Signal theory: Part 1

Ivan Markovsky\ Vrije Universiteit Brussel

Contents

1 Introduction

Welcome to the beautiful subject of signals and systems! These notes cover

1. the *behavioral approach* to system theory,
2. representations of *linear time-invariant systems*, and
3. *least-squares approximation and estimation*.

You will build upon previously studied concepts, such as *convolution*, *differential equation*, and *transfer function*, in order to obtain a deeper understanding of dynamical phenomena. We will make links among concepts that you studied before but did not know are connected. In particular, you will learn how to convert one representation of a system, *e.g.*, an impulse response, into another equivalent one, *e.g.*, a state space. The ability to switching from one representation to another is a powerful solution tool for analysis and design. A famous example that illustrates this point is the recursive solution of the optimal filtering problem of linear Gaussian systems, by what came to be known as the *Kalman filter*. Between 1940 and 1960, N. Wiener and others tried without success to solve the problem using the classical at this time *transfer function approach*. In 1960, R. Kalman found an elegant solution using the *state space approach*, which at this time was new. Moreover, in solving the optimal filtering problem, Kalman developed much of the state space theory for linear multivariable systems, which proved useful for solving many other signal processing and control problems.

Solution of real-life engineering problems involve multidisciplinary skills. We will use concepts from

- *linear algebra*, such as matrix, subspace, basis, rank;
- *probability theory*, such as probability density function, mean, variance; and
- *optimization*, such as local and global solution, first and second order optimality conditions.

When needed, we will review them, so that in addition to system theory, you will refresh and expand your knowledge of linear algebra, probability, and optimization.

The word "theory" in the title does not mean that the course is useful for theoretical analysis only. You will learn to apply the theory in practice. For this purpose, we will use MATLAB. If you are not already familiar with the programming environment MATLAB or you need to refresh your knowledge, you can follow the "Introduction to MATLAB" course offered to EIT Bruface students.

The teaching and learning method that we will use is nicely summarized by the saying:

> "I hear, I forget; I see, I remember; I do, I understand." Chinese philosopher

Independent of the teaching method, the only way to understand anything is to use it. For signal theory, this means solving problems and working on (mini)projects, where the theory is applied. Unfortunately, the traditional teaching method is lecturing, which encourages passive absorption of information. We will follow a different method: 1. before coming to classes, you will prepare by reading this document and external sources, 2. in class, we will discuss difficult concepts in order to clarify what you did not understand by self reading, 3. in class, you will use the material to solve the problems from Sections 6–8 and learn how to present and defend your ideas. Submitting the solutions by October 30, you earn a passing mark 10/20 for part 1. (The actual mark above 10 will be determined by the final exam.)

2 The behavioral approach

The notion of a (mathematical) model is at the heart of science and engineering. Indeed, scientist build models for real-life phenomena and engineers use models to control the phenomena or make predictions about them. This section defines formally the notion of a model. Ohm's law, modeling a resistor, is used as an illustrative example. The general definition of a model is specialize for the case of dynamical models, *i.e.*, models of phenomena where the observed variables evolve in time. All subsections finish with a summary and references.

2.1 Definition of a system

Throughout mathematical engineering (control, signal processing, machine learning, ...), we use models of real-life *phenomena*. The nature of real-life is that it is infinitely complex. On the other hand, in order to be practically useful, a model should give a simple and accurate description of the phenomenon. A model of the reality often implies abstraction and simplification — we focus only on some aspects of the phenomenon and ignore others. Even those aspects that the model describes may be described approximately rather than exactly.

It is important to distinguish the model from the real-life phenomenon. The model can be a physical object, a piece of software, or a mathematical object. In signal processing and control, we use mathematical models. This means that the model is defined by equations.

> **Example (Resistor modeled by Ohm's law):** The real-life phenomenon is a resistor subject to a voltage or current source. The model of the phenomenon is defined by Ohm's. The model gives a relation between the current through and voltage across the resistor. Ohm's law is an idealization of the phenomenon for low-frequency and low-current sources. More complicated (nonlinear dynamic) models are used for high-frequencies and high-currents.
>
> **Question:** Give other examples of models.
>
> - What is the real-life phenomenon and what is the model?
> - When does the model adequately describe the phenomenon?

In order to define a mathematical model for a phenomenon, we first select the variables of interest. They are called *manifest variables* in order to distinguish them from other variables, called *latent variables*, that may be needed to describe the dynamical process but are not directly observable or are not of interest.

A *measurement experiment* involves observing the q manifest variables: $w_1, \ldots, w_q$. We collect them in a vector w and call the observed vector w an *outcome* of the experiment. Let $\mathscr{U}$ be the set, called the *universum*, of all possible outcomes from an experiment. The data

$$\mathscr{D} = \{ w^1, \ldots, w^N \},$$

collected from N experiments is a subset $\mathscr{D} \subset \mathscr{U}$ of the universum.

> **Example (Resistor modeled by Ohm's law):** The manifest variables are the current i through the resistor and voltage v across the resistor. A measurement experiment is a tuple (i, v). Since both i and v are real numbers, the universum $\mathscr{U}$ is the 2-dimensional real vector space $\mathbb{R} \times \mathbb{R} = \mathbb{R}^2$ ($\times$ denotes the Cartesian product of two sets).

The model defined by Ohm's law has no latent variables. A more complicated model of a resistor that aims to capture the dependence on the source frequency or the environmental temperature would have the frequency or the temperature as a latent variable.

Question: For the examples you gave in the answer to the previous question, what are the manifest variables and what is the universum? Are there latent variables?

Models may be obtained in practice using already existing scientific knowledge about the phenomenon. This is called *modeling from first principles*. In engineering the relevant scientific knowledge comes from physics. For example, in circuit theory, you learned how to derive model of electrical circuits, knowing what elements constitute the circuit and how they are connected.

A fundamentally different approach of obtaining models is by collecting data from the phenomenon and matching the empirically observed data by the prediction of a model. This modeling approach is called *data-driven modeling* or, in the context of dynamical systems, *system identification*. Depending on how much prior information about the phenomenon is used along with the data for the derivation of the model, data-driven methods are further split into *black-box* and *gray-box*. Black-box methods use no information about the internal structure of the phenomenon, while gray-box models assume that the true structure of the phenomenon is known and estimate from the data only a few parameters. In this classification, modeling from first principles is refered to as *white-box* modeling, *i.e.*, both the structure and the parameters are obtained without resort to empirical data.

Example (Resistor modeled by Ohm's law): Since we used knowledge from physics (Ohm's law) in modeling the resistor, the model is obtained from first principles and is according to the classification a white-box model. In the black-box modeling approach, we would, first, postulate a relation for i and v and, then, estimate the parameters of this relation from data. Note that, if the postulated relation is linear, we are back to Ohm's law. In this case, the to-be-estimated model parameter is the resistance R, which is unknown in the black-box approach but is given in the white-box approach.

Question: Give example of white-, gray-, and black-box models you've seen in other courses.

A model $\mathscr{B}$ is an exclusion rule: it postulates which outcomes are possible and which are not. As a mathematical object, the model is a subset of the universum, *i.e.*, $\mathscr{B} \subset \mathscr{U}$. The notation $w \in \mathscr{B}$ means that:

- the w is a *possible outcome* of the model $\mathscr{B}$,
- $\mathscr{B}$ is an *exact model* for the data point w,
- $\mathscr{B}$ is *unfalsified* by the data w.

Example (Resistor modeled by Ohm's law): Ohm's law postulates which outcomes $(i,v) \in \mathbb{R}^2$ are possible. These are the ones that satisfy the relation $i = Rv$, where R, the resistance, is a parameter of the model. Therefore, by definition the model is the set

$$\mathscr{B} = \{\, (i,v) \mid v = Ri \,\}.$$

Question: For the examples you gave in the answer to the previous question, how does the model restrict the possible outcomes from the universum?

The definition of a system as a set is general. It covers

- dynamic as well as static systems,
- nonlinear as well as linear systems,
- time-varying as well as time-invariant systems,
- discrete-time as well as continuous-time, and
- finite automata as well as continuous systems.

Indeed, any type of phenomena that can be quantified by observed variables fits in the set definition of a system.

We will refer to $\mathscr{B}$ also as the *behavior* of the model. The model is *exact* for the collected data $\mathscr{D}$ from the phenomenon if $\mathscr{D} \subset \mathscr{B}$. Otherwise, the model is *approximate*.

Example (Resistor modeled by Ohm's law): The model is not the function $v = Ri$, the set $\mathscr{B} = \{(i,v) \mid v = Ri\}$. This is an important distinction, because the function and the set are different mathematical objects. The set is more general than the function and may be described in different ways, *e.g.*, the same model may be defined as $\mathscr{B} = \{(i,v) \mid i = Av\}$, where the parameter is the admittance $A = 1/R$.

Example (Planetary motion modeled by Kepler's laws): Another example, which illustrates the advantage and generality of the behavioral definition of a model as a set is planetary motion. Let $w(t)$ be the vector giving the position of the plant in space at time t. Kepler proposed a model for the position w of a planet moving around the sun. His model is completely specified by the three famous laws (the laws of Kepler):

$$\mathscr{B} = \{ w \mid w \text{ satisfies the three laws of Kepler.} \}$$

Question: How was the basic notion of a model defined in other subjects that you've studied? Are these definitions related to each other and are they related to the behavioral definition?

Summary

- A mathematical model is an abstract object. It is used to mimic some features of a real-life phenomenon. The model and the modeled phenomenon are two distinct objects.
- Models are obtained from first principles (physical laws) or observed data (system identification). In practice, most often the two methods are mixed (gray-box modeling).
- Models describe observed variables $w_1, \ldots, w_q$ of the phenomenon. These variables are collected in a vector w.
- A system $\mathscr{B}$ is a set of possible outcomes; $w \in \mathscr{B}$ means that the outcome w is possible.

Notes and references

The definition of a system as a set was given in the three part paper [9] by J. C. Willems. This paper initiated what came to be known as *the behavioral approach* to systems and control. For more information about the behavioral approach, see

`https://en.wikipedia.org/wiki/Behavioral_modeling`

2.2 Dynamical systems

Dynamical systems have as outcomes one-dimensional signals, *i.e.*, functions of one independent variable, referred to as "time". In the previous section, a system was defined as a subset $\mathscr{B}$ of a universum set $\mathscr{U}$. In the context of dynamical systems, $\mathscr{U}$ is a set of functions $w : \mathbb{T} \to \mathbb{W}$, denoted by $\mathbb{W}^{\mathbb{T}}$. The sets $\mathbb{W}$ and $\mathbb{T} \subseteq \mathbb{R}$ are called, respectively, *signal space* and *time axis*. The signal space is the set where the system variables take on their values and the time axis is the set where the time variable takes on its values.

A system $\mathscr{B} \subseteq \mathbb{W}^{\mathbb{T}}$ is a set of functions from the universum set $\mathscr{U} = \mathbb{W}^{\mathbb{T}}$. As with any set, the behavior $\mathscr{B}$ can be specified in different ways, *e.g.*, by an equation $f(w) = 0$ that $w \in \mathscr{B}$ has to satisfy

$$\mathscr{B} = \{ w \in \mathbb{W}^{\mathbb{T}} \mid f(w) = 0 \}.$$

Example (Convolutional models): Consider the class of systems (called convolution al models) that are describe by the equation $y = h \star u$, where u is the input, y is the output, and h the impulse response of the system (a parameter that specifies the system). The behavior of a convolution model is

$$\mathscr{B} = \{ \left[\begin{smallmatrix} u \\ y \end{smallmatrix}\right] \mid y = h \star u \}.$$

The convolution equation $y = h \star u$ is a special case of $f(w) = 0$, for $f = y - hstaru$ and $w = \left[\begin{smallmatrix} u \\ y \end{smallmatrix}\right]$.

Question: Give other examples of behavioral equations $f(w) = 0$.

Of interest are systems with special properties. In the behavioral setting, a property of the system is defined in terms of the behavior and then translated to equivalent statements in terms of particular representations. Similarly, the statement that w is a trajectory of the system, *i.e.*, $w \in \mathscr{B}$, is translated to more convenient characterizations for numerical verification in terms of representations.

In the classical theory, system properties are often defined on the representation level; *i.e.*, a property of the system is defined as a property of a particular representation. (For example, controllability is defined as a property of a state space representation.) This has the drawback that such a definition might be representation dependent and therefore not a genuine property of the system itself.

It is more natural to work instead the other way around.

1. Define the property in terms of the behavior $\mathscr{B}$.

2. Find the implications of that property on the parameters of particular representations of the system.

The way of developing system theory as a sequence of the steps above is characteristic for the behavioral approach.

Question: List properties of the system you have studied before. Are they defined in terms of a model representation or in terms of the model behavior?

A static system $\mathscr{B}$ is *linear* when the universum set $\mathscr{U}$ is a vector space and the behavior $\mathscr{B}$ is a subspace. Analogously, a dynamical system $\mathscr{B}$ is linear when the signal space $\mathbb{W}$ is a vector space and $\mathscr{B}$ is a subspace of $\mathbb{W}^{\mathbb{T}}$.

Question: Is the behavioral definition of a linear system, given above, equivalent to the classical one that you've used before?

The universum set $\mathbb{W}^{\mathbb{T}}$ of a dynamical system has special structure that is not present in the static case. For this reason dynamical systems are richer in properties than static systems. Next, we restrict ourselves to the case when the time axis is either $\mathbb{T} = \mathbb{N}$ or $\mathbb{T} = \mathbb{Z}$ and define the property of time-invariance.

A system $\mathscr{B}$ is *time-invariant* if $\mathscr{B} \subseteq \sigma\mathscr{B}$, where σ is the backward shift operator

$$(\sigma w)(t) := w(t+1)$$

and $\sigma\mathscr{B} := \{\, \sigma w \mid w \in \mathscr{B} \,\}$. In the case $\mathbb{T} = \mathbb{Z}$, a system $\mathscr{B}$ is time-invariant if $\mathscr{B} = \sigma\mathscr{B}$. Time-invariance requires that if a time series w is a trajectory of a time-invariant system, then all its backward shifts $\sigma^t w$, $t > 0$, are also trajectories of that system.

Question: Compare the behavioral definition of a time-invariant system to the classical one that you've used before.

The class of linear time-invariant systems with $\mathbb{W} = \mathbb{R}^{\mathtt{w}}$ is denoted by $\mathscr{L}^{\mathtt{w}}$. Next, we discuss representations of the class $\mathscr{L}^{\mathtt{w}}$. In addition, we will focus on the discrete-time case.

Summary

- A dynamical system is set of trajectories. In the discrete-time case the trajectories are time series

$$w = \big(\ldots, w(-1), w(0), w(1), \ldots\big).$$

- In the behavioral approach, properties of the system are first defined in terms of the model behavior and then translated to equivalent statements in terms of model representations.
- Linearity means that the model is a subspace, *i.e.*, if $w \in \mathscr{B}$ then $\alpha w \in \mathscr{B}$ for all $\alpha \in \mathbb{R}$.
- Time-invariance means that the model contains all possible shifts of its trajectories, *i.e.*, if $w \in \mathscr{B}$, then $\sigma^\tau w \in \mathscr{B}$, for all τ.

Notes and references

This section is an adapted version of [5, Section 7.1].

3 Representations of linear time-invariant systems

The class of linear time-invariant systems $\mathscr{L}^{\mathtt{w}}$ is important in practice as well as in theory. Even though in reality systems are rarely linear time-invariant, they can be approximated around an operation point by a linear time-invariant system. Such local approximation is often sufficient to solve design problems such as stabilization.

In addition to approximation of complicated systems, the class $\mathscr{L}^{\mathtt{w}}$ is the basis for understanding of nonlinear system theory. Indeed, methods for analysis and design of nonlinear and time-varying systems are based on or are an extension of methods for linear time-invariant systems. For these reasons, we focus on the class of the linear time-invariant systems. In this section, we study two different ways to represent a system $\mathscr{B} \in \mathscr{L}^{\mathtt{w}}$: kernel of a polynomial operator and state space.

