift using the feedback gains L:

Theorem 1.4.1 The state feedback law (1.10) allows to assign the eigenvalues of the closed-loop matrix $A - BL$ to arbitrary complex conjugate locations if and only if the open-loop system is reachable.

When the state vector is not measureable, it is natural to try to find an estimate $\hat{x}_t$ of x_t and use this estimate when computing the control action. If we would know the initial state x_0, we could simply estimate future states by simulating the system,

$$\begin{aligned} \hat{x}_{t+1} &= A\hat{x}_t + Bu_t \\ \hat{y}_t &= C\hat{x}_t + Du_t \end{aligned}$$

When x_0 is not known, one can instead use an observer, which corrects the state estimate based on discrepancies between the estimated output $\hat{y}_t$ and the measured y_t. The classical observer is

$$\begin{aligned} \hat{x}_{t+1} &= A\hat{x}_t + Bu_t + K(y_t - \hat{y}_t) \\ \hat{y}_t &= C\hat{x}_t + Du_t \end{aligned}$$

for some constant gain matrix $K \in \mathbb{R}^{n \times p}$. The estimation error $e_t = x_t - \hat{x}_t$ then evolves as

$$\begin{aligned} e_{t+1} &= x_{t+1} - \hat{x}_{t+1} = \\ &= Ax_t + Bu_t - A\hat{x}_t - Bu_t - KC(x_t - \hat{x}_t) = \\ &= (A - KC)e_t \end{aligned}$$

Thus, the error dynamics of the observer are characterized by the eigenvalues of the matrix $A - KC$, which we can alter using proper selection of the observer gain K. The next result characterizes when we can assign arbitrary error dynamics to the observer.

Theorem 1.4.2 For the linear system (1.1), there exists $K \in \mathbb{R}^{n\times p}$ so that the n eigenvalues of $A - KC$ can be assigned to arbitrary real or complex conjugate locations if and only if the system is observable.

The combination of an observer, and static linear feedback from the observed states results in the following output feedback controller:

$$\begin{aligned}\hat{x}_{t+1} &= A\hat{x}_t + Bu_t + K(y_t - C\hat{x}_t)\\ u_t &= -L\hat{x}_t + l_r r_t\end{aligned}$$

or, equivalently,

$$\begin{aligned}\hat{x}_{t+1} &= (A - BL - KC)\hat{x}_t + Ky_t\\ u_t &= -L\hat{x}_t + l_r r_t\end{aligned}$$

and the closed-loop system, when this controller is applied to (1.1), is

$$\begin{aligned}\begin{bmatrix} x_{t+1}\\ \hat{x}_{t+1}\end{bmatrix} &= \begin{bmatrix} A & -BL\\ KC & A - BL - KC\end{bmatrix}\begin{bmatrix} x_t\\ \hat{x}_t\end{bmatrix} + \begin{bmatrix} Bl_r\\ 0\end{bmatrix} r_t\\ y_t &= \begin{bmatrix} C & -DL\end{bmatrix}\begin{bmatrix} x(t)\\ \hat{x}_t\end{bmatrix} + Dl_r r_t\end{aligned}$$

The closed-loop dynamics is easier to analyze in terms of the system state and the estimation error:

$$\begin{aligned}\begin{bmatrix} x_{t+1}\\ e_{t+1}\end{bmatrix} &= \begin{bmatrix} A - BL & BL\\ 0 & A - KC\end{bmatrix}\begin{bmatrix} x_t\\ e_t\end{bmatrix} + \begin{bmatrix} Bl_r\\ 0\end{bmatrix} r_t\\ y(t) &= \begin{bmatrix} C - DL & DL\end{bmatrix}\begin{bmatrix} x_t\\ \hat{x}_t\end{bmatrix} + Dl_r r_t\end{aligned}$$

Since the system matrix is block-diagonal, its eigenvalues equal those of its diagonal blocks, i.e. those of $A - BL$ and of $A - KC$, respectively. Note that the error dynamics is not reachable when we consider the reference as an input to the closed-loop system.

1.5 Discrete-time descriptions of continuous-time systems

Most models of the physical world are based on ordinary differential equations, and in many cases their behavior is linear close to a fixed operating point. This makes for a compelling argument to study control of continuous-time linear systems, as is done in most introductory control courses. In a similar way, discrete-time linear systems arise naturally when we want to use a computer to control physical systems.

Sampling and reconstruction of continuous-time signals

Digital hardware cannot directly deal with continuous-time (analog) signals. Instead, we will have to sample sensor signals to create a discrete-time representation, and reconstruct continuous-time control signals from a computed command sequence. There are many ways of performing sampling and reconstructions, but we will only consider uniform sampling and zero-order hold reconstruction.

Consider a continuous-time cosine signal with frequency f (rad/sec)

$$y(t) = \cos(2\pi f t).$$

Sampling this signal every h seconds results in the discrete-time representation

$$y_k = \cos(2\pi fhk).$$

It is common to refer to $f_s = 1/h$ as the sampling frequency. Since k is integer-valued, we note that

$$y_k = \cos(2\pi fhk + 2\pi nk) = \cos(2\pi(f + nh^{-1})hk)$$

for any positive integer n. In addition, since $\cos(x) = \cos(-x)$, we also have that $\cos(2\pi fhk) = \cos(-2\pi fhk) = \cos(2\pi(-f + nh^{-1})hk)$. Hence, from the sampled signal, it will be impossible to know if the continuous signal has frequency f or $\pm f + nf_s$. This effect is called *aliasing*.

To avoid aliasing effects, we will have to sample sufficiently fast compared to the frequencies contained in the underlying analog signal. Specifically, assume that the analog signal contains frequencies in the interval $[0, f_{\max}]$. If we sample the signal with sampling frequency f_s, then aliasing will appear at frequencies $nf_s + [-f_{\max}, f_{\max}]$. To ensure that we can separate these from the original signal, we have to ensure that $f_{\max} < f_s - f_{\max}$, *i.e.* that $f_s \geq 2f_{\max}$. The critical frequency $f_s = 2f_{\max}$ is called the *Nyquist frequency* and provides a fundamental lower bound on the sampling rate. Taking one step back, this limitation is quite natural: it means that we have to take at least two samples per period of a sinusoidal signal with frequency $f_{\max}$.

The creation of a continuous-time signal (defined for all times) from a discrete-time sequence (defined only at sampling instances) is referred to as reconstruction. Most control systems use *zero-order hold* reconstruction, *i.e.* they create a continuous-time signal which is held constant between sampling instances:

$$u(t) = u_k \qquad \text{for } t \in [kh, kh + h).$$

Equivalent discrete-time system under periodic sampling and zero-order hold

We will now show how discrete-time linear systems arise naturally when we want to use a computer to control a physical system. Let the system be described by the continuous-time linear system

$$\begin{cases} \dot{x}(t) &= A_c x(t) + B_c u(t) \\ y(t) &= Cx(t) + Du(t) \end{cases} \tag{1.11}$$

and assume we sample the continuous state trajectory with sampling time h. If the control input is held constant between samples, then we can use the solution of (1.11) to determine $x(kh + h)$ in terms of $x(kh)$ and $u(kh)$. Specifically, we have

$$x(kh + h) = e^{A_c h} x(kh) + \int_{s=0}^{h} e^{A_c s} B_c u(kh)\, ds.$$

We can hence describe the behavior of (1.11) at the sampling instants via

$$\begin{aligned} x(kh + h) &= Ax(kh) + Bu(kh) \\ y(k) &= Cx(k) + Du(k) \end{aligned}$$

where $A = e^{A_c h}$ and $B = \int_{s=0}^{h} e^{As} B\, ds$. If we drop reference to physical time and only count sample instances, we get a model on the form

$$\begin{aligned} x_{k+1} &= Ax_k + Bu_k \\ y_k &= Cx_k + Du_k \end{aligned}$$

which has the precise form of the discrete-time linear systems that we study in this course. There are many alternative ways of determining discrete-time approximations of continuous-time dynamical systems. We will explore some of these in the exercises.

