Authors:

Heinz A Preisig

The ABC of Modelling

Lecture Notes

TKP4106 & 4135

Version:

2.0

Version Date:

2016-01-12

Printed:

2020-10-09

Address:

Chemical Engineering
Norwegian University of Science and Technology (NTNU)
7491 Trondheim, Norway

Heinz.Preisig@chemeng.ntnu.no

© 2015

The ABC of Modelling
Heinz A Preisig
Chemical Engineering
Norwegian University of Science and Technology (NTNU)
7491 Trondheim, Norway

Contents

Appendices

Part I

Abstraction of physical processes

Introduction

This course intends to introduce to modelling of physical-chemical processes. It aims at providing an overview and concatenate the material that has been part of the earlier curriculum: the introductory process engineering course, thermodynamics and transport as well as reaction kinetics. Why? because in order to model the process one needs them all: behaviour of capacities, flows, and changes of the handled material's nature. Also control is part of the description, as it changes the dynamic behaviour of the process to match the requested specifications. So dynamics are part of the game. This course breaks out of the stationary view that is traditional to chemical engineering mainly because it completes the picture and provides the necessary insight to achieve control over the process. With control being an enabling technology, this extension to the early curriculum material represents an essential knowledge component to a chemical engineer.

The course also makes use of the mathematics that is part of the standard math courses. In particular we use vectors and matrices to handle multi-dimensional objects; after all, we do not deal with only one species, but many and the description of a plant will in general include more than just a few capacity elements in the form of hold ups, material bodies, units and the likes. To give an example, a model of a distillation column for only a couple of species easily requires a couple of hundred differential equations and a multiple of algebraic equations. The dimensions are such that one has to resort to a higher abstraction level, in this case to capture multiple-dimensional objects in corresponding structures.

Another little headache is caused by the notation. It is complicated as we want to be precise and avoid misunderstandings. The used characters indicate the nature of the quantity, for example we use "n" for molar mass and vector n: $\hat{n}$ for a vector of molar masses. We use a high dot for the time derivative giving a measure for the change of the extensive quantity in a capacity. A hat is used to indicate a flow and a tilde for the transposition / reaction changing the nature of the extensive quantity. All of this will be introduced very carefully - step by step with the appropriate explanations being attached. In order to make the readers life easier, links to a comprehensive nomenclature section have been introduced, which become active in the electronic version of this book.

When discussing a particular subject, one intrinsically uses a set of context dependent terms. The likelihood, that the individual reader may not interpret the same way as the author does, can cause some nasty misunderstandings. So an effort is made to define some of the key terms in the form of a glossary added at the end.

1

Modelling, a core activity

Synopsis *Modelling is a core activity in engineering and science: it provides insight, understanding and models are great sand-boxes – any game can be played.*

Modelling is central to nearly all engineering and science activities and consequently the term *model* is used in a wide range of contexts. In chemical engineering and more specifically process systems engineering, models are the basis for *design* and *operations*. The use in *design* spreads the whole range of design, including crude design, plant design all the way to detailed unit design. In *operations* models are the basis for controller design on the basic level, the intermediate level where more advanced model-based control technologies are employed and also on the high level where the operation of whole plants and production sites are coordinated. Also on the top level, the production planning and all logistics operations are computed based on the models one has of the involved processes, storage and transportation systems.

1.1 Models in chemical engineering

1.1 shows a selection of common engineering activities in which models are being used as basic ingredients. Traditionally, the plant design is separated from the operations domain, latter often being split into three sub-domains, namely control, which usually deals with low-level control using basic controllers, medium-level on which the model-based controllers are often being used today, with Model-Predictive Controllers being a very successful technology. On the top level optimising control is a common model-based technology. Things that *just happen* are typically on the very low and the very high level. These are processes that are driven by events, such as a temperature boundary has been crossed or a switch has been changed. These processes are called discrete-event dynamic processes. In terms of using mechanistic models, the discrete-event domain is probably the area

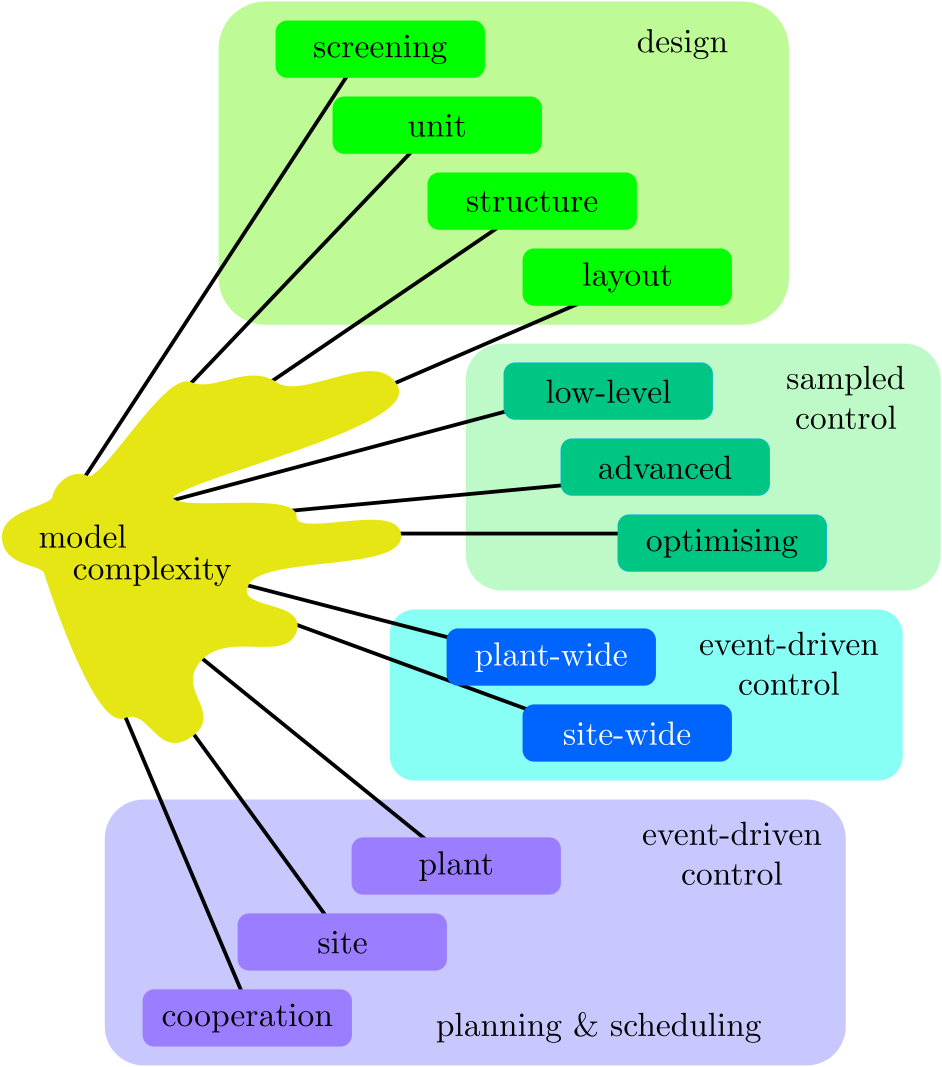


Figure 1.1: Different domains where models are used in design and operations

lagging somewhat behind the continuous processes. Here often empirical models come in use to construct start-up and shut-down procedures, which for supervisory control of batch processes is very much the same. On the plant or site level and the geographically disjunct level of a cooperation, the planning is done using rather simple and generic models. Issue is that essentially each of application uses a different model of the same plant, where the structure of the model is largely defined by the application. For example for the design of the intermediate-level controller one uses models of the plant focusing on the time-scales associated with this intermediate level, which commonly is a complex unit or a group of units. On the plant level it is usually sufficient to use crude steady-state models, which describe the relation of what is coming into the plant (raw products and energy) with what is coming out of the plant in terms of products and waste in a average over a section of the planning period.

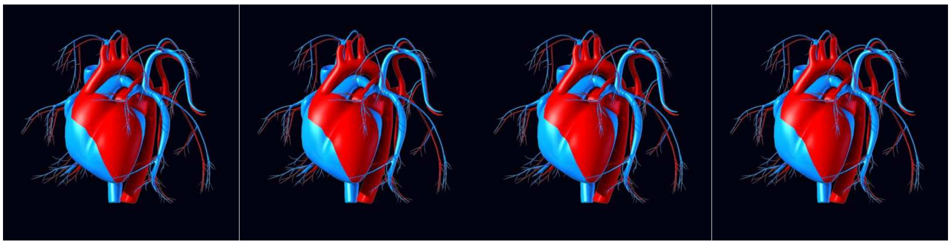


Figure 1.2: A computer-rendered heart model

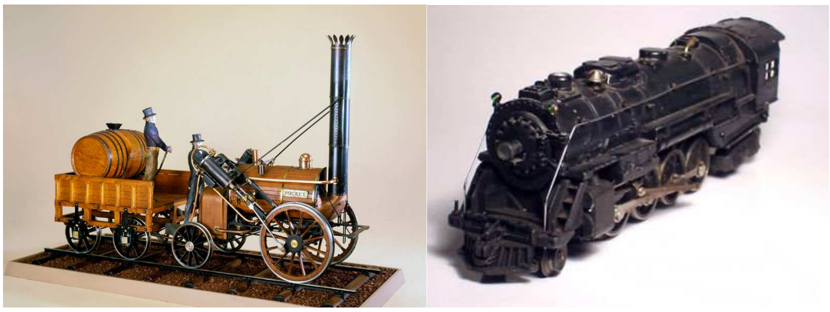


Figure 1.3: Model locomotives

1.2 What is a model

So what is a model? Looking into the dictionaries and similar sources, one finds nearly as many definitions as there are contexts in which the term model is being used. Thus giving a comprehensive definition of model is always going to be very abstract and generic (Apostel, 1960). Rather than providing an explicit definition of the term model, it may be more descriptive to talk about key properties or incentives for making models: The main property of a model and the modelled object is the fact that there must be some similarities defining the relation model-plant. Taking such a view immediately opens the term model to include very many different

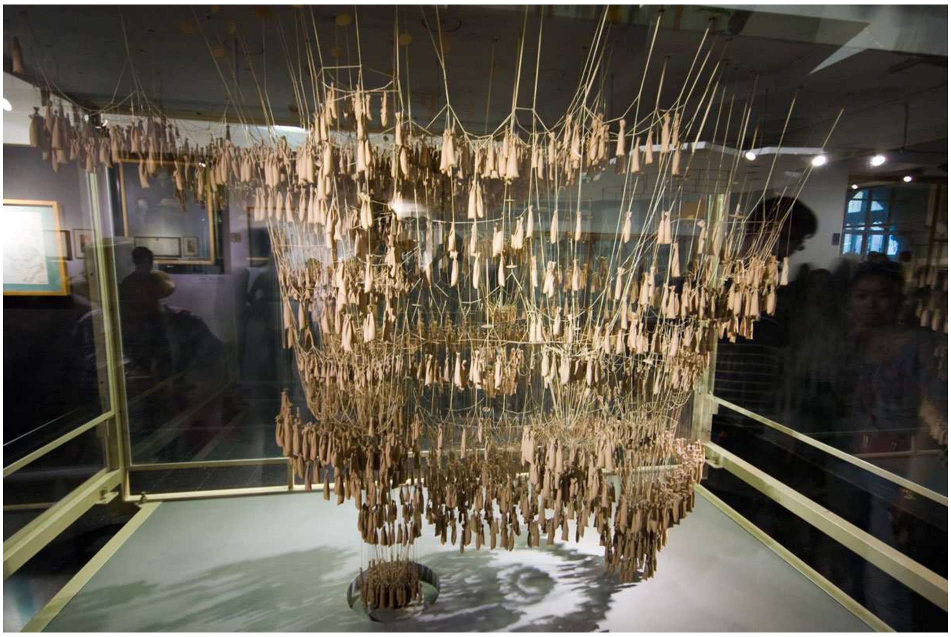


Figure 1.4: Gaudi: funicular model of Sagrada Familia: A method that Gaudi used to design buildings based on catanery arches (the arches that result from hanging chains – the same method used in many medieval cathedrals) – Gaudi used string and little sand-filled bags to load the strings. The result is a geometrical model that represents the structural elements upside down. This is a reconstruction, which is based on a photographs of the original. The original was lost when Gaudi's studio burned down. It is displayed in the 'museum' at Sagrada Familia. One of the most fascinating exhibition pieces.

models such as physical models of a physical object: model trains modelling a real-world train, a network of pipes, pumps and valves representing the current flow in a complex electrical current distribution network. But it also includes people being models for others: role models, stage players model other people behaviours and actions, archeotypes as models for ideas or mode of behaviours for people in a society, plastic models of organs, toys, miniatures of any kind, etc etc.

Making a model is done with having a purpose in mind (Apostel, 1960; Aris, 1978). Models are used to mimic behaviours, to map behaviours into objects. One can then manipulated these objects and do experiments, play with the model without having to "fool" with the real system, and not at least one can explore operational domains that the real-world system may not permit. Models give freedom to the mind, allow tampering and testing, playing with what could become real before it has real-world consequences beyond the use of modelling and simulation time. Models are thus central to any type of exploratory work that has the objective to exploit the object's behaviour.

If we define the context of physical-chemical-biological systems and further constrain the purpose to process design and operations, then the most common core applications are simulation and model inversion. Why? Because design, being the synthesis of an input/output behaviour is mostly done by trial, that is given an input keep on adjusting the plant's behaviour until the desired output is obtained and optimising a process is changing some of the inputs so as to achieve the "best" results by a given measure. Having said so, one seems to point to mathematical textures. Indeed today this is usually the preferred type of models, though physical models may certainly also be an option.

Since models are used so widely, it is no surprise that the use of the term *modelling* varies with the contexts and *models* represents are used to represent an equally wide a range of different "things". Here, we shall refer to modelling as the process of generating a mathematical construct that mimics the behaviour of the piece of world being modelled. The *piece of world* can nearly be anything, a processing plant, any part thereof, in any detail, a living species, microbes, green plant, a piece of rock, tectonic plate, really anything that exists, but also any artificial object such as an algorithm, a program, to mention just two.

The task of generating a mathematical model takes a number of major steps and we can identify at least three primary activities Figure 1.5:

1. **Primary mapping:** The first major step is to map the real-world object of interest, the "plant", into a mathematical object. The basis for this operation is some kind of theory, which is usually the subject of a specific discipline, such as fluid mechanics to model flows, material sciences and thermodynamics to model the material properties, to mention just two. The result of this operation is a set of equations, which, if dynamics are described, are either a set of ordinary differential equations (ODE)combined with a set of algebraic equations (AE) or a set of partial differential equations (PDE) combined with a set of algebraic equations (AE). The first type of model is a *differential algebraic model* (DAE model) whilst the second is a *partial differential algebraic model* (PDAE). The first one will be referred to as a lumped model whilst the second will be called a distributed model. The chosen structure represents the implementation of a first set of assumptions primarily considering time scales and length scales.

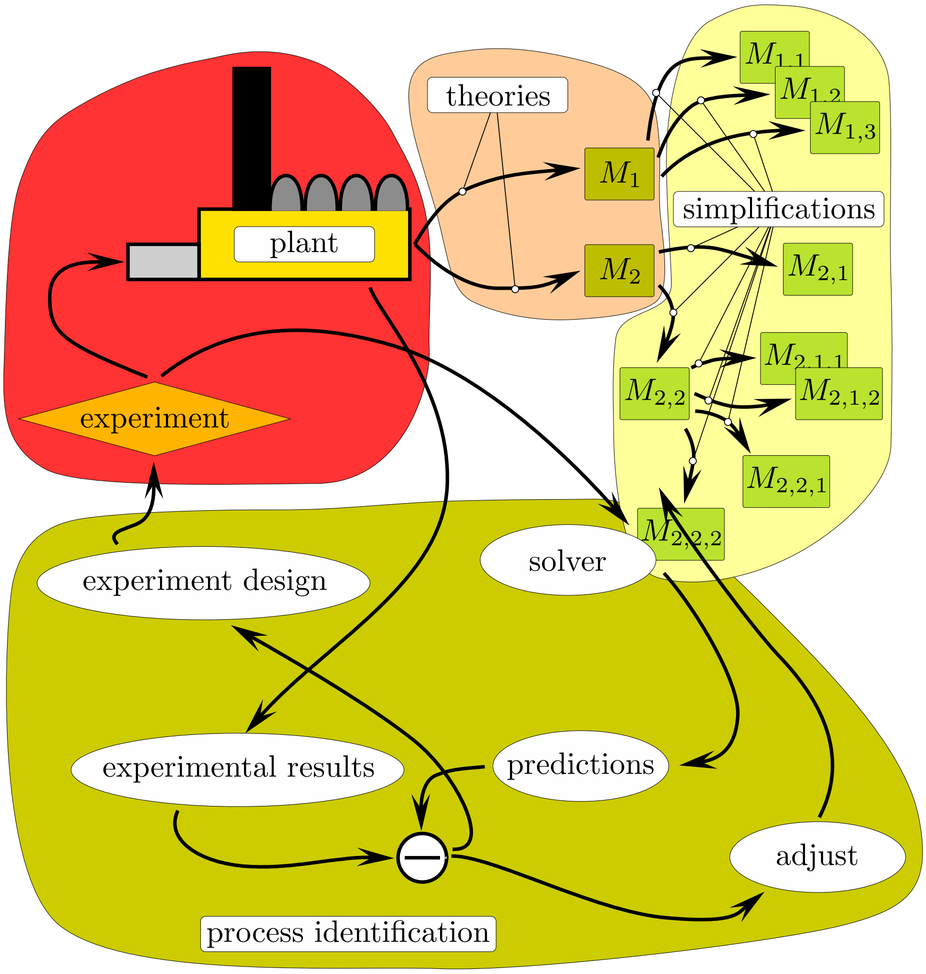


Figure 1.5: Modelling overview: three major domains 1. primary modelling: maps the world into a mathematical object using theory T 2. model simplification: simplifies to match the use of the model using simplification S 3. model identification: fits model to plant adjusting the model to minimize prediction-result mismatch.

2. **Model simplification:** Here, the model fidelity is adjusted. This adjustment is in all cases a simplification. We consider model refinement to be part of the primary domain. Simplifications are of the type to implement additional time scale and length scale assumptions. Often they are order-of-magnitude assumptions, which lead to simplifications of the model. Additionally pure mathematically motivated simplifications may be introduced, such as a polynomial approximation, of which linearisation is a common operation.

3. **Model fitting:** The third domain fits the available ”free” variables of the model such that the predictions obtained from the model match the experimental results best in the sense of a defined objective function and a measure for the mismatch. This is usually referred to as model identification or parameter identification. In the first case the structure of the model may change as well as the respective parameters, whilst in the second case the structure is fixed and only the parameters may change.

Models come in many different flavours: mechanistic descriptions that are based on the principles that form the foundation of science, which are mathematical constructs that capture a certain part of the nature of a natural system. The former is often referred to as a white box model indicating that one can ”see” the mechanics of the box whilst the latter are referred to as black box models, as there is no real mechanistic consideration behind the formulation of the mathematical object representing the modelled systems behaviour. Both boxes do nearly never exist in a pure form, but most often one makes use of a combination of the two approaches. Often the reason for resorting to black box models is simply the fact that one does not know enough about the mechanics of the process or it is far too complicated for the intended use. Another reason is complexity, as black box models are often very simple and particularly in computer applications lead to a lower computational complexity and thus reduced computation time.

2

Processes as a set of volumes

Synopsis *Physical processes occupy a spatial domain, which for the purpose of modelling we split into a set of interacting volumes. The modelled behaviour of the overall process is the concerted behaviour of the individually modelled volumes and interactions. Core is thus description on how the constituent "volumes" behave together with a description of the basic characteristics of the interactions.a Some things just happen, whilst others take for ever. Some volumes are pretty uniform, whilst others are not at all. We introduce lumped and distributed system and very fast systems where "things", for all practical purpose, just happen as well as reservoirs that are on the other end of the scale, are infinitely large and behave in the opposite way: their intensive properties remain constant over time. All this is exposed by looking at an example, a very basic, simple physical system: a cup of water, coffee or tea for that matter.*

2.1 About process' guts

How do we build models for processes? Being curious, we consider the process to be like the corpse of a living species of which we would like to understand its functioning. So we have a look at all of its parts and explore their respective function. Take for example a frog 2.1.

Figure 2.1: A frog

The external anatomy of the frog includes the head, the trunk and the two pairs of legs. The head is equipped with sensors: the eyes and the ears and is also taking up the food with the mouth. The tongue serves the purpose of acquiring the food. The front legs will help this process besides giving support for the trunk when sitting and moving. The hind leg provide the main means of locomotion. The digits of the four feet are connected by membranes that help to transfer the force from the extremity to the water. The cloaca located at the hind side of the corpse serves the purpose of expelling waste and seeds or eggs. The skin varies in texture over the body and is also equipped with different sensors such as tactile sensors and heat sensors. The skin is represents the boundary of the animal and separates it from its environment.

The frog lives in an environment: it mainly takes up food and expels waste, senses light, mechanical forces and pressure waves and it seeks interaction with other individuals of its species. All of which is to be considered when trying to understand the frog's way of life.

What is the "corpse" for the life sciences is the "plant" for the engineer. One may also use the term "system". The term "system" is very general whilst the term "plant" has become the de facto standard in control more or less independent of the specific application field. For the time being, let us remain general and use the term system. Later, when we get to discuss applications to a plant that actually produces a product of one or the other kind, we shall use the term plant.

Like the frog, any physical system is seen as a spatial domain bounded by its boundary. It is also always embedded in its environment, with which it shares its boundary. The environment affects the embedded system and thus the modelling of the system's behaviour requires the model of the immediate environment.

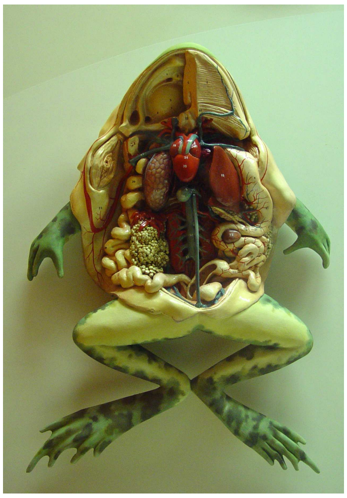


Figure 2.2: Frog anatomy model

If we want to understand more about the frog's functioning, we need to get to look inside and explore the functionality and interactions of all internal pars. So we go about dissecting the beast's corps and analyse each part that we find in isolation and explore on how it interacts with the parts it is connected to.

When opening the frog's corps we will first find the different organs, the heart with the small lungs behind on each side, the stomach, liver below and the intestines attached to the stomach. The kidneys are hidden behind the liver and several other organs are coming to the light as one moves others out of the way, for example the pancreas.