3.1 Kernel representation

Consider the vector difference equation

$$R_0 w(t) + R_1 w(t+1) + \cdots + R_l w(t+l) = 0, \qquad \text{where} \quad R_\tau \in \mathbb{R}^{g \times \mathtt{w}}. \tag{DE}$$

The left-hand-side is a linear combination of $\ell + 1$ consecutive samples of the time series w. Assuming that $R_l \neq 0$, the number of shifts l is called the *lag* of the equation. The set of solution of (DE) defines a dynamical system

$$\mathscr{B} = \{ w \in (\mathbb{R}^{\mathtt{w}})^{\mathbb{N}} \mid \text{(DE) holds} \}.$$

It is convenient to use polynomial matrix algebra to manipulate (DE). Operations on the system of difference equations are represented by operations on the polynomial matrix

$$R(z) := R_0 + R_1 z^1 + R_2 z^2 + \cdots + R_l z^l.$$

We denote by $\mathbb{R}^{g \times \mathtt{w}}[z]$ the set of $g \times \mathtt{w}$ polynomial matrices. The equation (DE) is compactly written in terms of a $g \times \mathtt{w}$ polynomial matrix R as $R(\sigma)w = 0$. Then, the system $\mathscr{B}$ induced by (DE) is

$$\ker\big(R(\sigma)\big) := \{ w \in (\mathbb{R}^{\mathtt{w}})^{\mathbb{N}} \mid R(\sigma)w = 0 \}. \tag{KER repr}$$

(KER repr) is a kernel representation of the system.

Any finite-dimensional linear time-invariant system $\mathscr{B} \in \mathscr{L}^{\mathtt{w}}$ admits a kernel representation, *i.e.*, there is a polynomial matrix $R \in \mathbb{R}^{\bullet \times \mathtt{w}}[z]$, such that $\mathscr{B} = \ker\big(R(\sigma)\big)$. The linearity of the system induced by (DE) follows from the linearity of (DE) with respect to w. The shift-invariance follows from the time-invariance of the coefficients $R_0, \ldots, R_l$, and the completeness follows from the fact that (DE) involves a finite number l shifts of the time series.

A kernel representation associated with a given $\mathscr{B} \in \mathscr{L}^{\mathtt{w}}$ is not unique. The non-uniqueness is due to:

1. linearly dependent equations, and
2. equivalence of the representations $\ker\big(R(\sigma)\big) = 0$ and $\ker\big(U(\sigma)R(\sigma)\big) = 0$, where $U \in \mathbb{R}^{g \times g}[z]$ is a unimodular matrix (a square polynomial matrix that has a polynomial inverse).

The minimal number of equations $\mathtt{p}$, the lag ℓ, and the total lag $\mathtt{n}$ (sum of the lags of the equations $R_i(\sigma)w = 0$) are invariants of $\mathscr{B}$. It turns out that $\mathtt{p}$ is equal to the number of outputs, called output cardinality, in an input/output representation. Correspondingly, the integer $\mathtt{m} := \mathtt{w} - \mathtt{p}$ is also an invariant of $\mathscr{B}$ and is called the input cardinality. It is equal to the number of inputs in an input/output representation. The total lag $\mathtt{n}$ is equal to the state dimension in a minimal state space representation of $\mathscr{B}$.

Summary

- Any finite dimensional linear time-invariant system admits a kernel representation (representation in terms of a linear constant coefficients difference or differential equation).
- The parameter of a kernel representation is a polynomial matrix R.
- The parameter R is not unique due to 1) redundant equations and 2) multiplication of R by a unimodular matrix.

Notes and references

This section is an adapted version of [5, Section 7.2].

3.2 State, input/output, and input/state/output representations

Modeling from first principles invariably requires the addition to the model of other variables apart from the ones that the model aims to describe (manifest variables). Such variables are called latent, and we denote them by l (not to be confused with the lag of a difference equation).

An important result, called the elimination theorem [8, Theorem 1], states that the behavior

$$\mathscr{B}(R,M) := \left\{ w \in (\mathbb{R}^{\mathtt{w}})^{\mathbb{N}} \;\middle|\; \exists\, l \in (\mathbb{R}^{\ell})^{\mathbb{N}}, \text{ such that } R(\sigma)w = M(\sigma)l \right\} \tag{LV repr}$$

induced by the latent variable equation

$$R(\sigma)w = M(\sigma)l \tag{LV eqn}$$

is linear time-invariant. The behavior $\mathscr{B}(R,M)$ is called manifest behavior of the latent variable system. The behavior of the manifest and latent variables together is called the full behavior of the system. The elimination theorem states that if the full behavior is linear time-invariant, then the manifest behavior is linear time-invariant; *i.e.*, by eliminating the latent variables, the resulting system is still linear time-invariant.

A latent variable system is *observable* if there is a map $w \mapsto l$, *i.e.*, if the latent variables can be inferred from the knowledge of the system and the manifest variables. The kernel representation is a special case of the latent variable representation for $R = I$.

State variables are special latent variables that specify the memory of the system. More precisely, latent variables x are called state variables if they satisfy the following axiom of state [10, Definition VII.1]:

$$(w_1,x_1),(w_2,x_2) \in \mathscr{B},\ t \in \mathbb{N}, \text{ and } x_1(t) = x_2(t) \quad \Longrightarrow \quad (w,x) \in \mathscr{B},$$

where

$$\big(w(\tau),x(\tau)\big) := \begin{cases} \big(w_1(\tau),x_1(\tau)\big) & \text{for } \tau < t \\ \big(w_2(\tau),x_2(\tau)\big) & \text{for } \tau \geq t. \end{cases}$$

A latent variable representation of the system is a state variable representation if there exists an equivalent representation whose behavioral equations are first order in the latent variables and zeroth order in the manifest variables. For example, the equation

$$\sigma x = A'x + B'v, \qquad w = C'x + D'v$$

defines a state representation. It is called state representation with a driving input because v acts like the input: v is free and, together with the initial conditions, determines a trajectory $w \in \mathscr{B}$. The system induced by (A', B', C', D') is

$$\mathscr{B}_{\mathrm{ss}}(A', B', C', D') := \left\{ w \in (\mathbb{R}^{\mathtt{w}})^{\mathbb{N}} \;\middle|\; \exists\, v \text{ and } x \in (\mathbb{R}^{\mathtt{n}})^{\mathbb{N}}, \text{such that } \sigma x = A'x + B'v,\ w = C'x + D'v \right\}.$$

Contrary to the classical approach, in the behavioral approach the model is not a priori defined in terms of input and output variables. Instead, the variables of interest, *i.e.*, the manifest variables are viewed on an equal footing. It is possible however to partition the manifest variables into inputs and outputs. It turns out that this partitioning is in general not unique.

> **Example (Resistor modeled by Ohm's law):** In the model of the resistor $\mathscr{B} = \{ (i, v) \mid v = Ri \}$, we can choose the current i as an input, in which case the voltage v is an output, or the voltage v as an input, in which case, the current becomes an output.

In general, a subset of variables can be selected as inputs of the model if the projection of the behavior on these variables is unrestricted. An input/output partitioning of the variables is obtained when the subset of free variables has a maximal cardinality (which is independent of the choice and is called the input cardinality).

There always exists an input/output partitioning of the variables of $\mathscr{B} \in \mathscr{L}^{\mathtt{w}}$. In a (minimal) kernel representation $\ker\big(R(\sigma)\big) = \mathscr{B}$, the choice of such a partitioning amounts to the selection of a full-rank square submatrix of R. The variables corresponding to the columns of R that form the full-rank submatrix are dependent variables and the other variables are free.

Let $\ker\big(R(\sigma)\big)$ be a minimal kernel representation of $\mathscr{B} \in \mathscr{L}^{\mathtt{w}}$. One can always find a permutation matrix $\Pi \in \mathbb{R}^{\mathtt{w}\times\mathtt{w}}$, such that $P \in \mathbb{R}^{\mathtt{p}\times\mathtt{p}}[z]$, defined by $R\Pi =: \begin{bmatrix} Q & -P \end{bmatrix}$, has a nonzero determinant and the rational polynomial matrix

$$G(z) := P^{-1}(z)Q(z) \in \mathbb{R}^{\mathtt{p}\times\mathtt{m}}(z) \tag{TF}$$

is proper. This requires selecting a submatrix P among all full-rank square submatrices of R that has determinant of maximal degree. Then the corresponding partitioning of w, $\mathrm{col}(u, y) := \Pi^{\top} w$, is an input/output partitioning. G being proper implies that u is not anticipated by y.

The difference equation

$$P(\sigma)y = Q(\sigma)u \tag{I/O eqn}$$

with an input/output partitioning Π is called an input/output equation, and the matrix G, defined in (TF), is called the transfer function of the system $\mathscr{B} := \ker\big(R(\sigma)\big)$.

The system $\mathscr{B} \in \mathscr{L}^{\mathtt{w}}$ induced by an input/output equation with parameters (P, Q) (and input/output partitioning defined by Π) is

$$\mathscr{B}_{\mathrm{i/o}}(P, Q, \Pi) := \{ w := \Pi\, \mathrm{col}(u, y) \in (\mathbb{R}^{\mathtt{w}})^{\mathbb{N}} \mid P(\sigma)y = Q(\sigma)u \}. \tag{I/O repr}$$

(I/O repr) is called an input/output representation of the system $\mathscr{B} := \mathscr{B}_{\text{i/o}}(P, Q, \Pi)$. If Π is the identity matrix $I_{\mathtt{w}}$, it is skipped in the notation of the input/output representation.

Apart from kernel, state, and input/output representations, any linear time-invariant system $\mathscr{B} \in \mathscr{L}^{\mathtt{w}}$ admits a representation by an input/state/output equation

$$\sigma x = Ax + Bu, \qquad y = Cx + Du, \qquad w = \Pi \operatorname{col}(u, y), \tag{I/S/O eqn}$$

in which both the input/output and the state structure of the system are explicitly displayed [8, Theorem 3]. The system $\mathscr{B}$, induced by an input/state/output equation with parameters (A, B, C, D) and Π, is

$$\mathscr{B}_{\text{i/s/o}}(A, B, C, D, \Pi) := \{ w := \Pi \operatorname{col}(u, y) \in (\mathbb{R}^{\mathtt{w}})^{\mathbb{N}} \mid \exists\, x \text{ such that } \sigma x = Ax + Bu,\ y = Cx + Du \}. \tag{I/S/O repr}$$

(I/S/O repr) is called an input/state/output representation of the system $\mathscr{B} := \mathscr{B}_{\text{i/s/o}}(A, B, C, D, \Pi)$.

An input/state/output representation is not unique. The minimal state dimension $\mathtt{n} = \dim(x)$ among all input/state/output representations of $\mathscr{B}$, however, is invariant (denoted by $\mathbf{n}(\mathscr{B})$). We denote the class of linear time-invariant systems with $\mathtt{w}$ variables, at most $\mathtt{m}$ inputs, and minimal state dimension at most $\mathtt{n}$ by $\mathscr{L}_{\mathtt{m}}^{\mathtt{w},\mathtt{n}}$.

Summary

- Any finite dimensional linear time-invariant system admits an input/state/output representation (representation in terms of a first order linear constant coefficients difference or differential equation).
- The input/output partition of the variables is defined by a permutation matrix. In general it is not unique.
- The parameters A, B, C, D are not unique: 1) the state dimension may be nonminimal, 2) a change of basis of the state vector results in an equivalent model.

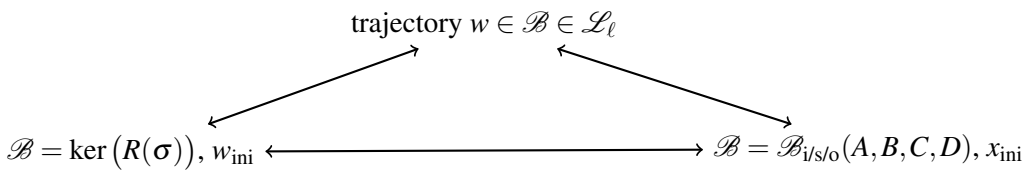
Notes and references

This section is an adapted version of [5, Section 7.4].

3.3 Links among the representations

We've seen in the previous sections that a finite-dimensional linear time-invariant system can be represented in both kernel and input/state/output form. Suppose that a model is given in a kernel form but the solution of the problem at hand is more naturally expressed in terms of an input/state/output representation. In this case, we need to convert a given representation of the model to an equivalent one.

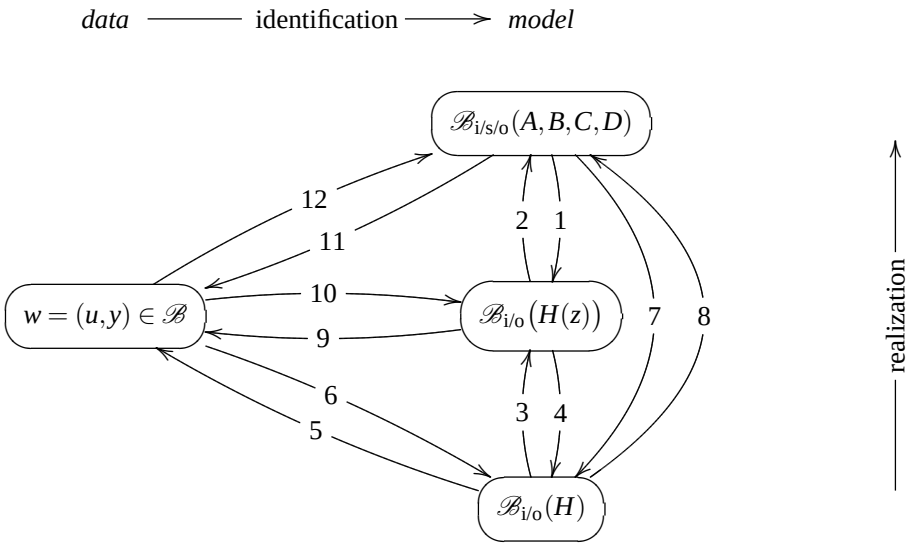
An exact trajectory $w \in \mathscr{B}$ carries information about the data generating system $\mathscr{B}$. Under certain assumptions, see [5, Section 8.3], w completely specifies $\mathscr{B}$, *i.e.*, there is a map $w \to \mathscr{B}$. It is called an exact (or deterministic) *identification problems*. The reverse map $\mathscr{B} \to w$ is the *simulation problems* and, of course, is not unique. Given a model, a trajectory of that model can be parameterized by an input and initial conditions.



A more general scheme that includes the impulse response representation is shown in Figure 1. This scheme summarizes much of the linear system theory:

- identification problems, which aims to derive a model from data,
- simulation problems, which derive data from a model, and
- realization problems, which derive a state space representation from transfer function or impulse response.

The three representations: impulse response, transfer function, and state space have increasing level of structure. Any linear system has an impulse response representation, however, only finite-dimensional systems have transfer function representations. A state space model is, in additional to finite dimensional, causal and explicitly displays the state of the system. Going from a less structured model to a more structured one is in general a harder problem. The realization problem, which requires state construction, is a nontrivial problem. Its solution is described in [6, Sections 6.5–8] and [4, Sections 2.2 and 3.1].



1. $H(z) = C(Iz-A)^{-1}B+D$
2. realization of a transfer function
3. Z or Laplace transform of $H(t)$
4. inverse transform of $H(z)$
5. convolution $y_{\text{d}} = H \star u_{\text{d}}$
6. exact identification
7. $H(0) = D$, $H(t) = CA^{t-1}B$ (discrete-time), $H(t) = Ce^{At}B$ (continuous-time), for $t > 0$
8. realization of an impulse response
9. simulation with input u_{d} and $x(0) = 0$
10. exact identification
11. simulation with input u_{d} and $x(0) = x_{\text{ini}}$
12. exact identification

Figure 1: There are 12 links among the four representations of a linear time-invariant dynamical system—input/state/output, transfer function, impulse response, and general trajectory. The links 6, 10, and 12 from a trajectory to parameters of the model are (exact/deterministic) /identification problems/. The reverse links 5, 9, and 11 are /simulation problems/. The links from 2 and 8 from transfer function and impulse response to a input/state/output representation are /realization problems/. The reverse links 1 and 7 are given by closed form expressions. The links 3 and 4 among transfer function and impulse response are given by the Z- (in discrete-time) / Laplace- (in continuous-time) transform and its inverse.

The realization of an impulse response can be viewed as a special exact identification problem. Indeed, the impulse response is a special trajectory of the model—the one obtained from the impulse input under zero initial condition. Identification from a general trajectory is an important subject, which is the topic for other course in your master program. We will touch upon it in the special case of of identification of autonomous systems, which as you will see is an equivalent problem to the realization problem.