Guidelines for choosing sampling time

We have already discussed how the Nyquist frequency provides a lower bound on the sampling time. However, this is a lower-bound and typically not enough to get reasonable quality in the desired time-responses, but significantly higher rates are often used. Sample time selection for control involves a compromise between computational load on the controller (since a new control signal should be computed at every sample interval) and effectiveness in tracking references and rejecting disturbances (since the controller cannot react until the next sampling instance after a disturbance hits the system). One is easily led to believe that faster sampling is better, but going to far can lead to accuracy loss since $A \approx I + hA_c + h^2/2A_c^2 + \cdots \to I$ as $h \to 0$.

The standard rule-of-thumb instead proposes to use a sampling time which results in $4-10$ samples per rise-time of the closed-loop system [2], or a sampling frequency $2\pi/h$ of $20-40$ times the desired closed-loop bandwidth ω_{BW} [3].

Using continuous-time insight to understand transient properties of discrete-time systems

One of the most popular control design techniques for linear system is based on pole placement. To choose the appropriate closed-pole locations, it is essential to have a good understanding on how the pole locations affect the transient response of the system. In traditional basic control courses, significant attention is therefore given to developing such an understanding for simple prototype systems. We will leverage on this understanding to develop guide lines for appropriate closed-pole locations of discrete-time systems.

■ **Example 1.4** Consider the mechanical system

$$m\ddot{x}(t) + d\dot{x}(t) + kx(t) = cF(t)$$

We will re-parameterize the model in terms of $\omega_0^2 = k/m$, $\zeta = d/2m\omega_0$ and $k = cm/\omega_0^2$ so that

$$\ddot{x}(t) + 2\zeta\omega_0\dot{x}(t) + \omega_0^2 x(t) = k\omega_0^2 F(t)$$

These model parameters have distinct influence on the transient response: ω_0 is the natural frequency of the system, ζ determines its damping and k its static gain. A damping close to zero gives an oscillatory response (the extreme case of $\zeta = 0$ corresponds to a harmonic oscillator) and one would typically like to strive for a damping of at least 0.5.

To develop corresponding guide lines for appropriate closed-pole locations of discrete-time systems, we will study where the poles of the discrete-time equivalent lie for different values of ζ and different sampling intervals. To this end, we notice that the system admits the state-space form

$$\dot{x}(t) = \begin{bmatrix} 0 & 1 \\ -\omega_0^2 & -2\zeta\omega_0 \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ k\omega_0^2 \end{bmatrix} u(t)$$

Hence, we can compute the corresponding discrete-time systems using the formulas derived above. Doing so for $\omega_0 h \in [0, \pi]$ and $\zeta \in \{0, 0.25, 0.5, 0.75, 1\}$ yields the results shown in Figure 1.4.

■

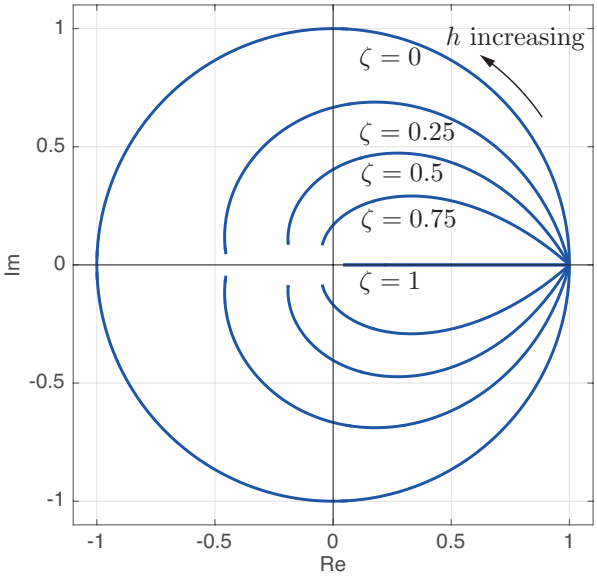

Figure 1.3: Pole locations of discrete-time equivalent of second-order system. Low damping yields poles close to the unit circle, while a damping of one results in poles on the positive real axis. Fast sampling gives poles close to 1 and slower sampling moves the poles further inside the unit circle.

1.6 Design example: level control of a double tank

To get some experience in control of discrete-time linear systems, we will consider the two-tank system illustrated in Figure 1.6. We would like to design a discrete-time control system which controls inflow q_i to the first tank in order to maintain a desired level h_2 of the second tank.

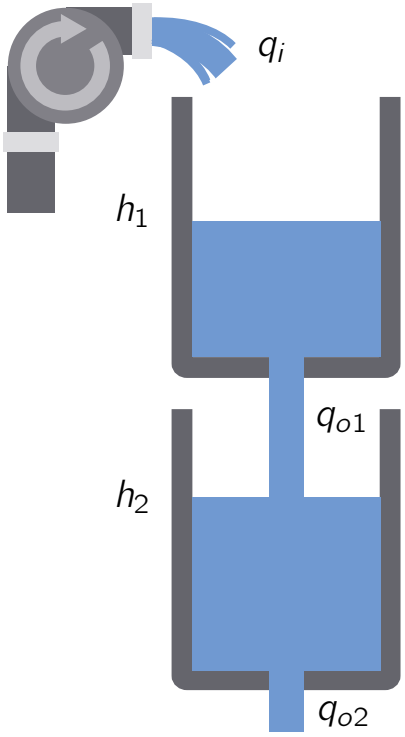

Figure 1.4: Double tank.

The tank dynamics can be described by the second-order system

$$\dot{x}(t) = \begin{bmatrix} -0.07 & 0 \\ 0.07 & -0.07 \end{bmatrix} x(t) + \begin{bmatrix} 0.18 \\ 0 \end{bmatrix} u(t)$$

where x_1 corresponds to the level in the upper tank, x_2 corresponds to the level in the lower tank, and u is the voltage applied to the pump (to generate the inflow in the upper tank).

To perform a discrete-time design, we will have to convert the continuous-time dynamics into a discrete-time equivalent. We will use periodic sampling of the lower tank level, and a zero-order-hold control (*i.e.* keep the input voltage constant between sampling instances), and can therefore use the formulas developed earlier in these notes. However, we must select the appropriate sampling interval. There are several rules-of-thumb for selecting sampling intervals. We will use the one proposed in [2]: select the sampling interval to allow for $4-10$ samples during a rise-time of the closed-loop system. Aiming for a rise-time of around 40 seconds suggests a sampling time in the range $4-10$ seconds. We will settle for $h = 5$ seconds, for which we find the corresponding discrete-time system described by

$$A = \begin{bmatrix} 0.7047 & 0 \\ 0.2466 & 0.7047 \end{bmatrix}, \qquad B = \begin{bmatrix} 0.7594 \\ 0.1252 \end{bmatrix}$$

Figure 1.5 shows the step responses of the continuous-time system and the corresponding discrete-time equivalent. Note that the discrete-time model predicts the state evolution of the continuous-time system exactly at the sampling instances.

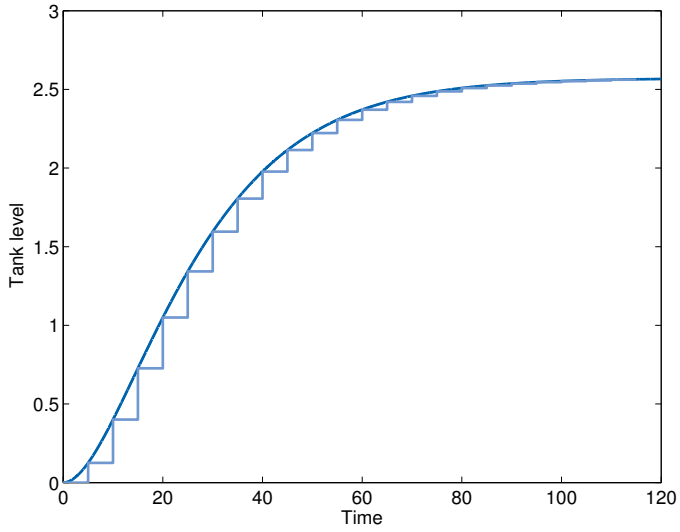

Figure 1.5: Step responses of continuous-time system (dark) and discrete-time system sampled with $h = 5$ seconds (light).