The organs are clearly identifiable entities that are coupled together, and the functioning of the frog is given by the concerted operation of the connected organs combined with the muscular structures, supported by the skeleton and wrapped into the skin.

On this level, the subdivision is based on a visible boundary, the skin of the organ.

The system may not be simple but may have internal structure. The structure is captured by sub-dividing the space being occupied by the system into smaller parts. The behaviour of the system is then captured by describing the behaviour of the individual sub-systems and their interactions. The approach taken to analyse and describe the functioning of a plant is essentially the same: one subdivides the plant based on the argument of "visible" boundaries. These define units, phases, particles or the like. The *visibility* is often characterised by a type of density, which changes discretely as one moves in the space of the macroscopic world. The *discontinuities* are usually taken as the boundaries defining the "parts" or subsystems.

2.1.1 Conceptual system parts

2.1.1.1 Distribution effects

When using intensive quantities to identify subsystems, one finds quickly that whilst they do change quite abruptly when crossing the boundary, they also change as one moves to a different location inside, though the variations may or may not be of relevant magnitude. The term *relevant* will be subject of many more discussions throughout the book. So we will come back to it later. The fact that the intensities may or may not be considered constant throughout the volume occupied by the (sub)-system gives raise to the definition of two conceptual systems: lumped and distributed systems.

> **Definition** – **Lumped system:** is a spatial domain in which the intensive quantities **do not** change with the position.
>
> **Definition** – **Distributed system:** is a spatial domain in which the intensive quantities **do** change with the position.

Defining a Cartesian co-ordinate systems, the variations may occur only in one or two dimensions, in which case one refers to a 1D or 2D distributed system, where the "D" stands for Dimension. Obviously there is also a 3D distributed system. In many cases it is also an advantage to adapt the co-ordinate system to the peculiarity of the geometry of the system, so for a cylindrical system the model equations may be simpler in a cylindrical co-ordinate system, which is the case if things do not change in the rotational co-ordinate but only in the radial and/or the axial direction. Similarly for a spherical geometry, a spherical co-ordinate system may be of advantage.

If the geometry is not cylindrical or spherical, one may consider a mapping of the geometrical space into one of the regular spaces retaining the angles, which makes mathematics easier. Such a transformation is called a ”conformal mapping” (Rudin, 1987).

2.1.1.2 Time-scale effects

Another criteria characterising systems is associated with the time-scale in which the model is drawn up. It is the application that determines on how well the model must mimic reality. If we are interested in using the model for designing a continuous plant that primarily operates in a stationary mode, then we need to put the emphasis on the stationary behaviour of the plant. In contrast, if we want to control a fast reactive system, we need to focus on the fast dynamics of the part in which the reaction takes place. Often, we also assume that things ”just happen”, i.e. in the time frame we are interested some things can be considered to occur instantaneously.

In general, the time scale always splits into three parts, The dynamic domain of a process description splits into three parts: one which one assumes constant, which map into our reservoirs, one which is very fast, so fast that it is assumed to happen in an instant, mapped into our event-dynamic systems, and the third one is in between, what we refer to as the dynamic part. So if we look at the time scale, every model contains these three parts.

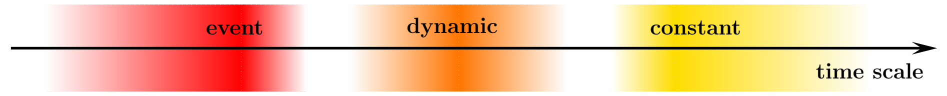


Figure 2.3: Three time domains of time scales: event - just happens, dynamics - everything happens, constant - nothing happens

Given a time scale split into **constant, dynamic, event dynamic** systems can be classified into three classes of dynamic behaviour:

> **Definition – Constant system:** does not change with time; thus all its intensive properties remain constant over time. It is consequently also infinite.
>
> **Definition – Dynamic system:** does change continuously in time. Its capacity may vary over the whole scale.
>
> **Definition – Event-dynamic system:** does change in an instance thus some properties change step-wise in time. This may be for intensive as well as for extensive quantities.

Constant systems are also called reservoirs mostly following thermodynamics’ terminology. They are always part of the environment in which the plant/system is embedded.

2.1.2 Abstraction to a graph

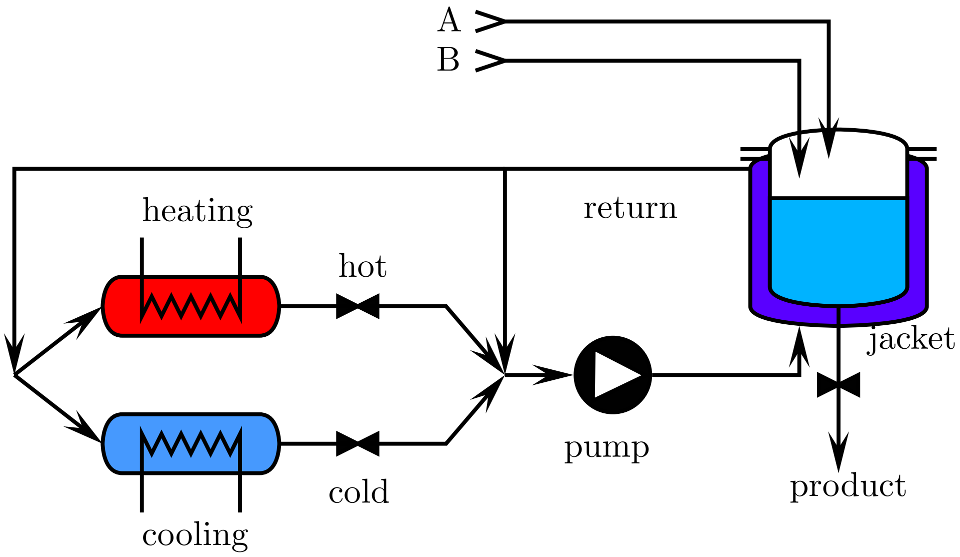


Figure 2.4: A batch plant with a heating cooling system attached to the jacket

To make the "guts" visible, the process is seen as a set of control volumes. The decisions taken when subdividing into sub-volumes are critical, because the chosen structure is determining the "contents" of the model, meaning what and how well the model describes the process. Any error in the choice of the model structure is expensive as the structure is the foundation of the model and has to be done at the very beginning of the whole modelling process. The choice determines what the person recognises as the essential parts of the model. Anything that is not considered at this level, will simply not be included in the model.

What is to be modelled is the first decision that must be taken. This object is embedded in an environment with which it interacts. Those parts of the world that are considered to not interact with the plant are left out. This decision defines the systems boundary, the overall scope of the model. It consists of the plant and a process-model-relevant universe in which the plant is embedded. Next, both the plant and the environment are split into sub-systems. As mentioned, as an argument for the splitting one uses commonly phase boundaries, that is step-like changes of an intensive property. This is done in three dimensions with the sub-systems being volumes.

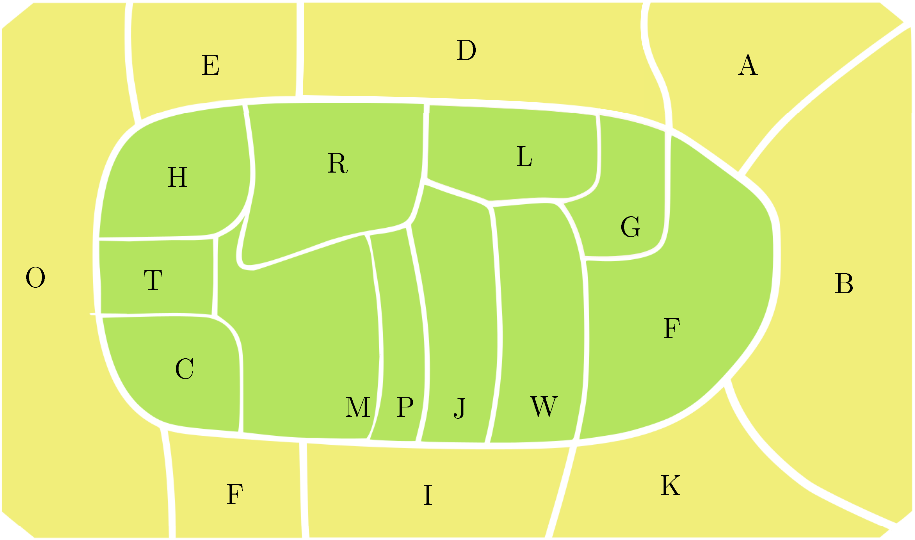


Figure 2.5: The process cut into bits and pieces

These volumes are often also referred to as *control-volumes*, reason of which will become evident later when we describe the behaviour with equations. Figure 2.5 shows our imaginary plant and its environment dissected. the plant is in 3D, whilst the picture is in 2D.

So far we have not indicated anything about what is in the system and what parts we consider to communicate with each other through their common boundaries. Figure 2.6 shows some of these details. The colours indicated domains or phases, where domains are connected spatial domains. The arrows crossing the boundaries indicate the communication of the two parts connected by the tail and the head of the arrow. As an example we show also different quantities being exchanged by colour-coding this information. We could think of black indicating mass flow, whilst red would be used for energy flow other than convective flow.

In a next step, the graph is ”exploded” showing now more clearly the structure one has decided to use for the representation.

The final stage of the primary abstraction process is to make assumptions about the nature of each individual part: discrete-event dynamic, lumped, 1D, 2D, 3D distributed.

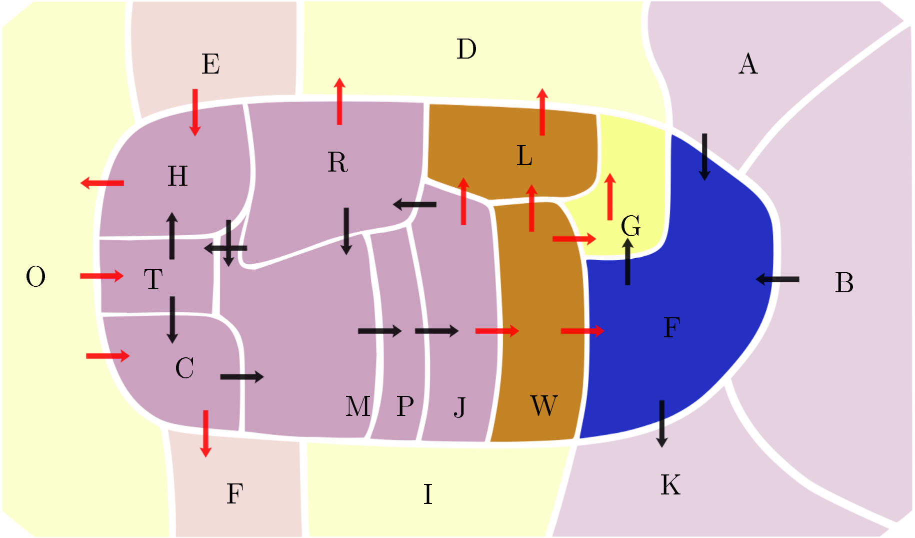


Figure 2.6: The process cut into bits and pieces exploded. Finally a graph with an abstract representation of the plant

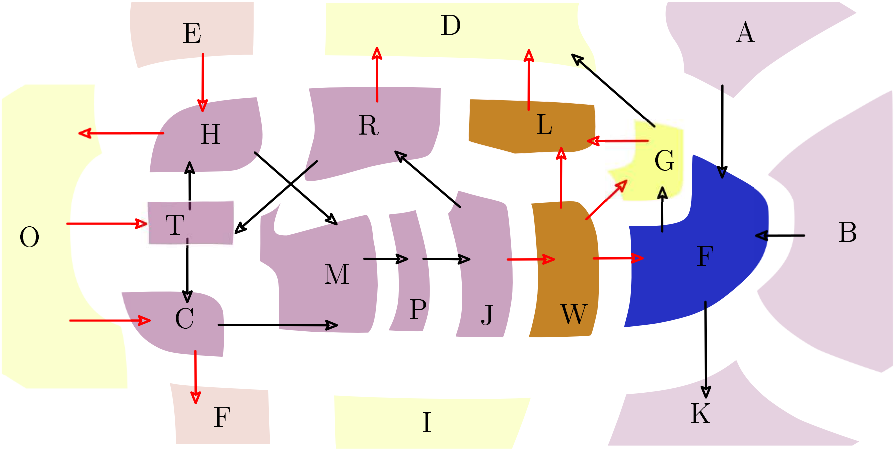


Figure 2.7: The exploded plant

2.2 Capturing system structure in a graph

The underlying structure of the model is a graph, a directed graph to more precise. Whilst this ”being a graph” is apparent from the visual representation, it offers a convenient abstraction, which enables a proper representation of the behaviour in terms of a directed graph with nodes that represent primitive or simple capacitive containers, that interacting with each other

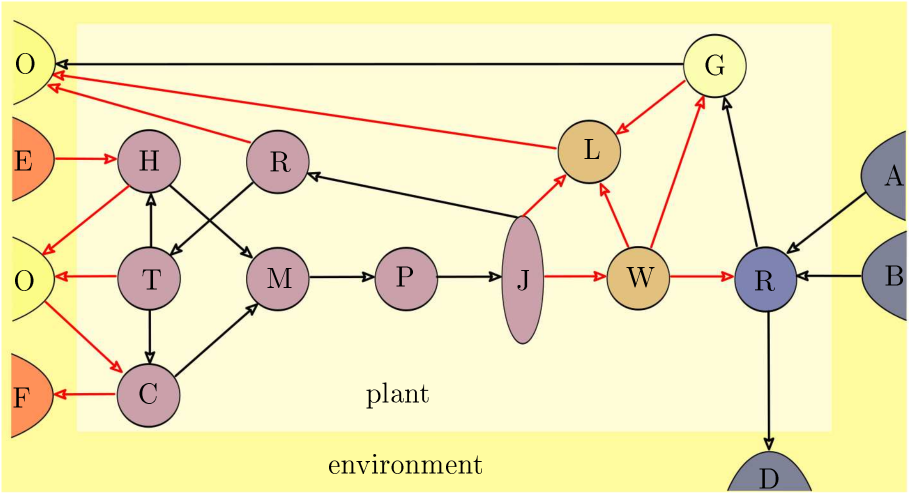


Figure 2.8: Finally a graph with an abstract representation of the plant where we used partially the graphical symbols as listed in Table 2.1, and Table 2.2

and the environment across the common boundaries. The graph is an abstraction designed to reflect the components of the process as viewed by the person modelling the process.

Choosing the conserved quantities as the "contents", which later we will refer to as tokens, the link to a mathematical description of the system becomes apparent: the conservation equations for each sub-system supplemented with the description of the exchange between sub-systems and the internal changes provide us with the basic description of the overall system's behaviour. How to establish these equations will be the main subjects of following chapters, which will show that establishing the mathematical description is a rather mechanical operation that can be done systematically. The key is really the choice of the process structure. It defines how well the process is being modelled. Examples will illustrate that it is not quite so easy to suggest a structure that captures the process characteristics that are relevant for the model application. Having a visual method to represent the model makes it much easier to discuss and comprehend the models' structure compared with having to extract this information from a set of mathematical equations.

The graphical representation we introduce has really only a few components, which may be adjusted to the particular need if required (Table 2.1). For connections (Table 2.2) we use different styles of lines and colours to distinguish between different extensive quantities and signals and different

Table 2.1: Graphical elements for systems

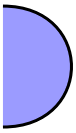
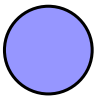
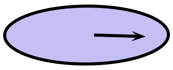
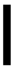
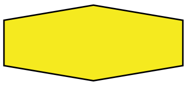
	**Reservoir** : an infinitely large source of extensive quanity with constant insensitive properties
	**Lumped system** : a capacity of finite volume with the intensive properties being uniform over the whole volume the system occupies
	**Distributed system** : a finite-sized volume with the intensive properties being a function of the position. A little co-ordinate system is shown to indicate the distribution in 1D, 2D or 3D with one, two and three arrows. The distribution effect on the surface is not shown here. It can be made visible in the connections.
●	**Event-dynamic system** : a very fast system, so fast that one assumes it to be in pseudo-steady state all the time.
	**Event-dynamic system** : a surface with no capacity, thus exhibiting event dynamics.
	**Information processing systems** : controllers are information processing systems that obtain information either from other information processing systems or from sensors linking them to the plant on the input side. On the output side they may connect to other information processing systems or stream manipulators such as valves and switches.

Table 2.2: Graphical elements associated with connections – they represent interactions between systems. They are used to signify transfer of extensive property such as mass, energy, work, heat etc, for each of which one chooses a different form in terms of line type and colour. Below a set of possible choices.

⟶	**Connection mass** : mass transfer, both for total mass or species. Mass transfer induces transfer of internal energy and volumetric work.
· · · ➤	**Connection heat** : heat transfer.
- - - ➤	**Connection work** : work transfer, often volume work or mechanical work.
· · · ➤	**Connection signal** : signal.
⋈	**Continuous manipulator** : flows may be manipulated continuously, for which we use a valve symbol.
•⟋ •	**Discrete manipulator** : flows may be switched on or off for which we use a switch symbol.
▭▷	**Observer** : a sensor device that extracts information about the state of the system.

morphologies of the extensive quantities. It should be mentioned here that the connections transferring any kind of extensive quantity are the result of making a fast-transport assumption on a physical transfer system, usually a distributed system, which we will discuss in more detail later in more detail (see 4.2).

2.3 How much detail

The plant structure is often complex in terms of ”many”, which motivates a hierarchical approach by which the plant is recursively sub-divided into smaller and smaller pieces. But then, how ”far” should one subdivide? How small should the smallest pieces of the space be that one considers? Well, as so many other things, the choice depends on what one wants, namely what one wants the model to describe, because it is the ”purpose” in the context of the application that defines the required detail. There is no point of being too accurate – the model must just be good enough for the application. Fine – but then what is accurate? Accuracy is only defined in the context of the application. The consequence of all of this is that the model has to be chosen first, then applied and only in terms of the performance of the application, a judgement of the model accuracy can be given. This leads to an overall iterative approach in which the model may either be refined in parts where there were too few details or one may drop some details as they are not relevant for the application. This implies that one will generate not only one model, but typically there will be a set of models, which describe the plant on different levels of detail. Since the ”level of detail” often refers to the degree with which the spatial domain is subdivided, one talks about ”granularity”. Thus a fine-granular model describes the process in more detail than the more crude one.

The separation into control volume and connection is probably the most critical step in the whole modelling process. Why? Because it determines the level of detail that is included in the model. If we take our frog as an example and open the frog, we find the primary organs, like heart, lungs, stomach, intestines, brain, eyes, muscles. If we want to know how any of these organs is working we have to open it up again and look into its internal structure. This process can be repeated through layers and layers of smaller and smaller structural components down to the molecular, atomic level. For example, the muscle will show several levels of detail as shown in 2.9.

This process of looking into and adding more details to the model is called ”refining”. It always requires adding structure to the model. This can only be done on this initial level of the modelling process. The finer the detail, the

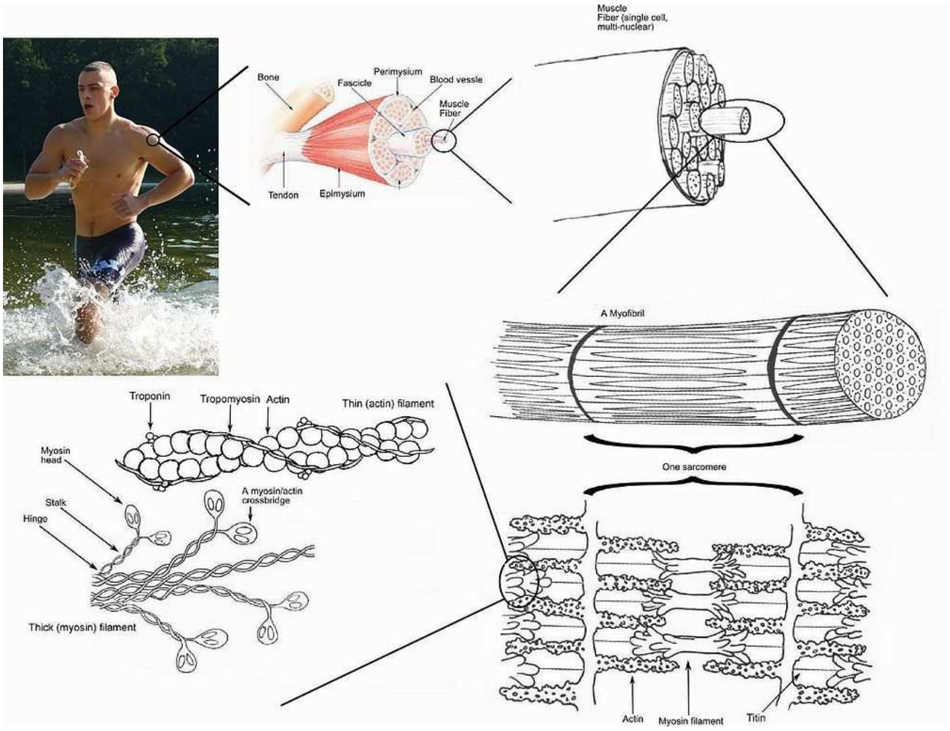


Figure 2.9: Structure of a muscle (source: Wikipedia)

finer the "granularity" of the model. The term "granularity" nicely captures the nature of the model in terms of the structural appearance of the process, and thus model one had in mind when deciding on the level of details to be described. Once a structure has been chosen, on the mathematical side, models can be simplified, not just once but recursively. Thus the choice of the initial structure is crucial to defining the level of detail being included.

The detailing is done on the length scales: on each level, smaller details are being brought to the light. In most cases this directly relates to the time scale as well, smaller also implies faster, though the dynamics depend on the capacity and not the size. However as the detailing is usually involving the same type of materials and thus specific capacity, the increasing detailing almost always implies a simultaneously zooming into the time scale.

Asking the question on how many details should be included, so the generic answer is: just as much as required by the application of the model. In practical terms this requires to "balance" the granularity such that where fast application-relevant processes occur, the granularity is fine, whilst where not much happens it is large.

2.4 Mapping the behaviour

The model describes the behaviour of the modelled system interpreting all spatial parts as capacities. The conservation principles are applied to these capacities thus describing the change of the conserved quantities in the system as a function of the exchange of the extensive quantities with the system's environment and in some cases transposition of extensive quantities from one type into another one. The latter captures the changes due to *chemical or biological reactions* (8.2), when one type of mass is changed into another one. Also phase changes belong to this class and the change of mass into energy and back.

For a *distributed system* (4.2) the accumulation is the integral over the changes in the conserved quantity in each point. Thus it is a volume integral over a density measure, whilst the exchange with the environment is an integral over the surface. More precisely, the flow is the integral of flux in normal direction as a function of the location on the boundary over the *whole* boundary. Emphasis is placed on the fact that the *whole* boundary must be included. This may sound somewhat trivial, but then in applications it is easy to overlook this fact and it is good practice to check very carefully if indeed all the streams crossing all the boundary elements are included.