Summary

- The links among the four model representations—general trajectory, impulse response, transfer function, and state space—are an important part of system theory—the identification, simulation, and realization problems.
- The four model representations have different degree of structure and generality. The least structured being a general trajectory and the most structured an input/state/output representation.
- A transition from a less structured to a more structured representation requires conditions and involves a nontrivial solution method. The most complicated problem, studied in this course, is the realization problem, which is the transition from an impulse response to a state space representation. Even more complicated identification problems will be the topic of other courses.

Notes and references

This section is an extract from Sections 6 and 7 of [5].

4 Stochastic signals and least-squares estimation

4.1 Stochastic signals

Addition of random variables $\leftrightarrow$ convolution of probability density functions

A discrete random variable X is characterized by the probabilities p_i of the events $X = i$,

$$p_i = \text{prob}(X = i).$$

We have that $0 \leq p_i \leq 1$ and $\sum p_i = 1$. The vector p of all probabilities is called the *probability density function* of X.

In this subsection, we consider the problem of finding the probability density function of the sum $X^{(1)} + X^{(2)}$ of two *independent identically distributed (i.i.d.)* random variables $X^{(1)}$ and $X^{(2)}$ with pdf's $p^{(1)}$ and $p^{(2)}$, *i.e.*, we aim to find the probabilities of the events $X^{(1)} + X^{(2)} = k$. An example of the sum of discrete random variables problem is dice trowing. In this case, $p_i^{(j)} = 1/6$, for $j = 1, 2$ and $i = 1, \ldots, 6$.

By the independence assumption, we have that

$$\text{prob}(X^{(1)} = i \text{ and } X^{(2)} = j) = p_i^{(1)} p_j^{(2)}.$$

Moreover, $X^{(1)} + X^{(2)} = k$ is a union of the mutually exclusive events $X^{(1)} = i$ and $X^{(2)} = k - i$, which probability is $p_i^{(1)} p_{k-i}^{(2)}$. Therefore,

$$\text{prob}(X^{(1)} + X^{(2)} = k) = \sum p_i^{(1)} p_{k-i}^{(2)} = (p^{(1)} \star p^{(2)})(k).$$

In the dice trowing example, $X^{(1)} + X^{(2)}$ takes on the values $2, 3, \ldots, 12$ with corresponding probabilities

$$\frac{1}{6}(1,1,1,1,1,1) \star \frac{1}{6}(1,1,1,1,1,1) = \frac{1}{36}(1,2,3,4,5,6,5,4,3,2,1).$$

Central limit theorem

What would be the distribution of a random variable which is the sum of $N \geq 2$ independent identically distributed random variables? The convolution of probabilities result allow us to make an experiment which would answer this question empirically:

```
n = 7; T = 20; N = 4;
box = zeros(1, T); box((round(T/2) - n):(round(T/2) + n)) = 1;
p = box; figure(1), stem(p), pause(1),
for i = 1:N,
  p = conv(p, box); stem(p), pause(1)
end
```

It turns out that as $N \to \infty$, the probability density function of the sum of independent random variables convergence to a Gaussian distribution. This result is called the *central limit theorem*. (The distribution of the variables need not be identical.)

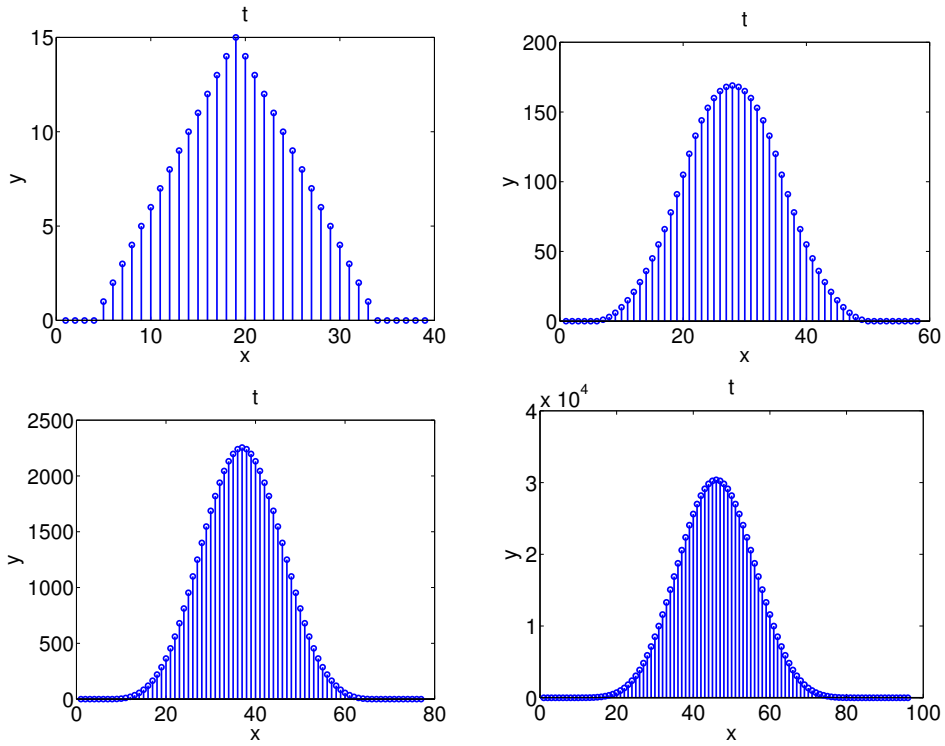


Figure 2: Empirical illustration of the central limit theorem: the convolutions of N rectangular pulses converges to a Gaussian distribution.

Gaussian distribution (second order process)

The univariate Gaussian distribution is

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2} \left(\frac{x-\mu_x}{\sigma_x} \right)^2 \right).$$

Note that this distribution is completely characterized by the mean μ and the variance σ^2. We use the notation $x \sim \mathbf{N}(\mu, \sigma^2)$ for a Gaussian random variable with mean μ and variance σ^2.

In the multivariable case, the Gaussian random vector $x \sim \mathbf{N}(\mu, V)$ has distribution

$$p(x) = \frac{1}{(2\pi)^{n/2}\sqrt{\det(V)}} \exp\left(-\frac{1}{2}(x-\mu)^\top V^{-1}(x-\mu) \right), \tag{G}$$

characterized by the mean vector μ and the *covariance matrix* V. With diagonal V, x is a collection of *uncorrelated random variables*.

An affine transformation $y = Ax + b$ of a Gaussian random vector $x \sim \mathbf{N}(\mu, V)$, results in another Gaussian random vector $y \sim \mathbf{N}(A\mu + b, AVA^\top)$. Of particular interest is a transformation A that de-correlates (whitens) x, *i.e.*

$$AVA^\top = D, \qquad \text{where } D \text{ is diagonal.}$$

Such a transformation can be found from the Cholesky decomposition of D

$$V = LDL^\top, \qquad L \text{ — lower triangular with ones on the diagonal.}$$

Then, A can be taken as L^{-1}.

Random processes

A discrete-time random process y is an infinite sequence of random variables. An independent sequence of Gaussian random variables (called white Gaussian process) is often used as a model for measurement noise and disturbance acting on a system. Another example is a sequence of random variables, where each variable is a function of the previous variable only. This process is called *Markov chain* and is used in queuing theory. Also autonomous linear time-invariant systems with random initial condition generates a Markov process.

The *mean of the process* is the sequence of the means and the *covariance matrix* (assuming zero mean process) is defined as

$$R_y = \big[\underbrace{\mathbf{E}\big(y(i)y(j)\big)}_{r_y(i,j)} \big] = \mathbf{E} \begin{bmatrix} y(1) \\ y(0) \\ \vdots \end{bmatrix} \begin{bmatrix} y(1) & y(0) & \cdots \end{bmatrix}$$

Note that the covariance matrix is infinite but any leading $N \times N$ sub-block is a positive semi-definite matrix. Since we will consider Gaussian processes, we will not need moments of order higher than two. The mean and the covariance matrix completely characterize a Gaussian process.

The process is called (wide sense) *stationary* if $\mathbf{E}(y)$ is a constant and $r_y(i,j)$'s depends only on $\tau = i - j$. In this case, R_y is a Toeplitz matrix. The function

$$r_y(\tau) = E\big(y(t)y(t-\tau)\big)$$

called *correlation function* of the process is symmetric, *i.e.*, $r_y(\tau) = r_y(-\tau)$, and has its maximum at the origin, *i.e.*, $r_y(0) \geq |r_y(\tau)|$, for all τ.

It can be shown that the correlation function is the convolution of the processes with its time-reversed signal

$$r_y = y \star \mathrm{rev}(y),$$

where "rev" is the time reversal operation

$$\big(\mathrm{rev}(y)\big)(t) = y(-t).$$

With probability one, a realization of a random process has no finite energy and therefore it has no Fourier transform. However, a random processes usually has finite average power and can be characterized by the average power spectral density. The Fourier transform of the correlation function

$$S_y(\omega) = \sum_{k=-\infty}^{\infty} r_y(k) e^{-\mathbf{i}\omega k}$$

is called the *spectral density* function.

The spectral density is positive

$$S_y(\omega) \geq 0, \qquad \text{for all } \omega$$

and for real processes, it is symmetric

$$S_y(\omega) = S_y(-\omega)$$

The theorem of Parseval

$$\sum_{t=-\infty}^{\infty} = \frac{1}{2\pi} \int_{-\pi}^{\pi} S_y(\omega) d\omega$$

states that the energy in time is equal to the energy in the frequency domain.

By definition a process is *ergodic* if its expectation can be computed by time averaging

$$r_y(\tau) = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} y(t) y(t-\tau).$$

Consider an LTI system with transfer function H and input u, which is a stochastic process with power spectral density S_x. Then, the output y of the system is also a stochastic process. Its power spectral density S_y is

$$S_y(\omega) = |H(\omega)|^2 S_u(\omega).$$

In order to show this, first note that

$$r_y(\tau) = \int \int h(\alpha) h(\beta) r_u(\tau + \alpha - \beta) d\alpha d\beta$$

Then, using

$$S_y(\omega) = \int r_y(\tau) e^{-\mathbf{i}\omega\tau} d\tau$$

and changing variables $\lambda = \tau + \alpha - \beta$, we have

$$\begin{aligned}
S_y(\omega) &= \int \int \int h(\alpha) h(\beta) r_u(\tau + \alpha - \beta) e^{-\mathbf{i}\omega\tau} d\alpha d\beta d\tau \\
&= \int \int \int h(\alpha) h(\beta) r_u(\tau + \alpha - \beta) e^{-\mathbf{i}\omega(\lambda - \alpha + \beta)} d\alpha d\beta d\lambda \\
&= \int h(\alpha) e^{-\mathbf{i}\omega\alpha} d\alpha \int h(\beta) e^{-\mathbf{i}\omega\beta} d\beta \int r_u(\lambda) e^{-\mathbf{i}\omega\lambda} d\lambda \\
&= H(-\omega) H(\omega) S_u(\omega) \\
&= |H(\omega)|^2 S_u(\omega)
\end{aligned}$$

Another characterization of the output y of an LTI system with stochastic process as an input is in terms of the cross-power spectral density S_{uy}

$$S_{uy}(\omega) = H(\omega)S_u(\omega)$$

A white Gaussian noise has power spectrum $S_u(\omega) = 1$, for all ω. The problem of designing a filter which has input white Gaussian process and as an output a process with spectral density S_y is called *shaping filter design problem*. The inverse problem: designing a filter which has input a process with spectral density S_u and output a white Gaussian process is called *whitening filter design problem*. The solution of both problems is given by the spectral factorization of the correlation function.

Wiener-Khintchine theorem

Consider a random process y. The Fourier transform of the autocorrelation function r

$$S_y = F(r_y).$$

is by definition the power spectral density of y. The Wiener-Khintchine theorem says that

$$S_y := F(y)F^*(y) = |F(y)|^2$$

The proof

$$S_y = F(y)F^*(y) = F(y)F\big(\mathrm{rev}(y)\big) = F\big(y \star \mathrm{rev}(y)\big) = F(r_y)$$

is based on the following three properties:

- $F(y \star y) = F(y)F(y)$,
- $F\big(\mathrm{rev}(y)\big) = F^*(y)$, and
- $y \star \mathrm{rev}(y) = r_y$.

Direct computation requires $O(n^2)$ flops. Using the FFT, the computational cost is $O(n\log(n))$.

Estimation principles

- **Conditional expectation** Consider two random variables x and y with joint pdf $p_{x,y}$. We want to infer x from observations of y. The estimator is given in the form $\widehat{x} = h(y)$. If x and y are independent, nothing can be said about x, seeing y. The optimal estimator turns out to be the conditional expectation $\widehat{x} = \mathbf{E}(x|y)$ of x, given y. An important special case is linear estimators and Gaussian pdf.
- **Maximum likelihood** Consider a parameter dependent pdf $p(x,\theta)$. Substituting the observed data in $p(x,\theta)$ and viewing the resulting function $L(\theta)$ in θ as the "likelihood" for the occurrence of the data, given the parameters θ, we obtain the maximum likelihood estimation principle: maximize the likelihood for all admissible parameter values $\theta \in \Theta$.

Spectral estimation

The spectral estimation problem aims to determine the spectral content of a signal (random process) based on finite set of observations. The spectral density describes the distribution of power of the signal with frequency. A "physical way of doing it" is:

1. filter with a sufficiently narrow band-pass filter centered at $f = f_0$ and measure the power of the output.
2. divide the power by the filter width,
3. repeat the process for different f_0

This method is traditionally based on the Fourier transform. Since 1980 new "modern" ("high resolution" approaches) were developed. In the last 10 years compressive sensing is taking over.

Note that the task of estimating the spectral density based on finite data is an ill-posed (impossible to solve) problem. At best we can estimate a subset of the most significant values, which are assumed to be the M leading coefficients. If the process exhibits strong correlations for lags $k > M$, the results are heavily biased.

The "impossibility problem" can be resolved by postulating specific forms of the spectral density, e.g., rational form. The process is parameterized by a finite (and small) number of coefficients. The spectral estimation problem becomes the one of parameter estimation. It is important however that the model is an accurate representation of the true spectral density.

Nonparameteric methods include:

- periodogram

$$\widehat{S}(\omega) = \frac{1}{N}\left|\sum_{t=1}^{T} y(t)e^{-\mathbf{i}\omega t}\right|^2$$

  Square the absolute value of the DFT on the process.

- correlogram

$$\widehat{S}(\omega) = \sum_{k=-(N-1)}^{N-1} \widehat{r}_y(k)e^{-\mathbf{i}\omega k}$$

  Apply the DFT on the estimated autocorrelation sequence.

$$\widehat{r}_y(k) = \frac{1}{T}\sum_{t=k+1}^{N} y(t)y(t-k), \qquad \text{for } 0 \le k \le T-1$$

  For negative lags $\widehat{r}_y(-k) = \widehat{r}_y(k)$, $k = 0, \ldots, T-1$.

4.2 Least-squares estimation

Definition and geometric interpretation

The least-squares method for solving approximately an overdetermined system $Ax = y$ of equations is defined as follows. Choose x such that the 2-norm of the residual (equation error) $e(x) := y - Ax$ is minimized. A minimizer

$$\widehat{x}_{\text{ls}} := \arg\min_x \| \underbrace{y - Ax}_{e(x)} \|_2 \tag{1}$$

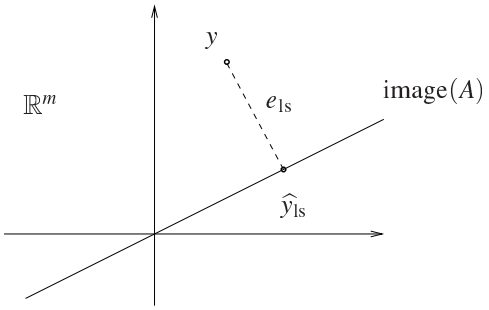


Figure 3: Geometric interpretation of the least-squares approximation problem (1): projection of y onto image(A).

is called a *least-squares approximate solution* of the system $Ax = y$. (Note that in general $\widehat{x}_{\text{ls}}$ is not a solution of $Ax = y$.)