Initially, we assume that we can measure both system states (*i.e.* the two tank levels). We want to design a state-feedback

$$u_t = -Lx_t = -\begin{bmatrix} l_1 & l_2 \end{bmatrix} x_t$$

which makes the closed-loop dynamics slightly faster, *i.e.* moves the closed-loop poles closer to the origin. We will aim at placing both system poles at $z = 0.5$. To find the corresponding feedback gains, we will have to ensure that the characteristic equation

$$\lambda(z) = \det(zI - (A - BL)) = (z - 0.5)^2$$

which leads to the system of linear equations

$$\begin{aligned} -1.4094 + 0.7594l_1 + 0.1252l_2 &= -1 \\ 0.4966 - 0.5351l_1 + 0.0990l_2 &= 0.25 \end{aligned}$$

from which we find

$$l_1 = 0.5022, \qquad l_2 = 0.2237$$

Since we only want to measure the level of the lower tank, we construct a state observer whose poles are slightly faster than the ones used in our state feedback design. We thus consider the output

$$y_t = Cx_t = \begin{bmatrix} 0 & 1 \end{bmatrix} x(t)$$

For purposes of illustration, we choose to place the observer poles (the eigenvalues of $A - KC$) at $z = 0.4$. A similar calculation as for the state feedback design results in the estimator gain matrix

$$K = \begin{bmatrix} 0.3765 \\ 0.6095 \end{bmatrix}$$

Combining the state estimator and feedback from the estimated states, we find the output feedback controller on the form

$$\begin{aligned} \hat{x}_{t+1} &= A\hat{x}_t + Bu_t + K(y_t - C\hat{x}_t). \\ u_t &= -L\hat{x}_t \end{aligned}$$

Figure 1.6 shows the open loop response (dark) and the controlled response to an initial disturbance in the lower tank level. As can be seen, the output feedback has rendered the closed-loop dynamics faster than the open-loop.

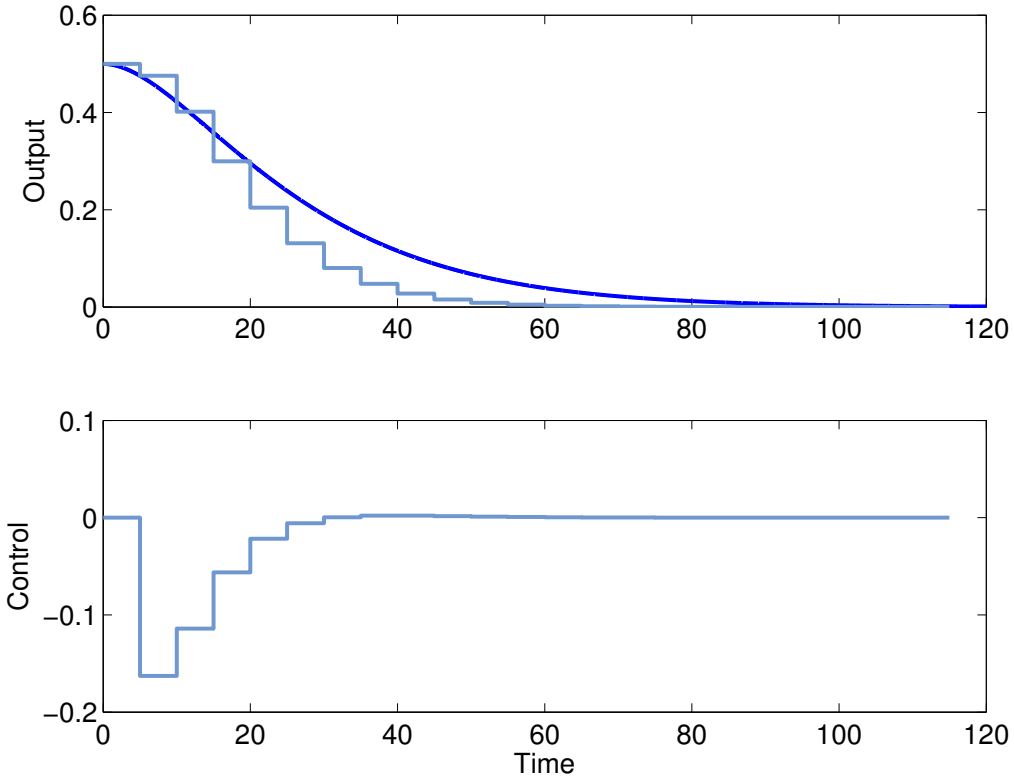

Figure 1.6: Initial value responses of uncontrolled system (dark) and closed-loop (light) with the designed output feedback controller.

1.7 Input-output properties of discrete-time linear systems*

Although most of this course considers state-space models, one can develop considerable insight into the properties of discrete-time linear systems using input-output models. To this end, this section contains a brief introduction to the z-transform, transfer functions for discrete-time linear systems, and a few words about their frequency responses.

The z-transform

A discrete-time signal $s(k)$ can be represented by a list of real numbers (or vectors), $\{s_0, s_1, \dots\}$. The analysis of discrete-time signals and systems is sometimes simplified by using the *z-transform*:

Definition 1.7.1 The *z-transform* $\mathcal{Z}(s)$ of the discrete-time signal $s(k) = \{s_0, s_1, \dots\}$ is

$$\mathcal{Z}(s) = \sum_{k=0}^{\infty} s_k z^{-k}$$

Note that the z-transform maps the space of discrete-time signals to the space of functions over the complex plane (or, rather, over a subset of the complex plane for which the summation converges). To emphasize this, we will write the z-transform of a signal $s(k)$ as $S(z)$. The next example derives the z-transform for some common signals.

Signal	Transform	Name
$\alpha s_1(k)+\beta s_2(k)$	$\alpha S_1(z)+\beta S_2(z)$	Linearity
$s(k+1)$	$zS(z)-zs(0)$	Forward shift
$s(k-1)$	$z^{-1}S(z)+s(-1)$	Backward shift
$s(k)=\sum_{l=0}^{k} x_l$	$S(z)=\frac{z}{z-1}X(z)$	Accumulation
$\lim_{k\to\infty} s(k)$	$\lim_{z\to 1}(z-1)Y(z)$	Final-value theorem
$\lim_{k\to 0} s(k)$	$\lim_{z\to\infty} S(z)$	Initial-value theorem
$s(k)=\sum_{l=0}^{k} x[l]y[k-l]$	$S(z)=X(z)Y(z)$	Convolution

Table 1.1: Key properties of the z-transform.

■ **Example 1.5** The unit step

$$\mathbf{1}(k)=\begin{cases}1 & \text{if } k\geq 0\\ 0 & \text{if } k<0\end{cases}$$

has z-transform

$$\mathcal{Z}(\mathbf{1})=\sum_{k=0}^{\infty} z^{-k}=\frac{1}{1-z}$$

with region of convergence $|z|<1$. The geometric sequence

$$s(k)=a^k\mathbf{1}(k)$$

has z-transform

$$\mathcal{Z}[s]=\frac{z}{z-a}$$

with region of convergence $|z|>|a|$. ■

From the definition, one can also derive many important properties of the z-transform, some of which are summarized in Table 1.1.

Transfer functions of discrete-time linear systems

We can now apply the z-transform to the state-space model

$$\begin{aligned}x_{k+1}&=Ax_k+Bu_k\\ y_k&=Cx_k+Du_k\end{aligned}$$

with initial state x_0. For convenient notation, we define $X(z)=\mathcal{Z}[x(k)]$, $U(z)=\mathcal{Z}[u(k)]$ and $Y(z)=\mathcal{Z}[y(k)]$. Then, by the linearity and forward-shift property of the z-transform

$$\begin{aligned}zX(z)-zx_0&=AX(z)+BU(z))\\ Y(z)&=CX(z)+DU(z)\end{aligned}$$

from which we find

$$Y(z)=C(zI-A)^{-1}zx_0+(C(zI-A)^{-1}B+D)U(z):=C(zI-A)^{-1}zx_0+G(z)U(z)$$

The first term in this expression is the z-transform of the free system response, while the second one is the z-transform of the driven response. If $x_0=0$, then $Y(z)=G(z)U(z)$ where $G(z)$ is called the transfer function of the system. We define it below for easy reference

Definition 1.7.2 The *transfer function* of the discrete-time linear system (A,B,C,D) is

$$G(z) = C(zI-A)^{-1}B+D.$$

Recall that the inverse of a matrix is the ratio between its adjugate matrix and its determinant, *i.e.* $M^{-1} = \text{adj}(M)/\det(M)$. This allows us to re-write the transfer function expression as

$$G(z) = \frac{C\text{adj}(zI-A)B}{\det(zI-A)} + D$$

Hence, the characteristic polyonial of the system matrix A appears in the denominator of the transfer function (or transfer matrix elements, in case there are many inputs or outputs). When the linear system has a single input and a single output, then

$$G(z) = \frac{B(z)}{A(z)} = \frac{b_0 z^m + \cdots + b_{m-1}z + b_m}{z^n + a_1 z^{n-1} + \cdots + a_n}$$

In the absence of cancellations between numerator and denominator, the poles of the transfer functions are exactly the eigenvalues of the system matrix A in the state-space description.