2.4.1 Mapping a container – as simple as it goes

For the moment, let us step back from the distributed systems as this primarily seems to lead into the issue of how to describe the flow across the boundary. So let us be generous in the sense of ignoring some of the details and take a good look at a really simple system: a glass of water or any other container, for example a tank, if one prefers a technical application. In any case a container that is being filled with water from a pipe for further use.

2.10 shows the graph of the simple system. As we decided to ignore details, we leave out any thermal or mechanical effects, so that the process of filling in water is well described by a dynamic mass balance: The change of the mass is equal to the amount of water poured into the container per unit time. Defining the symbol m_W for the mass in the container W, and the flow by the symbol $\hat{m}_{R|W}$ the differential change in the mass over time dt is:

$$d m_W = \hat{m}_{R|W} \, dt$$

which can be rewritten as a differential equation, making it look more fa-

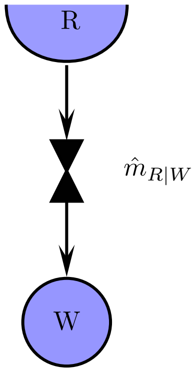


Figure 2.10: Simplest version, no fringes, only the water in the glass is modelled and the inflow of water from a reservoir

miliar:

$$\frac{d\,m_W}{d\,t} = \hat{m}_{R|W}$$

This equation defines the mass m_W as the state of the container and the flow $\hat{m}_{R|W}$ acts here as the input.

We will always write the balances in the same form, namely the accumulation on the left-hand side is equal to the transfer across the boundaries and the internal changes. The accumulation term is the differential change of the state with time whilst the flow across the boundary is the same token as the state flowing, with the flow indicated by the decorator " ˆ " and the boundary identifier " source | sink ", whilst the internal changes will be indicated using the decorator " ˜ ".

So what is the "system" here? Well the container - and the environment is the pipe with the tap from which water is flowing with a given rate. But then looking at this balance equation, the container itself has absolutely no effect; it does not seem to appear in any form. Nothing in the model reflects even the existence of the container. Agreed, there is mass accumulated IN the container. So the model has the property of being able to hold and accumulate mass inside.

Some more of the container properties appear if we ask for the level in the container. To introduce this bit of information, we have to add two more relations, namely one that links the mass to the volume (V) and one that relates the volume with the level (h). The first one introduces a mass density ρ:

$$V_W := \rho^{-1}\, m_W$$

The second one describes the geometry of the fluid body contained in the container. Since in general the cross-sectional area of the container (A_W) changes with the height h_W, the computation of the volume of fluid in the container is the integral:

$$V_W := \int_0^{h_W} A_W(\tau)\, d\tau$$

The three equations form a differential algebraic equation (DAE) system, which can be solved directly using a corresponding integrator, but which in most textbooks would be converted into a simple differential equation in the level. This can be done by changing the state space from mass to level. Mathematically this is achieved through differentiation and substitution. Assuming the density is constant, differential change in the volume-mass relations is:

$$\begin{aligned} dV_W &:= d\left(\rho^{-1}\, m_W\right) = \rho^{-1}\, dm_W \\ &:= \frac{\partial\, V_W(h_W)}{\partial\, h_W}\, dh_W \quad := A_W(h_W)\, dh_W \end{aligned}$$

Substitution then yields :

$$\begin{aligned} \frac{d\, h_W}{d\, t} &= A_W^{-1}(h_W)\, \frac{d\, V_W(h_W)}{d\, t} \\ &= A_W^{-1}(h_W)\, \rho^{-1}\, \frac{d\, m_W}{d\, t} \\ &= A_W^{-1}(h_W)\, \rho^{-1}\, \hat{m}_W \end{aligned}$$

which is an ordinary differential equation in the level as a function of the inflow of mass. Since we needed to differentiate only once, the above DAE system is called a DAE of (differential) index 1 (Brenan et al., 1989). The ordinary differential equation can be solved if the area is given as a function of the height.

2.4.2 And it may overflow: events

So again, now what is this model describing? It gives an accumulator of mass, but then if we compare it with our original physical system, there will be a point where the container, however big, is full as no water is taken out. The above model though knows nothing about the container being full. The mass and thus volume and level will just simply keep on growing as water keeps on flowing in.

In order to resolve this problem, an additional flow must be introduced, which comes into action as the maximum level has been reached (2.11). We

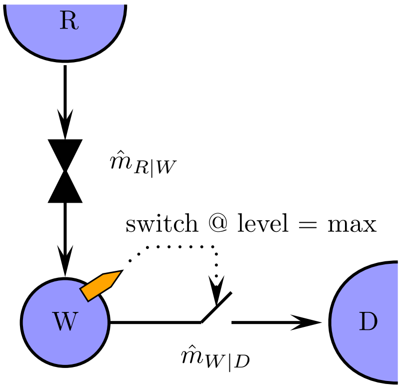


Figure 2.11: In order to take care of the possible overflow, the model is extended with an discrete-event-dynamic component, namely an outflow, which is switched on as the level reaches the maximum.

may express this using a switch-type of representation, that may look like:

$$\hat{m}_{W|D} := \begin{cases} \hat{m}_{R|W} & \text{if} \quad h >= h_{max} \\ 0 & \text{otherwise} \end{cases}$$

The balance equation is extended by this additional flow:

$$\frac{d\,m_W}{d\,t} = \hat{m}_{R|W} - \hat{m}_{W|D}$$

This switching behaviour being the result of having reached a state condition is called a discrete-event dynamic behaviour. "Discrete-event" because it is an event that triggers the discrete change.

At this point it is convenient to switch to another measure of mass, namely number of moles n:

$$\frac{d\,n_W}{d\,t} = \hat{n}_{R|W} - \hat{n}_{W|D} \,,$$

reason being that the physical properties usually are given in the molar representation. The link between mass in kg and in moles is simply the molecular mass.

2.4.3 Extending to warm fluid content

Now let us raise the level of complexity by assuming the water to be warmer than the room in which the container is located. We remain "generous" by assuming that the physical properties of the involved materials are not a function of the temperature. In order to include the thermal effect of losing heat to the environment, we have to include at least one heat loss, but possibly several when we look closer (2.12). Why several? Well, because

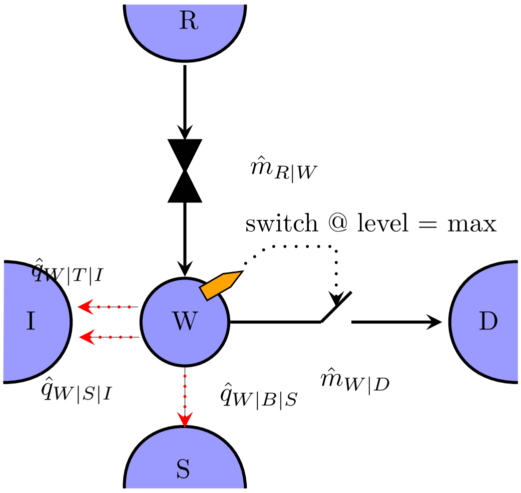


Figure 2.12: Having warm water in the container and being interested in the temperature adds the need for an energy balance and the description of the main heat losses.

the container interacts differently with the environment through different parts of its boundary. If the container is open, the heat loss through this part will be different than through the side walls, which in turn also will be different through the bottom, depending on what the bottom interacts with. So the problem now proliferates quite quickly. The *energy balance* (9.1), which we can argue reduces to an enthalpy balance that includes the change of the enthalpy of the water being equal to the enthalpy entering and leaving the body of water due to inflow and potentially also outflow and the various streams extracting energy in the form of heat. The model is now augmented with the enthalpy balance:

$$\frac{d\,H_W}{d\,t} := \hat{H}_{R|W} - \hat{H}_{W|D} - \hat{q}_{W|T|I} - \hat{q}_{W|S|I} - \hat{q}_{W|B|S} \tag{2.1}$$

where the three heat flows $\hat{q}_{W|T|I}$, $\hat{q}_{W|S|I}$, $\hat{q}_{W|B|S}$ for the side, the top and the bottom.

Having introduced enthalpy H and conductive heat flow $\hat{q}$ we must relate these quantities to the state, namely enthalpy and mass. Since enthalpy is an Euler homogeneous function of degree 1 (Callen, 1985), the enthalpy can be written:

$$H := \frac{\partial\,H}{\partial\,n}\,n := h\,n$$

with h being the partial molar enthalpy.

Heat is flowing downhill the temperature gradient. So a simple representation is known as *Newton's law of cooling*:

$$\hat{q}_k := -{k^q}_k \left(T_E - T\right),$$

where k is the stream index and the subscript E stands for the environment assuming that the flow is modelled as going from the system to its environment in all cases. The c is a conductivity parameter characterising the property of the boundary times the size of the boundary, thus the area.

Two new variables have been introduced, namely the partial molar enthalpy h and the temperature T. The two are linked over the relation:

$$h := \int_{T_{ref}}^{T} c_p(\tau)\, d\tau$$

with $h(T)$ being the heat capacity, which in this context is a function of the temperature. The reference temperature appears because one measures enthalpies relative to a standard as the absolute value is not known. From this equation we need the temperature, because it appears in the heat transfer equation. The problem of extracting T is obviously in general not trivial. But again let us be generous and assume that the specific heat capacity h is constant. This makes the task of computing the temperature from the enthalpy and the h easy. So if we now look at the process, then it is driven by the mass flows but also by the room temperature.

2.4.3.1 Warm fluid in a high-capacity container

Think of replacing the container with a cup and the above water with coffee – assume a espresso size, it is quite obvious that the coffee stays warmer if the cup is warmed up before the coffee is added as compared to pouring it into a cold cup. Industrial espresso machines have therefore a heating surface for keeping the cups hot. So clearly what we have so far not considered is the effect of the container beyond it representing a resistance to the heat flow. Adding the container as a capacity and assuming that it is open, which a cup usually is, the heat transfers on the sides and the bottom must be refined by adding a capacity effect and again, we raise the level of complexity of our model. The mathematical representation of such systems will be done later. For now we shall resort to the model in the form of a topology.

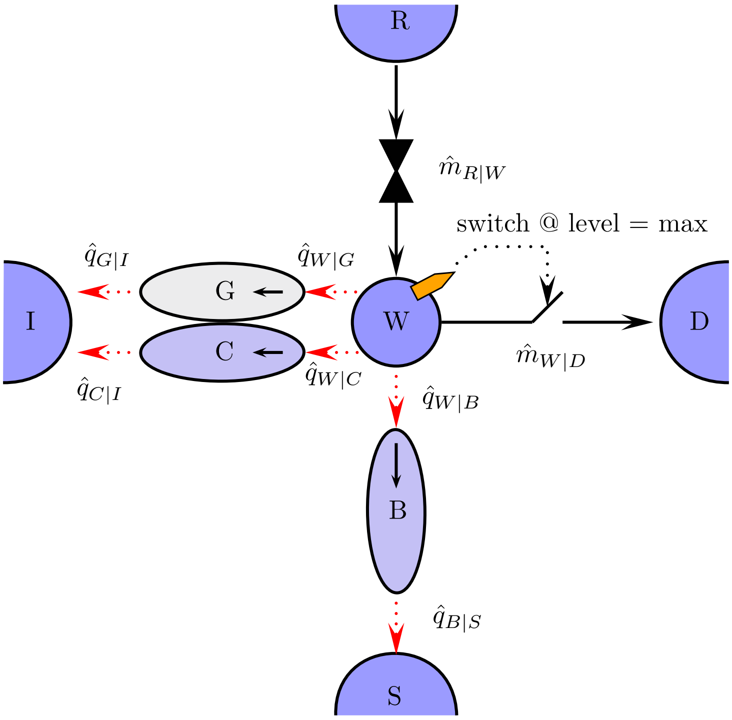


Figure 2.13: Having warm water in the container and being interested in the temperature adds the need for an energy balance and the description of the main heat losses. Adding a significant capacity of the container increases the complexity a level more.

2.4.3.2 Warm fluid in a high-capacity container with lid

Things get even more interesting if we introduce a heater in the fluid body and cover the container with a lid. Now we have the situation that the fluid will eventually boil and the main heat loss will probably occur through the lid. The mechanism is that the fluid will evaporate, travel through the gas phase and then condense on the lid. If it condenses in drops, things will look different than if it condenses as a film. The condensed material will then flow back to the liquid phase.

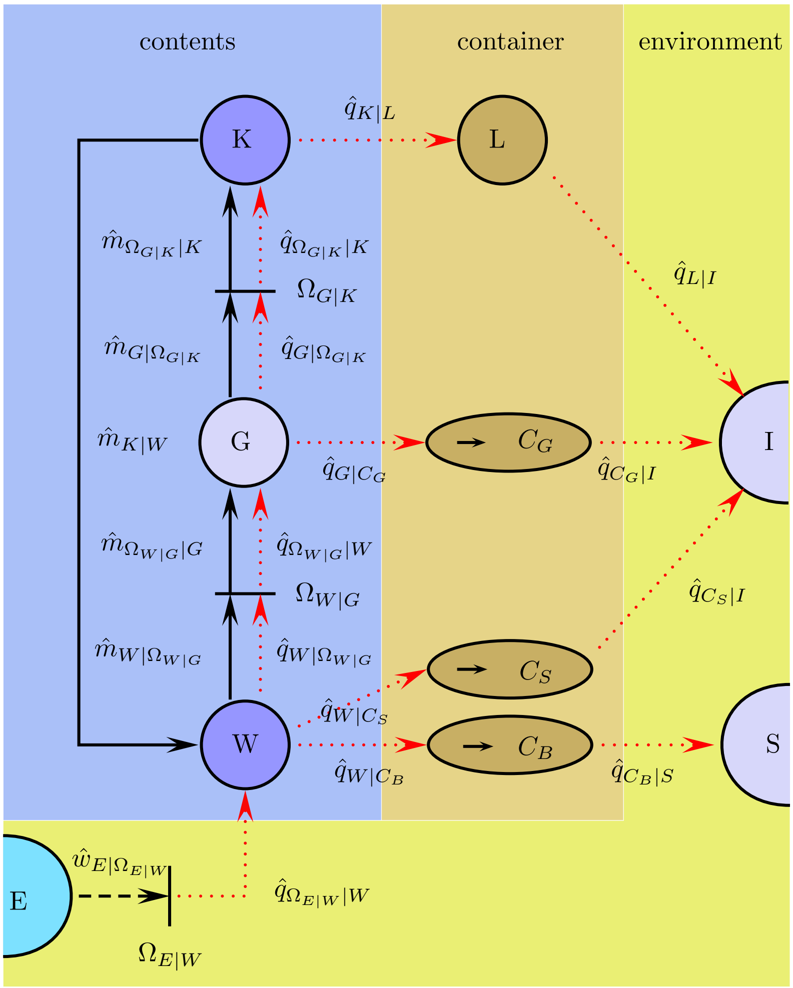

contents
container
environment
K
$\hat{q}_{K|L}$
L
$\hat{m}_{\Omega_{G|K}|K}$
$\hat{q}_{\Omega_{G|K}|K}$
$\Omega_{G|K}$
$\hat{q}_{L|I}$
$\hat{m}_{G|\Omega_{G|K}}$
$\hat{q}_{G|\Omega_{G|K}}$
$\hat{m}_{K|W}$
G
C_G
I
$\hat{q}_{G|C_G}$
$\hat{q}_{C_G|I}$
$\hat{m}_{\Omega_{W|G}|G}$
$\hat{q}_{\Omega_{W|G}|W}$
$\Omega_{W|G}$
$\hat{q}_{C_S|I}$
$\hat{m}_{W|\Omega_{W|G}}$
$\hat{q}_{W|\Omega_{W|G}}$
C_S
$\hat{q}_{W|C_S}$
W
C_B
S
$\hat{q}_{W|C_B}$
$\hat{q}_{C_B|S}$
$\hat{w}_{E|\Omega_{E|W}}$
E
$\hat{q}_{\Omega_{E|W}|W}$
$\Omega_{E|W}$

Figure 2.14: A tank with water being heated up. As one gets over a certain temperature, the heat losses are mainly due to the condensation on the lid.

3

Network communicating tokens

Synopsis *At one point, one needs to decide on what balances must be written. A simple model of a sedimentation process is used to illustrate how one decides on what component mass balances must be written. Adding a thermostat to a part of the system extends this discussion to where energy balances must be established.*

3.1 What to include

Models are established for a certain application, which intrinsically define the time scale in which one will work. This in turn will provide the arguments for defining the universe in which the application and thus the model lives. It provides the main argument to make the split between the active plant and its environment, both of which make up the model universe. Also the next step in the modelling process, namely the breaking up of the model universe into pieces is also based on the same arguments though augmented by structural considerations, such as phases, equipments or the like. These are the first and most important decisions being taken in the overall modelling process, simply because they define the application domain. The process of subdividing the plant is recursive and can be extended in depth to any desired level, at least in principle, whilst extending the application domain brings one back right to the beginning.

Once one has made the decision on the granularity of the model, the basic graph such as shown in Figure 2.7 is established. This graph shows the distribution of the capacities for the physical part of the plant, thus the accumulation of conserved quantities, and the transfer of these same quantities between the various parts. It is a directed graph with the direction indicating the reference co-ordinate system for each transfer. Besides, there is an attached graph, which represents the information-driven part: the control system.

On the next level, we introduce the dynamic assumptions being made about all systems. As we will discuss in more detail later, the transfers of extensive quantities are seen as physical systems that transfer very fast and exhibit no capacity effects (3.5). For the capacities one has first to judge if it is to be seen as having a negligible capacity, thus if it is to be seen as an event-dynamic system, or if it so large and thus slow that it can be viewed as a reservoir. All others are seen as dynamic systems. Secondly for the two dynamic systems, one has to consider if they are to be modelled as distributed system or if one can assume uniform conditions across the spatial domain the system occupies. The latter is essentially a judgement on the internal transport dynamics. For a lumped system the internal transport dynamic is in the event-dynamic time scale, whilst for a distributed system the internal transport dynamic is not negligible, both in the light of the application (see 3.5). The result of these considerations gives the modified graph of the type as shown in Figure 2.8. This graph represents all the components that are physically present, for which reason we refer to it as the *physical topology*.

By defining the dynamics and the distribution effects, one defines the nature of the involved mathematics, namely if the core conservation equations are algebraic for event-dynamic and constant systems, ordinary differential equations for the lumped systems and partial differential equations for the distributed systems. Question then remains what conserved quantities must be considered in each part of the directed graph; and what does each part conserve and exchange with its neighbours and what is being undergoing any transposition of any kind including reactions. The ”what” is the subject of this chapter together with the concise mathematical representation of the balances for networks as they are represented by these directed graphs.

3.2 Network representation

In order to describe more complex systems we suggested to model the process by first splitting it up into control volumes (see 2.1). So the ”guts” of the process are being exposed and split into different control volume each of which can be considered to be of one or the other type (see 2.2). When writing the equations representing the plant, which starts with the balance equations for each capacity, one collects all the streams crossing the boundary of the system. The latter is really the reason one calls it a ”control volume” using the term ”control” in an accounting context. The flows are always coming from somewhere, a system of one or the other kind, and are always going somewhere, again a system of one or the other kind.

The control volumes, together with the transfers of extensive quantities form a network represented as a directed graph and we can use the graph theory to find a compact description. For the discussion it is sufficient to look at a simple network as it is shown in figure 3.1. The extensive quantity we want to balance is Φ and we look at the system a. Then it has one arrow out, which is labelled with $\hat{\Phi}_{a|b}$. The label has as its index the definition of the reference co-ordinate introduced by defining the arrow, namely it comes from a and it goes to b.

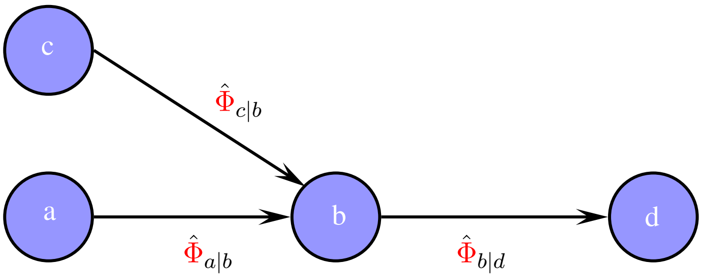


Figure 3.1: A sample network

$$
\begin{aligned}
\dot{\Phi}_a &= -\hat{\Phi}_{a|b} \\
\dot{\Phi}_b &= +\hat{\Phi}_{a|b} + \hat{\Phi}_{c|b} - \hat{\Phi}_{b|d} \\
\dot{\Phi}_c &= -\hat{\Phi}_{c|b} \\
\dot{\Phi}_d &= +\hat{\Phi}_{b|d}
\end{aligned}
$$

To take a physical example to help us interpreting, take the extensive quantity Φ being the internal energy and the effort π being the temperature, then indeed the energy contents increases in system a if the temperature in b is larger than in a, thus $T_b > T_a$, with the energy coming from system b.

This handling of signs we can formalise, which turns out to be a healthy thing to do, because sign errors are very common in these equations and making things a ”rule”, makes the writing of the equation procedural essentially removing the problem.

For this purpose we introduce a quantity, which we call the direction indicator α, which for a tail of an arrow is -1 and for a head $+1$. The direction indicator is to be indexed with the connection and the system it is connected to. So for the connection $\hat{\Phi}_{a|b}$ the tail is $\alpha_{a,a|b} := -1$ and for the head $\alpha_{b,a|b} := +1$. The formalised balance equations then are:

$$\begin{aligned}
\dot{\Phi}_a &= \alpha_{a,a|b}\,\hat{\Phi}_{a|b}\\
\dot{\Phi}_b &= \alpha_{b,a|b}\,\hat{\Phi}_{a|b} + \alpha_{b,c|b}\,\hat{\Phi}_{c|b} + \alpha_{b,d|b}\,\hat{\Phi}_{d|b}\\
\dot{\Phi}_c &= \alpha_{c,c|b}\,\hat{\Phi}_{c|b}\\
\dot{\Phi}_d &= \alpha_{d,b|d}\,\hat{\Phi}_{b|d}
\end{aligned}$$

So defining the stream list

$$\mathcal{F} := [a|b, c|b, b|d]$$

then we can write the balance for system b very compactly:

$$\dot{\Phi}_b = \sum_{m} \alpha_{b,m}\,\hat{\Phi}_{m}, \quad m \in \mathcal{F} \tag{3.1}$$

Defining the vectors:

$$\underline{\mathbf{\Phi}} := \left[\Phi_a, \Phi_b, \Phi_c, \Phi_d\right]^T \qquad \underline{\hat{\mathbf{\Phi}}} := \left[\hat{\Phi}_{a|b}, \hat{\Phi}_{c|b}, \hat{\Phi}_{b|d}\right]^T$$

we can write the whole system of equation into a matrix equation:

$$\underline{\dot{\mathbf{\Phi}}} = \underline{\underline{\mathbf{F}}}\,\underline{\hat{\mathbf{\Phi}}}$$

with $\underline{\underline{\mathbf{F}}}$ being the incidence matrix of the directed graph 3.1:

$$\underline{\underline{\mathbf{F}}} := \begin{bmatrix} -1 & 0 & 0\\ +1 & +1 & -1\\ 0 & -1 & 0\\ 0 & 0 & +1 \end{bmatrix}$$

The incidence matrix is easily constructed from the incidence list, which is a list of tuples one for each connection (arc) in the graph. The first value in the tuple is the source node and the second the sink node, thus in our notation the first index and the second index separated by the $|$. The graph is not quite completely specified by the incidence list, as nodes that are not connected are then not included and one needs to add a list of unconnected nodes as a minimal information.