Denote by $\widehat{y}_{\text{ls}} := A\widehat{x}_{\text{ls}}$ the projection of y on the image of A and let

$$e_{\text{ls}} := \widehat{y}_{\text{ls}} - A\widehat{x}_{\text{ls}}.$$

We call $\widehat{y}_{\text{ls}}$ the *least-squares approximation* of y and e_{ls} is the *approximation error* or *residual*.

Let a_i be the ith row of A. We refer to the vector $\text{col}(a_i, y_i)$ as a "data point". We have,

$$\begin{aligned} A\widehat{x}_{\text{ls}} = \widehat{y}_{\text{ls}} \quad &\iff \quad \begin{bmatrix} A & \widehat{y}_{\text{ls}} \end{bmatrix} \begin{bmatrix} \widehat{x}_{\text{ls}} \\ -1 \end{bmatrix} = 0 \\ &\iff \quad \begin{bmatrix} a_i & \widehat{y}_{\text{ls},i} \end{bmatrix} \begin{bmatrix} \widehat{x}_{\text{ls}} \\ -1 \end{bmatrix} = 0, \quad \text{for } i = 1, \ldots, m \end{aligned}$$

so that for all i, $(a_i, \widehat{y}_{\text{ls},i})$ lies on the subspace perpendicular to $(\widehat{x}_{\text{ls}}, -1)$. $(a_i, \widehat{y}_{\text{ls},i})$ is an the least-squares approximation of the i data point $\text{col}(a_i, y_i)$.

$$(a_i, \widehat{y}_{\text{ls},i}) = (a_i, \widehat{y}_{\text{ls},i}) + (0, e_{\text{ls},i}),$$

and $(0, e_{\text{ls},i})$ is the least-squares approximation error. Note that $e_{\text{ls},i}$ is the vertical distance from (a_i, y_i) to the subspace.

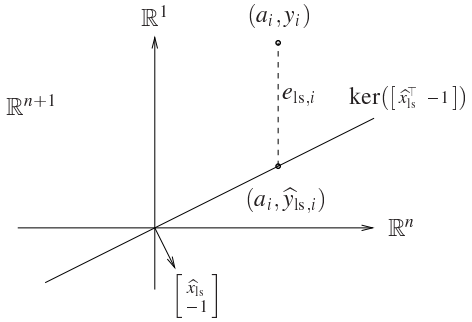


Figure 4: Geometric interpretation of the least-squares approximation in the (data space) $\mathbb{R}^{n+1}$.

Analytical formula for the least-squares approximation

Assume that $m \geq n = \text{rank}(A)$, *i.e.*, A is full column rank. Next, we show that

$$\widehat{x}_{\text{ls}} = (A^\top A)^{-1} A^\top y.$$

The residual norm is

$$\|e(x)\|_2^2 = \|y - Ax\|_2^2 = (y - Ax)^\top (y - Ax) = x^\top A^\top Ax - 2y^\top Ax + y^\top y.$$

In order to minimize it over x, we set the gradient with respect to x equal to zero

$$\nabla_x \|e(x)\|_2^2 = \nabla_x (x^\top A^\top Ax - 2y^\top Ax + y^\top y) = 2A^\top Ax - 2A^\top y = 0.$$

This gives the linear equation $A^\top Ax = 2A^\top$ in x, called *normal equation*.

By A full column rank, we have that $A^\top A$ is nonsingular, so that

$$\widehat{x}_{\text{ls}} = (A^\top A)^{-1} A^\top y$$

is the *unique least-squares approximate solution*.

Notes:

- $A_{\text{ls}} := (A^\top A)^{-1} A^\top$ is a left-inverse of A
- $\widehat{x}_{\text{ls}}$ is a linear function of y (given by the matrix A_{ls})
- If A is square, $\widehat{x}_{\text{ls}} = A^{-1} y$ (*i.e.*, $A_{\text{ls}} = A^{-1}$)
- $\widehat{x}_{\text{ls}}$ is an exact solution if $Ax = y$ has an exact solution
- $\widehat{y}_{\text{ls}} := A\widehat{x}_{\text{ls}} = A(A^\top A)^{-1} A^\top y$ is a least-squares approximation of y

Orthogonality principle

The matrix

$$\Pi_{\text{image}(A)} := A(A^\top A)^{-1} A^\top \in \mathbb{R}^{m \times m}$$

is the *orthogonal projector* onto the subspace $\mathscr{L} := \text{image}(A)$. Suppose that the columns of A form an orthonormal basis for $\mathscr{L}$. Then, $\Pi_{\text{image}(Q)} := AA^\top$.

The least-squares residual vector

$$e_{\text{ls}} := y - A\widehat{x}_{\text{ls}} = \underbrace{\left(I_m - A(A^\top A)^{-1} A^\top\right)}_{\Pi_{(\text{image}(A))^\perp}} y$$

is orthogonal to $\text{image}(A)$

$$\langle e_{\text{ls}}, A\widehat{x}_{\text{ls}} \rangle = y^\top \left(I_m - A(A^\top A)^{-1} A^\top\right) A\widehat{x}_{\text{ls}} = 0. \tag{2}$$

It can be shown (see exercises) that the orthogonality condition (2) is a necessary and sufficient condition for $\widehat{x}_{\text{ls}}$ being a least squares approximate solution to $Ax = b$.

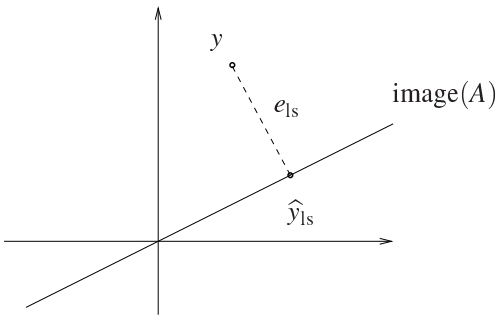


Figure 5: A geometric interpretation of the least-squares approximation in the (data space) $\mathbb{R}^{n+1}$.

Weighted least-squares

Given a positive definite matrix $W \in \mathbb{R}^{m\times m}$, define the weighted 2-norm $\|e\|_W^2 := e^\top We$. and the weighted least-squares approximate solution

$$\widehat{x}_{W,\mathrm{ls}} := \arg\min_x \|y - Ax\|_W^2.$$

It can be shown (see exercises) that

$$\widehat{x}_{W,\mathrm{ls}} = (A^\top WA)^{-1}A^\top Wy.$$

Also the least-squares orthogonality principle holds for the weighted least-squares problem as well by replacing the inner product $\langle e, y\rangle$ by the weighted inner product $\langle e, y\rangle_W := e^\top Wy$.

5 Concluding remarks

Summary

- We use stochastic processes to model measurement noise and disturbances acting on the system.
- A stochastic process is a sequence of random variables. It is characterized by the (joint) probability density function.
- A Gaussian process has probability density function of the form (G) on page 15. It is completely characterized by the first two moments: mean and covariance.
- Linear functions preserve the Gaussian distribution, but change the covariance matrix.
- The process is stationary when it is zero mean and the covariance is a function of one argument (the difference of the indeces).
- The process is ergodic when the expectation operation can be computed by averaging over time.
- The Wiener-Khintchine theorem justifies the Fourier transform of the correlation (the spectral density) as having the meaning of distribution of power over frequencies.
- The maximum likelihood estimation principle applied to a Gaussian random variable leads to the least-squares estimation problem.

- Another reason (the real one!) for considering Gaussian distributed random variables is that the in this and only this case, the maximum likelihood estimator has an analytical expression.
- The least squares estimation problem has the geometrical interpretation of projection on a subspace (the model). The approximation error is orthogonal to the model (orthogonality principle).

Some parallels between deterministic and stochastic signals and systems

deterministic		stochastic
2-norm approximation	—	maximum-likelihood estimation for Gaussian distribution
time-invariance	—	stationarity
impulse response	—	correlation function
frequency response	—	spectral density

Notes and references

The material collected in this section is sampled from many different references. I recommend the following ones [1, 2, 7].

6 Problems

6.1 Linear algebra

Rank of a matrix and solution of linear equations

- Without using a computer, determine whether the matrix $\left[\begin{smallmatrix}1&2&3\\1&2&4\\1&2&5\end{smallmatrix}\right]$ is invertible and what its rank is.
- Using a computer, solve the system $\left[\begin{smallmatrix}1&2&3\\4&5&6\\7&8&9\end{smallmatrix}\right] x = \left[\begin{smallmatrix}1\\1\\1\end{smallmatrix}\right]$.
- Given the following four vectors: $v_1 = \left[\begin{smallmatrix}1\\1\\1\end{smallmatrix}\right]$, $v_2 = \left[\begin{smallmatrix}1\\2\\3\end{smallmatrix}\right]$, $v_3 = \left[\begin{smallmatrix}3\\5\\7\end{smallmatrix}\right]$, $v_4 = \left[\begin{smallmatrix}0\\1\\t\end{smallmatrix}\right]$. For which value(s) of t is the subspace, spanned by v_1 and v_2, the same as the subspace, spanned by v_3 and v_4?

Fast method for computing A^{100}

How many scalar multiplications requires the direct computation of A^{100} as $\underbrace{A\cdots A}_{100}$ for a 2×2 matrix A? \
Suggest a faster method. Using the method, find a good approximation of $\left[\begin{smallmatrix}-1/4&1/4\\-3/2&1\end{smallmatrix}\right]^{100}$.

Solution: The matrix product AB, for $A,B\in\mathbb{R}^{2\times 2}$, requires 8 multiplications, so that A^{100} requires 99×8 multiplications. A fast method is obtained by the eigenvalue decomposition $A = V\Lambda V^{-1}$ of A, because

$$A^{100} = V\Lambda^{100}V^{-1} = V\begin{bmatrix}\lambda_1^{100} & \\ & \lambda_2^{100}\end{bmatrix}V^{-1},$$

requires only $2\times 99+12$ multiplications once the eigenvalue decomposition is computed. In the example the eigenvalues of A are $1/2$ and $1/4$. We have $(1/2)^{100} < 10^{-30}$ and $(1/4)^{100} < 10^{-60}$, so that $A^{100}\approx 0$. □

Eigenvalues and eigenvectors of a companion matrix

The companion matrix related to the 3rd order polynomial

$$a(\lambda) := (\lambda-1)(\lambda-2)(\lambda-3) = -6+11\lambda-6\lambda^2+\lambda^3$$

is the 3×3 matrix

$$C_a := \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 6 & -11 & 6 \end{bmatrix}. \tag{3}$$

1. Find 3 linearly independent eigenvectors of C_a.
2. Show that the eigenvalues of C_a are 1, 2, and 3.
3. Based on 1 and 2 conjecture and prove general properties of the eigenvalues and eigenvectors of an $n\times n$ companion matrix C_a related to an nth order polynomial a.

Solution: The companion matrix related to the polynomial

$$a(\lambda) := a_0 + a_1\lambda + \cdots + a_{n-1}\lambda^{n-1} + \lambda^n$$

of degree n is the $n\times n$ matrix

$$C(a) := \begin{bmatrix} 0 & 1 & & \\ & & \ddots & \\ & & & 1 \\ -a_0 & -a_1 & \cdots & -a_{n-1} \end{bmatrix}.$$

(All missing elements are zeros.) Assuming that the eigenvalues $\lambda_1,\ldots,\lambda_n$ of $C(a)$ are simple, 1. the n linearly independent eigenvectors of $C(a)$ are

$$v_i := \begin{bmatrix} 1 & \lambda_i & \lambda_i^2 & \cdots & \lambda_i^{n-1} \end{bmatrix}^\top, \qquad \text{for } i = 1,\ldots,n,$$

2. the eigenvalues $\lambda_1,\ldots,\lambda_n$ are the roots of the polynomial a. □

Orthogonality of left and right eigenvectors

A left eigenvector of a matrix A, related to an eigenvalue λ_i, is an $1\times n$ vector w_i, such that

$$w_i A = \lambda_i w_i.$$

Find the left eigenvectors of (3) and verify that they are orthogonal to the corresponding (right) eigenvectors v_i

$$Av_i = \lambda_i v_i.$$

Show that w_i, v_i can be chosen, so that $w_i v_i = 1$.

Computation time for the discrete Fourier transform

1. Find a matrix representation F of the discrete Fourier transform of signal $x \in \mathbb{R}^n$.

2. What is the number of multiplications needed to compute $\widehat{x} = Fx$ by matrix-vector multiplication? Compare this number with the $n\log_2(n)$ multiplications needed for the same computation by the fast Fourier transform.

3. Using Matlab's `tic` and `toc` functions, measure the computation times of the matrix-vector multiplication and the fast Fourier transform methods for the computation of the discrete Fourier transform of x. (Use a random x.) Repeat the experiment for different size n of x and plot the results. Do the empirical observations match the theoretical predictions?

Computation time for the convolution operation

1. Find a matrix representation M_h of the convolution $h \star x$ of $h \in \mathbb{R}^n$ and $x \in \mathbb{R}^n$.

2. What is the number of multiplications needed to compute $y = M_h x$ by matrix-vector multiplication?

3. Propose a fast method for convolution based on the fast Fourier transform. What is the computational cost of this method?

4. Measure the computation times of the matrix-vector multiplication and the fast Fourier transform methods for the computation of the convolution

$$y = \exp_\lambda \star x, \qquad \text{where } h(t) = \exp_\lambda(t) = e^\lambda t.$$

   (Use a random x.) Repeat the experiment for different size n and plot the results. Do the empirical observations match the theoretical predictions?

5. For the special case of exponential h, can you propose another fast method?

Identification of a linear function

Let $\mathbb{R}^n$ be the \$n\$-dimensional real vector space. A linear function f from $\mathbb{R}^m$ to $\mathbb{R}^n$ has a matrix representation $f(x) = Ax$, where A is an $m \times n$ real matrix. The problem is to find A from observed data

$$\mathscr{D} := \{ (x^1, y^1), \ldots (x^N, y^N) \}, \qquad \text{where } y^i = f(x^i). \tag{4}$$

1. Give conditions under which it is possible to find A from $\mathscr{D}$. Describe a computational method that does the job. Implement the method in a language of your choice and test it on an example.

2. If you can choose the points $x^1, \ldots, x^N$ how many and what points would you choose?

3. If the y^i's are corrupted by additive zero mean, uncorrelated, Gaussian noise e^i, *i.e.*, in (4)

$$y^i = f(x^i) + e^i, \qquad \text{for } i = 1, \ldots, N,$$

   how would you estimate A from $\mathscr{D}$?

Kroneker product

Let vec be the "matrix vectorization" operator defined by

$$vec(\begin{bmatrix} a_1 & \cdots & a_n \end{bmatrix}) := \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$$

and let $A \otimes B$, where A is an $m \times n$ matrix, be a matrix product (called the Kroneker product) operator defined by

$$A \otimes B := \begin{bmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{bmatrix}.$$

Consider the function

$$f : \text{vec}(X) \mapsto \text{vec}(AXB)$$

.

1. Show that f is linear.
2. Show that $B^\top \otimes A$ is a matrix representation of f.

6.2 Difference equations / LTI systems

Scalar equation (autonomous system)

Consider all sequences $y = \big(y(1), y(2), \dots\big)$ that satisfy the difference equation

$$y(t) = a_1 y(t-1) + \cdots + a_n y(t-n), \qquad \text{for } t = 1, 2, \dots \tag{5}$$

for some given $a_1, \dots, a_n \in \mathbb{R}$. A particular solution y of (5) is determined by the "initial conditions"

$$x_{\text{ini}} := \big(y(-n+1), y(-n+2), \dots, y(0)\big).$$

1. Give a closed form expression for the solution y of (5), in terms of $a_1, \dots, a_n$ and x_{ini}.
2. Under what conditions on $a_1, \dots, a_n$ are all solutions y of (5) bounded?
3. For given $a_1, \dots, a_n$, explain how to select the initial conditions x_{ini}, for which the corresponding solution y is bounded.

Equation with two variables (single-input single-output system)

Consider the linear time-invariant system defined by the difference equation

$$y(t) = a_1 y(t-1) + \cdots + a_n y(t-n) + b_0 u(t) + b_1 u(t-1) + \cdots + b_m u(t-m), \qquad \text{for } t = 1, 2, \dots \tag{6}$$

for some given $a_1, \dots, a_n \in \mathbb{R}$ and $b_0, b_1, \dots, b_m \in \mathbb{R}$. u is the "input" and y is the "output" of the system. A particular output y is determined by the input u and the initial conditions

$$x_{\text{ini}} := \big(y(-n+1), y(-n+2), \dots, y(0), u(-m+1), u(-m+2), \dots, u(0)\big).$$

1. Under what conditions on $a_1, \ldots, a_n$ and $b_0, b_1, \ldots, b_m$, all bounded inputs u generate bounded outputs (independent of the initial condition)?