The frequency response of a discrete-time linear system

As we saw in the discussion about Nyquist sampling, a discrete-time sinusoid takes the form

$$u(k) = \cos(\omega_s k + \theta_0)$$

where $\omega_s = 2\pi fh$ is the angular frequency measured in radians per sample. Note that the lowest possible rate of variation for this signal is $\omega_s = 0$, which corresponds to a constant. The highest rate of variation happens for $\omega_s = \pi$, when the signal alternates sign at each time step, *i.e.* $u(k) = (-1)^k$, which happens for $\omega_s = \pm\pi$. Thus, the interesting range for discrete-time sinusoidals is $[-\pi,\pi]$; outside this range everything repeats periodically in ω_s.

By a similar calculation as for continuous-time systems, one can show that the output of an asymptotically stable linear systems driven by an input $u_k = \cos(\omega_s k)$ satisfies

$$y_k = |G(e^{i\omega_s})|\cos(\omega_s k + \arg\, G(e^{i\omega_s}))$$

once the transient has died out. Hence, one refers to $G(e^{i\omega_s})$ for $\omega_s \in [0,\pi]$ as the frequency response of G. As in the continuous-time case, one typically visualizes the frequency response using Bode or Nyquist diagrams. However, in contrast to the continuous-time case, there are no simple rules for drawing these diagrams by hand (since $G(e^{i\omega_s})$ is irrational in ω_s), but one has to resort to numerical computations.

2. Stability and invariance of nonlinear systems

In the previous chapter we discussed the stability of autonomous linear systems on the form

$$x_{t+1} = Ax_t.$$

We noted that the state trajectory $\{x_t\}$ may essentially exhibit three qualitatively different behaviors: it may converge to zero; it may diverge in the sense that $\|x_t\| \to \infty$ as $t \to \infty$; or it may exhibit a sustained oscillation (in which the state stays bounded, but neither converges to the origin nor diverges). We also demonstrated that the stability properties of an autonomous linear system are completely characterized by its eigenvalue structure: if all eigenvalues satisfy $|\lambda_i| < 1$, then all trajectories will converge asymptotically to zero; if the eigenvalues satisfy $|\lambda_i(A)| \leq 1$ and the eigenvalues on the unit circle are distinct, then there will be initial values from which the system maintains a sustained oscillation; and in all other cases (some eigenvalues have magnitude greater than one, or multiple eigenvalues located at the same position on the unit circle), there are initial values from which the state grows unbounded.

Assessing stability of nonlinear and constrained systems is much more involved. Nevertheless, there is a powerful and flexible framework which dates back to the work of A. M. Lyapunov [5]. The framework is based on energy considerations and gives useful geometrical insight into the system dynamics. In addition, when applied to linear systems, it results in an alternative necessary and sufficient test for stability with many powerful extensions.

2.1 Stability concepts

Consider the nonlinear system

$$x_{t+1} = f(x_t) \tag{2.1}$$

with equilibrium point x^{eq} for which

$$x^{\text{eq}} = f(x^{\text{eq}}).$$

We are interested in studying the following three stability concepts.

Definition 2.1.1 — Stability concepts. The system (2.1) is

- *globally asymptotically stable* if, for every x_0, it holds that

$$x_t \to x^{\text{eq}} \text{ as } t \to \infty;$$

- *locally asymptotically stable* near x^{eq} if there is a constant $\delta > 0$ such that

$$\|x_0 - x^{\text{eq}}\| \leq \delta \Rightarrow x_t \to x^{\text{eq}} \text{ as } t \to \infty; \text{ and}$$

- (locally) *stable* if for every (small) $\varepsilon > 0$, there exists $\delta > 0$ such that

$$\|x_0 - x^{\text{eq}}\| \leq \delta \Rightarrow \|x_t - x^{\text{eq}}\| \leq \varepsilon \text{ for all } t \geq 0.$$

In a globally asymtotically stable system, all trajectories converge to the equilibrium points; in a locally asymptotically stable system, trajectories that start sufficiently close to the equilibrium point also converge to it; in a locally stable system, trajectories that start close enough to the equilibrium are guaranteed to stay bounded (but need not converge); See Figure 2.1. Clearly, a globally asymptotically stable system is also locally asymptotically stable and locally stable, and a locally asymptotically stable system is also locally stable.

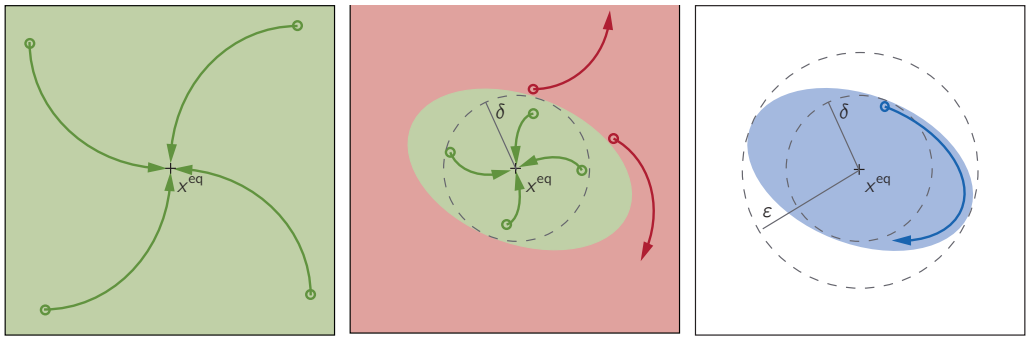

Figure 2.1: Pictorial illustration of the three stability concepts.

■ **Example 2.1** A simple example of a globally asymptotically stable system is

$$x_{t+1} = 0.5x_t.$$

No matter what initial value we choose, the state vector will converge to the origin.

Locally asymptotically stable systems appear, for example, when we try to stabilize unstable systems with a bounded control. Consider the unstable system

$$x_{t+1} = 2x_t + u_t$$

with the linear feedback $u_t = -1.5x_t$ saturated at ± 1

$$u_t = \text{sat}(-1.5x_t) = \max(-1, \min(1, -1.5x_t)).$$

The closed-loop system

$$x_{t+1} = 2x_t - \text{sat}(1.5x_t)$$

is only locally asymptotically stable. For initial values in the region $|x| \leq 2/3$ the control will not saturate, and the state will evolve according to

$$x_{t+1} = 0.5x_t$$

and hence converge to the origin. If $x_0 \geq 2/3$, the system will evolve as

$$x_{t+1} = 2x_t - 1$$

Trajectories of this system converge if $x_0 < 1$, stay constant if $x_0 = 1$ and diverge if $x_0 > 1$. A analogous argument holds for $x_0 < -2/3$. Hence, the origin is a locally asymptotically stable equilibrium point with a region of attraction $X_{\text{roa}} = \{x \mid |x| < 1\}$. Note that the two additional equilibria $x = \pm 1$ introduced by the feedback are not locally stable (in fact, they are unstable).

Finally, an example of a locally stable system is

$$x_{t+1} = -x_t$$

which will maintain a sustained oscillation and evolve as $\{x_0, -x_0, x_0, -x_0, \dots\}$. ∎

2.2 Lyapunov stability

The basic idea of Lyapunov stability is to introduce an abstract energy measure and show that the energy does not increase along system trajectories. With the appropriate definition of energy measure, we will be able to guarantee stability, asymptotic stability, regions of local attraction, and much more. In some cases, these conditions are both necessary and sufficient. The next theorem takes the first step towards this aim.