This then also lets us interpret the abstraction of the physical topology:

> **Definition – Physical topology:** The physical containment as a directed graph, in which the nodes are the capacities and the arcs are the connections transferring intensive quantity.

Adding the control units, adds a sub-graph which uses signals as connections and has signal processing units as nodes. The physical topology and the control topology are connected with uni-directional arcs which from the physical topology to the control topology require an observer unit for the state of the observed capacity and on the other side it is a signal that connects from the control topology to the physical topology at a steady-state unit representing a stream-manipulating element, such as a valve.

3.3 Flowsheets vs abstract topologies

Flowsheets are used to schematically represent plants for various purposes. They are used in plant design to reflect the structure of the plant with various levels of detail matching the design purpose. Each of the design levels will usually generate a new flow sheet which will increase in detail and will be more specific as one progresses in the design process. The flow sheet is then used to be extended into technical drawings, which are used for the engineering of the plant. Thereafter the same drawings will be used for maintenance. Another version is used by the operators. Today the latter is an electronic version, which is being displayed on the operators monitors.

Figure 3.2 shows a P&ID diagram of a distillation column. The grey-filled circular components are represent controllers with P for pressure, L for level and T for temperature. The C stands in all cases for controller. The Figure 3.3 shows a possible representation of the same distillation, assuming six stages for the whole column. The feed is on stage 3, from the top. The boiler is controlled over the temperature in the lower part of the column. The reflux is controlled over a flow controller and the two products are controlled over the levels in the distillate drum and the bottom of the column. The topology has the same information as the P&ID diagram, but in addition shows the assumptions made about the dynamics of the individual parts and the nature of the interactions. The models of the two heat exchangers is extremely simple: a lumped system for each of the two sides interacting through the wall by the means of heat exchange.

3.4 Tokens

The basic graph is established on the background knowledge of what each node and each arc thus also represents what the individual parts of the plant contain and exchange. The model is to mimic the model universe in the form of a network. On the high level, we refer to what is moving and changing in the network as *tokens*. So we use the picture of the network to

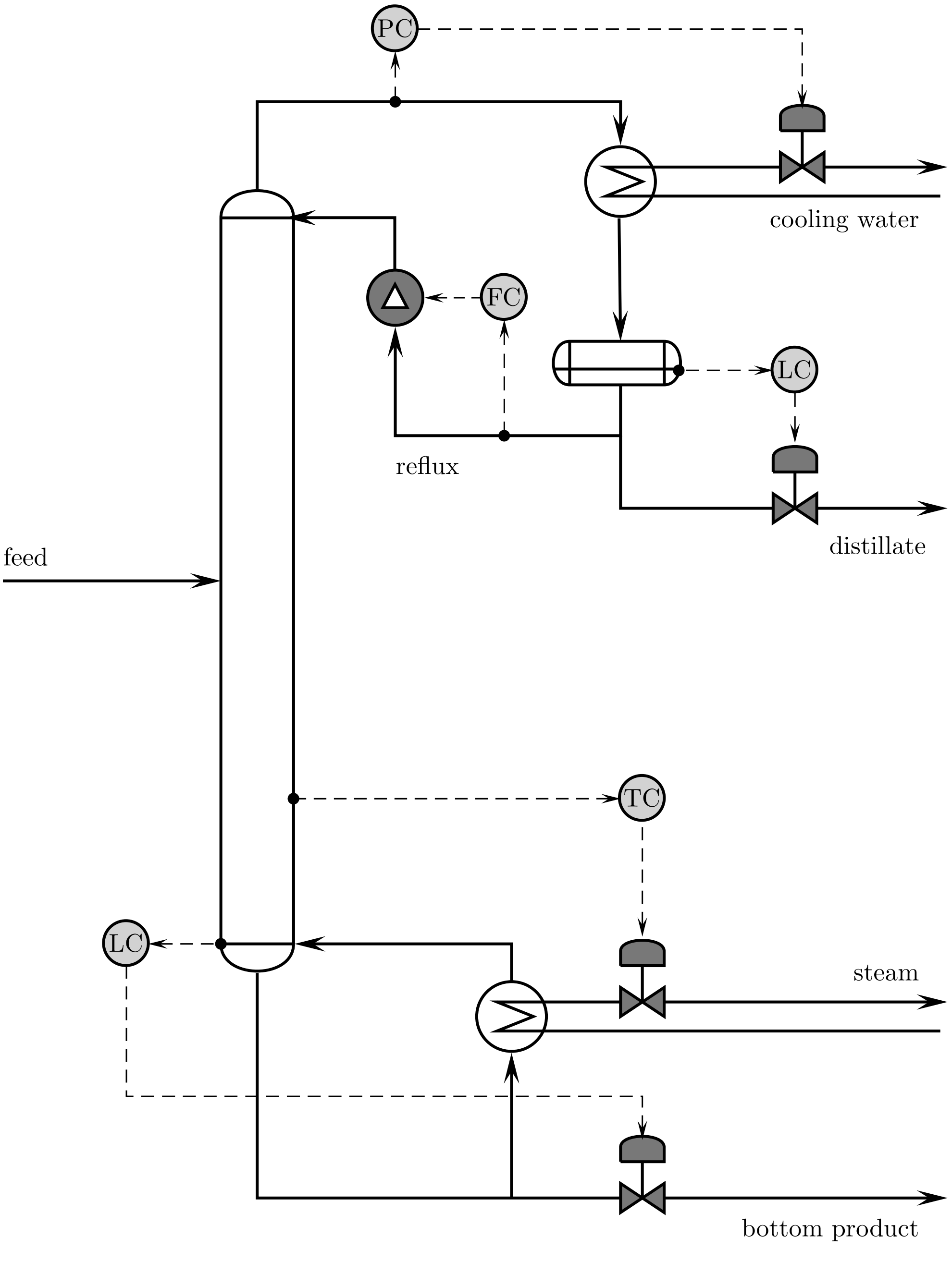


Figure 3.2: A P&ID diagram of a distillation column

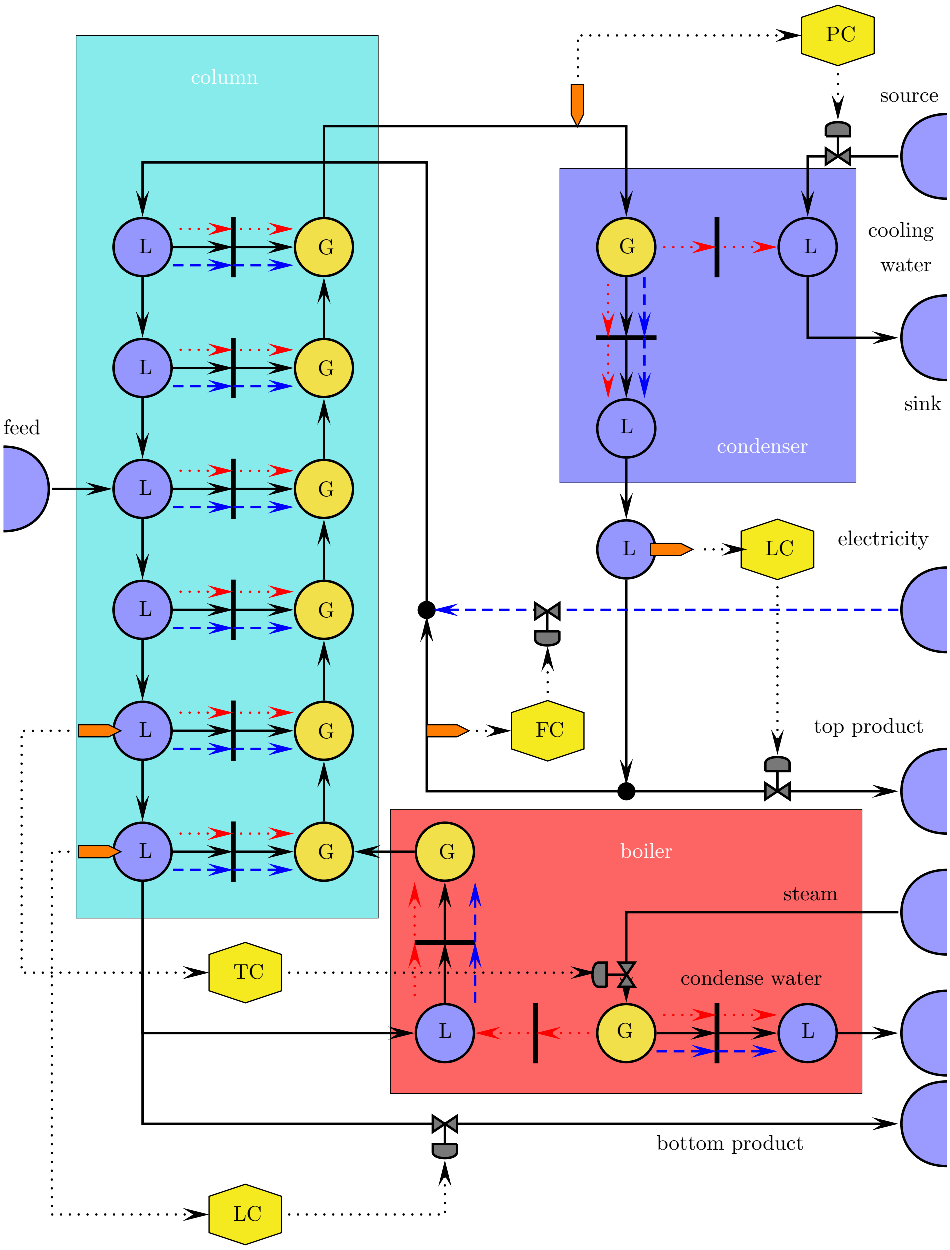


Figure 3.3: A simple topology of a distillation column

accumulate, transfer and modify tokens. Together with the dynamics, these tokens must reflect the behaviour of the contents. So for tokens we choose the base quantities that represent the behaviour of the physical system, namely the conserved quantities for the physical processing part: *mass*, *energy* and *(linear) momentum* to list the main ones. For the information processing part we simply use a token *information*.

3.4.1 Colouring graphs

Tokens can be introduced in different ways. One can inject them into nodes, which most logically will be reservoirs in the environment, after all things have to come from somewhere also at the very beginning. Having it injected in a source node one can ask the question on if there exists a transfer system for this token that is attached to the current node. If the answer is yes and one marks this transfer system for its ability to transfer the said token and transfers the token, it is added to the node that is attached through the same transfer system. This process is continued thereby filling the connected network with the tokens. In graph terms, this defines a sub-graph for the said token. We call this procedure ”colouring algorithm”.

> **Definition – Token domain:** is a connected graph within which the token is being transferred.

The dual approach is to start with a transfer system and define its ability to transfer a token. The two connected nodes must then also contain that token. Defining the ability to transfer which token for all systems also yields the desired information. The first approach of injecting the token into a source node has the advantage of automatically resulting into the token sub-graphs or token sub-networks, whilst from the user’s point of view it seems somewhat artificial and tedious. The second approach does not automatically yield the sub-network information, whilst seemingly being more attractive for the user. By defining which of the connections, represented as arcs, transfers a particular token we also define that this token must be present in the two nodes the arc connects. In this way we can ”fill” the physical topology with the tokens. Whilst this concept is very simple, nature adds a little spice, in that certain tokens induce others. Thus whilst each connection only transfers one particular token, it will also transfer the induced token. The most common example is mass, that induces energy flow, because mass ”carries” energy by its mere existence.

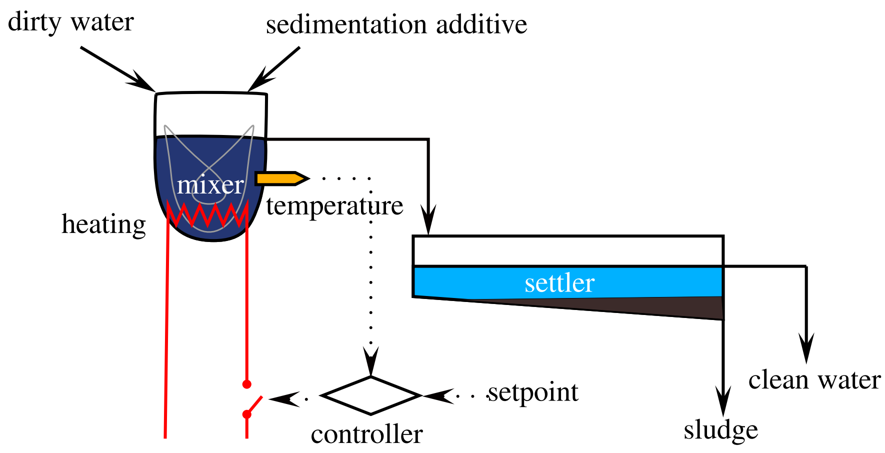


Figure 3.4: A simple sedimentation plant that uses an additive to improve sedimentation

3.5 Network representation of a model universe

We build the network representation of process models from the very beginning up to the level on where we can write the basic conservation equations. We use a simple sedimentation plant for our ”guinea pig”: Dirty water, that is, water and some suspended material is entering a mixed tank, where a sedimentation additive is added. The mixture is then passed on to a settler, where the sediments settle and the clean water is taken from the top. It is assumed that the additive is fixing itself completely to the sediments and that the separation is complete, thus consequently the produced water is pure. (3.4). We further assume that the additive is working best at a particular temperature. Since the temperature of the feed changes and the tank looses heat to the environment, we add a heater to the tank and control it by measuring the temperature.

The first step in the analysis is to define an abstraction of the process. In this case, the definition of the universe is rather straightforward and so is the definition of what is the environment and the plant. Since the quality of the description is here not of relevance, we generate a very low-granular model by assuming only simple lumped systems. We will come back to more complex behaviours later in . The first graph, (Figure 3.5), shows all considered plant components as nodes and the communication paths as arcs. We have decided to have one lump for the liquid in the tank (L), three nodes for the settler, (I) representing the part where the fluid is entering the settler, (T) the clear top water phase and (B) the bottom sludge phase. For the control system we introduced the stirrer (M), the sensor (O), the

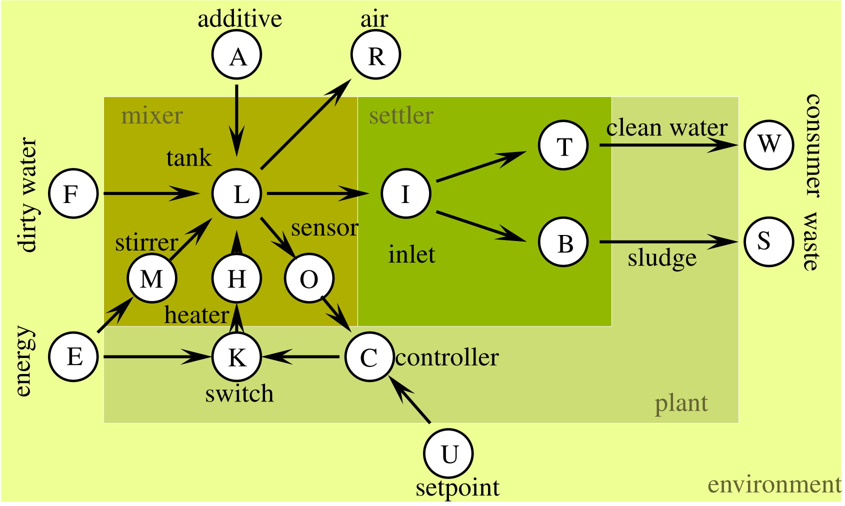


Figure 3.5: The basic underlying graph

controller (C) and the switch (K).

Second we characterise the dynamics of each node, which leads to Figure 3.6. Here we have assumed that all flow systems are lumped and that all elements associated with controlling the plant exhibit event-dynamic behaviour. The assumptions for the fluid systems is that all are lumped systems, whilst for the control-related components all are assumed to exhibit event-dynamic behaviour. The environment consists exclusively of reservoirs both for resources as well as products.

3.5.1 Token - mass

Having the physical topology in place, we start to inject the tokens. We do that by defining for each connection what token it transfers. Each transfer system can only transfer one type of token. In our example we have connection for mass, the convective streams: $F|L$, $A|L$, $L|I$, $I|T$, $I|B$, $T|W$, $B|S$ where we used the notation of source|sink for each connection. The list of tuples, is the incidence list of the arcs that transfer mass. For energy the incidence list is $H|L$, $L|O$, $L|R$, $E|M$, $M|L$, $E|K$, $K|H$. This then defines that mass is present in the nodes F, L, A, I, T, B, W, S and energy in H,L,O,R,M,K,E, F, R and information in O,K,C,U. Figure 3.7 shows the graphs for the three token domains, namely mass, energy and information.

Having defined the mass transfer network, the mass balances are readily established by considering all the streams that cross the surface of each

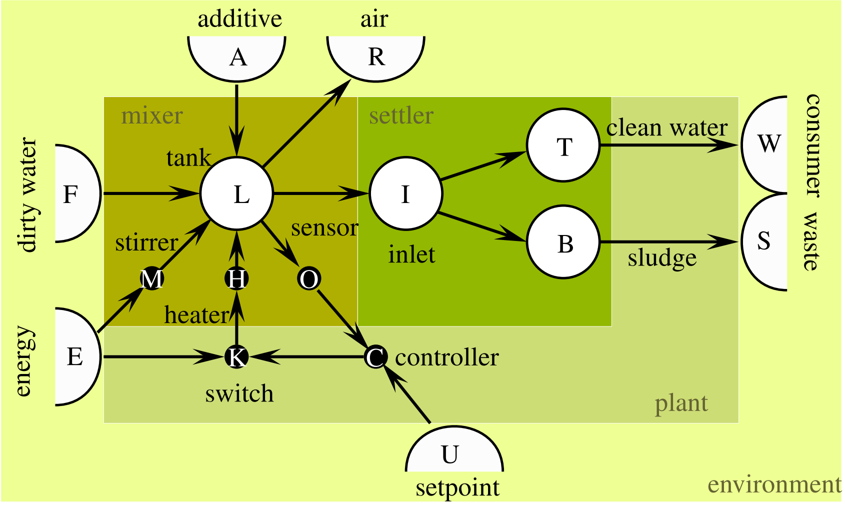


Figure 3.6: The graph with the added dynamic assumptions

control volume. For an arbitrary non-reactive system s we can write:

$$\dot{m}_s := \sum_{\forall m} \alpha_{s,m}\, \hat{m}_m$$

where the $\dot{m}_s$ is the time derivative of the mass of system s with respect to time, thus the accumulation of mass in the control volume V_s. The $\alpha_{s,m}$ represents the reference co-ordinate system for the flow m seen from system s. So for an arrow pointing inwards $\alpha_{s,m} := +1$ whilst for the opposite, namely the arc pointing outwards, it will be $\alpha_{s,m} := -1$. If we take the whole network into consideration, and assume no reactions in the system, we get a matrix equation:

$$\underline{\dot{\mathbf{m}}} := \underline{\underline{\mathbf{F}}}^m\, \underline{\hat{\mathbf{m}}}$$

here the $\underline{\hat{\mathbf{m}}}$ is the vector of masses in the control volumes and $\underline{\underline{\mathbf{F}}}^m$ is the incidence matrix of the directed graph, whilst $\underline{\hat{\mathbf{m}}}$ is the vector of mass flows leaving and entering the control volumes that make up the plant. So if we use the same notation for identifying the flows, the mass flow vector is defined as:

$$\underline{\hat{\mathbf{m}}} := [\hat{m}_{F|L}, \hat{m}_{A|L}, \hat{m}_{L|I}, \hat{m}_{I|B}, \hat{m}_{T|W}, \hat{m}_{B|S}]^T$$

The incidence matrix has as row identifier the systems and for the column

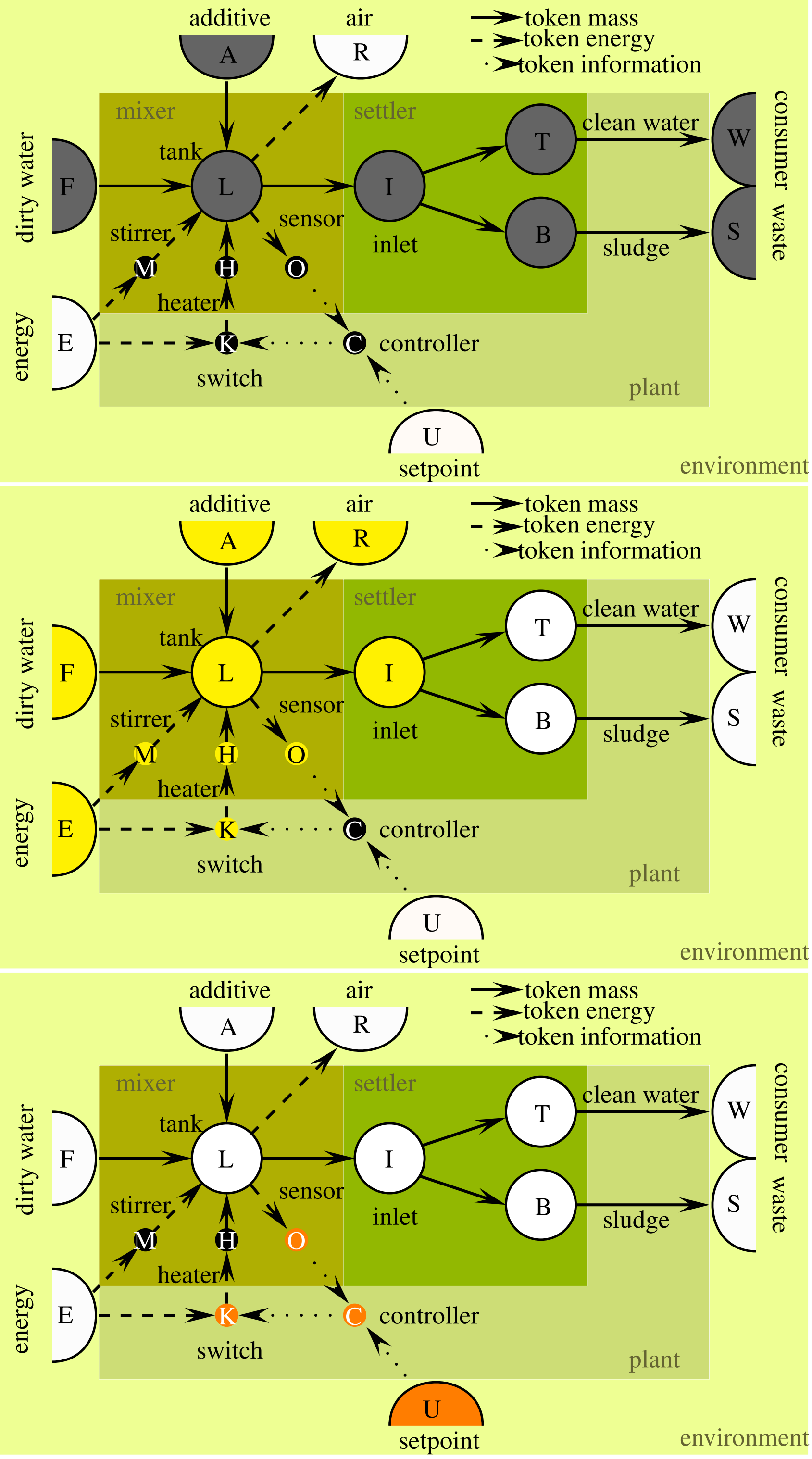


Figure 3.7: The three token domains: mass, energy and information, form top to bottom

identifiers again the duple defining the source and sink of a flow.