2. If the conditions of 1 are not satisfied, for a given x_{ini}, explain how to select u, so that y is bounded.

3. If the conditions of 1 are not satisfied, explain how to select u, so that y is bounded for all x_{ini}.

Elimination of transient by setting initial conditions

Consider a stable single-input single-output linear time-invariant system $\mathscr{B}$. Choose the initial conditions x_{ini}, so that the transient response associate with an input $u(t) = \sin(\omega t)$ is eliminated.

Decomposition into stable and anti-stable subsystems

Consider a state space representation $\sigma x = Ax + Bu$ of a linear time-invariant system. Let $\lambda(A)$ be the set of eigenvalues of A and let $\lambda(A) = \lambda_{\text{s}} \cup \lambda_{\text{a}}$, where $|\lambda| \leq 1$, for all $\lambda \in \lambda_{\text{s}}$ and $|\lambda| > 1$, for all $\lambda \in \lambda_{\text{a}}$. In this problem, you will construct a change of basis, such that the state space representation of the system in the new basis decouples the anti-stable dynamics from the stable one:

$$\sigma \begin{bmatrix} x_{\text{s}} \\ x_{\text{a}} \end{bmatrix} = \begin{bmatrix} A_{\text{s}} & A_{\text{sa}} \\ 0 & A_{\text{a}} \end{bmatrix} \begin{bmatrix} x_{\text{s}} \\ x_{\text{a}} \end{bmatrix} + \begin{bmatrix} B_{\text{s}} \\ B_{\text{a}} \end{bmatrix} u. \tag{7}$$

1. Use the Schur decomposition (implemented in the function `schur` of Matlab) to compute the stable/anti-stable decomposition (7).

2. Use the stable/anti-stable decomposition to generate bounded trajectories of the system.

3. Use the stable/anti-stable decomposition to elimination of transient due to applying an input $u(t) = \sin(\omega t)$ on the interval [0,∞).

Multiple poles (problem from [3])

Consider the autonomous system represented by a difference equation

$$y(t+2) - 2ay(t+1) + a^2 y(t) = 0.$$

(Its characteristic polynomial has both roots equal to $\lambda = a$.)

1. Show that both $y(t) = a^t$ and $y(t) = ta^t$ are solutions.

2. Find the trajectory generated from the initial conditions $y(0) = 1$ and $y(1) = 0$.

Solution **SOL**

1. To check $y(t) = a^t$, we have

$$a^{t+2} - 2a \cdot a^{t+1} + a^2 \cdot a^t = a^{t+2} - 2a^{t+2} + a^{t+2} = 0.$$

Thus, $y(t) = a^t$ is a solution. Checking $y(t) = ta^t$ we have

$$\begin{aligned}(t+2)a^{t+2} - 2a(t+1)a^{t+1} + a^2 ta^t &= (t+2)a^{t+2} - 2(t+1)a^{t+2} + ta^{t+2} \\ &= (t+2-2t-2+t)a^{t+2} = 0.\end{aligned}$$

Thus, $y(t) = ta^t$ is also a solution.

2. A second-order linear difference equation has two degrees of freedom in its general solution, *i.e.*, two linearly independent solutions can form a fundamental set of solutions. We have found two solutions a^t and ta^t. It is easy to prove that these two solutions are linear independent. Because we can't find two constant c_1 and c_2 at least one of which is nonzero to satisfy

$$c_1 a^t + c_2 ta^t = 0$$

for all $t = 0, 1, 2, \ldots, N$. Thus, any solution $y(t)$ can be expressed as a linear combination of the fundamental set of solutions a^t and ta^t.

$$y(t) = c_1 a^t + c_2 ta^t,$$

where c_1 and c_2 are constant. Using the conditions $y(0) = 1$ and $y(1) = 0$, we find c_1 and c_2

$$\begin{aligned} y(0) = c_1 a^0 = 1 \qquad &\Longrightarrow \qquad c_1 = 0 \\ y(1) = c_1 a + c_2 a = a + c_2 a = 0 \qquad &\Longrightarrow \qquad c_2 = -1. \end{aligned}$$

Thus, the solution to this equation is

$$y(t) = a^t - ta^t.$$

[3], Chapter 2, Problem 2

A bank offers 7% annual interest. What would be the overall annual rate if the 7% interest were compounded quarterly?

Solution **SOL**

Let $y(t)$ denote the amount in the account at the beginning of season t and the bank pays interest at the end of each season. If the 7% interest were compounded quarterly, then the quarter interest is 7/4% and the account balance is governed by

$$y(t+1) = (1 + 7/4\%)y(t).$$

For an year, suppose $y(t)$ is the amount in the account at the beginning of this year, then $y(t+4)$ is the amount in the account at the beginning of following year. In order to know the overall annual rate, we should specify the relationship between $y(t)$ and $y(t+4)$. We have

$$\begin{aligned} y(t+4) &= (1+7/4\%)y(t+3) = (1+7/4\%)^2 y(t+2) \\ &= (1+7/4\%)^3 y(t+1) = (1+7/4\%)^4 y(t). \end{aligned}$$

Thus, the annual rate is

$$(1+7/4\%)^4 - 1 = 7.1859\%.$$

[3], Chapter 2, Problem 5

Find the second order linear homogeneous difference equation which generates the sequence 1, 2, 5, 12, 29, 70, 169. What is the limiting ratio of consecutive terms?

Solution **SOL**

By observing the sequence we find the relationship among three consecutive terms is

$$\begin{aligned} 5 &= 2\cdot 2+1 \\ 12 &= 2\cdot 5+2 \\ 29 &= 2\cdot 12+5 \\ 70 &= 2\cdot 29+12 \\ 169 &= 2\cdot 70+29 \end{aligned}$$

This relation can be written as a second-order linear homogeneous difference equation

$$y(t+2) = 2y(t+1)+y(t).$$

By dividing both sides of the equation by $y(t+1)$, the equation becomes

$$\frac{y(t+2)}{y(t+1)} = 2+\frac{y(t)}{y(t+1)}.$$

Here $\frac{y(t+1)}{y(t)}$ is the ratio of two consecutive term. When $t\to\infty$, the ratio converges to a constant which defines as a. Therefore

$$\frac{y(t+2)}{y(t+1)} = \frac{y(t+1)}{y(t)} = a, \qquad t\to\infty.$$

The limiting ratio a satisfy an equation

$$a = 2+\frac{1}{a} \quad\Longrightarrow\quad a^2 = 2a+1 \quad\Longrightarrow\quad a = 1\pm\sqrt{2}.$$

However, all terms of the sequence are positive so that the ratio of consecutive terms is positive. Therefore, the limiting ratio $a = 1+\sqrt{2}$.

[3], Chapter 2, Problem 10

Consider the second order difference equation

$$y(t+2)-2ay(t+1)+a^2y(t) = 0.$$

Its characteristic polynomial has both roots equal to $\lambda = a$.

1. Show that both $y(t) = a^t$ and $y(t) = ta^t$ are solutions.

2. Find the solutions of this equation that satisfies the auxiliary conditions $y(0) = 1$ and $y(1) = 0$.

Solution **SOL**

1. To check $y(t) = a^t$, we note that $y(t+2) = a^{t+2}$ and $y(t+1) = a^{t+1}$

$$a^{t+2} - 2a \cdot a^{t+1} + a^2 \cdot a^t = a^{t+2} - 2a^{t+2} + a^{t+2} = 0.$$

Thus, $y(t) = a^t$ is a solution. Checking $y(t) = ta^t$ we have

$$\begin{aligned}(t+2)a^{t+2} - 2a(t+1)a^{t+1} + a^2 ta^t &= (t+2)a^{t+2} - 2(t+1)a^{t+2} + ta^{t+2} \\ &= (t+2-2t-2+t)a^{t+2} = 0.\end{aligned}$$

Thus, $y(t) = ta^t$ is also a solution.

2. A second-order linear difference equation has two degrees of freedom in its general solution, i.e., two linearly independent solutions can form a fundamental set of solutions. We have found two solutions a^t and ta^t. It is easy to prove that these two solutions are linear independent. Because we can't find two constant c_1 and c_2 at least one of which is nonzero to satisfy $c_1 a^t + c_2 ta^t = 0$ for all $t = 0, 1, 2, \ldots, T$. Thus, any solution $y(t)$ can be expressed as a linear combination of the fundamental set of solutions a^t and ta^t.

$$y(t) = c_1 a^t + c_2 ta^t,$$

where c_1 and c_2 are constant. Using the conditions $y(0) = 1$ and $y(1) = 0$, we find c_1 and c_2

$$\begin{aligned} y(0) = c_1 a^0 = 1 \qquad &\Longrightarrow \qquad c_1 = 0 \\ y(1) = c_1 a + c_2 a = a + c_2 a = 0 \qquad &\Longrightarrow \qquad c_2 = -1. \end{aligned}$$

Thus, the solution to this equation is

$$y(t) = -ta^t$$

Newton cooling ([3], Chapter 2, Problem 19)

A thermometer reading 21°C, which has been inside a house for a long time, is taken outside. After one minute the thermometer reads 15°C; after two minutes it reads 11°C. What is the outside temperature? (According to Newton's law of cooling, an object of higher temperature than its environment cools at a rate that is proportional to the difference in temperature.)

Solution: Let $y(t)$ be the reading of the thermometer at time t and let $\bar{u}$ be the environmental temperature. From Newton's law of cooling, we have that

$$\frac{\mathrm{d}}{\mathrm{d}t} y = a\big(\bar{u}s - y\big)$$

for some unknown constant $a \in \mathbb{R}$, $a > 0$, which describes the cooling process. Integrating the differential equation, we obtain an explicit formula for y in terms of the constant a, the environmental temperature $\bar{u}$, and the initial condition $y(0)$

$$y(t) = e^{-at} y(0) + (1 - e^{-at})\bar{u}, \qquad \text{for } t \geq 0 \tag{8}$$

The problem is to find $\bar{u}$ from (8) given that $y(0) = 21$, $y(1) = 15$, and $y(2) = 11$. Substituting the data in (8), we obtain a nonlinear system of two equations in the unknowns $\bar{u}$ and $f := e^{-a}$

$$\begin{cases} y(1) = fy(0) + (1-f)\bar{u} \\ y(2) = f^2 y(0) + (1-f^2)\bar{u} \end{cases} \tag{9}$$

We may stop here and declare that the solution can be computed by a method for solving numerically a general nonlinear system of equations.

System (9), however, can be solved without using "nonlinear" methods. Define Δy to be the temperature increment from one measurement to the next, *i.e.*, $\Delta y(t) := y(t) - y(t-1)$, for all t. The increments satisfy the homogeneous differential equation $\frac{\mathrm{d}}{\mathrm{d}t}\Delta y(t) = a\Delta y(t)$, so that

$$\Delta y(t+1) = e^{-a}\Delta y(t) \qquad \text{for } t = 0, 1, \ldots \tag{10}$$

From the given data we evaluate

$$\Delta y(0) = y(1) - y(0) = 15 - 21 = -6, \qquad \Delta y(1) = y(2) - y(1) = 11 - 15 = -4.$$

Substituting in (10), we find the constant $f = e^{-a} = 2/3$. With f known, the problem of solving (9) in $\bar{u}$ is linear, and the solution is found to be $\bar{u} = 3°\mathrm{C}$. □

Embedded statics ([3], Chapter 4, Problem 21)

Suppose a system is described by a set of equations of the form

$$\begin{bmatrix} T \\ 0 \end{bmatrix} x(t+1) = \begin{bmatrix} C \\ D \end{bmatrix} x(t) + \begin{bmatrix} u(t) \\ v(t) \end{bmatrix},$$

where $x(t)$ is an \$n\$-dimensional vector, T and C are $m \times n$ matrices, D is an $(n-m) \times n$ matrix, and $u(t)$ and $v(t)$ are m and \$(n-m)\$-dimensional vectors, respectively. Assume that the $n \times n$ matrix $\left[\begin{smallmatrix} T \\ D \end{smallmatrix}\right]$ is nonsingular. Following the steps below, it is possible to convert this system to state vector form.

1. Define $y = Tx$ and show that with this definition, and the lower part of the system equation, one may express $x(t)$ in the form

   $$x(t) = Hy(t) - Gv(t)$$

   Give an explicit definition of G and H.

2. Show that the top part of the original system can be written in the state vector form

   $$y(t+1) = Ry(t) + Bv(t) + u(t)$$

   and give expressions for R and B. Note that x can be recovered from y using part 1.

3. Apply this procedure to the following example.

   > Samuelson proposed a model of the national economy based on the following assumptions. National income $Y(k)$ is equal to the sum of consumption $C(k)$, investment $I(k)$, and government expenditure $G(k)$. Consumption is proportional to the national income of the preceding year; and investment is proportional to the increase in consumer spending of that year over the preceding year. In equation form, the Samuelson model is

   $$\begin{aligned} Y(k) &= C(k) + I(k) + G(k) \\ C(k+1) &= mY(k) \\ I(k+1) &= \mu\big(C(k+1) - C(k)\big) \end{aligned}$$

6.3 Transition among LTI model representations

$(A,B,C,D) \mapsto$ **impulse response**

Find the impulse response of the linear time-invariant system

$$\mathscr{B}(A,B,C,D) := \{\, (u,y) \mid \text{there is } x, \text{ such that } \sigma x = Ax + Bu,\ y = Cx + Bu \,\}.$$

$(A,B,C,D) \mapsto$ **transfer function**

Find the transfer function of the linear time-invariant system $\mathscr{B}(A,B,C,D)$.

$(A,B,C,D) \mapsto (A',B',C',D')$

Let $\mathscr{B}(A,B,C,D) = \mathscr{B}(A',B',C',D) = \mathscr{B}$ be two minimal state space representations of the system $\mathscr{B}$. Find the change of basis transformation V, which maps the parameters A,B,C to A',B',C'.

Harmonic oscillator

Find a state space representation of the harmonic oscillator $\frac{\mathrm{d}^2}{\mathrm{d}t^2}y = -\alpha y$, where $\alpha > 0$.

6.4 Behavioral approach

Definition of stability

Using the behavioral approach (a system is a set of trajectories), define the stability property.

Step function output

Write down a representation of an autonomous linear time-invariant system that can give a step function output.

Solution: The continuous-time system, defined by

$$\frac{\mathrm{d}}{\mathrm{d}t}y = 0$$

and the discrete-time system, defined by

$$\sigma y = y$$

have constant output $y(0)e^{0t}$ and $y(0)1^t$, for $t \geq 0$. □

Sine function output

Write down a representation of an autonomous LTI system that can give a sine with frequency ω output.

Solution: A sine with frequency ω can be written as $ce^{\mathbf{i}\omega t} + \bar{c}e^{-\mathbf{i}\omega t}$, for some $c \in \mathbb{C}$. For the continuous-time system, we have

$$p(s) = (s - \mathbf{i}\omega)(s + \mathbf{i}\omega) = s^2 + \omega^2 \quad \Longrightarrow \quad \frac{\mathrm{d}}{\mathrm{d}t}y + \omega^2 y = 0.$$

For the discrete-time system, we have

$$p(z) = (z - e^{\mathbf{i}\omega})(z + e^{-\mathbf{i}\omega}) = z^2 - 2\cos(\omega)z + 1 \quad \Longrightarrow \quad \sigma^2 y - 2\cos(\omega)\sigma y + y = 0.$$

□

$y \stackrel{?}{\in} \mathscr{B}(A,C)$

Check if $y_\mathrm{d} = (2,3,5,9,17,33,64)$ is a possible output of the system defined by the state space representation

$$x(t+1) = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} x(t), \quad y(t) = \begin{bmatrix} 1 & 1 \end{bmatrix} x(t). \tag{SS}$$

If y_d is not a possible output of the system, suggest a way of correcting it, so that the corrected signal is.