Theorem 2.2.1 If there exists a continuous function $V(x)$ whose sublevel sets

$$\mathcal{L}_V(\alpha) = \{x \mid V(x) \leq \alpha\}$$

are bounded for every value of α and

$$\Delta V(x) = V(f(x)) - V(x) \leq 0$$

for all x, then all trajectories of (2.1) are bounded.

Proof.

$$\begin{aligned} V(x_t) &= V(x_t) - V(x_{t-1}) + V(x_{t-1}) - V(x_{t-2)}) + \cdots + V(x_1) - V(x_0) + V(x_0) = \\ &= V(x_0) + \sum_{k=0}^{t-1} \Delta V(x_k) \leq V(x(0)). \end{aligned}$$

Hence, every trajectory lies in the set

$$\mathcal{L}_V(V(x_0)) = \{x \mid V(x) \leq V(x_0)\}$$

which is bounded by assumption. ■

This result tells us that if we can guarantee that V is non-increasing along system trajectories, then once the state enters a level set of V, it never leaves this set. The assumption of bounded level sets ensures that the state remains bounded, and hence that the system is stable; see Figure 2.2.

By imposing a few additional conditions on V, we will be able to obtain conditions that guarantee asymptotic stability. To this end, we introduce the following definitions.

Definition 2.2.1 A function $V : \mathbb{R}^n \mapsto \mathbb{R}$ is *positive semidefinite* if
(a) $V(x) \geq 0 \qquad \forall x.$
V is *positive definite* if it, in addition to (a), also satisfies the conditions

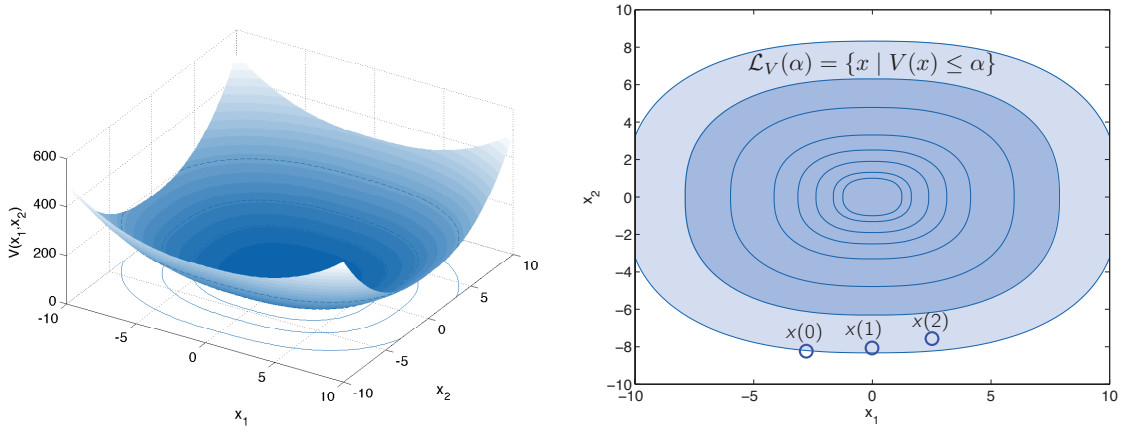

Figure 2.2: The left figure shows a Lyapunov function and some of its level sets; the right figure illustrates the requirement that once the state vector enters a level set, it will not leave. If the level set is bounded, this implies that the state will also be bounded.

(b) $V(0) = 0$ if and only if $x = 0$
(c) all sublevel sets of V are bounded

These definitions are not universal. For example, condition (c) is sometimes replaced by the equivalent requirement that $V(x) \to \infty$ as $\|x\| \to \infty$, and they differ from definitions used in some other fields of mathematics. Still, the definitions are natural generalization of the corresponding notions for quadratic functions $V(x) = x^T Px$. If P is a positive definite matrix, then $V(x)$ is a positive definite function, and if P is positive semidefinite, then so is V.

We can now state a first Lyapunov theorem for asymptotic stability.

Theorem 2.2.2 If there exists a continuous function $V : \mathbb{R}^n \mapsto \mathbb{R}$ such that
(a) $V(x)$ is positive definite, and
(b) $V(f(x)) - V(x) \le -l(x)$ for some positive semidefinite $l(x)$,
then $l(x_t) \to 0$ as $t \to 0$. If $l(x)$ is positive definite, then $x_t \to 0$ as $t \to \infty$.

Proof. Summing up the inequality in condition (b) over time yields

$$\sum_{k=0}^{\infty} V(x_{k+1}) - V(x_k) \le -\sum_{k=0}^{\infty} l(x_k)$$

Hence

$$\sum_{k=0}^{\infty} l(x_k) \le V(x_0) - \lim_{k\to\infty} V(x_k)$$

Now, since $V(x_t) \ge 0$ by condition (a), and $\{V(x_t)\}$ is a decreasing sequence by condition (b), the right-hand side will converge to a finite limit. By Cauchy's convergence criterion (*e.g.*, [1, Theorem 2.7.2], convergence of the infinite sum implies that $l(x_t) \to 0$ as $t \to \infty$. Since $l(x)$ is positive definite, $l(x) = 0$ implies that $x = 0$. The proof is complete. ■

By adding additional assumptions on $V(x)$, one can also derive bounds on how quickly the state converges to the origin. We will not make the general extension here, but discuss these type of results only for the case of linear systems, which we will consider next.

Lyapunov stability of linear systems

Theorem 2.2.3 The autonomous linear system

$$x_{t+1} = Ax_t \tag{2.2}$$

is asymptotically stable if and only if, for every positive definite matrix Q, there exists a unique positive definite matrix P satisfying the *Lyapunov equation*

$$A^T PA - P + Q = 0 \tag{2.3}$$

Proof. Let Q be an arbitrary positive semidefinite matrix and assume that (2.3) admits a positive definite solution P . The Lyapunov function candidate $V(x) = x^T Px$ then satisfies

$$\begin{aligned} V(x_{t+1}) - V(x_t) &= x_{t+1}^T Px_{t+1} - x_t^T Px_t = \\ &= x_t^T \left(A^T PA - P\right) x_t = -x_t^T Qx_t \end{aligned}$$

Since $x^T Qx$ is a positive definite function, Theorem 2.2.2 guarantees asymptotic stability.

If, on the other hand, the system (2.2) is asymptotically stable, then $|\lambda_i(A)| < 1$ for all i and

$$P = \sum_{k=0}^{\infty} \left(A^T\right)^k QA^k$$

exists and satisfies the Lyapunov equation (2.3).

Finally, to show uniqueness assume that both P and P' satisfy the Lyapunov equation. Then

$$A^T(P - P')A - (P - P') = 0$$

Repeated application of this relationship yields

$$P - P' = A^T(P - P')A = \cdots = \lim_{k\to\infty} (A^T)^k (P - P')A^k = 0$$

where the last equality follows from stability of A. Thus, $P' = P$, i.e. P is unique. □.

If one is only concerned about asymptotic stability, Theorem 2.2.3 is very simple to use: just pick any positive definite matrix Q (for example, the identity matrix $Q = I$) and solve the Lyapunov equation. The system (2.2) is asymptotically stable if and only if the solution P is positive definite. Note that the Lyapunov equation is linear in the elements of the matrix P, and since P is symmetric, the Lyapunov equation yields a system of $n(n+1)/2$ linear equations. Although this indicates that one could solve Lyapunov equation using a standard linear equation solver, there are more efficient numerical routines for solving Lyapunov equations.

The theorem does *not* state that asymptotically stable linear systems admit unique Lyapunov functions. The theorem states that the solution P to the Lyapunov equation is unique *for a given* Q. Different Q matrices typically give different solutions P and hence different Lyapunov functions. As the next result shows, different choices for Q allow for different bounds on the convergence rates of trajectories.