	$F\vert L$	$A\vert L$	$L\vert I$	$I\vert T$	$I\vert B$	$T\vert W$	$B\vert S$
L	$+1$	$+1$	-1				
I			$+1$	-1	-1		
T				$+1$		-1	
B					$+1$		-1

Again the top heading is the incidence list for the mass transfer graph, a subgraph of the physical topology. In more complex plants we will have several of such mass transfer systems.

3.5.2 Species

In the next stage, we want to identify what species are present where in the network. This provides us with the information on what species mass balances we must establish. For this we *refine* the definition of mass by considering individual species or reaction-invariant combinations of species. We again use the colouring algorithm (3.4.1) and start with specifying what species are present in the plant and then seek reservoirs that represent sources for the species. Once these species are added to the respective reservoirs, we define where these species will be present in the network.

Unidirectional flows: In general all transfer may go either way - the directionality of the arcs is merely introducing a reference direction. It is though often desirable to limit the flow of mass in one direction by assuming the pressure gradient to point always in this one direction. Introducing such a constraint reduces the number of species balances to be written, but also imposes a certain physical constraint, which not always may be satisfied in reality. So whilst convenient, it must be kept in mind as a simplification, particularly when one deals with safety and hazard problems.

Species water : Let us have a closer look at the species water. It has its natural source in the dirty water supply. In a large network one would find the way backwards along the unidirectional flows from an arbitrary node. Changing the initial node all possible sources can be identified. As said in this case the source is quite obvious. The water is then ”flowing” from the source to the mixer, into the settler and out into the two connected reservoirs. The sequence, we call colouring is shown for water in Figure 3.8

Species additive and dirt : So we repeat the procedure for all species in the system. In this case it is dirt and additive. We assume that the

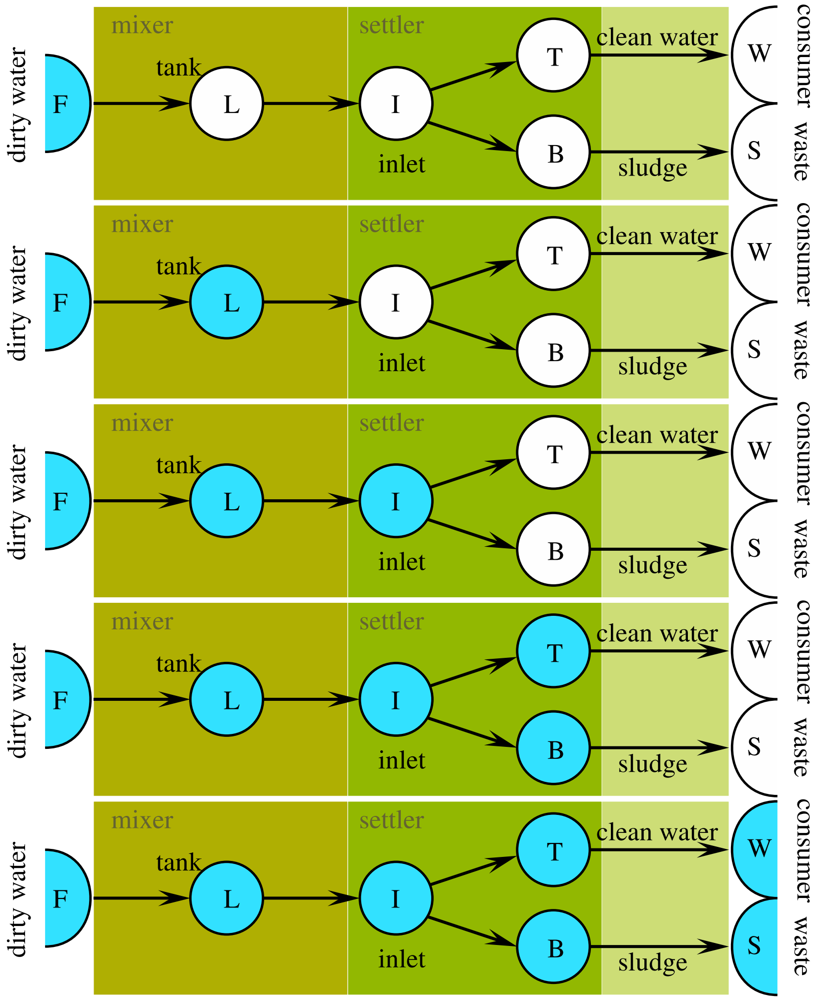


Figure 3.8: Injecting the species water in the dirty water supply reservoir

additive is completely attaching to the solid in the sludge so it does not appear in the product stream, so the product is pure water.

Species topologies : The result is a set of species topologies which are coloured versions of the physical topology. This makes each species topology a layer on top of the physical topology. If there is reaction, then there also exist a link between the layers in the capacities where reactions are taking place, with the reactions representing the link between the layers. Whilst when injecting the generic mass token into the physical topology the property of unidirectional mass transfer is not applied, observing the species distribution is essential for the limited directionality of the flow.

For each species we can draw up a species topology and a corresponding set of conservation equations. For the additive, species A, this then is:

$$\dot{\underline{\mathbf{n}}}_A := \underline{\underline{\mathbf{F}}}^m{}_A \, \underline{\mathbf{n}}_A$$

If we define the flow vectors to have the dimension of the species present in the plant, then the matrix $\underline{\underline{\mathbf{F}}}^m{}_A$ is the same as we defined above, $\underline{\underline{\mathbf{F}}}^m$.

If we write all species balances and wrap them into a model for the species then we will write the balance:

$$\dot{\underline{\mathbf{n}}} := \underline{\underline{\mathbf{F}}}^m \, \underline{\mathbf{n}}$$

where we now have the stack of vectors:

$$\underline{\mathbf{n}} := \begin{bmatrix} \underline{\mathbf{n}}_L^T & \underline{\mathbf{n}}_I^T & \underline{\mathbf{n}}_T^T & \underline{\mathbf{n}}_B^T \end{bmatrix}^T$$

	$F\|L$	$A\|L$	$L\|I$	$I\|T$	$I\|B$	$T\|W$	$B\|S$
L	$+\underline{\underline{\mathbf{I}}}_{3,2}$	$+\underline{\underline{\mathbf{I}}}_{3,2}$	$-\underline{\underline{\mathbf{I}}}_{3,3}$				
I			$+\underline{\underline{\mathbf{I}}}_{3,3}$	$-\underline{\underline{\mathbf{I}}}_{3,1}$	$-\underline{\underline{\mathbf{I}}}_{3,3}$		
T				$+\underline{\underline{\mathbf{I}}}_{1,1}$		$-\underline{\underline{\mathbf{I}}}_{1,1}$	
B					$+\underline{\underline{\mathbf{I}}}_{3,3}$		$-\underline{\underline{\mathbf{I}}}_{3,3}$

The component mass balances are thus block-matrix/vector equations. Here we have chosen to define equal length species vectors for each flow. So the relation between $\underline{\underline{\mathbf{F}}}^n$ and $\underline{\underline{\mathbf{F}}}^m$ is:

$$\underline{\underline{\mathbf{F}}}^n := \underline{\underline{\mathbf{F}}}^m \otimes \underline{\underline{\mathbf{I}}}_3$$

where ⊗ represents the block-wise-Kronecker product, known as Khatri-Rao product. It is possible to use a minimal definition, namely to have only those species in the vectors that are present (Preisig (2010)).

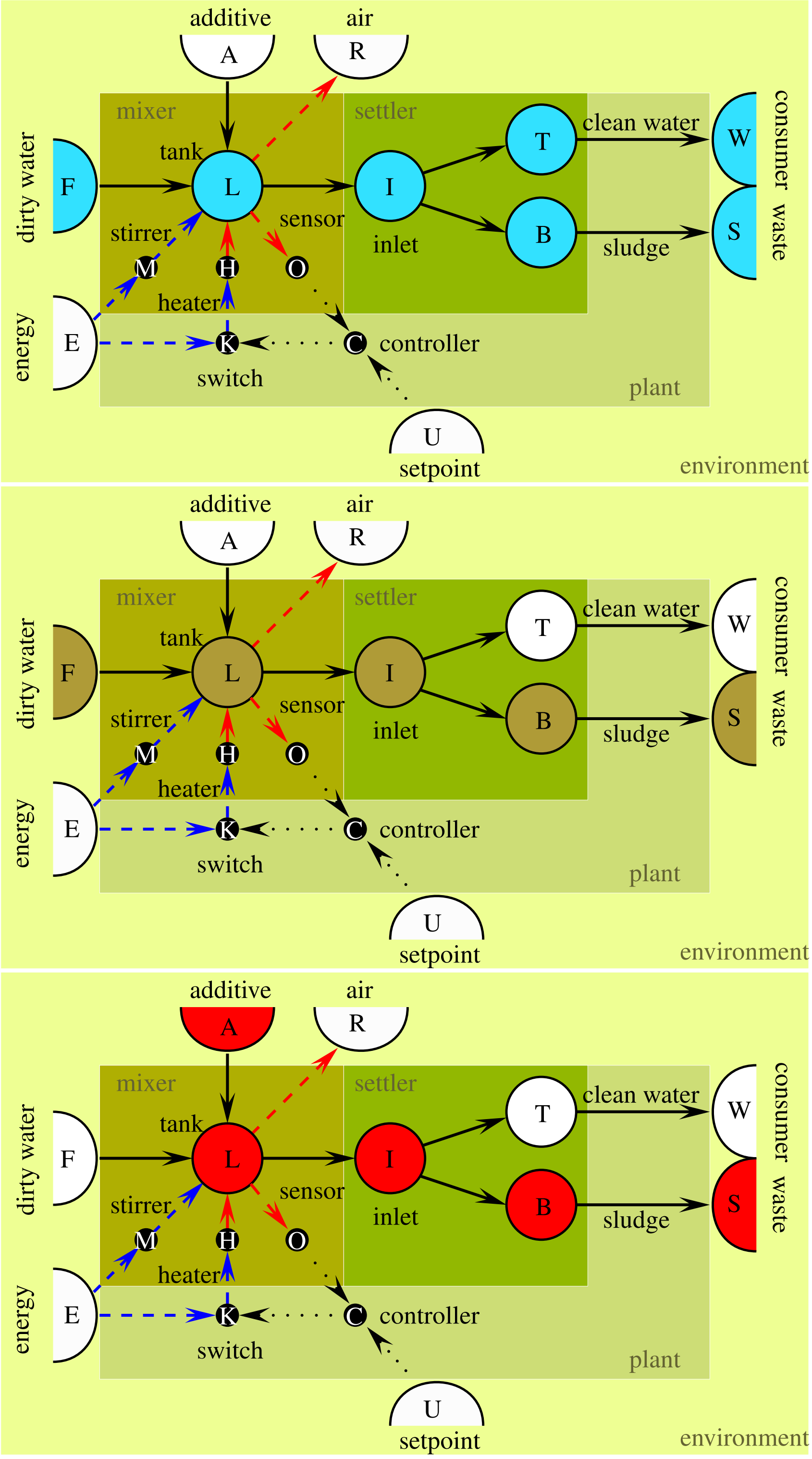


Figure 3.9: The three species topologies: water, sediments and additives

3.5.3 Token - energy

Energy is one of the fundamental tokens that we consider in our process model. Introducing the token, as suggested above, yields the energy transfer domain as shown in the middle picture of Figure 3.10. The domain was identified by assigning the token energy to be present on either side of an arc that transfers energy in the form of heat, radiation or any form of work, besides the streams that induce energy, namely mass. The topology defines for which part of the overall network the energy balances are to be established.

In Chapter 9 we shall discuss in more detail what knowledge is required for the energy balance. Besides the different forms of energy, it will also about the different forms of energy transfer, which does not only include the heat streams and the work stream, but also the mass streams. Mass carries internal energy, besides that is moving, which adds kinetic energy and potential energy, latter induced by the gravitational field. Mass has a finite density and thus a finite volume. Moving mass across a system's boundary thus implies a flow of volume work: mass is ”injected” or ”ejected”. This adds volume work to the balance. All other forms of energy transfer are to be added too, of which the main ones are heat and mechanical work. Also the volume of the system itself may shrink or expand as the result of the various flows crossing the boundary, which implies that the system does positive volume work if it is expanding or is subject to volume work if it is shrinking. Since mass induces energy, one can have mass balances in isolation but not energy balances in system that exchange mass.

Unidirectional flows : discussing the species topologies, we introduced a concept of ”unidirectional” flows for the reasons of reducing the number of equations though with the costs of constraining the applicability of the model to conditions where indeed the flows are going in the pre-defined direction. If we apply this concept also to the energy balance, we find that whilst we have to consider the mass flow inflows, we can ignore the outflows in terms of dissipating the token energy. Consequence being that we can remove the energy balance for the inlet of the settler in our case. To make things more visible we also marked the two mass streams relevant for the energy balance of the liquid in the tank (L) with a yellow shadow.

So energy must account for several flows, namely mass, heat and radiation and any type of work that is exerted on the system. To make these facts visible, we detail the energy flows by specifying their specific nature, i.e. heat, work and even more detail if desirable. Examples for the latter can be the type of work like friction, volume work. So this results in a detailed

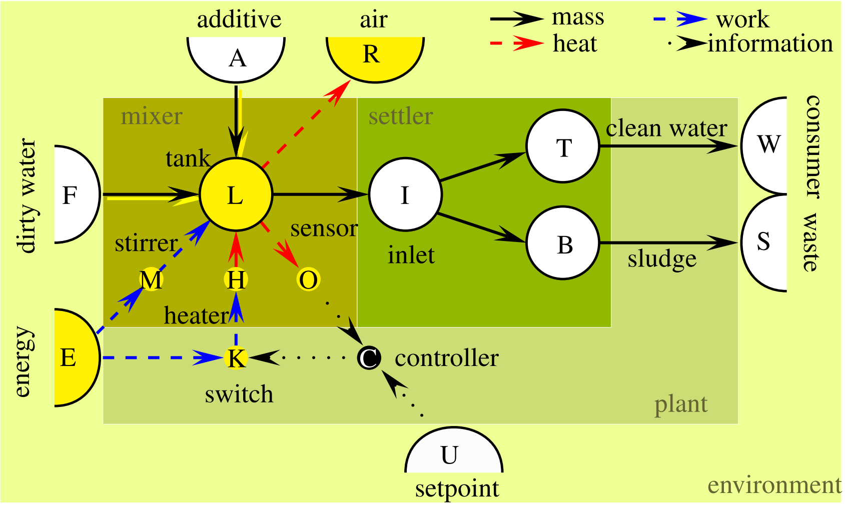


Figure 3.10: Energy topology: where to write the energy balance

topology for reflecting the energy household for the plant Figure 3.10. The detailed colouring of the graph is instrumental for the formulation of the energy balances. Denoting total energy with E for an arbitrary system s, we get the balance

$$\frac{d\,E_s}{d\,t} := \underline{\underline{\mathbf{F}}}^m{}_s\,\underline{\hat{\mathbf{E}}} + \underline{\underline{\mathbf{F}}}^q{}_s\,\underline{\hat{\mathbf{q}}} + \underline{\underline{\mathbf{F}}}^w{}_s\,\underline{\hat{\mathbf{w}}}$$

The mass flow matrix $\underline{\underline{\mathbf{F}}}^m{}_s$ is the appropriate submatrix of the above defined mass flow matrix $\underline{\underline{\mathbf{F}}}^m$, thus:

	$F\vert L$	$A\vert L$	$L\vert I$	$I\vert T$	$I\vert B$	$T\vert W$	$B\vert S$
L	+1	+1	−1				
M							
K							
H							
O							

The heat flow network matrix $\underline{\underline{\mathbf{F}}}^q$ is for our case:

	$H\vert L$	$L\vert O$	$L\vert R$
L	+1	−1	−1
M			
K			
H	−1		
O		+1	

and the work flow matrix is:

	$E\|M$	$M\|L$	$E\|K$	$K\|H$
L		$+1$		
M	$+1$	-1		
K			$+1$	-1
H				$+1$
O				

4

Capacities – Balances are the core

Synopsis *The behaviour of the individual communicating capacities form the core of the model. In general the capacities are of distributed nature where the volume-normed conserved quantity is, besides time, also a function of the position, the spatial coordinates. In contrast, lumped systems are only a function of time.*

4.1 A global system's view

The exchange of extensive quantity between the plant and its embedding environment induces changes both in the plant and in the environment. Since the environment is constant, the change is measured in terms of extensive quantities of the plant by accounting for all conserved extensive quantities in each subsystem of the plant. The basic accounting rule is simple:

accumulation:: change of extensive quantity per unit time
=
flow of extensive quantity (in - out) per unit time

So for each subsystem the rule states that: *what is coming in on extensive quantity must either go out or it is accumulated inside.*

This basic system behaviour can be nicely derived by analysing a system before and after an applied change. In 4.1 the green balance surface includes the system and a small volume. On the left the small volume is being added, whilst on the right a small volume, not necessarily the same, is extracted. We analyse the system by assuming that the balance equations hold for the quantity being associated with the different volumes. Being interested in the continuous behaviour we attempt to find the behaviour of the overall system as the time changes. Let Φ be a generic conserved quantity. The

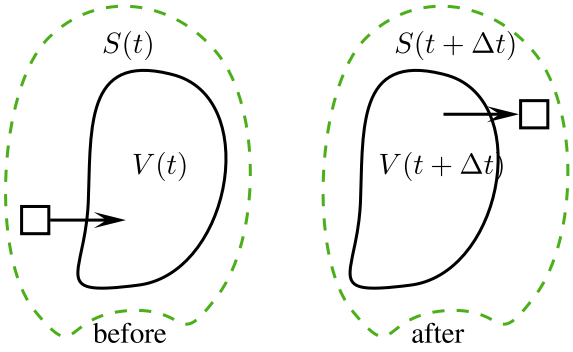


Figure 4.1: Deriving a system's behaviour

change of the system with time we can represent as:

$$\begin{aligned}
\frac{d\,\Phi_S}{d\,t} &:= \lim_{\Delta t \to 0} \frac{\Phi_S(t+\Delta t) - \Phi_S(t)}{\Delta t} \\
&:= \lim_{\Delta t \to 0} \frac{(\Phi_V(t+\Delta t) + \Phi_{V_{out}}) - (\Phi_V(t) + \Phi_{V_{in}})}{\Delta t} \\
&:= \lim_{\Delta t \to 0} \frac{(\Phi_V(t+\Delta t) - \Phi_V(t)) + (\Phi_{V_{out}} - \Phi_{V_{in}})}{\Delta t} \\
&:= \lim_{\Delta t \to 0} \left(\frac{\Phi_V(t+\Delta t) - \Phi_V(t)}{\Delta t} + \frac{\Phi_{V_{out}}}{\Delta t} - \frac{\Phi_{V_{in}}}{\Delta t} \right) \\
&:= \frac{d\,\Phi_V}{d\,t} + \hat{\Phi}_{out} - \hat{\Phi}_{in} \\
&:= 0
\end{aligned}$$

Which gives us the differential balance law:

$$\frac{d\,\Phi_V}{d\,t} = \hat{\Phi}_{in} - \hat{\Phi}_{out} \tag{4.1}$$

The transfer of extensive quantity may occur in different form due to the different morphologicy the extensive quantity can assume. For example for energy transfer occurs in the form of mass transfer in the form of internal energy, or we may transfer energy in the form of radiation, conductive heat diffusion, mechanical work flow or the like.

Balancing gets slightly more involved when the balanced extensive quantity can transpose into another extensive quantity. Common transpositions are due to interaction of chemical or biological species changing into something else, which usually is referred to as chemical or biological reaction (see 8.2). Moreover mechanical work can transpose into heat and or electrical energy into heat, etc. Also a species may change phase taking up or releasing some energy in one or the other form. These transpositions link different conservation laws but do never affect the observation that the basic conserved quantities do satisfy the conservation principle, thus mass, energy, linear and rotational momentum is always conserved.