Solution: $y_\mathrm{d} = \big(y_\mathrm{d}(1), \ldots, y_\mathrm{d}(T)\big) \in \mathscr{B}(A,C)$ if and only if there is $x(0)$, such that $y_\mathrm{d}(t) = CA^t x(0)$, for $t = 1, \ldots, T$. Written in a matrix form, this condition is the following system of linear equations

$$\begin{bmatrix} y_\mathrm{d}(1) \\ y_\mathrm{d}(2) \\ \vdots \\ y_\mathrm{d}(T) \end{bmatrix} = \underbrace{\begin{bmatrix} C \\ CA \\ \vdots \\ CA^{T-1} \end{bmatrix}}_{\mathscr{O}} x(0). \tag{$*$}$$

For the given example, we have

$$\begin{bmatrix} 2 \\ 3 \\ 5 \\ 9 \\ 17 \\ 33 \\ 64 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 4 \\ 1 & 8 \\ 1 & 16 \\ 1 & 32 \\ 1 & 64 \end{bmatrix} \begin{bmatrix} x_1(0) \\ x_2(0) \end{bmatrix},$$

which has no solution, so that $y_\mathrm{d} \notin \mathscr{B}(A,C)$. A way of correcting y_d, so that it becomes an output of $\mathscr{B}(A,C)$, is to solve (3) in a least-squares sense, and define the corrected output $\widehat{y} = \mathscr{O}\widehat{x}(0)$, where $\widehat{x}(0) = (\mathscr{O}^\top \mathscr{O})^{-1} \mathscr{O}^\top y_\mathrm{d}$ is the least-squares approximate solution of (3). $\widehat{y}$ has the property that $\|y_\mathrm{d} - \widehat{y}\|_2$ is minimized over all outputs $\widehat{y}$ of $\mathscr{B}(A,C)$. □

$y \stackrel{?}{\in} \ker(P(z))$

Check if $y_\mathrm{d} = (2,3,5,9,17,33,64)$ is a possible output of the system defined by the difference equation

$$2y(t) - 3y(t+1) + y(t+2) = 0. \tag{KER}$$

If y_d is not a possible output of the system, suggest a way of correcting it, so that the corrected signal is.

Solution: $y_\mathrm{d} = \big(y_\mathrm{d}(1), \ldots, y_\mathrm{d}(T)\big) \in \ker(P(z))$ if and only if

$$P_0 y(t) + P_1 y(t+1) + \cdots + P_n y(t+n) = 0, \qquad \text{for } t = 1, \ldots, T-n.$$

Written in a matrix form, this condition is the following system of linear equations

$$\begin{bmatrix} P_0 & P_1 & \cdots & P_n \end{bmatrix} \begin{bmatrix} y(1) & y(2) & \cdots & y(T-n) \\ y(2) & y(3) & \cdots & y(T-n+1) \\ \vdots & \vdots & \cdots & \vdots \\ y(n+1) & y(n+2) & \cdots & y(T) \end{bmatrix} = 0. \tag{$**$}$$

For the given example, we have

$$\begin{bmatrix} 2 & -3 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 3 & 5 & 9 & 17 \\ 3 & 5 & 9 & 17 & 33 \\ 5 & 9 & 17 & 33 & 64 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & -1 \end{bmatrix} \neq 0,$$

so that $y_{\mathrm{d}} \notin \ker(P(z))$. An ad hock way of correcting y_{d} in the specific example is to set $y_{\mathrm{d}}(7) = 65$. □

$\mathscr{B}(A,C) \stackrel{?}{=} \ker(P(z))$

Check if the system defined by (21) is the same system as the one defined by (KER).

Solution: We need to verify that the eigenvalues $z_1 = 1$ and $z_2 = 2$ of A are the same as the roots of $P(z)$. Indeed,

$$(z-1)(z-2) = 2 - 3z + z^2,$$

so that the two representations define the same system. □

$\mathscr{B}(A_1,C_1) + \mathscr{B}(A_2,C_2)$

Let $\mathscr{B}$ be the system obtained by adding the outputs of two autonomous LTI systems $\mathscr{B}_1$ and $\mathscr{B}_2$ of orders n_1 and n_2. Is $\mathscr{B}$ linear time-invariant? What is its order?

Solution: $\mathscr{B}$ is linear because the sum of two subspaces is a subspace. $\mathscr{B}$ is also time-invariant because, for any $y \in \mathscr{B}$, it follows that $y = y_1 + y_2$, where $y_1 \in \mathscr{B}_1$ and $y_2 \in \mathscr{B}_2$, but $\sigma^\tau y_1 \in \mathscr{B}_1$ and $\sigma^\tau y_2 \in \mathscr{B}_2$ for all τ, so that

$$\sigma^\tau y = \sigma^\tau y_1 + \sigma^\tau y_2 \in \mathscr{B}.$$

The order of $\mathscr{B}$ is

$$n = n_1 + n_2 - \text{"\# of common poles of } \mathscr{B}_1 \text{ and } \mathscr{B}_2\text{"}.$$

□

Selecting bounded trajectories of an unstable linear time-invariant system

1. Let $\mathscr{B} \in \mathscr{L}^1_{0,\ell}$, explain how to find the "stable subbehavior" of $\mathscr{B}$, *i.e.*,

$$\mathscr{B}_{\text{stable}} = \{\, w \in \mathscr{B} \mid w \text{ is bounded} \,\}.$$

2. Explain how to find the "stable subbehavior" of an open linear time-invariant system $\mathscr{B} \in \mathscr{L}^2_{1,\ell}$.

Polynomial multiplication, convolution, and differential/difference operators

There is a natural correspondence among \$m+1\$-dimensional vectors

$$a = \begin{bmatrix} a_0 & a_1 & \cdots & a_m \end{bmatrix}^\top,$$

m-th order polynomials

$$a(z) = a_0 + a_1 z + \cdots + a_m z^m,$$

and \$(m+1)\$-taps sequences

$$a = (a_0, a_1, \ldots, a_m).$$

With some abuse of notation we denote with the same symbol a the vector, the polynomial, and the sequence. For given $a \in \mathbb{R}^{m+1}$ and a natural number n, define the $(m+n+1) \times (n+1)$ matrix

$$\mathscr{M}_n(a) := \begin{bmatrix} a_0 & & \\ a_1 & \ddots & \\ \vdots & \ddots & a_0 \\ a_m & & a_1 \\ & \ddots & \vdots \\ & & a_m \end{bmatrix}.$$

1. Show that the product of two polynomials $c(z) = a(z)b(z)$ corresponds to the matrix-vector product $c = \mathscr{M}_n(a)b$, where n is the degree of b.

2. Show that the convolution of two sequences—a with $m+1$ taps and b with $n+1$ taps, *i.e.*,

$$c = a \star b, \qquad c(i) := \sum_{k=0}^{m} a_k b_{i-k}, \quad \text{for } i = 0, \ldots, m+n,$$

   where $b_k = 0$ for $k < 0$ and $k > n$, corresponds to the matrix-vector product $\mathscr{M}_n(a)b$.

3. Show that the action of a difference operator defined by a polynomial a of degree n on a sequence b with $m+1$ taps, *i.e.*,

$$c = a(\sigma)b, \qquad c(i) = a_0 b_i + a_1 b_{i+1} + \cdots + a_m b_{i+m}, \quad \text{for } i = 0, \ldots, n-m,$$

   corresponds to the matrix-vector product $\mathscr{M}_{m-n}^{\top}(a)b$.

Kernel of $\mathscr{M}^{\top}(a)$

For a polynomial a of degree n, let $\lambda(a) = (\lambda_1, \ldots, \lambda_n)$ be the set of its roots. Define the (extended) Vandermonde matrix

$$V_i(\lambda) := \begin{bmatrix} 1 & \cdots & 1 \\ \lambda_1 & \cdots & \lambda_n \\ \vdots & & \vdots \\ \lambda_1^{i-1} & \cdots & \lambda_n^{i-1} \end{bmatrix}.$$

1. Assuming that $\lambda_i \neq \lambda_j$, for all $i \neq j$, show that

$$\ker\big(\mathscr{M}_i^{\top}(a)\big) = \text{image}\big(V_i\big(\lambda(a)\big)\big).$$

2. Show that if $\lambda_1 = \lambda_2$, then $(1, \lambda_1, \lambda_1^2, \ldots)$ and $(0, 1\lambda_1, 2\lambda_1^2, \ldots)$ are in the kernel of $\mathscr{M}_i^{\top}(a)$.

3. Guess the general form of the kernel of $\mathscr{M}_i^{\top}(a)$ in case of roots of arbitrary multiplicity.

Two autonomous behaviors of a single-input single-output system

Consider the single-input single-output linear time-invariant system

$$\mathscr{B} = \{ w = \left[\begin{smallmatrix} u \\ y \end{smallmatrix}\right] \mid p(\sigma)y = q(\sigma)u \}.$$

Two subsets of $\mathscr{B}$ are the following scalar autonomous linear time-invariant behaviors

$$\mathscr{B}_u = \{ w = \left[\begin{smallmatrix} u \\ 0 \end{smallmatrix}\right] \mid 0 = q(\sigma)u \} \qquad \text{and} \qquad \mathscr{B}_y = \{ w = \left[\begin{smallmatrix} 0 \\ y \end{smallmatrix}\right] \mid p(\sigma)y = 0 \}.$$

A particular output y is determined by the input u and the initial conditions

$$x_{\text{ini}} := \big(y(-n+1), y(-n+2), \ldots, y(0), u(-m+1), u(-m+2), \ldots, u(0)\big).$$

1. Explain how to find the initial conditions x_{ini} for a given $w \in \mathscr{B}$.
2. Construct an example (*i.e.*, specify a single-output linear time-invariant system), where a nonzero input u and an initial conditions x_{ini}, give a zero output $y = 0$.
3. Construct an example (*i.e.*, specify a single-output linear time-invariant system), where the input $u(t) = \sin(\omega_u t)$ and an initial conditions x_{ini}, give an output $y(t) = c\sin(\omega_y t)$, with $\omega_u \neq \omega_y$.

6.5 Least-squares estimation

Existence and uniqueness of the least squares approximate solution

Consider the system of linear equations $Ax = b$, where $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$ are given, and $x \in \mathbb{R}^n$ is unknown. A least squares approximate solution of the system is a solution to the optimization problem

$$\text{minimize} \quad \text{over } \widehat{x} \in \mathbb{R}^n \quad \|A\widehat{x} - b\|_2. \tag{LS}$$

When does a solution exists and when it is unique. Under the assumptions of existence and uniqueness, derive a closed form expression for the least squares approximate solution. Characterize all least squares approximate solutions in the general case. Solve the least-squares approximation problem

$$\begin{bmatrix} 1 & 2 \\ 2 & 4 \\ 4 & 8 \end{bmatrix} x = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

Orthogonality principle for least-squares estimation

Show that

1. $\widehat{x}$ being a least squares approximate solution of the system $Ax = b$, and
2. $\widehat{x}$ being such that $b - A\widehat{x}$ is orthogonal to the span of the columns of A,

are equivalent. (This result is known as the orthogonality principle for least squares approximation.)

Weighted least-squares approximate solution

For a given positive definite matrix $W \in \mathbb{R}^{m\times m}$, define the weighted 2-norm

$$\|e\|_W = e^\top We.$$

The weighted least-squares approximation problem is

$$\text{minimize} \quad \text{over } \widehat{x} \in \mathbb{R}^n \quad \|A\widehat{x} - b\|_W. \tag{WLS}$$

When does a solution exist and when is it unique? Under the assumptions of existence and uniqueness, derive a closed form expression for the least squares approximate solution.

Orthogonality principle for weighted least-squares estimation

Formulate and prove a similar property to the orthogonality of $A\widehat{x}-b$ and $\text{span}(A)$ for the weighted least-squares problem, *i.e.*, generalize the orthogonality principle to (WLS).

6.6 Stochastic signals

Finite Markov chain model ([3, Chapter 7])

The weather in a certain city can be characterized as being either sunny, cloudy, or rainy. If it is sunny one day. then sun or clouds are equally likely the next day. If it is cloudy, then there is a fifty percent chance the next day will be sunny, a twenty-five percent chance of continued clouds, and a twenty-five percent chance of rain. If it is raining, it will not be sunny the next day, but continued rain or clouds are equally likely.

Denoting the three types of weather by S, C, and R, respectively, this model can be represented by an array of transition probabilities:

	S	C	R
S	1/2	1/2	0
C	1/2	1/4	1/4
R	0	1/2	1/2

This array is read by going down the left column to the current weather condition. The corresponding row of numbers gives the probabilities associated with the next weather condition. The process starts with some weather condition and moves, each day, to a new condition. There is no way, however, to predict exactly which transition will occur. Only probabilistic statements, presumably based on past experience, can be made.

We model the weather dynamics by a stochastic system, called a finite Markov chain. Let $x_1(t)$ be the probability that day-t is sunny, $x_1(t)$ be the probability that day-t is cloudy, and $x_3(t)$ be the probability that day-t is cloudy. Obviously $x(t)$ is nonnegative and its elements sum to one. (Such a vector is called a stochastic vector.)

1. Express the weather dynamics, described by the transition probabilities

$$P = \begin{bmatrix} .5 & 0.5 & 0 \\ 0.5 & 0.25 & 0.25 \\ 0 & 0.5 & 0.5 \end{bmatrix},$$

as a deterministic linear time-invariant system for the vector of probabilities x.

2. What conditions needs to satisfy the matrix P so that $x(t)$ is a stochastic vector for all $t > 0$, provided $x(0)$ is a stochastic vector?

3. A limiting distribution of the Markov chain model is a distribution x_∞ such that $x(t) \to x_\infty$, as $t \to \infty$ (independent of the initial conditions). Is there a limiting distribution of the weather model? If so, find it.

Solution **SOL**

$x_\infty := \begin{bmatrix} 0.4 & 0.4 & 0.2 \end{bmatrix}$

Periodogram computation via FFT

The periodogram method for spectral estimation is

$$\widehat{\phi}(\omega) = \frac{1}{T} |\sum_{\tau=1}^{T} y(\tau) e^{-\mathbf{i}\omega\tau}|^2.$$

1. Explain how to compute $\widehat{\phi}(\omega_k)$, where

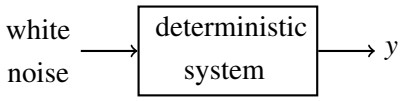
$$\omega_k = \frac{2\pi}{T} k, \qquad \text{for } k = 0, 1, \ldots, T-1,$$

   using matrix-vector product. How many multiplication operations are required by this method?

2. Explain how to use the fast Fourier transform for fast computation of $\widehat{\phi}$ for $\omega = \omega_k$.

Covariance structure of an ARMA model

An auto-regressive moving-average (ARMA) system is a deterministic linear time-invariant system, "driven" by white noise input e.



In the scalar case, it can be represented by a difference equation

$$y(t) + \sum_{\tau=1}^{n} a_\tau y(t-\tau) = \sum_{\tau=0}^{m} b_\tau e(t-\tau),$$

with $b_0 = 1$. Prove that the auto-covariance r of y satisfies the equation

$$r(k) + \sum_{\tau=1}^{n} a_\tau r(k-i) = 0.$$

Solution **SOL**

Multiplying of the difference equation with $y(t-k)$ and taking expectation yields

$$r(k) + \sum_{\tau=1}^{n} a_\tau r_{k-i} = \sum_{\tau=0}^{m} b_\tau \mathbf{E}(e(t-\tau)y(t-k))$$

Since, $y(t) = \sum_{\tau=0}^{\infty} h(\tau)e(t-\tau)$ we have

$$\mathbf{E}(e(t-\tau)y^*(t-k)) = \sigma^2 h(\tau-k)$$

For $k < m$,

$$r(k) + \sum_{\tau=1}^{n} a_\tau r(k-i) = 0.$$

Yule-Walker equation

Yule-Walker equation is a method for estimation of the coefficients of an AR process

$$y(t) + \sum_{\tau=1}^{n} a_\tau y(t-\tau) = e(t).$$

Using the results from the previous problem, show that

$$\begin{bmatrix} r(0) & r(-1) & \cdots & r(-n) \\ r(1) & r(0) & \ddots & \vdots \\ \vdots & & \ddots & r(-1) \\ r(n) & \cdots & & r(0) \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} \sigma^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

(This equation is known as the Yule-Walker equation and forms the basis of many parametric estimation methods.) Explain how to estimation the coefficients a if the autocorrelation r was known.