Theorem 2.2.4 Consider the autonomous linear system () and assume that the Lyapunov function () admits a positive solution P for a given positive matrix Q. Then

$$\|x_t\|_2^2 \leq \frac{\lambda_{\max}(P)}{\lambda_{\min}(P)} \left(1 - \frac{\lambda_{\max}(Q)}{\lambda_{\min}(P)}\right)^t \|x_0\|_2^2$$

Proof. The proof is based on the fact that for any positive definite matrix M,

$$\lambda_{\min}(M)\|x\|_2^2 \leq x^T Mx \leq \lambda_{\max}(M)\|x\|_2^2.$$

Now, a solution to the Lyapunov equation guarantees that $V(x) = x^T Px$ satisfies

$$V(x_{t+1}) - V(x_t) = x_t^T Q x_t \leq \lambda_{\max}(Q)\|x_t\|_2^2 \leq \frac{\lambda_{\max}(Q)}{\lambda_{\min}(P)} V(x_t),$$

i.e. $V(x_{t+1}) \leq \rho V(x_t)$ with $\rho = 1 - \lambda_{\max}(Q)/\lambda_{\min}(P)$. By repeated application of this inequality

$$V(x_t) \leq \rho^t V(x_0)$$

and, in terms of the state vector,

$$\|x_t\|_2^2 \leq \frac{\lambda_{\max}(P)}{\lambda_{\min}(P)} \left(1 - \frac{\lambda_{\max}(Q)}{\lambda_{\min}(P)}\right)^t \|x_0\|_2^2$$

■

Evaluating quadratic costs via Lyapunov functions

From the proof of Theorem 2.2.2, a solution to the Lyapunov equation (2.3) guarantees that

$$\sum_{k=0}^{\infty} x_k^T Q x_k \leq x_0^T P x_0 - \lim_{k\to\infty} x_k^T P x_k$$

Since it also guarantees asymptotic stability, the final term in this expression will vanish. Hence,

$$\sum_{k=0}^{\infty} x_k^T Q x_k \leq x_0^T P x_0$$

Thus, by solving the Lyapunov equation we are able to estimate the infinite sum $\sum_{k=0}^{\infty} x_k^T Q x_k$ for any initial value x_0. A common performance criterion for linear systems is the energy contained in some output $y = Cx$, *i.e.*

$$\sum_{k=0}^{\infty} y_k^T y_k = \sum_{k=0}^{\infty} x_k^T C^T C x_k$$

Unfortunately, Theorem 2.2.3 does not apply to this case since $C^T C$ is only positive semidefinite. However, as the next result illustrates, it is possible to allow for semidefinite Q matrices, provided that they satisfy an additional observability assumption.

Theorem 2.2.5 Let (A, C) be detectable. Then

$$x_{t+1} = Ax_t$$

is asymptotically stable if and only if there exists a unique matrix $P = P^T \geq 0$ which satisfies the Lyapunov equation

$$A^T PA - P + C^T C = 0 \tag{2.4}$$

If (A, C) is observable, then the unique solution to (2.4) is positive definite.

Proof. As in the proof of the basic Lyapunov theorem, if the system is asymptotically stable, then

$$P = \sum_{k=0}^{\infty} (A^T)^k C^T C A^k = \lim_{l\to\infty} \mathcal{O}_l^T \mathcal{O}_l$$

exists and satisfies the Lyapunov equation. Uniqueness follows by the same arguments as in Theorem 2.2.3. If the system is observable, the observability matrix has rank equal to the system order n, which implies that P is positive definite.

Conversely, assume that P satisfies the Lyapunov equation (2.4). Let λ be en eigenvalue of A with associated eigenvector v. Multiplying (2.4) with v^* from the left and v from the right yields

$$(1-|\lambda|)v^*Pv = \|Cv\|_2^2 \tag{2.5}$$

By Theorem 1.3.5, $Cv = 0$ implies that $|\lambda| < 1$. If $Cv \neq 0$, then the right-hand side of (2.5) is positive. As $v^*Pv \geq 0$, we must then have $|\lambda| < 1$. Thus, all eigenvalues of A are inside the unit circle and the system is asymptotically stable. The proof is complete. ■

We will make use of this theorem to prove stability of optimal control laws later in the course.

2.3 Positively invariant sets

A limitation with asymptotic stability is that it disregards important transient properties of the state evolution. In addition to stability, it is often important to guarantee that the state does not violate given constraints. To this end, we will study the concept of *positively invariant sets*.

Definition 2.3.1 The set $\mathcal{I} \subseteq \mathbb{R}^n$ is *positively invariant* under the dynamics

$$x_{t+1} = f(x_t)$$

if

$$x_t \in \mathcal{I} \Rightarrow x_k \in \mathcal{I} \text{ for all } k \geq t$$

The definition states that if the state vector enters a positively invariant set at time t, then it will remain in this set for all future times. This property will be useful for guaranteeing constraint satisfaction. If, for example, we can prove that the set $\mathcal{S} = \{x \mid y_{\min} \leq Cx \leq y_{\max}\}$ is positively invariant, then we are sure that the limits on $y = Cx$ will never be violated.

A given constraint set $\mathcal{S}$ may not, in general, be positively invariant. In this case, it is important to be able to estimate the largest positively invariant set contained in $\mathcal{S}$, since it describes the initial values for which the state vector will remain in $\mathcal{S}$. To this end, we make the following definition.

Definition 2.3.2 The non-empty set $\mathcal{I}_\infty(\mathcal{S})$ is *the maximally positively invariant set contained in* $\mathcal{S}$ under $x_{t+1} = f(x_t)$ if it is positively invariant and contains all positively invariant sets contained in $\mathcal{S}$.

Positively invariant sets are defined for autonomous systems, *i.e.* systems without any external inputs. In this course, we will consider control systems and try to optimize the controlled input so that the closed loop system satisfies given constraints. In this case, it will sometimes be more natural to consider the concept of *control invariant sets*, defined next:

Definition 2.3.3 The set $\mathcal{C} \subseteq \mathbb{R}^n$ is *positively control invariant* under the dynamics

$$x_{t+1} = f(x_t, u_t)$$

and control constraint $u_t \in \mathcal{U}$ for $t \geq 0$ if

$$x_t \in \mathcal{C} \Rightarrow \exists \{u_t, u_{t+1}, \dots\} \text{ such that } u_t \in \mathcal{U} \text{ and } x_t \in \mathcal{C}\ \forall t \geq 0$$

In words, this definition states that if x_t belongs to a control invariant set, then there exists a control sequence which satisfies the control constraints and makes the state stay in the control invariant set for all future times.

Like for invariant sets, a given $\mathcal{S} \subseteq \mathbb{R}^n$ may not be control invariant. We will then be interested in finding the largest subset of $\mathcal{S}$ which is control invariant. We will use the following definition.

Definition 2.3.4 The non-empty set $\mathcal{C}_\infty(\mathcal{S}, \mathcal{U})$ is the maximally positively control invariant set contained in $\mathcal{S}$ under the dynamics $x_{t+1} = f(x_t, u_t)$ and the control constraint $u_t \in \mathcal{U}$ if it is positively control invariant and contains all positively control invariant sets contained in $\mathcal{S}$.

As indicated by the wording in Definition 2.3.2, maximal positively invariant sets are unique.

Invariant sets from Lyapunov functions

An important observation in our study of Lyapunov functions was their level sets are positively invariant: if the system state enters a level set of a valid Lyapunov function it will never leave. We state this result without proof here.

Theorem 2.3.1 Let $V : \mathbb{R}^n \mapsto \mathbb{R}$ satisfy

$$\Delta V(x) = V(f(x)) - V(x) \leq 0 \qquad \text{for all } x$$

Then, its level sets

$$\mathcal{L}_V(\alpha) = \{x \mid V(x) \leq \alpha\}$$

are positively invariant under the dynamics $x_{t+1} = f(x_t)$.

We have shown that stable linear systems $x_{t+1} = Ax_t$ admit quadratic Lyapunov functions $V(x) = x^T Px$ where $P \succ 0$ satisfies the Lyapunov equation $A^T PA - P + Q = 0$ for some $Q \succ 0$. The level sets of quadratic functions

$$\mathcal{L}_V(\alpha) = \left\{x \mid x^T Px \leq \alpha\right\}$$

define ellipsoids. Thus, stable linear systems admit invariant sets that are ellipsoids. Theorem 2.3.1 also applies to unstable systems. However, if we solve the Lyapunov equation for an unstable A, the solution P will be indefinite and its level sets will be unbounded.

With a given constraint set $\mathcal{S}$, the Lyapunov argument only allow us to guarantee invariance of the level sets fully contained in $\mathcal{S}$. In this case, the invariant sets induced by quadratic Lyapunov functions are typically neither unique nor maximal.