4.2 Zooming into the distributed description

In the above analysis we kind of assumed that the different parts of the plant are lumped and thus they are characterised by uniform intensive quantities. In nature that is usually not the case and things are distributed in all spatial dimensions, which though does not inhibit us to formulate the conservation principle over a control volume: It simply states, in mathematical form, the fact that the net flow across the boundary is compensated by the accumulation of the balanced extensive quantity inside the system. So defining a vectorial extensive quantity $\underline{\boldsymbol{\Phi}}$ and a net flow across the boundary with $\hat{\underline{\boldsymbol{\Phi}}}$ we can write:

$$\dot{\underline{\boldsymbol{\Phi}}} = \hat{\underline{\boldsymbol{\Phi}}} \tag{4.2}$$

The extensive quantity changing with time and spatial co-ordinate $\underline{\mathbf{r}} := \begin{bmatrix} r_x & r_y & r_z \end{bmatrix}^T$. is a point property, which in contrast to the extensive quantity we signify as $\underline{\varphi}$. the accumulation term is the integral of the point property over the volume:

$$\underline{\boldsymbol{\Phi}}(t) := \int_0^{V(t)} \underline{\varphi}(V;t)\, dV = \iiint_{\underline{\mathbf{r}}} \underline{\varphi}(\underline{\mathbf{r}};t)\, d\underline{\mathbf{r}}$$

which gives us an expression for the extensive quantity and not yet its time derivative. The result can be obtained by applying the generalised Leibniz rule (see B.1), which essentially is fluid mechanics' Reynolds theorem:

$$\dot{\underline{\boldsymbol{\Phi}}} := \frac{d}{dt} \int_0^{V(\underline{\mathbf{r}};t)} \underline{\varphi}(V;t)\, dV := \Big(\underline{\varphi}(\underline{\mathbf{r}};t)\, \dot{V}(\underline{\mathbf{r}};t) \Big)_S - 0 + \int_V \frac{\partial\, \underline{\varphi}(V;t)}{\partial\, t}\, dV$$

The first non-zero term on the left is the change of size of the system, which is the integral over the boundary S of the density and the volume change. The volume change on its own would be:

$$\dot{V}(\underline{\mathbf{r}};t)_S := \int_S \underline{\mathbf{v}}^T\, \underline{\mathbf{n}}\, dS$$

and the associated change of the accumulation term:

$$\Big(\underline{\varphi}(\underline{\mathbf{r}};t)\, \dot{V}(\underline{\mathbf{r}};t) \Big)_S := - \int_S \underline{\varphi}(S;t)\, \underline{\mathbf{v}}^T\, \underline{\mathbf{n}}\, dS$$

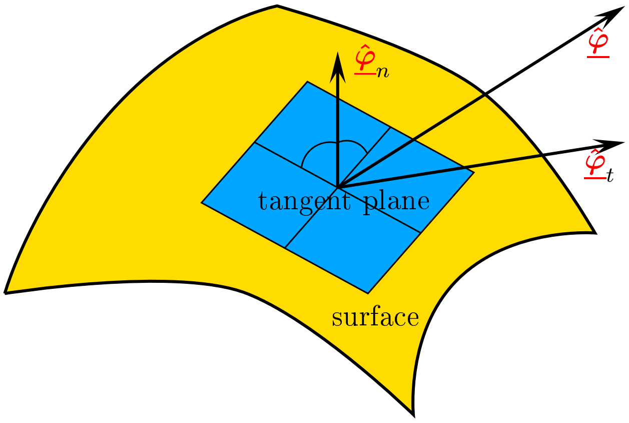

$\hat{\underline{\varphi}}_n$
$\hat{\underline{\varphi}}$
$\hat{\underline{\varphi}}_t$
tangent plane
surface

Figure 4.2: The flow vector across the red surface $\hat{\underline{\varphi}}$ is projected onto the outwards-pointing normal direction $\hat{\underline{\varphi}}_n$ which is the part making a contribution to the flow across the boundary, as the remaining part is projected onto the tangential plane $\hat{\underline{\varphi}}_t$, which makes no contribution to the flow across the boundary

Dropping the variable lists for readability, substitution yields:

$$\frac{d}{dt}\int_0^{V(t)} \underline{\varphi}(V;t)\,dV := -\int_S \underline{\varphi}\,\underline{\mathbf{v}}^T\,\underline{\mathbf{n}}\,dS + \int_V \frac{\partial\,\underline{\varphi}}{\partial\,t}\,dV$$

Now let us focus on the right-hand side of the balance, namely $\hat{\underline{\mathbf{n}}}$: The flow across the boundary is the integral of the local flow over the surface of the volume. Since the surface is of arbitrary geometry and not just a cube this integration requires a little bit more effort. Let us have a look at a piece of the surface.

The projection of the flow onto the normal direction is the dot product of the vector with the normal vector $\underline{\mathbf{n}}$ or, which is identical: the scalar product:

$$\hat{\underline{\varphi}}_n := \hat{\underline{\varphi}} \cdot \underline{\mathbf{n}} := \hat{\underline{\varphi}}^T\,\underline{\mathbf{n}}$$

So if the surface is not moving, the net flow of the extensive quantity across the surface S is the integral:

$$\hat{\underline{\mathbf{\Phi}}} := -\int_S \hat{\underline{\varphi}}^T\,\underline{\mathbf{n}}\,dS$$

whereby special attention is to be paid to the direction: the normal vector points outwards, which is the opposite of the system's egocentric view requiring a change of sign of the resulting flow. If the boundary is moving with a velocity $\underline{\mathbf{v}}$, then a second term must be added which accounts for the increase in volume, again it is the normal direction that makes the contribution and the sign must be negative to adhere to the system-egoistic point of view:

$$\underline{\hat{\boldsymbol{\Phi}}} := -\int_S \underline{\hat{\varphi}}^T \, \underline{\mathbf{n}} \, dS - \int_S \underline{\varphi} \, \underline{\mathbf{v}}^T \, \underline{\mathbf{n}} \, dS$$

So the balance would then read:

$$-\int_S \underline{\varphi} \, \underline{\mathbf{v}}^T \, \underline{\mathbf{n}} \, dS + \int_V \frac{\partial \, \underline{\varphi}}{\partial \, t} \, dV = -\int_S \underline{\hat{\varphi}}^T \, \underline{\mathbf{n}} \, dS - \int_S \underline{\varphi} \, \underline{\mathbf{v}}^T \, \underline{\mathbf{n}} \, dS$$

which simplifies to:

$$\int_V \frac{\partial \, \underline{\varphi}}{\partial \, t} \, dV = -\int_S \underline{\hat{\varphi}}^T \, \underline{\mathbf{n}} \, dS$$

Applying Gauss' divergence theorem transmogrifies the surface integral into a volume integral:

$$\int_V \frac{\partial \, \underline{\varphi}}{\partial \, t} \, dV = -\int_V \frac{\partial}{\partial \, \underline{\mathbf{r}}} \, \underline{\hat{\varphi}} \, dV \tag{4.3}$$

which implies that:

$$\frac{\partial \, \underline{\varphi}}{\partial \, t} = -\frac{\partial}{\partial \, \underline{\mathbf{r}}} \, \underline{\hat{\varphi}} \tag{4.4}$$

Substitution of the linear-in-conductivity, isotropic linear-in-gradient-driven transport law gives:

$$\begin{aligned}
\frac{\partial \, \underline{\varphi}}{\partial \, t} &= \frac{\partial}{\partial \, \underline{\mathbf{r}}} \, \underline{\underline{\mathbf{C}}} \, \frac{\partial \, \underline{\boldsymbol{\pi}}}{\partial \, \underline{\mathbf{r}}} \\
&= \underline{\underline{\mathbf{C}}} \, \frac{\partial}{\partial \, \underline{\mathbf{r}}} \, \frac{\partial \, \underline{\boldsymbol{\pi}}}{\partial \, \underline{\mathbf{r}}}
\end{aligned} \tag{4.5}$$

which is a second-order partial differential equation describing the transport of extensive quantity as a function of the second-order derivative of the effort variable, which for mass is the chemical potential. This model is thus the second law of Fick but with the effort variable being the chemical potential instead of the traditionally used concentration.

4.3 An alternative quicky one

The same result can also be found by continuing the reasoning on 4.1: we take another limit, but now to a zero volume:

$$\begin{aligned}
\lim_{\underline{\mathbf{r}} \to 0} \frac{d \, \Phi_V}{d \, t} &= \lim_{\underline{\mathbf{r}} \to 0} \left(\hat{\Phi}_{in} - \hat{\Phi}_{out} \right) \\
\frac{\partial \, \Phi_V}{\partial \, t} &= -\frac{\partial}{\partial \, \underline{\mathbf{r}}} \, \hat{\Phi}
\end{aligned}$$

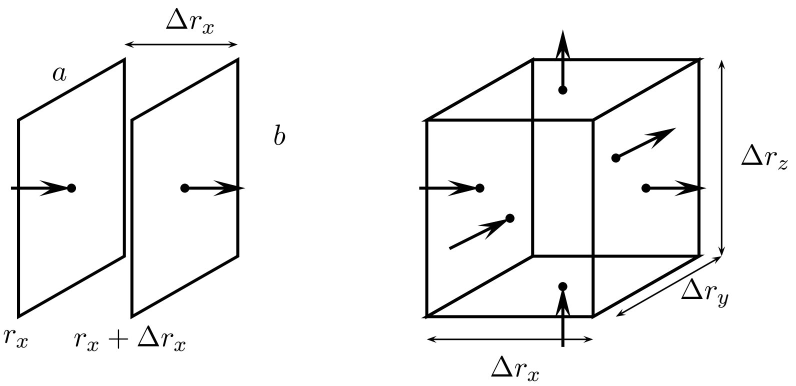


Figure 4.3: A 1D and a 3D unit zell example

This is kind of quick and dirty because no consideration has been given to the fact that the process could change shape.

4.4 Yet another approach

The derivation of the same equation using a shell balance can be nearly seen as the chemical engineering standard. The idea is to balance over a small but finite dimensional unit cell of the system being described. The "unit cell" is thereby taken as being representative for the whole of the process in terms of geometric distribution (4.3). Taking the most simple case as a demonstration, the conservation of an extensive quantity is drawn up over the volume enclosed between r_x and $r_x + \Delta r_x$ assuming the dimensions in r_y and r_z are given as a and b respectively. Then volume piece being balanced is $\Delta V = \Delta r_x \, a \, b$ and the surfaces are $a \, b$ through which the flow is passing. Defining the conserved extensive quantity Φ and the flux $\hat{\varphi}$ then the balance reads:

$$\frac{d\,\Phi}{d\,t} = a\,b\,\left(\hat{\varphi}_{r_x} - \hat{\varphi}_{r_x+\Delta r_x}\right)$$

The flux at the second position can be approximated by a first variation which is the Taylor series truncated after the linear term:

$$\hat{\varphi}_{r_x+\Delta r_x} \approx \hat{\varphi}_{r_x} + \left(\frac{\partial\,\hat{\varphi}_{r_x}}{\partial\,r_x}\right)_{r_x} \Delta r_x$$

Substitution gives:

$$\frac{d\,\Phi}{d\,t} = a\,b\,\left(\hat{\varphi}_{r_x} - \left(\hat{\varphi}_{r_x} + \left(\frac{\partial\,\hat{\varphi}_{r_x}}{\partial\,r_x}\right)_{r_x} \Delta r_x\right)\right)$$

simplifying to:

$$\frac{d\,\Phi}{d\,t} = -\Delta V \left(\frac{\partial\,\hat{\varphi}_{r_x}}{\partial\,r_x} \right)_{r_x}$$

with ΔV being the volume of the unit cell. Taking the limit of $\Delta V \to 0$ again gives for the differential volume the partial differential equation in the density:

$$\lim_{\Delta V \to 0} \frac{d\,\Phi/\Delta V}{d\,t} := \frac{\partial\,\varphi}{\partial\,t} = -\frac{\partial\,\hat{\varphi}}{\partial\,\underline{\mathbf{r}}}$$

whereby φ is a density of the Φ.

4.5 "Lumpy" Boundaries

When fractioning the overall volume into smaller volumes, one generates surface elements that separate adjacent systems. In most cases, one is not interested in the flux, but rather in the total flow across such a surface element, which is one reason for which one "lumps" the boundary. Secondly, one may have more than one type of interaction between two adjacent systems, for example there may be a heat flow through a non-porous physical wall and flow through an opening in the same physical wall, which allows the two systems to interact via heat transfer through the wall and mass transfer through the hole. The lumping thus primarily splits the boundary into local boundary elements that may be classified with regard to the type of extensive quantity being transferred, a concept that is directly coupled to the typed thermodynamic walls (open, closed, adiabatic, etc.).

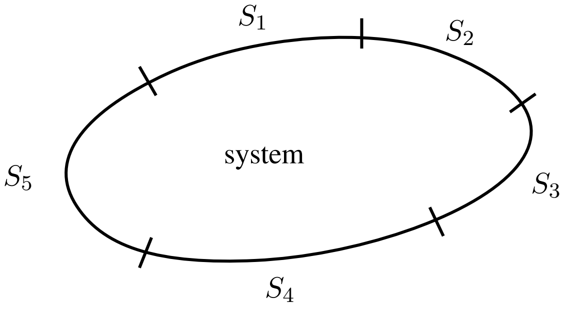


Figure 4.4: System with "lumpy boundary"

The cumulative flow through a piece of boundary is simply the integral over the respective boundary element S_i :

$$\underline{\hat{\mathbf{\Phi}}}_{S_i} := \int_{S_i} \underline{\hat{\phi}}^T\, \underline{\mathbf{n}}\, dS\,,$$

This integral measures the flow in the direction relative to the normal vector of the boundary, where by convention the normal vector points away from the boundary. In the abstraction process, the systems are pictorially pulled apart and represented as circles, or other graphical objects depending on the type of system (Figure 2.8).

The flow through the common piece of boundary between two systems is mapped into a *connection*, which introduces a unique coordinate system against which the actual flow between the connected systems is measured 5.2. This information is captured in a notation $< a > |S_i| < b >$. $< a >$ is the place holder for the system where the origin of the reference co-ordinate system is located, whilst the place holder $< b >$ is denoting system at the other end, the sink is located. The name of the common boundary piece S_i is placed between two vertical bars on either side guarded by the two systems. The reference co-ordinate, being introduced for each connection, is denoted by $\alpha \in \{-1, 0, +1\}$ where the "+1" indicates a head of a connection arrow, a "−1" a respective tail and a "0" no connection. Obviously, a flow must always be defined between two systems, that is, flow may not just disappear or appear into or from the void. The sum of the control volumes is thus always closed representing the process-relevant.

The integral balance equation for a system with stationary boundaries, that is $\underline{\mathbf{v}}_S := \underline{\mathbf{0}}$ reads more compactly when lumping the flows for the boundary elements:

$$\underline{\dot{\mathbf{\Phi}}}_S := \sum_{\forall c} \alpha_{S,c} \, \underline{\hat{\mathbf{\Phi}}}_c$$

We can cast this into a vector equation by using the network representation (9.1):

$$\underline{\dot{\mathbf{\Phi}}}_S := \underline{\underline{\mathbf{F}}}_S \, \underline{\hat{\mathbf{\Phi}}}_s \,, \tag{4.6}$$

with

$$\underline{\underline{\mathbf{F}}}_S := \left[\left[\alpha_{S,c} \, \underline{\underline{\mathbf{I}}} \right]_{\forall c} \right]_{\forall S} \,,$$

a block diagonal matrix with identity blocks weighted with the respective reference coordinate and

$$\underline{\hat{\mathbf{\Phi}}}_s := \left[\underline{\hat{\mathbf{\Phi}}}_{S,c} \right]_{\forall c} \,,$$

a stack of all flow vectors. The row and the column sums of the connection matrix are zero.

4.6 Lumped systems

Lumped systems are assuming fast internal transfer, which we will derive in details in 14. For now we assume without showing the details that this results in the intensive properties to be uniform in the space the systems occupies. The representation for non-reactive system is thus again:

$$\underline{\dot{\mathbf{\Phi}}}_S := \underline{\underline{\mathbf{F}}}_S \, \underline{\hat{\mathbf{\Phi}}}_s \, . \tag{4.7}$$

5

What makes it move

Synopsis *What drives the exchange, why are things "dynamic". We introduce the main driving forces in space, being the differences of temperature, pressure and chemical potentials. A sense of directionality is given to each connection.*

If we couple two systems together by establishing a connection that enables the transfer of extensive quantity, say for example mass or heat, the two systems will likely not be in equilibrium and will experience an exchange of extensive quantity until an equilibrium is achieved. For example opening a valve blocking a tube connection between two tanks containing a liquid, will result in the exchange of this fluid until the two levels are the same and thus a balance is achieved between the two pressures on both sides of the pipe.

5.1 Effort & driving force

Nature tends to level and minimise energy and to maximise entropy. The abstract graph representing the overall structure of the process models introduces capacities and transfer as the two main components for the physical part of the plant. The graph serves as the basis for assembling the mathematical model of the process providing mathematical models for the various capacities and the transfers. Whilst the fundamental functions are used to describe the state of capacities, specifically the internal energy and entropy, the connections are described by the transfer equations for the fundamental conserved quantities. Choosing the energy representation, the internal energy is a function of its canonical variables namely entropy, volume and species mass. The change of internal energy is thus

$$dU = \left(\frac{\partial U}{\partial S}\right)_{V,\underline{\mathbf{n}}} dS + \left(\frac{\partial U}{\partial V}\right)_{S,\underline{\mathbf{n}}} dV + \left(\frac{\partial U}{\partial \underline{\mathbf{n}}^T}\right)_{S,V} d\underline{\mathbf{n}}$$

The change in the extensive quantities is driven by the intensive quantities:

$$
\begin{aligned}
\left(\frac{\partial U}{\partial S}\right)_{V,\underline{\mathbf{n}}} &=: T \\
\left(\frac{\partial U}{\partial V}\right)_{S,\underline{\mathbf{n}}} &=: -p \\
\left(\frac{\partial U}{\partial \underline{\mathbf{n}}^T}\right)_{S,V} &:= \mu
\end{aligned}
$$

which is a proposition going far back in the history of science. Gibbs introduced vector analysis based on the theories of the two mathematicians William Rowan Hamilton and Hermann Grassmann. He defined the thermodynamic equilibrium by maximising a Hamiltonian. Today, the Hamiltonian based approaches lead the way towards a more advanced description of open thermodynamic systems (Grmela and Öttinger (1997); Öttinger and Grmela (1997); Jongschaap and Öttinger (2004); Mrugala (2000); Favache (2009); Favache et al. (2010); Rajeev (2008)). The approach assumes a continuous behaviour in a continuous spatial domain. Thus classical field theory concepts apply making the transport of entropy S, Volume V and species mass $\underline{\mathbf{n}}$ a function of the gradient of the above conjugate variables, namely temperature T, pressure p and chemical potential μ. The intensive variables temperature, pressure and chemical potential thus take a very special role in the model framework in that the gradient, or in a discrete environment the difference is the driving force for the transport of extensive quantity. Bond graph theory introduces the term *effort variables* for them Breedveld (1984) and flow for the respective conjugates. It is thus the difference in the effort variables, or in the continuous domain, the gradient that drives the exchange of the conjugate extensive quantity.

5.2 Transport laws

5.2.1 A conceptual derivation

Transfer laws are most commonly introduced as *constitutive equations* based on empirical considerations. Kinetic theory, though provides a more analytical entrance to the fact that it is the gradient of the effort variables that drives the flow of extensive quantities Chen (2005). The simple analysis applies to anything that can be viewed as a type of gas, particles that move freely in a space. The conceptional thought of the derivation is that the particles carry energy and only exchange energy when they interact with each other or another object like the wall of the containment. It also assumes that the main part of the energy is translational energy and that rotation, oscillations etc are not significant as this is the case for mono-atomic gases. In such a "gas", the particles move more or less randomly and the average velocity over all particles is zero. On the micro scale, the carriers move the free path length before they interact. So for the analysis, we look at a imaginary surface and account for the carriers that pass through the surface within the time-span it takes to cover the free path length (Figure 5.1). We

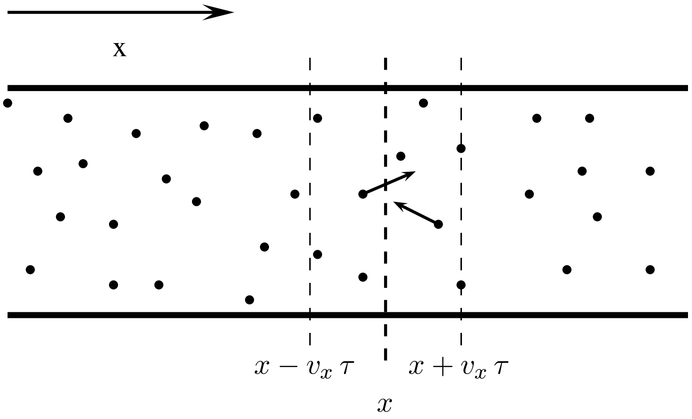


Figure 5.1: A pipe filled with gas, cold on one side and hot on the other. Gas molecules pass through the surface carrying energy.

shall have a look at the derivation of the transfer law based on the above-described picture. To make things more concrete, we shall use a "gas" a mono-atomic gas, which largely satisfies all the set conditions.

Let e be the specific energy of a single particle, and n the number of particle crossing in one direction then the heat is the net balance of the energy going back and fourth:

$$\hat{q} := 1/2\ (n\,e\,v_x)_{x-v_x\,\tau} - 1/2\ (n\,e\,v_x)_{x+v_x\,\tau}$$

for which we used the equal partition principle, so an equal number of carriers passes through the surface in positive, 1/2 of them, and negative x-direction, another half.

Expanding the second term into a Taylor expansion around the imaginary interface, and truncating after the linear term yields:

$$\begin{aligned}(n\,e\,v_x)_{x+v_x\,\tau} &\approx x\,(n\,e\,v_x)_x + \left(\frac{\partial\,n\,e\,v_x}{\partial\,x}\right)_x\ v_x\,\tau \\ (n\,e\,v_x)_{x-v_x\,\tau} &\approx x\,(n\,e\,v_x)_x + \left(\frac{\partial\,n\,e\,v_x}{\partial\,x}\right)_x\ (-v_x\,\tau)\end{aligned}$$

which then gives for the heat flow:

$$\hat{q} := -\left(\frac{\partial\, n\, e\, v_x}{\partial\, x}\right)_x v_x\, \tau$$

Since $n\, e =: u$ the net specific internal energy of all particles crossing the surface, we can write:

$$\begin{aligned}
\hat{q} &:= -\left(\frac{\partial\, u\, v_x}{\partial\, x}\right)_x v_x\, \tau \\
&:= -\tau\, v_x^2 \left(\frac{\partial\, u}{\partial\, x}\right)_x \\
&:= -\tau\, v_x^2 \left(\frac{\partial\, u}{\partial\, T}\right)_x \frac{d\, T}{d\, x}
\end{aligned}$$

The derivation neglects the dependency of the velocity on the temperature, which simplifies the derivation significantly by neglecting the coupling of heat and mass transfer. The derivative $\frac{\partial u}{\partial T}$ is the specific heat capacity per volume c_p of the carriers passing net through the surface. The mean path length $v\, \tau$ and the x-velocity component of the carriers can be obtained from Maxwell's gas theory. Latter can be readily related to the mean velocity of the carriers in that again the equal partition principle applies splitting the kinetic energy in equal parts in the three directions. Thus $v_x = v/3$ and from the gas theory one gets for a mono-atomic gas the average kinetic energy the relation:

$$\langle e \rangle := \frac{m\, v^2}{2} := \frac{3}{2}\, k_B\, T$$

which for room temperature of 300 K gives a mean velocity of about 1000 m/s for Helium and around 500 m/s for a pseudo mono-atomic gas "Air". The mean free path length for is

$$\lambda := \frac{k_B\, T}{\pi\, d^2\, p}$$

which for 300 K and atmospheric pressures gives around $0.14\, \mu m$. Yielding for the average time between collision in the order of $10^{-10}\, s$.