Central limit theorem (Matlab)

Observe the smoothing effect of the convolution operation by convolving a rectangular pulse, centered at 0 with itself N times, where $N = 1, 2, \ldots, N$. Observe also that by proper normalization the resulting signal converges to a Gaussian distribution. Can you explanation these empirical observations?

7 Mini-projects

7.1 Behavior of a second order autonomous system

Consider a two-dimensional autonomous dynamical system described by the differential equations

$$\begin{matrix} \frac{dx(t)}{dt} \\ \frac{dy(t)}{dt} \end{matrix} = \begin{bmatrix} a & 1 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} x(t) \\ y(t) \end{bmatrix},$$

where the parameter a is a real number.

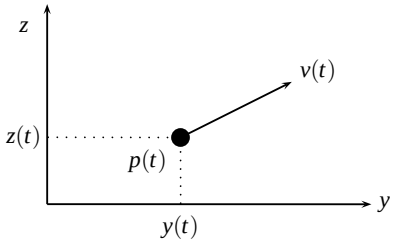
1. Write down the general solution of the differential equations for $x(t)$ and $y(t)$. Discuss.
2. The eigenvalue analysis of the matrix $\begin{bmatrix} a & 1 \\ 0 & -1 \end{bmatrix}$ is a natural tool for solving such an autonomous linear dynamical system, and analyzing its stability. What are the eigenvalues and eigenvectors?
3. What are the equilibrium points of the system? Try to understand what happens when a goes from $-\infty$ to $+\infty$. Is the stability of the point $(0,0)$ affected?

4. There are five interesting cases for different values of a, namely $a < -1$, $a = -1$, $-1 < a < 0$, $a = 0$ and $a > 0$. Discuss (qualitatively) the stability of the point $(0,0)$ in these five cases. Can you give a (geometrical) interpretation to the eigenvalues and eigenvectors in these cases? Try to sketch the trajectories of $(x(t), y(t))$ in the (x,y) plane for the five cases (for different initial conditions $(x(0), y(0))$).

7.2 Dynamics of free fall

Trajectory of a freely falling object

Consider a free falling object in a 2-dimensional gravitational field. Let $p(t)$ be the position of the object at time t. We choose a reference moment of time $t = 0$ to be the moment when the object starts its free fall and an orthogonal coordinate system in the plane of motion with vertical axis along the negative of the gravitational force and a perpendicular horizontal axis at the ground level. The horizontal displacement of the object at time t



with respect to the origin is denoted by $y(t)$ and the vertical displacement by $z(t)$. More notation used is:

$p(t) = \begin{bmatrix} y(t) \\ z(t) \end{bmatrix}$	—	object's position and its coordinates at time t
$v(t)$	—	object's velocity at time t
p_{ini}, v_{ini}	—	initial (time $t = 0$) position and velocity
$x(t)$	—	state (position and velocity) at time t
m	—	object's mass
g	—	gravitational constant
$m\mathbf{g}$	—	gravitational force ($\mathbf{g}$ is a vector in the "negative vertical" direction with norm equal to g)

By the second law of Newton, for any $t > 0$, the position of the object is described by the differential equation

$$m\ddot{p} = m\mathbf{g}, \qquad \text{where} \quad p(0) = p_{\text{ini}} \text{ and } \dot{p}(0) = v_{\text{ini}}. \tag{11}$$

Find an analytic expression for the trajectories p.

Free fall with friction

In this exercise, we relax the assumption that the object is in vacuum. The net effect of the air on the object is a force f, acting on the object. The model equation (11) becomes

$$m\ddot{p} = m\mathbf{g} + f, \qquad \text{where} \quad p(0) = p_{\text{ini}} \text{ and } \dot{p}(0) = v_{\text{ini}}. \tag{12}$$

Without wind and turbulence, the force f is due to friction with the air and can be approximated by a linear function of the velocity, *i.e.*, we take

$$f = -\gamma v, \tag{13}$$

where γ is a constant depending on the physical properties of the environment as well as the size and shape of the object. Repeat the previous exercise for the case when a friction force (24) is present. Experiment with different values for the mass m and the friction constant γ.

Throwing an object as far as possible

How far can you throw an object? In order to formulate the question mathematically, assume that you can give a maximal initial velocity to the object (by accelerating it for a period of time, after which period the object is freely falling). The question then is what direction should the initial velocity have in order for the object to reach as far as possible when it lands on the ground. The problem is considerably simpler when $z_{\text{ini}} = 0$, *i.e.*, the object is thrown from the ground and there is no friction. Assume that this is the case. Once you come up with an answer, use your free falling simulation function to compute and plot the trajectory. Try some alternative admissible trajectories to make sure that your solution gives the best result.

7.3 Prediction using a model

1. *State space approach* Given a state space representation of a discrete-time autonomous system~$\mathscr{B}(A,C)$ of order `n` and a finite, $T \geq$ `n` samples long, trajectory

   $$y_{\text{p}} := \big(y(1),\ldots,y(T)\big)$$

   of that system, find the next T_{f} samples

   $$y_{\text{f}} := \big(y(T+1),\ldots,y(T+T_{\text{f}})\big)$$

   of the given trajectory, *i.e.*, find $y(T+1),\ldots,y(T+T_{\text{f}})$ such that

   $$y := \big(y(1),\ldots,y(T),y(T+1),\ldots,y(T+T_{\text{f}})\big)$$

   is a trajectory of $\mathscr{B}(A,C)$.

2. *Polynomial approach* Solve the problem of 1 using a polynomial representation of the system $\mathscr{B}(P) = \mathscr{B}(A,C)$. Assume that the highest power coefficient of P is I.

3. *Simulation example* Implement your solutions of 1 and 2 in Matlab and apply them on the data in the file

   `http://homepages.vub.ac.be/~imarkovs/elec3035/tutorial2.mat}`

   (use the command `load` to read the data) The file contains the following variables: `yp`, `a`, `c`, `p`, `Tf`, and `yp_noisy`, which correspond to col $\big(y_{\text{p}}(1),\ldots,y_{\text{p}}(T)\big)$, $y_{\text{p}}(t) \in \mathbb{R}$, A, C, col$(p_n,\ldots,p_0)$, T_f, and a noise corrupted version of `yp` needed in 4.

4. *Simulation example with noisy sequence y_p* Apply your solutions of 1 and 2 on the noise corrupted version of the data sequence y_{p}, stored in the variable `yp_noisy`. Compare the obtained predictions with the exact trajectory y_{f}, obtained in 3.

Solution **SOL**

Solution:

1. *State space approach* A trajectory y of an autonomous system $\mathscr{B}(A,C)$ is completely specified by an initial condition $x(1)$, so the problem of predicting the future part y_f of the trajectory y from its given past y_p is equivalent to the problem of determining the initial condition $x(1)$ of y from y_p.

   From the general expression of a response of an autonomous system

   $$y(t_1) = CA^{t_1-t_2}x(t_2)$$

   we have a system of equations for the unknown initial condition $x(1)$

   $$\underbrace{\begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(T) \end{bmatrix}}_{y_\mathrm{p}} = \underbrace{\begin{bmatrix} C \\ CA \\ \vdots \\ CA^{T-1} \end{bmatrix}}_{\mathscr{O}_\mathrm{p}} x(1). \tag{14}$$

   In order to be able to determine $x(1)$ uniquely from y_p, the matrix $\mathscr{O}_\mathrm{p}$, called an extended observability matrix of the system $\mathscr{B}(A,C)$, should have full column rank. Note that $\mathscr{O}_\mathrm{p}$ is $T\mathtt{p} \times \mathtt{n}$, where $\mathtt{p} = \operatorname{row}\dim(C)$ is the number of outputs. Under the assumption $T \geq \mathtt{n}$ (*i.e.*, "enough data is given") the matrix $\mathscr{O}_\mathrm{p}$ has the right dimension for being full column rank.

   The condition $T\mathtt{p} \geq \mathtt{n}$ is necessary but not sufficient for $\mathscr{O}_\mathrm{p}$ to be full column rank. The extended observability matrix $\mathscr{O}_\mathrm{p}$ depends on the system parameters A and C, so an extra condition is needed on the matrices A and C, *i.e.*, on the state space representation of the system. This condition is so important that it is given the name *observability*.

   Provided that we have enough data $T \geq \mathtt{n}/\mathtt{p}$ and the representation $\mathscr{B}(A,C)$ is observable, we can uniquely determine $x(1)$ from y_f via

   $$x(1) = (\mathscr{O}_\mathrm{p}^\top \mathscr{O}_\mathrm{p})^{-1}\mathscr{O}_\mathrm{p} y_\mathrm{p}.$$

   Then we predict y_f using $y(t) = CA^{t-1}x(1)$:

   $$\underbrace{\begin{bmatrix} y(T+1) \\ \vdots \\ y(T_\mathrm{f}) \end{bmatrix}}_{y_\mathrm{f}} = \underbrace{\begin{bmatrix} CA^{T} \\ \vdots \\ CA^{T+T_\mathrm{f}-1} \end{bmatrix}}_{\mathscr{O}_\mathrm{f}} x(1).$$

   The final answer is

   $$y_\mathrm{f} = \mathscr{O}_\mathrm{f}(\mathscr{O}_\mathrm{p}^\top \mathscr{O}_\mathrm{p})^{-1}\mathscr{O}_\mathrm{p} y_\mathrm{p} = \mathscr{O}_\mathrm{f}\left(\sum_{\tau=0}^{T-1} CA^\tau (A^\tau)^\top C^\top\right)^{-1} \mathscr{O}_\mathrm{p} y_\mathrm{p}.$$

2. *Polynomial approach*

In this case we consider the polynomial representation

$$P_0 y(t) + P_1 y(t+1) + \cdots + P_{\ell-1} y(t+\ell-1) + y(t+\ell) = 0, \qquad \text{for all } t \in \mathbb{Z}.$$

Because of the assumption that the highest power coefficient P_ℓ is I, we can find for each t, $y(t)$ as a linear combination of $y(t-1), \ldots, y(t-\ell)$

$$y(t) = -\big(P_0 y(t-\ell) + P_1 y(t-\ell+1) + \cdots + P_{\ell-1} y(t-1)\big). \tag{15}$$

Assuming that $T \geq \ell$, we can apply this formula recursively and "extend" y_{p} to y_{f}, *i.e.*, we simulate the response y_{f} corresponding to the initial conditions $y(T), y(T-1) \ldots, y(T-\ell+1)$, which is the end part of y_{f}. For this to be possible, we need $T \geq \ell$. It can be shown that this condition follows from the assumption $T \geq \mathtt{n}$.

3. *Simulation example* Here is Matlab code for the prediction methods described in 1 and 2.

`sspredict.m`

`polpredict.m`

The obtained simulation result with the given exact data is shown in Figure 6, left. Up to numerical errors, the results obtained by the two methods are the same.

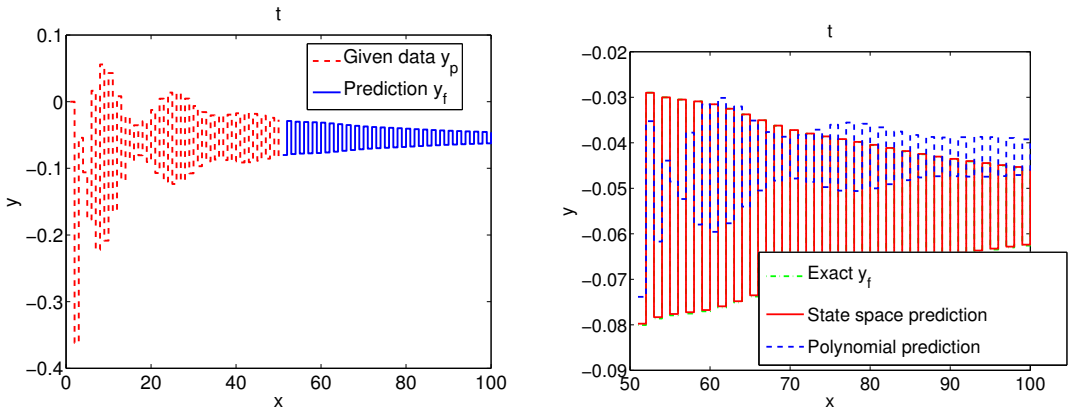


Figure 6: Simulation result for problems 3 and 4.

4. *Simulation example with noisy sequence y_p*

The predictions from noisy data are shown in Figure 6, right. The state space method gives much better result than the polynomial method. The explanation for this is that in 3 we use the whole sequence y_{p} in order to predict y_{f}, while in 4 we use only the last 16 samples (and ignore the other 34 samples). Note that in the noisy case, the system (14) is incompatible and we compute the least squares approximate solution. The least squares approximation is the reason for the good performance of the method of 3. The polynomial approach can also be modified so that it uses all given data and does least squares approximation, so the difference in the performance of the two methods is not in the approach (polynomial vs state space) but in the particular way we solved the problem in the two cases.

□

7.4 Autonomous system identification

1. Given a scalar sequence
$$y_{\mathrm{d}} := \big(y_{\mathrm{d}}(1),\dots,y_{\mathrm{d}}(T)\big)$$
and a natural number n, the aim of the considered identification problem is to find an autonomous system, such that the given sequence is a trajectory of that system. Under what conditions a solution exists. If a solution exists, when it is unique.

2. *Simulation example* Implement your solution in Matlab and apply it on the sequence
$$y_{\mathrm{d}} = (1,2,4,7,13,24,44,81).$$
What is the data generation rule?

3. *Prediction without model* Consider a prediction problem without a given model. Under what conditions it is possible to predict y_{f} from y_{p} (see Problem 7.3) without having a model? Implement your method in Matlab\ and test it on the exact and noisy sequences.

Solution **SOL**

Solution:

1. We aim to determine parameters $P_0, P_1, \dots, P_\ell$, so that
$$P_0 y_{\mathrm{d}}(t) + P_1 y_{\mathrm{d}}(t+1) + \cdots + P_\ell y_{\mathrm{d}}(t+\ell) = 0, \qquad \text{for } t = 1,\dots,T-\ell \tag{16}$$
holds. Using matrix vector notation, the system of equations (16) can be written in a matrix form as
$$\begin{bmatrix} P_0 & P_1 & \cdots & P_\ell \end{bmatrix} \begin{bmatrix} y_{\mathrm{d}}(1) & y_{\mathrm{d}}(2) & y_{\mathrm{d}}(3) & \cdots & y_{\mathrm{d}}(T-\ell) \\ y_{\mathrm{d}}(2) & y_{\mathrm{d}}(3) & y_{\mathrm{d}}(4) & \cdots & y_{\mathrm{d}}(T-\ell+1) \\ \vdots & \vdots & \vdots & & \vdots \\ y_{\mathrm{d}}(\ell+1) & y_{\mathrm{d}}(\ell+2) & y_{\mathrm{d}}(\ell+3) & \cdots & y_{\mathrm{d}}(T) \end{bmatrix} = 0. \tag{17}$$
Assuming that the highest power coefficient of P is I, we can rewrite (17) as
$$\underbrace{\begin{bmatrix} P_1 & P_2 & \cdots & P_\ell \end{bmatrix}}_{P} \underbrace{\begin{bmatrix} y_{\mathrm{d}}(2) & y_{\mathrm{d}}(3) & \cdots & y_{\mathrm{d}}(T-\ell+1) \\ y_{\mathrm{d}}(3) & y_{\mathrm{d}}(4) & \cdots & y_{\mathrm{d}}(T-\ell+2) \\ \vdots & \vdots & & \vdots \\ y_{\mathrm{d}}(\ell+1) & y_{\mathrm{d}}(\ell+2) & \cdots & y_{\mathrm{d}}(T) \end{bmatrix}}_{H} = -\underbrace{\begin{bmatrix} y_{\mathrm{d}}(1) & y_{\mathrm{d}}(2) & \cdots & y_{\mathrm{d}}(T-\ell) \end{bmatrix}}_{Y_{\mathrm{d}}}. \tag{18}$$
If (18) is solvable, then the autonomous system defined by
$$y(t) + P_1 y(t+1) + \cdots + P_{\ell-1} y(t+\ell-1) + y(t+\ell) = 0, \qquad \text{for all } t \in \mathbb{Z}$$
has as a trajectory the given sequence y_{d}. For uniqueness we need H to be full row rank. The identification procedure is to compute the solution of the system $PH = Y_{\mathrm{d}}$, *i.e.*,
$$P = Y_{\mathrm{d}} H^\top (HH^\top)^{-1}.$$

2. *Simulation example* A Matlab function implementing the method of 1 is:

   ident.m

   Applied on the given data it gives

$$p = \begin{bmatrix} 1 & -1 & -1 & -1 \end{bmatrix}.$$

   Since in this case the system (18) is solvable and A is full column rank, p is an exact solution and is unique. Therefore the data generating rule is

$$y(t+3) = y(t+2) + y(t+1) + y(t), \qquad y(-2) = 1,\ y(-1) = 0,\ y(0) = 0.$$

   The sequence y_{d} is obtained from $\mathscr{B}(p)$ under initial condition $(1,0,0)$.