Maximal positively invariant sets for linear systems with linear constraints

When the dynamics is linear and the state and control constraints are defined by linear inequalities, the maximally invariant sets are also defined by linear inequalities. The inequalities defining these sets can be computed directly by the definition of invariance.

Theorem 2.3.2 Assume that a polyhedral constraint set

$$\mathcal{S} = \{x \mid Sx \leq s\}$$

is given. The maximally positively invariant set contained in $\mathcal{S}$ under the dynamics $x_{t+1} = Ax_t$ is

$$\mathcal{I}_\infty(\mathcal{S}) = \{x \mid Qx \leq q\}$$

where

$$Q = \begin{bmatrix} S \\ SA \\ SA^2 \\ \vdots \end{bmatrix} \qquad q = \begin{bmatrix} s \\ s \\ s \\ \vdots \end{bmatrix}$$

Proof. The proof follows directly from expressing the definition of positive invariance, *i.e.* $x_t \in \mathcal{S}$ for all $t \geq 0$, in terms of the initial state $x_t = A^t x_0$:

$$\begin{aligned} &x_0 \in \mathcal{S} \Leftrightarrow Sx_0 \leq s, \\ &x_1 \in \mathcal{S} \Leftrightarrow Sx_1 \leq s \Leftrightarrow SAx_0 \leq s, \\ &x_2 \in \mathcal{S} \Leftrightarrow Sx_2 \leq s \Leftrightarrow SA^2x_0 \leq s, \quad \text{etc.} \end{aligned}$$

■

The next example illustrates the ideas that we have discussed so far.

■ **Example 2.2** Consider the linear system $x_{t+1} = Ax_t$ with

$$A = \begin{bmatrix} 1.5 & -0.9 \\ 1.0 & 0.0 \end{bmatrix}$$

and state constraints

$$x_t \in \mathcal{S} = \{x \mid |x_i| \leq 1\} \qquad \text{for all } t = 0, 1, \ldots$$

We are interested in finding a positively invariant set $\mathcal{I}(\mathcal{S})$ contained in $\mathcal{S}$. We first use the Lyapunov function approach and thus solve the Lyapunov equation

$$A^T PA - P + I = 0$$

to find the solution

$$P = \begin{bmatrix} 27.9 & -19.9 \\ -19.9 & 23.6 \end{bmatrix}.$$

Every level set of $V(x) = x^T Px$ fully contained is $\mathcal{S}$ is invariant. The largest such level set is $\mathcal{L}_V(9.406)$ which is shown in Figure 2.2 (left). A few state trajectories validate that the set is invariant, and reasonably tight, since trajectories with initial values slightly outside the proposed set violate the state constraints. ■

A limitation of this result is that it expresses the invariant set in terms of semi-infinite matrices Q and q. This is partially due to the problem itself: for some systems, the maximally invariant set simply does not admit a finite polyhedral representation. However, in most practical cases, the maximally invariant set can be described by a finite number of inequalities and we can detect when we can stop adding constraints. To understand why and how, it is useful to give an alternative perspective on positively invariant sets. To this end, we introduce the following definition.

Definition 2.3.5 The predecessor set of $\mathcal{S} \subseteq \mathbb{R}^n$ under the dynamics $x_{t+1} = f(x_t)$ is the set

$$\text{pre}(\mathcal{S}) = \{x \mid f(x) \in \mathcal{S}\}$$

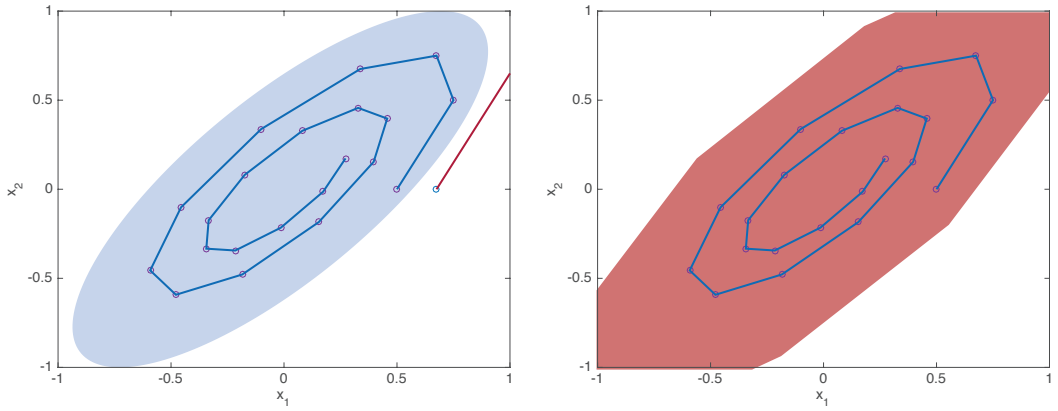

Figure 2.3: Positively invariant set guaranteed by a quadratic Lyapunov function (left) and two trajectories indicating that the set is reasonably tight. The maximal positively invariant set (right) is polyhedral and slightly larger than the invariant set computed using the Lyapuov fuction approach.

In words, $\text{pre}(\mathcal{S})$ is the set of states that will evolve into $\mathcal{S}$ in one step. With this definition, we can now construct the set of states which are guaranteed to stay in $\mathcal{S}$ for at least k time steps, denoted $\mathcal{I}_k(\mathcal{S})$. Clearly, $\mathcal{I}_0(\mathcal{S}) = \mathcal{S}$ and

$$\mathcal{I}_k(\mathcal{S}) = \mathcal{I}_{k-1}(\mathcal{S}) \cap \text{pre}(\mathcal{I}_{k-1}(\mathcal{S}))$$

Applying this formula recursively for $k = 1, 2, \ldots$ generates a sequence of sets $\mathcal{I}_k(\mathcal{S}) \subseteq \mathcal{I}_{k-1}(\mathcal{S}) \subseteq \mathcal{I}_{k-2}(\mathcal{S}) \subseteq \cdots \subseteq \mathcal{S}$, which converges to $\mathcal{I}_\infty = \lim_{k\to\infty} \mathcal{I}_k(\mathcal{S})$. If we detect that $\mathcal{I}_k(\mathcal{S}) = \mathcal{I}_{k-1}(\mathcal{S})$, then iteration has converged and we have found the maximal invariant set. When the iteration convergences in a finite number of steps, we say that the maximal invariant set is *finitely generated*. The smallest k such that $\mathcal{I}_{k+1}(\mathcal{S}) = \mathcal{I}_k(\mathcal{S})$ is called the *determindeness index*.The next result gives conditions for when the maximal positively invariant set is finitely generated.

Theorem 2.3.3 Consider the discrete-time linear system $x_{t+1} = Ax_t$. If (a) $|\lambda_i(A)| < 1$ for $i = 1, \ldots, n$ and (b) $\mathcal{S}$ is bounded and cotains 0 in its interior, then $I_\infty(\mathcal{S})$ is finitely generated.

Proof. The proof follows from Theorem 4.1 in [4]. ■

Although all asymptotically stable linear systems have maximal positively invariant sets which are finitely generated, the determindeness index (and hence the complexity of the representation) depends on both the dynamics and the constraints. The next example illustrates this fact.

■ **Example 2.3** Let us consider the mechanical system studied in Example 1.4 with $\omega_0 = 1$ with sampling time $h = 0.25$. We assume that the constraint set is $\mathcal{S} = \{x \mid \|x\|_\infty \leq 10\}$. If $\zeta = 1$, the dynamics is well-damped and $\mathcal{I}_\infty(\mathcal{S}) = \mathcal{I}_3(\mathcal{S})$; see Figure 2.4(left). When $\zeta = 0.1$, the dynamics becomes more oscillatory, and the determinedness index increasest to 5, Figure 2.4(middle). Making the system even more oscillatory increases the determinedness index even further.