So for the heat transfer we can then write:

$$\hat{q} := -\frac{\tau\, c_p\, v}{3}\, \frac{d\, T}{d\, x} := -k^q\, \frac{d\, T}{d\, x}$$

with k the heat conductivity. This gives us the insight that it is the gradient in the temperature that drives the transfer.

The property ”heat conductivity” is a macroscopic property and does not apply to the nano scale.

We have stated that this picture can be used for different types of ”gases”. So the same picture applies to metals in which it is the electrons that are the major carriers. The electrons form an ”electron gas”. For non-metallic materials, one talks about the transport of Phonons. Phonons are multiples of the product of characteristic frequency and the Planck constant not to be mixed up with Photons. Though they are pictorially similar.

5.2.2 Generic gradient laws

The mathematical models for the transport of extensive quantity is thus a function of the gradient of the effort variables. For any transport of extensive quantity $\hat{\varphi}$ driven by the effort variable π mapped into a co-ordinate system $\underline{\mathbf{r}}$ the transfer laws take usually the form:

$$\hat{\varphi} = f\left(\frac{\partial\,\pi}{\partial\,\underline{\mathbf{r}}}, \underline{\underline{\underline{\mathbb{C}}}}(\underline{\mathbf{r}})\right) \tag{5.1}$$

whereby the expression $\frac{\partial \pi}{\partial \underline{\mathbf{r}}}$ is the gradient of the effort and $\underline{\underline{\underline{\mathbb{C}}}}$ is the conductivity tensor. If the conductivity is a function of the spatial co-ordinate, then the transport process is called *an-isotropic*. Otherwise, the transfer system is called *isotropic*.

The ”transfer laws” (5.1) can take different forms. Since it is the gradient of the effort variables that drive the transport, at equilibrium no gradient exists, thus the effort variables level out over the two system and become the same. For the transfer description this implies that its value must be zero as the two effort variables become the same. Thus for any transport of extensive quantity $\hat{\varphi}$ driven by the effort variable π it must be true that:

$$\hat{\varphi} = f\left(\frac{\partial\,\pi}{\partial\,\underline{\mathbf{r}}} = 0, \underline{\underline{\underline{\mathbb{C}}}}(\underline{\mathbf{r}})\right) = 0$$

Typically the laws are linear in the gradient with the constant conductivity being the most common case.

5.2.2.1 Isotropic behaviour

For a scalar extensive quantity in an isotropic medium the typical transfer law is:

$$\hat{\varphi} := -c\,\frac{\partial\,\pi}{\partial\,\underline{\mathbf{r}}}\,, \tag{5.2}$$

and in the case of a vector of effort variables $\underline{\boldsymbol{\pi}}$:

$$\underline{\hat{\varphi}} := -\underline{\underline{\mathbf{C}}}\,\frac{\partial\,\underline{\boldsymbol{\pi}}}{\partial\,\underline{\mathbf{r}}}\,, \tag{5.3}$$

with $\underline{\underline{\mathbf{C}}}$ being a diagonal matrix with the diagonal being the constant conductivities associated with each individual effort variable.

5.2.2.2 An-isotropic behaviour

For a scalar effort variable and an anisotropic medium with constant conductivity:

$$\hat{\varphi} := -\underline{\mathbf{c}}^T(\underline{\mathbf{r}})\,\frac{\partial\,\pi}{\partial\,\underline{\mathbf{r}}}\,. \tag{5.4}$$

with $\underline{\mathbf{c}}$ a constant conductivity for each co-ordinate. Obviously the most complex case would be description where the conductivity not is different in each direction, but is also varying with the position. Such problems are surprisingly common in mass diffusion.

5.3 Common simple transfer laws

The transport system is driven by the changes of the effort variables at its boundaries. These transfer systems are usually distributed systems, such as walls, filters, membranes, pipes etc. Making assumptions about their dynamic behaviour, namely that they are fast compared to the changes on their boundary, yields simple transfer laws. This analysis we will do later 10.1.

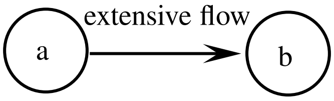


Figure 5.2: Two lumped capacities linked with a transfer of extensive quantity

For the moment, we shall just provide some simple, but very commonly used results from making these behaviour assumptions, thereby following the historical development of the past. The transfer model make the flow of the extensive quantity a function of the difference of the intensive state of the two coupled systems and some conductivity parameter. The most simple versions are linear or simple non-linear functions of the difference in the effort variables, thus the driving force:

$$\begin{aligned}
\text{conductive heat transfer} \qquad & \hat{q}_{a|b} := -{k^q}_{a|b} \, (T_b - T_a) \\
\text{diffusional mass transfer} \qquad & \hat{n} := -{k^n}_{a|b} \, (\mu_b - \mu_a) \\
\text{volume flow} \qquad & \hat{V} := -{k^V}_{a|b} \operatorname{sign}(p_b - p_a) \sqrt{|p_b - p_a|}
\end{aligned}$$

The volume transfer is the most complicated of the three. Why this is the case shall see later (10.3). Note that the transfer laws are directional, that is, the sign indicates the flow relative to the arrow in the graphical representation and indicated in the indices, which read ”from a to b”. All three are written in a ”negative gradient” form. The volume flow is introducing the sign through the sign function, because the square root function is the positive root in all cases.

Keeping the flow definition directional is essential for the formulation of the balances, which is the reason that 5.2 shows the transfer with an arrow. The arrows really introduce a reference co-ordinate system for each flow. If the effective flow goes in the direction of the arrow, the flow is positive and otherwise it is negative.

Systems are seen from an egocentric point of view, thus extensive quantity leaving a system are accounted for negatively and extensive quantity entering add to the contents of the extensive quantity accumulated in the system. Also, what is going out of one system is exactly coming in in the connect system. Systems exchanging extensive quantity are thus immediate connected neighbours.

5.4 Linear transfer laws in networks

Earlier we introduced the representation of the model in the form of a network, which results in a concise description as a directed graph (see 3.2). The simple transfer laws for diffusion of heat and mass are linear in the effort variables, which stimulates the thought of a similar concise description of a network with such transfers.

These transfers of extensive quantity are driven by the *negative difference in the effort variables* π that is the driving force. So for two coupled systems a and b the driving force is the negative of the difference between the value of the effort variable in b and minus the value of the effort variable in a. So the flow takes the form $\hat{\Phi}_{a|b} := -k_{a|b} \, (\pi_b - \pi_a)$. This flow does appear in both balances, namely in the balance of the system a and in the balance of the system b. So in 3.2 we wrote for the sample system:

$$\begin{aligned}
\dot{\Phi}_a &= -\hat{\Phi}_{a|b} \\
\dot{\Phi}_b &= +\hat{\Phi}_{a|b} + \hat{\Phi}_{c|b} - \hat{\Phi}_{b|d} \\
\dot{\Phi}_c &= -\hat{\Phi}_{c|b} \\
\dot{\Phi}_d &= +\hat{\Phi}_{b|d}
\end{aligned}$$

which when substituted becomes:

$$\begin{aligned}
\dot{\Phi}_a &= +k_{a|b}\,(\pi_b - \pi_a) \\
\dot{\Phi}_b &= -k_{a|b}\,(\pi_b - \pi_a) - k_{c|b}\,(\pi_b - \pi_c) + k_{b|d}\,(\pi_d - \pi_b) \\
\dot{\Phi}_c &= +k_{c|b}\,(\pi_b - \pi_c) \\
\dot{\Phi}_d &= -k_{b|d}\,(\pi_d - \pi_b)
\end{aligned}$$

In the concise form the balances are:

$$\underline{\dot{\boldsymbol{\Phi}}} = \underline{\underline{\mathbf{F}}}\,\underline{\hat{\boldsymbol{\Phi}}}$$

and we remind: with $\underline{\underline{\mathbf{F}}}$ being the incidence matrix of the directed graph 3.1:

$$\underline{\underline{\mathbf{F}}} := \begin{bmatrix} -1 & 0 & 0 \\ +1 & +1 & -1 \\ 0 & -1 & 0 \\ 0 & 0 & +1 \end{bmatrix}$$

If we now take a closer look at the list of transfers, we recognise that we can write them equally concise because the transpose of the F-matrix multiplied with the effort variables associated with the systems, provide us with the vector of discrete gradients: the differences in the effort variables and thus the driving forces:

$$\underline{\hat{\boldsymbol{\Phi}}} := -\underline{\underline{\mathbf{K}}}^{q}\,\underline{\underline{\mathbf{F}}}^{T}\,\underline{\boldsymbol{\pi}}\,, \tag{5.5}$$

where the $\underline{\underline{\mathbf{K}}}^{q}$ is a diagonal matrix with the heat conductivity parameters times interface areas in the diagonal.

6

State, input and output

Synopsis *"The state is the minimal information required to predict the future, given the current input." A statement made by R E Kalman. The state is the core of the description whilst the input describes what drives the system, which always involves the environment. The output is the observation, namely what we "see", which in all cases is information about the state in one or the other form.*

6.1 The state

6.1.1 The Concept *State*

The state is a central object in the discussion of dynamic systems. In the context of mathematical system theory, the term *state* is the essence of state space theory. Though the term *state* is for most intuitively interpretable, it is quite difficult to provide a precise definition and different people and different disciplines found quite different wordings :

System theory evolved from control-related subjects in the 50ties driven by the quest to get a generic view of systems. The systems view, as it was established by people like Norbert Wiener, was further developed with the state forming the core.

- Kalman (1963), one of the leading people in mathematical system theory, wrote in 1963: "The state is to be regarded always as an abstract quantity. Intuitively speaking, the state is the minimal amount of information about the past history of the system which suffices to predict the effect of the past upon the future."

- Kailath (1980): The state provides a "sufficient statistic" so to say, that enables us to calculate the future response to a new input without worrying about previous inputs. Note also that more than one past input can lead to the same state. Therefore, the state is really a minimal sufficient statistic. It contains just enough information, no less and no more, to enable us to calculate the future responses without further reference to the old history of inputs and responses as in more colloquial usage, the knowledge of the state vector at any time specifies the state or condition of the system at that time.

- Kailath also discusses a mathematical derivation which is due to Nerode (1958) specifying the meaning of the term state more precisely.

As a chemical engineer talking about states, it is interesting to compare our common understanding of the term state, as it is used in thermodynamics, with the usage in mathematical system theory :

- Denbigh (1971) uses it as a primitive without really explaining its meaning.
- Falk and Jung (1959) state, which translated reads : "In an axiomatic description the state plays the rôle of a not precisely defined basic object of the theory. The only condition is that the states are distinguishable objects."

The latter definition is extremely abstract and reflects a little bit of the difficulty that people have to define terms precisely capturing all aspects of their common usage. But, what are the conclusions? There are in principle three aspects of the term state that one should keep in mind :

1. states are distinguishable objects - no two states are the same
2. states are independent of the history - exact differentials
3. states contain all the information about the condition (state) of the system at a given time.

Given dynamic equations, it is most easy to identify the states, as they are the ones that are differentiated with time, because it is the state that changes with respect to time in a dynamic system. It carries the information required to compute the future state given the inputs at the current time. The main part of these statements is that the state contains all information about the system. Such systems are also called ”Markov systems”. If this is not the case and history of the process is required, then one refers to a it as a ”non-Markov system”.

6.1.2 Basic state - fundamental state

The taken approach to modelling physical-chemical-biological systems always requires us to formulate component mass balances, energy balances and momentum balances, which provide the information about the dynamic behaviour of the process. This immediately defines the state of the system, namely the component mass, the energy and the momentum. For macroscopic systems for which rotation is usually negligible only the linear momentum conservation is required, the dimension is (number of species + 1 energy + 3 momentum). Very often we also have a very fast momentum transfer as pressure travels with the speed of sound and chemical-biological processes are usually slower, giving rise to make pseudo steady-state assumptions for the momentum balances. Consequence being, that the most common systems are of the dimension (number of species + 1 energy) with the pressure being computed from the stationary momentum balances or the mechanical energy balances. 10.3

The mathematical object ”state” is though not unique. In fact one can define an infinite number of states by defining transformations, so-called similarity transformations (see C.1.1), without changing the input - output behaviour. The Figure 6.1 shows the transmution from one state space represention into another one.

The dimensionality of the space does not change, only its nature changes. Also it contains exactly the same information. It is only the ”form” in which the information is given that changes. Figure 6.1 shows on top (A) the original nonlinear system in block diagram. In the middle, a linear transformation of the state has been done, by multiplying the original state $\underline{\mathbf{x}}$ with the invertible matrix $\underline{\underline{\mathbf{T}}}$. The two other functions have not changed. On the bottom (C) the linear transformation has been included into the two functions.

6.2 Input

Having taken the physical-mechanistic approach, the input to a system is the exchange of the extensive quantity of the system with its direct environment, if we ignore some special cases like magnetic fields. In the first place the inputs are thus the extensive quantities crossing the system's surface. On the next level, it is the effort variables, or better the difference in the effort variables on both side of the common interface of the coupled systems that constitute the input. Since these effort variables are again a function of the state, the circle closes, which we will discuss later in more detail later in 15.2.

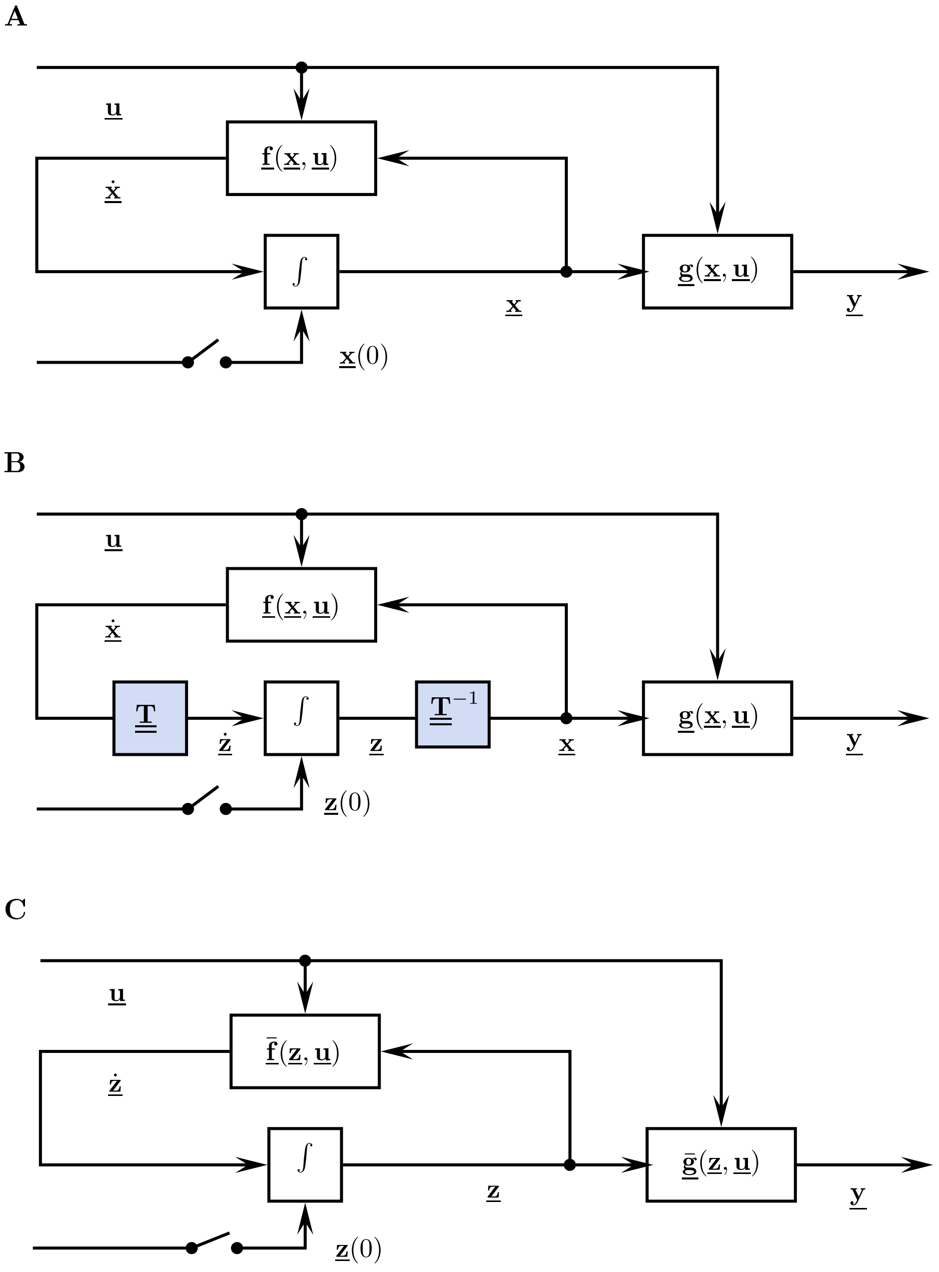

A
$\underline{\mathbf{u}}$
$\underline{\mathbf{f}}(\underline{\mathbf{x}}, \underline{\mathbf{u}})$
$\dot{\underline{\mathbf{x}}}$
$\int$
$\underline{\mathbf{g}}(\underline{\mathbf{x}}, \underline{\mathbf{u}})$
$\underline{\mathbf{x}}$
$\underline{\mathbf{y}}$
$\underline{\mathbf{x}}(0)$
B
$\underline{\mathbf{u}}$
$\underline{\mathbf{f}}(\underline{\mathbf{x}}, \underline{\mathbf{u}})$
$\dot{\underline{\mathbf{x}}}$
$\underline{\underline{\mathbf{T}}}$
$\int$
$\underline{\underline{\mathbf{T}}}^{-1}$
$\underline{\mathbf{g}}(\underline{\mathbf{x}}, \underline{\mathbf{u}})$
$\dot{\underline{\mathbf{z}}}$
$\underline{\mathbf{z}}$
$\underline{\mathbf{x}}$
$\underline{\mathbf{y}}$
$\underline{\mathbf{z}}(0)$
C
$\underline{\mathbf{u}}$
$\bar{\underline{\mathbf{f}}}(\underline{\mathbf{z}}, \underline{\mathbf{u}})$
$\dot{\underline{\mathbf{z}}}$
$\int$
$\bar{\underline{\mathbf{g}}}(\underline{\mathbf{z}}, \underline{\mathbf{u}})$
$\underline{\mathbf{z}}$
$\underline{\mathbf{y}}$
$\underline{\mathbf{z}}(0)$

Figure 6.1: State is not unique, all three have the same input/output behaviour

Whilst what has been said is quite logical, inputs are often defined in quite a different way, apparently also motivated by other considerations. It is often some practicalities that motivate the definition of ”input”. For the newcomer this can be quite puzzling. An example is quite easily constructed: Take a stirred tank reactor where we focus on the effect of the stirrer and its control. The stirrer is connected to a motor providing the energy to drive the blades through the fluid. The motor passes the energy on to the spindle, which in turn looses some to the bearings and the main effect, namely the energy transfer to the blades being pushed through the fluid and transferring momentum and energy to the fluid. The fluid may be reactive causing the viscosity to increase, which leads to an increased resistance for the blades, requiring more energy from the shaft and the motor. Depending on the system boundaries, one could choose the energy input to the motor, the stress measurement on the shaft providing an indirect measure for the torque acting on the shaft or the speed of rotation, or the energy being passed by the blades. Which one is chosen is dependent on the available measurement and very much the process boundary one considers, but also on the models one has of the different sub-processes.

The term ”input” is most commonly used in the operation domain. From a physical and control point of view it is best seen as the quantity that drives the system. So if we do include control into our considerations, then a typical input to the controlled system is a set-point to a controller. The controller, in turn generates a signal that manipulates the process behaviour. In physical terms this is always associated with affecting the flow of the extensive quantities between systems. The manipulating element has thereby the nature of a ”valve”, which in a volumetric convective flow usually is a valve and the controller changes the position between the cone and the seat of the valve thereby changing the cross section available for the flow and thus also the resistance. The flow of heat cannot be controlled directly, but in this case the temperature of the environment must be changed, which again cannot be done directly. The temperature is a state-dependent quantity of a system and can thus only be affected indirectly by changing the energy flow across the boundary of the corresponding system. This can be mass flow of different temperature or energy in the form of heat from a heater, to mention two examples. In this latter case, the input is the temperature, thus the state of the coupled system. So in physical systems we have two types of basic inputs, one being the change of the transfer property, the ”valve”, the other being the driving force by indirectly changing the state of the environment, which in turn again can only be done by changing a flow accross the boundary of the environment system.

6.3 Output

The output is what can be observed in the process. Using the term ”observed” instead of ”measured” is done in consideration of the fact that one has rarely direct access to the conserved quantities and also for the intensive quantities. Also almost all measurement methods do not measure the desirable quantity directly, but measure it indirectly by measuring a property that is dependent on the quantity being observed. For example a temperature measurement is not measuring the temperature, but some secondary effect like a resistance, which in turn can also not be measured directly but for constant potential it is the current that is measured.

Quantities of interest are the states in terms of the conserved quantities, but also derived states, mainly the densities, namely the extensive states normed by the volume, but also the effort variables (5.1). Usually we will refer to the derived states as ”secondary” states. Quite frequently we are also interested in flows. Since it is connections that often represent flows, it is the state of the flow that we attempt to assess. Flows that are represented as connections are implementing the assumption of a very fast underlying transfer system, thus eliminating its capacity effect as discussed in detail in 10.

6.4 View on the system

A plant communicates with its environment in two principle ways: one in which the flow of extensive quantity can be manipulated and one in which it is given in the sense one has no means of controlling the flow (6.2). The element that is manipulated by the controller is shown with a valve symbol. In general two sets of observations / measurements will be available, one from the plant, indicating the state of the plant and which is usually also the object of control, whilst the other provides state information of the environment, usually the part that is close to the connection affecting the plant.

The flows that are not manipulated are often disturbances or simple loads, namely flows that come from upstream with the downstream having no control. The flows that one can manipulate serve often the purpose to apply corrective actions so as to keep the process on the defined trajectory. The picture is generic and can be applied to any process.