3. *Prediction without model* Yes, provided

   - the given data y_{p} is exact (this guarantees existence of solution of (18)), and
   - the matrix A is full row rank (this is an implicit condition on y_{p}).

   Under these assumptions we can identify the data generating system from the data and apply the solutions of Problem 7.3. In the exact of exact data the method "works well", however, the noisy case is more challenging. (I will propose it as a 3rd year project.)

□

8 Sample exam questions

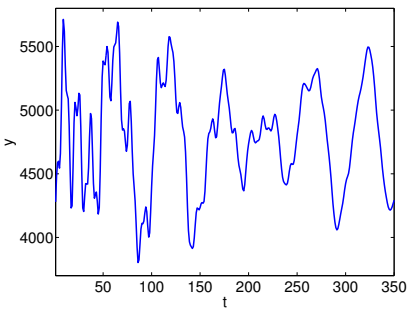
- *System's order* (2 points)
  The signal $y(t) = e^t + e^{2t} + e^{3t}$ is a response of an autonomous LTI system. What can you say about the order $\mathtt{n}$ of the system?

  (a) $\mathtt{n} = 1$ (b) $\mathtt{n} = 2$ (c) $\mathtt{n} = 3$ (d) $\mathtt{n} \geq 3$

  Solution: (d) □

- *Is the data generating system LTI?* (4 points)
  A colleague of yours shows you the signal on the right and says:

  > "I think the data generating system is not linear time-invariant, because the response of such a system is a sum of terms that are exponentially decaying, exponentially growing, or periodic while the behavior of the given signal is more complicated: it is obviously not periodic and it is neither exponentially decaying nor exponentially growing."



  Do you agree? If so, how would you make the argument rigorous? If not, what is wrong with the argument and how would you prove that it is wrong? You can assume that you have the observed signal numerically and a computer available to process the data.

Solution: The argument is wrong because

1. the LTI system may have an (unobserved) output, in which case the output may not be a sum-of-exponentials.
2. Even for an autonomous LTI system, however, the order of the system is not bounded in the argument so that any finite signal can be made an exact output of an autonomous LTI system by increasing the order.

The minimum order can be found computationally from the signal by a realization algorithm. □

- *From state space to transfer function*
  Find the transfer function H of a discrete-time system defined by a state-space representation with parameters
  $$A = \begin{bmatrix} 1/2 & 0 \\ 0 & 1/4 \end{bmatrix}, \quad B = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 \end{bmatrix}, \quad D = 1. \tag{19}$$
- *From state space to impulse response*
  Find the impulse response h of the state space model with parameters (19).
- *Order of linear time-invariant system*
  - What is the order of the state space model with parameters (19)?
  - What is the order of the transfer function model computed in problem 1?
- *Detecting static relation*
  - Given a trajectory $w = \big(w(1), \ldots, w(T)\big)$ of unknown discrete-time linear time-invariant system, how can you check if the system is static?
  - Describe a computational method that does the job.
  - Write Matlab code that implements the method.
- *Addition of autonomous systems*
  Consider the system $\mathscr{B}$ obtained by adding the outputs of two autonomous LTI systems of orders n_1 and n_2. Is $\mathscr{B}$ linear time-invariant? What is its order?
- The identification methods considered in the course use as data a trajectory $y = \big(y(1), \ldots, y(T)\big)$ and produce as an output the model parameters.
  1. Describe the methods.
  2. Extend the methods to use *two* given trajectories
     $$y^{(1)} = \big(y^{(1)}(1), \ldots, y^{(1)}(T_1)\big) \quad \text{and} \quad y^{(2)} = \big(y^{(2)}(1), \ldots, y^{(2)}(T_2)\big)$$
     of the system instead of one trajectory. Discuss both the exact data case and the noisy data case.
  3. Apply your method on the data $y^{(1)} = (0,1,1)$ and $y^{(2)} = (1,0,2)$, assuming that the data is exact and the data generating system is of second order.

- *True or false? If, false, give a counterexample.*
  In all subquestions A is an $m \times n$ matrix, x is an $n \times 1$ vector, and b is an $m \times 1$ vector.

  1. For any given A and b, $Ax = b$ always has a solution x.

     Solution: False. Counter example: for $A = \left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]$, $b = \left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]$, $Ax = b$ has no solution. □

  2. For any given A, $Ax = 0$ always has a unique solution x.

     Solution: False. Counter example: for $A = \begin{bmatrix}1 & 0\end{bmatrix}$, $Ax = 0$ has nonunique solution $x = \left[\begin{smallmatrix}0\\\alpha\end{smallmatrix}\right]$, $\alpha \in \mathbb{R}$. □

  3. For any given A and b with $m < n$, $Ax = b$ always has a solution.

     Solution: False. Counter example: for $A = \begin{bmatrix}0 & 0\end{bmatrix}$ and $b = 1$, $Ax = b$ has no solution. □

  4. For any given A and b with $m > n$, there is a unique solution x of the problem $\min_x \ \|Ax - b\|_2$.

     Solution: False. Counter example: for $A = \left[\begin{smallmatrix}1 & 0\\0 & 0\end{smallmatrix}\right]$ and $b = \left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]$, $\min_x \ \|Ax - b\|_2$ has a nonunique solution $x = \left[\begin{smallmatrix}1\\\alpha\end{smallmatrix}\right]$, $\alpha \in \mathbb{R}$. □

- *True or false? If, false, give a counterexample.*
  In all subquestions $\mathscr{B}$ is a dynamical system, defined by $\mathscr{B} = \{\, y \mid \frac{\mathrm{d}}{\mathrm{d}t}x = f(x),\ y = g(x),\ x(0) \in \mathbb{R}^{\mathrm{n}} \,\}$.

  1. If f and g are linear functions, any $y \in \mathscr{B}$ is a linear function.

     Solution: False. Counter example: $\mathscr{B} = \{\, y \mid \frac{\mathrm{d}}{\mathrm{d}t}x = x,\ y = x,\ x(0) \in \mathbb{R} \,\}$ has solutions $y(t) = x(0)e^t$. □

  2. If f and g are linear functions, $\mathscr{B}$ is a linear system.

     Solution: True. □

  3. If f is not a linear function, $\mathscr{B}$ is not a linear system. (Hint: consider a system with unobservable state.)

     Solution: False. Counter example: $\mathscr{B} = \{\, y \mid \frac{\mathrm{d}}{\mathrm{d}t}x = \left[\begin{smallmatrix}x_1^2\\x_2\end{smallmatrix}\right],\ y = x_2,\ x(0) \in \mathbb{R}^2 \,\}$ has solutions $y(t) = x_2(0)e^t$. □

  4. If g is not a linear function, $\mathscr{B}$ is not a linear system.

     Solution: True. □

  5. If $g = Cx + b$, where $b \neq 0$, $\mathscr{B}$ is not a linear system.

     Solution: True. □

- *Differential equation with a constant term.*
  Consider the system $\mathscr{B} = \{\, y \mid \frac{\mathrm{d}}{\mathrm{d}t}y = y + 1,\ y(0) \in \mathbb{R} \,\}$.

  1. Find the trajectories y of $\mathscr{B}$.

     Solution: $y(t) = \big(y(0) + 1\big)e^t - 1$, for $y(0) \in \mathbb{R}$. □

2. Find $a, b, c \in \mathbb{R}$, such that $\mathscr{B} = \{ y \mid \frac{\mathrm{d}}{\mathrm{d}t}x = ax,\ y = cx + b,\ x \in \mathbb{R} \}$.

   Solution: $a = 1$, $c = 1$, $b = -1$. □

- *Exact identification.*
  Find a recursive relation $p(\sigma)y = 0$ of minimal order for the sequence $y = (0, 1, 2, 7, 20)$.

  Solution: $\begin{bmatrix} 1 & 2/3 & -1/3 \end{bmatrix} \begin{bmatrix} 0 & 1 & 2 \\ 1 & 2 & 7 \\ 2 & 7 & 20 \end{bmatrix} = 0$, so that $p(z) = 1 + 2/3z - 1/3z^2$. □

- *Free fall in a gravitational field.*
  Your friend likes playing angry birds and is curious about the trajectory of a "bird" after being thrown.

  – Your friend: "The bird is a free falling object in a gravitational field. If y_1 is its horizontal position and y_2 is its vertical position, $y = \left[\begin{smallmatrix} y_1 \\ y_2 \end{smallmatrix}\right]$ can be viewed as a response of a dynamical system."

  – You: "I see what you mean. The system dynamics is ... This is ... system. An analytical formula for the trajectory y is ..." *(fill in the ellipses)*

  Solution: The position y of the object is described by the second law of Newton

  $$m\ddot{y} = m\mathbf{g}, \qquad y(0) = y_{\text{ini}}, \quad \dot{y}(0) = \dot{y}_{\text{ini}}, \tag{20}$$

  where

$\dot{y}(t)$	is the object's velocity at time t
y_{ini}, $\dot{y}_{\text{ini}}$	is the initial position and velocity
m	is the object's mass
g	is the gravitational constant
$m\mathbf{g}$	is the gravitational force ($\mathbf{g}$ is a vector in the "negative vertical" direction with length g)

  The system $\mathscr{B} = \{ y \mid (20) \text{ holds} \}$ is continuous-time, second order, linear time-invariant, with constant input $m\mathbf{g}$. It can be represented in a state space form as

  $$\dot{x} = Ax, \qquad y = Cx, \qquad x(0) = x_{\text{ini}}, \tag{21}$$

  where

  $$x := \begin{bmatrix} y_1 \\ \dot{y}_1 \\ y_2 \\ \dot{y}_2 \\ x_5 \end{bmatrix}, \quad x_{\text{ini}} := \begin{bmatrix} y_{1,\text{ini}} \\ \dot{y}_{1,\text{ini}} \\ y_{2,\text{ini}} \\ \dot{y}_{2,\text{ini}} \\ -mg \end{bmatrix}, \quad A := \begin{bmatrix} 0 & 1 & & & \\ 0 & 0 & & & \\ & & 0 & 1 & \\ & & 0 & 0 & 1 \\ & & & & 0 \end{bmatrix}, \quad \text{and} \quad C := \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}. \tag{22}$$

  The constant input $m\mathbf{g}$ in (20) is modeled by the extra state variable x_5, which satisfies the equation $\dot{x}_5(t) = 0$, with initial condition $x_5(0) = -mg$. Then, the trajectory y of (21) is given by

  $$y(t) = Ce^{At}x_{\text{ini}}. \tag{23}$$

  □

– Your friend: "But this model is not realistic because it ignores the friction with the air. The friction force is proportional to the velocity. Can you incorporate friction into the model?"

– You: "Yes, the model then is ... This is ... system." *(fill in the ellipses)*

Solution: The model with friction is

$$\mathscr{B} = \{\, y \mid m\ddot{y} = m\mathbf{g} + f, y(0) = y_{\text{ini}},\ \dot{y}(0) = \dot{y}_{\text{ini}} \,\}$$

where

$$f = -\gamma\dot{y}, \tag{24}$$

is the friction force. In the state space representation (21), the A matrix changes to

$$A := \begin{bmatrix} 0 & 1 & & & & \\ 0 & -\gamma/m & & & & \\ & & 0 & 1 & & \\ & & 0 & -\gamma/m & 1 & \\ & & & & & 0 \end{bmatrix}. \tag{25}$$

The trajectory y is given by the same formula (23) with the new A. □

- *True or false? If, false, give a counterexample.*
  In all subquestions A is an $m \times n$ matrix, x is an $n \times 1$ vector, and b is an $m \times 1$ vector.

  1. For any given A and b, $Ax = b$ never has a solution x.

     Solution: False. Counter example: for $A = 1$, $b = 1$, $Ax = b$ has no solution. □

  2. For any given A, $Ax = 0$ never has a unique solution x.

     Solution: True. □

  3. For any given A and b with $m > n$, $Ax = b$ never has a solution.

     Solution: False. Counter example: for $A = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $b = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $Ax = b$ has no solution. □

  4. For any given A and b with $m = n$, $Ax = b$ never has infinitely many solutions.

     Solution: False. □

- *True or false? If, false, give a counterexample.*

  1. If the input of an LTI system is equal to zero, then the output is equal to zero.
  2. If a response (a particular output) of an LTI system is exponentially growing, then the system is unstable.
  3. If a response of an autonomous LTI system is exponentially growing, then the system is unstable.
  4. If a response of an autonomous LTI system is not exponentially growing, then the system is stable.
  5. The response of an autonomous LTI system can not be a linearly growing function $y = (1, 2, 3, 4, ...)$.

- *Exact identification.*
  Find a recursive relation $p(\sigma)y = 0$ of minimal order for the sequence $y = (0,1,2,3,4,5)$. What is special about the model?

  Solution: $\begin{bmatrix} 1 & 2/3 & -1/3 \end{bmatrix} \begin{bmatrix} 0 & 1 & 2 \\ 1 & 2 & 7 \\ 2 & 7 & 20 \end{bmatrix} = 0$, so that $p(z) = 1 + 2/3z - 1/3z^2$. □

- *Inverse problems.*
  Your friends are studying for a signals and systems exam. They are confused about the conditions for existence of an "inverse function" and an "inverse system". Explain as clearly as you can the problem to them.

  - Your friend: "By definition the *left inverse* a_L of a satisfies the equation $a_L a = 1$, the *right inverse* a_R of a satisfies the equation $aa_R = 1$, and the *inverse* a^{-1} satisfies the equation $a^{-1}a = aa^{-1} = 1$. Applying these definitions to a real number $a \in \mathbb{R}$, clearly, the left inverse is the same as the right inverse and the inverse. Moreover, the inverse exists if and only if a is nonzero. I have a problem, however, when I use these concepts for matrices. For a square matrix, is there a difference between left inverse, right inverse, and inverse, and what are the conditions for their existence? Is it sufficient that the matrix is nonzero?
  - You: ...
  - Your friend: "And what about non-square matrices? Do the three definitions differ and when they exist?"
  - You: ...
  - Your friend: "When I try to apply the definition to a system, it is even more confusing. When is a SISO LTI system invertible?"
  - You: ...
  - Your friend: "This is helpful, but do you have any idea why I need to study "inverse problems" (finding the inverse) does it have any relevance to signal processing or is it just a mathematical problem."
  - You: ...

9 References

[1] T. Kailath, A. H. Sayed, and B. Hassibi. *Linear Estimation*. Prentice Hall, 2000.

[2] S. Kay. *Modern spectral estimation: Theory and applications*. Prentice-Hall, 1988.

[3] D. G. Luenberger. *Introduction to Dynamical Systems: Theory, Models and Applications*. John Wiley, 1979.

[4] I. Markovsky. *Low Rank Approximation: Algorithms, Implementation, Applications*. Springer, 2012. (Citations: 126 in GS).

[5] I. Markovsky et al. *Exact and Approximate Modeling of Linear Systems: A Behavioral Approach*. SIAM, 2006. (Citations: 97 in GS).

[6] E. D. Sontag. *Mathematical control theory: Deterministic finite dimensional systems*. Springer-Verlag, 1998.

[7] P. Stoica and R. Moses. *Spectral Analysis of Signals*. Prentice Hall, 2005.

[8] J. C. Willems. “From time series to linear system—Part I. Finite dimensional linear time invariant systems”. In: *Automatica* 22.5 (1986), pp. 561–580.

[9] J. C. Willems. “From time series to linear system—Part I. Finite dimensional linear time invariant systems, Part II. Exact modelling, Part III. Approximate modelling”. In: *Automatica* 22, 23 (1986, 1987), pp. 561–580, 675–694, 87–115.

[10] J. C. Willems. “Paradigms and puzzles in the theory of dynamical systems”. In: *IEEE Trans. Automat. Contr.* 36.3 (1991), pp. 259–294.