To illustrate the impact of the constraints on $\mathcal{I}_\infty(\mathcal{S})$, we keep $\zeta = 0.1$ but shift the constraint set to be $\mathcal{S} = \{x \mid -5 \leq [x]_i \leq 15\}$. This situation is similar to shifting the equilibrium point of the system to $(-5, -5)$. In this case, the detminedness index jumps to 16, see Figure 2.4 (right). ■

2.4 Positively control invariant sets

Positively invariant sets are defined for autonomous systems, *i.e.* systems without any external inputs. In this course, we will consider control systems and try to optimize the controlled input

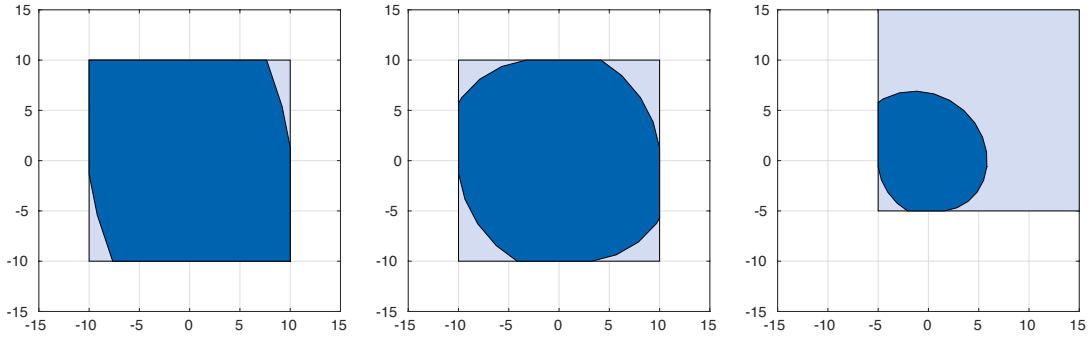

Figure 2.4: The determinedness index depends on both the system dynamics and the constraints. The constraint set $\mathcal{S}$ is in light blue and $\mathcal{I}_\infty(\mathcal{S})$ in dark blue. Decreasing the damping of the system increases the index (middle), while shifting the constraint set may alter both size, shape and determinedness index of the invariant set (right).

so that the closed loop system satisfies given constraints. In this case, it will sometimes be more natural to consider the concept of *control invariant sets*, defined next:

Definition 2.4.1 The set $\mathcal{C} \subseteq \mathbb{R}^n$ is *positively control invariant* under the dynamics

$$x_{t+1} = f(x_t, u_t)$$

and control constraint $u_t \in \mathcal{U}$ for $t \geq 0$ if

$$x_t \in \mathcal{C} \Rightarrow \exists \{u_t, u_{t+1}, \dots\} \text{ such that } u_t \in \mathcal{U} \text{ and } x_t \in \mathcal{C}\ \forall t \geq 0$$

In words, this definition states that if x_t belongs to a control invariant set, then there exists a control sequence which satisfies the control constraints and makes the state stay in the control invariant set for all future times. Like for invariant sets, a given $\mathcal{S} \subseteq \mathbb{R}^n$ may not be control invariant. We will then be interested in finding the largest subset of $\mathcal{S}$ which is control invariant.

Definition 2.4.2 The non-empty set $\mathcal{C}_\infty(\mathcal{S}, \mathcal{U})$ is the maximally positively control invariant set contained in $\mathcal{S}$ under the dynamics $x_{t+1} = f(x_t, u_t)$ and the control constraint $u_t \in \mathcal{U}$ if it is positively control invariant and contains all positively control invariant sets contained in $\mathcal{S}$.

Maximal positively control invariant sets are unique and can be computed in a similar manner as positively invariant sets. Let $\mathcal{S} = \{x \mid Sx \leq s\}$ be given and consider the linear system

$$x_{t+1} = Ax_t + Bu_t$$

under control constraints $Mu_t \leq m$ for all t. Then the maximally control invariant set $\mathcal{C}_\infty(\mathcal{S}, \mathcal{U})$ is the set of x_0 for which the inequalities

$$\begin{bmatrix} S & 0 & 0 & 0 & \cdots \\ 0 & M & 0 & 0 & \cdots \\ SA & SB & 0 & 0 & \cdots \\ 0 & 0 & M & 0 & \cdots \\ SA^2 & SAB & SB & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \end{bmatrix} \begin{bmatrix} x_0 \\ u_0 \\ u_1 \\ u_2 \\ \vdots \end{bmatrix} \leq \begin{bmatrix} s \\ m \\ s \\ m \\ s \\ \vdots \end{bmatrix}$$

admits a solution $(x_0, u_0, \dots)$. This set is a polyhedron in $\mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \times \cdots$, and the maximal control invariant set is the projection of this polyhedron onto its first n coordinates.

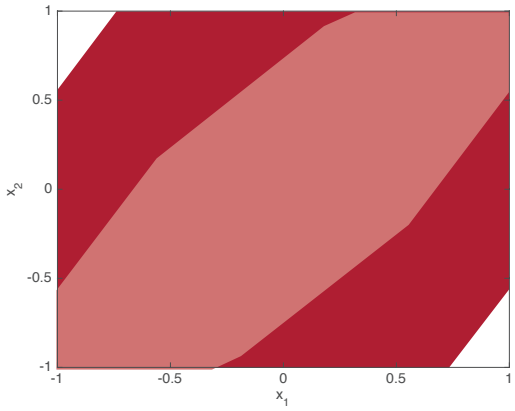

Figure 2.5: Maximally control invariant set (dark) and invariant set (light) for the linear system (2.7).

By a slight redefinition of predecessor sets, we can structure the controlled invariant set computations as we did for the invariant sets. Specifically, we use the following definition

Definition 2.4.3 The predecessor set of $\mathcal{S} \subseteq \mathbb{R}^n$ under the dynamics $x_{t+1} = f(x_t, u_t)$ and control constraints $u_t \in \mathcal{U}$ is

$$\text{pre}(\mathcal{S}) = \{x \mid \exists u \in \mathcal{U} \text{ such that } f(x,u) \in \mathcal{S}\}$$

We can now compute $\mathcal{C}_\infty(\mathcal{S}, \mathcal{U})$ recursively, analogously to how we computed $\mathcal{I}_\infty(\mathcal{S})$. To this end, let $\mathcal{C}_0(\mathcal{S}, \mathcal{U}) = \mathcal{S}$ and proceed with the iteration

$$\mathcal{C}_k(\mathcal{S}, \mathcal{U}) = \mathcal{C}_{k-1}(\mathcal{S}, \mathcal{U}) \cap \text{pre}(\mathcal{C}_{k-1}(\mathcal{S}, \mathcal{U})) \tag{2.6}$$

The recursion generates a sequence of sets which converges to $\mathcal{C}_\infty(\mathcal{S}, \mathcal{U}) = \lim_{k\to\infty} \mathcal{C}_k(\mathcal{S}, \mathcal{U})$. The maximally control invariant set is finitely generated if there is a k such that $\mathcal{C}_k(\mathcal{S}, \mathcal{U}) = \mathcal{C}_{k-1}(\mathcal{S}, \mathcal{U})$.

■ **Example 2.4** We now add a controlled input to the autonomous system used in the earlier examples and consider the linear system

$$x_{t+1} = \begin{bmatrix} 1.5 & -0.9 \\ 1.0 & 0.0 \end{bmatrix} x_t + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u_t \tag{2.7}$$

We add the control constraint that $|u_t| \leq 1$ and compute the maximally control invariant set via the recursive computations (2.6). As one can expect the control invariant set is larger than the invariant set for the autonomous system and even with the magnitude restrictions on u we are now able to guarantee constraint satisfaction for a large portion of initial values within the state constraints. ■

Bibliography

[1] S. Abbott. *Understanding Analysis*. Springer Verlag, 2015 (cited on page 30).

[2] K. J. Åström and B. Wittenmark. *Computer controlled systems: theory and design*. 2nd. Prentice Hall, 1990 (cited on pages 19, 21).

[3] Gene Franklin, J. D. Powell, and Abbas Emami-Naeini. *Feedback Control of Dynamic Systems (5th Edition)*. 5th edition. Prentice Hall, 2005 (cited on page 19).

[4] E. G. Gilbert and K. T. Tan. "Linear systems with state and control constraints: the theory and application of maximal output admissible sets". In: *IEEE Transactions on Automatic Control* 36.9 (Sept. 1991), pages 1008–1020 (cited on page 36).

[5] A. M. Lyapunov. "The General Problem of the Stability of Motion". Russian. Doctoral dissertation. Univ. Kharkov, 1892 (cited on page 27).