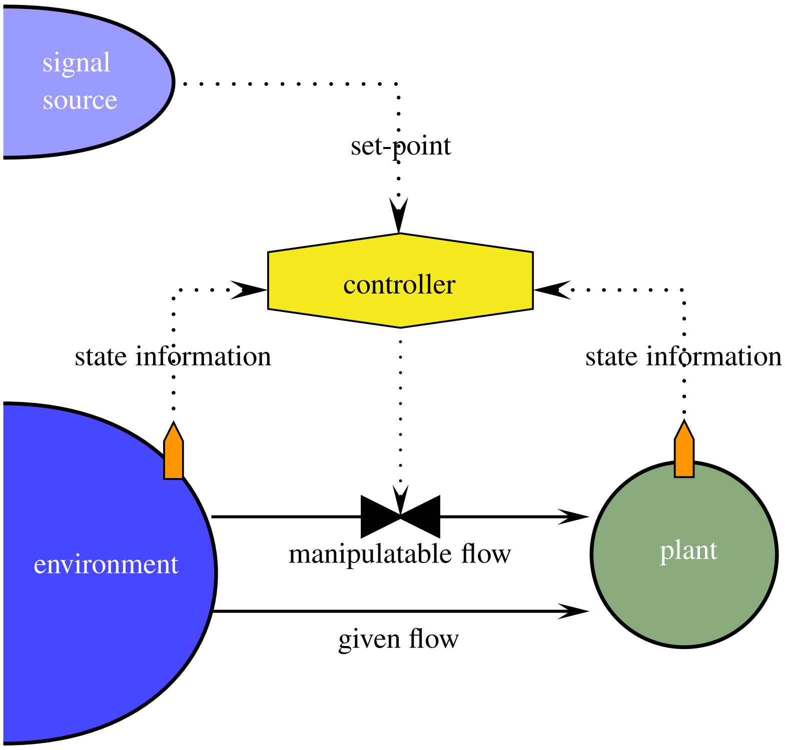


Figure 6.2: Plant under control

7

The two extremes: stationary and event-dynamics

Synopsis *If it is stable, the system will, after receiving a disturbance, converge to it's natural steady state.*

The conservation principles are linear differential equations, and for lumped system this reduces to linear ordinary differential equations. Since natural systems are conservative, they are also stable and will reach a natural steady state after having been pushed away from their stationary position. The conversion is exponential (see C.1.5). In many processes, the steady state is of special interest mainly because it provides information about the energy and mass household when running the process at the desired stationary operating point.

7.1 The analysis

With the conservation principles for a non-reactive system s being:

$$\underline{\dot{\mathbf{\Phi}}}_s = \underline{\underline{\mathbf{F}}}_s \, \underline{\hat{\mathbf{\Phi}}}$$

For a network of systems this is stacked up and we obtain a block-matrix equation. Let the systems in the network be labelled with $1, 2, 3, \ldots$ and the streams for simplicity with $a, b, c, \ldots$:

$$\begin{bmatrix} \underline{\dot{\mathbf{\Phi}}}_1 \\ \underline{\dot{\mathbf{\Phi}}}_2 \\ \vdots \end{bmatrix} = \begin{bmatrix} \underline{\underline{\mathbf{F}}}_{1,a} & \underline{\underline{\mathbf{F}}}_{1,b} & \underline{\underline{\mathbf{F}}}_{1,c} & \cdots \\ \underline{\underline{\mathbf{F}}}_{2,a} & \underline{\underline{\mathbf{F}}}_{2,b} & \underline{\underline{\mathbf{F}}}_{2,c} & \cdots \\ \vdots & \vdots & \vdots & \ldots \end{bmatrix} \begin{bmatrix} \underline{\hat{\mathbf{\Phi}}}_a \\ \underline{\hat{\mathbf{\Phi}}}_b \\ \vdots \end{bmatrix}$$

The blocks are not necessarily square, which is apparent if one keeps in mind that in a topology with different species, not all species are everywhere in the network. Let us have a look at an example in the form of a mixing plant

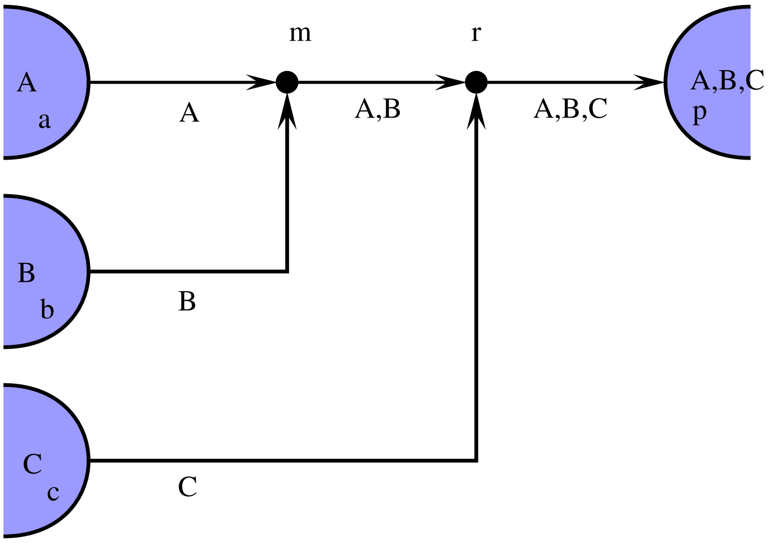


Figure 7.1: A simple mixing plant at steady state

similar to I.4. The topology for the stationary behaviour is as in Figure 7.1

The plant has two dynamic components, m, r, and four reservoirs a, b, c, p and 5 streams $a|m, b|m, c|r, m|r, r|p$. The streams have different components in them. Adopting the notation $\dot{n}_{\text{species,system}}$ for the systems and for the flows $\hat{n}_{\text{species,stream}}$ the accumulation terms are stacked as well as the five streams are stacked up:

$$\underline{\dot{\mathbf{n}}} := \left[\begin{array}{c} \underline{\dot{\mathbf{n}}}_m \\ \hline \underline{\dot{\mathbf{n}}}_r \end{array}\right] := \begin{bmatrix} \begin{bmatrix} \dot{n}_{A,m} \\ \dot{n}_{B,m} \end{bmatrix} \\ \begin{bmatrix} \dot{n}_{A,r} \\ \dot{n}_{B,r} \\ \dot{n}_{C,} \end{bmatrix} \end{bmatrix} \qquad \underline{\hat{\mathbf{n}}} := \begin{bmatrix} \underline{\hat{\mathbf{n}}}_{a|m} \\ \underline{\hat{\mathbf{n}}}_{b|m} \\ \underline{\hat{\mathbf{n}}}_{c|r} \\ \underline{\hat{\mathbf{n}}}_{m|r} \\ \underline{\hat{\mathbf{n}}}_{r|p} \end{bmatrix} := \begin{bmatrix} \begin{bmatrix} \hat{n}_{A,a|m} \end{bmatrix} \\ \begin{bmatrix} \hat{n}_{B,b|m} \end{bmatrix} \\ \begin{bmatrix} \hat{n}_{C,c|r} \end{bmatrix} \\ \begin{bmatrix} \hat{n}_{A,m|r} \\ \hat{n}_{B,m|r} \end{bmatrix} \\ \begin{bmatrix} \hat{n}_{A,r|p} \\ \hat{n}_{B,r|p} \\ \hat{n}_{C,r|p} \end{bmatrix} \end{bmatrix}$$

The $\underline{\underline{\mathbf{F}}}$ is a corresponding block matrix as shown below.

flows →		$a\vert m$	$b\vert m$	$c\vert r$	$m\vert r$		$r\vert p$		
↓ system ↓		A	B	C	A	B	A	B	C
m	A	1			-1				
m	B		1			-1			
r	A				1		-1		
r	B					1		-1	
r	C			1					-1

The model has 5 equations and 8 variables. Thus 3 variables must be defined. Since the balances equations do not interact, it is one of each of the species in any of the streams. Let us choose the variable $\hat{n}_{A,a|m}, \hat{n}_{B,b|m}, \hat{n}_{C,r|p}$. In the next step the linear set of equations is transmogrified into the normal form $\underline{\underline{\mathbf{A}}}\,\underline{\mathbf{x}} = \underline{\mathbf{b}}$ to prepare it for a standard solver. This is achieved by using selection matrices (see A.1.1) to split the variables into the known and the unknown variables:

$$\underline{\underline{\mathbf{S}}}_n := \begin{bmatrix} \mathbf{1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \mathbf{1} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{1} \end{bmatrix}$$

Correspondingly for the unknown variables:

$$\underline{\underline{\mathbf{S}}}_u := \begin{bmatrix} 0 & 0 & \mathbf{1} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{1} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathbf{1} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \mathbf{1} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{1} & 0 \end{bmatrix}$$

The two selection matrices stacked together are a permuted identity matrix. The stacked set of balance equations is modified:

$$\begin{aligned} \underline{\mathbf{0}} &= \underline{\underline{\mathbf{F}}}\,\underline{\hat{\mathbf{n}}} \\ &:= \underline{\underline{\mathbf{F}}}\,\underline{\underline{\mathbf{S}}}_u^T\,\underline{\underline{\mathbf{S}}}_u\,\underline{\hat{\mathbf{n}}} + \underline{\underline{\mathbf{F}}}\,\underline{\underline{\mathbf{S}}}_n^T\,\underline{\underline{\mathbf{S}}}_n\,\underline{\hat{\mathbf{n}}} \end{aligned}$$

Defining:

$$\begin{aligned} \underline{\underline{\mathbf{A}}} &:= \underline{\underline{\mathbf{F}}}\,\underline{\underline{\mathbf{S}}}_u^T \\ \underline{\mathbf{x}} &:= \underline{\underline{\mathbf{S}}}_u\,\underline{\hat{\mathbf{n}}} \\ \underline{\mathbf{b}} &:= \underline{\underline{\mathbf{F}}}\,\underline{\underline{\mathbf{S}}}_n^T\,\underline{\underline{\mathbf{S}}}_n\,\underline{\hat{\mathbf{n}}} \end{aligned}$$

We get the equations in standard form:

$$\underline{\mathbf{0}} = \underline{\underline{\mathbf{A}}}\,\underline{\mathbf{x}} + \underline{\mathbf{b}}$$

8

Processes convert material to new products

Synopsis *Most plants that are not of pure mechanical nature produce a product through a conversion or transposition of material. These material changes include first of all reactions, but also energetically weaker interactions like agglomeration or absorption. Also phase changes can be put into this category. Reactive systems though are of special interest for which the species mass balances are extended with a reaction term.*

8.1 Internal dynamics – conversion of extensive quantity

Reactions are about changing the nature of the material in a spatial domain, in a control volume, the system. Reactions are an interaction of extensive quantities thereby changing nature of the system. This requires an extension of our conservation principles 4.2 adding an additional term that describes the interaction of extensive quantities, namely a production term. Again to lift out the different nature of the interaction, we specially decorate the term using a ˜.

$$\underline{\dot{\mathbf{n}}} = \underline{\hat{\mathbf{n}}} + \underline{\tilde{\mathbf{n}}} \tag{8.1}$$

Having stated that the extensive quantities are being conserved, does this additional term now break the conservation laws? Obviously not. The production term is used to describe the internal change in a system, though strictly maintaining the conservation of the sum of "involved" extensive quantities. Whilst the total mass is conserved, reactions are ongoing in the system thereby changing the nature of the mass inside the system boundaries. The vector of extensive quantities defined here, must include all "involved" extensive quantity, such that if we sum all the mass up, the production term disappears. We get the mass from the molar vector as a weighted sum, where the weights are the molecular masses of the species.

Defining a vector of molecular masses $\underline{\boldsymbol{\lambda}}$, it must be true, that

$$\begin{aligned} \underline{\boldsymbol{\lambda}}^T\, \underline{\dot{\mathbf{n}}} &= \underline{\boldsymbol{\lambda}}^T\, \underline{\hat{\mathbf{n}}} + \underline{\boldsymbol{\lambda}}^T\, \underline{\tilde{\mathbf{n}}} \\ &= \underline{\boldsymbol{\lambda}}^T\, \underline{\hat{\mathbf{n}}} \\ \dot{m} &= \hat{m} \end{aligned}$$

Which says that the accumulation of total mass is the net flow of the total mass across the system boundary.

8.2 Atoms, species and reactions

Nature has a building box from which it can construct new items using existing ones. The Greeks and also India's early philosophers declared that all starts with atoms, nature is built from atoms. Much was being said, investigated and found giving more insight into the world of atoms since then.

But, whilst we know some more about the nature of matter, this concept of smallest building element still serves the very useful purpose as atoms are the principle building blocks from which we assemble chemical species. We use the notation H_2O for water thereby indicating that a water molecule consists of two hydrogen atoms and an oxygen atom, whilst methane is CH_4 as it consists of one carbon atom and four hydrogen atoms. The formal representation of a reaction is usually given in the form *reactants* → *products* or if the reaction is considered to go both ways *reactants* $\rightleftharpoons$ *products*. 8.1 depicts such a reaction in the form of 3D models showing the electron clouds of the involved molecules, the formal representation and the involved atoms.

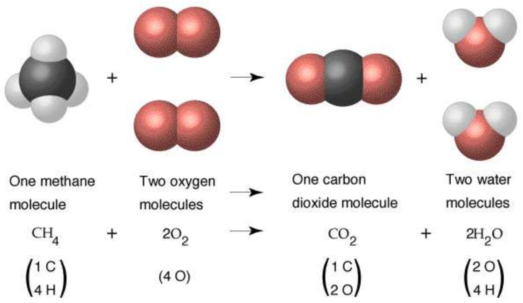


Figure 8.1: A sample reaction

In order to write mathematical equations, we use a different notation, representing a species as the sum of number of moles of species. Let $\mathtt{A}_i$ represent the i^{th} kind of atom taken from the list of atoms and let $\mathtt{S}$ be a species which consists of a atoms of $\mathtt{A}_1$ and b atoms of $\mathtt{A}_2$ then we write for the species $\mathtt{S}$, then $\mathtt{S} = a\mathtt{A}_1 + b\mathtt{A}_2$. So in general we write for species j:

$$\mathtt{S}_j = c_{i,j}\, \mathtt{A}_i \tag{8.2}$$

The weighting factors $c_{i,j}$ are integers corresponding to the number of atoms of the kind i to be present in the species j. The equation thus represents a "sum formula" for chemical species. So this representation maps between molecular species and the constituent atoms. Examples are water: $H_2O = 2\,H + 1O$ or ethanol $C_2H_6O = 2C + 6H + 1O$. The representation thus does not include any structural information of the molecules in discussion. So for our species C_2H_6O could also be dimethyl ether.

Processes produce chemicals by either extracting them from a mixture or by combining or re-combining the constituents of molecules. The constituents may be the basic building blocks, namely atoms or they may be groups of atoms. Apparently the first is more general than the second because the groups consist of atoms. Building a molecule can thus be seen as a recursive building-block system in which one constructs groups from atoms and molecules from groups.

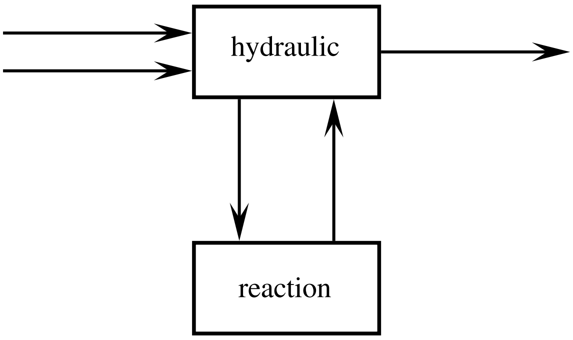


Figure 8.2: Separation of hydraulic and reaction

The physical environment in which molecules are "built" is in a physical space, thus a volume in which the constituents to the reaction are entering or being present when the process starts. Thus there is some kind of geometrical arrangement of the molecules to be considered besides the actual transposition that is taking place when the constituents come in physical contact with each other. So there is a mechanism that makes it possible for the species to meet and there is a mechanism which constitutes the reaction itself. This view can be abstracted by separating the getting-together-process from the reaction process, which in a flow system is a hydraulic subsystem such as a mixing tank, to give an explicit example. The reaction subsystem then takes from this hydraulic subsystem exactly what it requires to perform the reaction. It is in the reaction subsystem that the constituents meet and undergo the reactions in stoichiometric ratios. So any control volume in which a reaction takes place could be split into a hydraulic subsystem and a reaction subsystem as shown in 8.2.

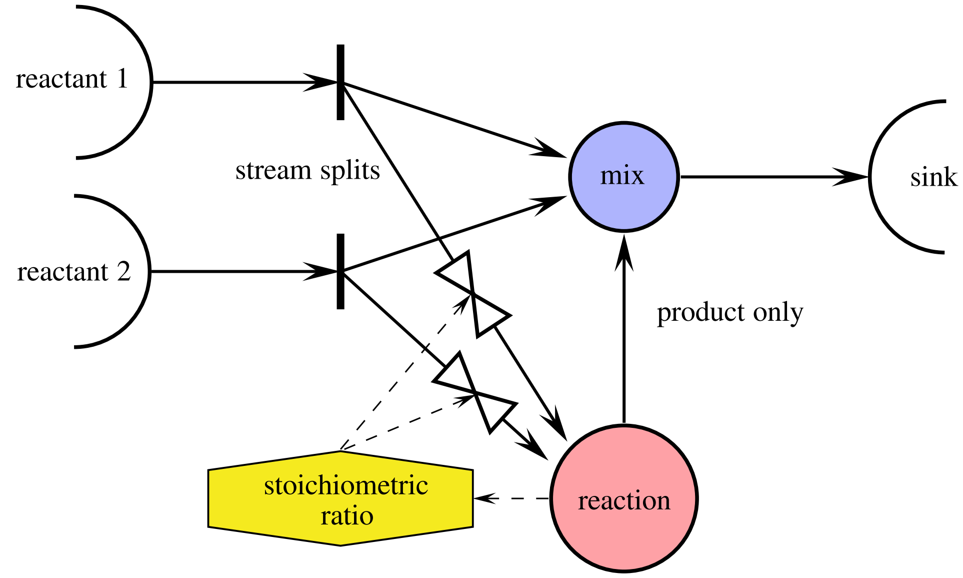


Figure 8.3: Abstraction of separating hydraulic and reaction into two separate systems

If we analyse it in more detail using the graphical representation we introduced earlier, things look rather complex as we need first to separate the reactants required by the reaction box in exact amounts from the feed streams. This is achieved by controlling the two streams to the ”molecule factory” labelled with ”reaction” using a ratio controller. The two resulting streams are fed to the reaction, where the molecule factory does its job producing the product. The thus generated product stream is fed into the mixing (hydraulic) system where it is mixed with the remaining streams from the feeds. The exit stream is then a mix of the products and the un-reacted reactants (8.3). This Gedankenexperiment can be applied to any system. Chemical species come into the reaction zone through one or the other transport mechanisms, where they become stock for the molecule factory. The factory operates like a shop floor: the building blocks are geometrically arranged to expose the connection points so they can undergo the desired transposition. Problem though being that we lack Maxwell demons (ref) as workers. So we really have no absolute control over this critical process. Instead molecules move in a potential field that they generate through their existence and interactions. This leaves the option of different arrangements, which nearly invariably also result in different products, since these potential fields have different valleys, saddles and hills, and thus different local minima. So one actually almost always has to consider families of reactions yielding not only the desired products but also side products defining the need for down-stream separation processes.

Coming back to the molecule factory, as long as we exclude nuclear reactions, the basic building blocks are always conserved. If we formulate a generic lumped system with a reaction for the species $\mathtt{A}_i \in \mathtt{A}$ with a set of inflows and an outflow, the model takes the form:

$$\dot{\underline{\mathbf{n}}} = \sum_{\forall m} \hat{\underline{\mathbf{n}}}_m - \hat{\underline{\mathbf{n}}}_{out} + \tilde{\underline{\mathbf{n}}}$$

So the accumulation of the species captured in the vector $\dot{\underline{\mathbf{n}}}$, using the dot decorator $\dot{}$ for the time derivative, is the sum over all the m input streams $\hat{\underline{\mathbf{n}}}_m$, where the stream is indicated by the hat decorator $\hat{}$, minus the outflow and adding the production of all the species in the process denoted by $\tilde{\underline{\mathbf{n}}}$ using the $\tilde{}$ as a decorator. Note that summing the above vector equation over the species gives on the left-hand-side the total molar mass being accumulated, which implies that the sum over the production terms must be zero as mass is being conserved.

If we now use the above-indicated mapping from the species to the atoms, we can get a representation of the reactor's behaviour in terms of atoms. For this purpose we have to distinguish between the molar vector of atoms and the molar vector of species. Let the decorator a be used for the atoms and the decorator s for the species present in the process. So $\underline{\mathbf{n}}^a$ is the vector of moles of atoms and $\underline{\mathbf{n}}^s$ is the vector of moles of process species. The transformation is then:

$$\underline{\mathbf{n}}^a := \underline{\underline{\mathbf{C}}}\,\underline{\mathbf{n}}^s \tag{8.3}$$

in which the $\underline{\underline{\mathbf{C}}}$ is the coefficient matrix $[c_{i,j}]$ with i being the atom index and j being the species index. Transforming the reactor's behaviour equation we get:

$$\underline{\underline{\mathbf{C}}}\,\dot{\underline{\mathbf{n}}}^s = \sum_{\forall m} \underline{\underline{\mathbf{C}}}\,\hat{\underline{\mathbf{n}}}^s_m - \underline{\underline{\mathbf{C}}}\,\hat{\underline{\mathbf{n}}}^s_{out} + \underline{\underline{\mathbf{C}}}\,\tilde{\underline{\mathbf{n}}}^s \tag{8.4}$$

$$\dot{\underline{\mathbf{n}}}^a = \sum_{\forall m} \hat{\underline{\mathbf{n}}}^a_m - \hat{\underline{\mathbf{n}}}^a_{out} + \underline{\underline{\mathbf{C}}}\,\tilde{\underline{\mathbf{n}}}^s \tag{8.5}$$

Since atoms are not being generated and do not disappear in a reaction scheme, the last term must result in a vector of zeros:

$$\underline{\underline{\mathbf{C}}}\,\tilde{\underline{\mathbf{n}}}^s = \underline{\mathbf{0}} \tag{8.6}$$

Molecules are always a combination of atoms. Thus the matrix $\underline{\underline{\mathbf{C}}}$ is not a square matrix. It will always have more columns than rows, which implies that there are some additional relations, here being the sought sum-reactions. Question then is on how many reactions are forming the defined species. The answer comes from linear algebra. The 8.6 is a linear homogeneous system with the variables being the reaction rates. There are more reaction rates than atomic species. The rank of the matrix provides the information on how many independent reactions there are and the vectors spanning the null space give the ratios between the reaction rates, thus the stoichiometric ratios. The literature on this subject is surprisingly rich, which gives an indication that the subject is not quite so well understood than what is being commonly assumed. The references Aris (1963, 1965) is seen as seminal for the subject though earlier contributions Truesdell and Toupin (1960) many of which one finds cited in Bjornbom (1977) indicating that the subject goes back to Gibbs. A very recent contribution Higgins and Whitaker (2011) re-iterates on the basic ideas and tries to bring the subject closer to the reader.

8.2.1 Finding a minimal set of reactions

The procedure to find the rank and a set of column vectors spanning the null space is well documented in the literature (Strang (2009)). A brief summary is located in the dedicated appendix (A.1.3).

It is best illustrated using an example. The set of molecules is

$$\mathcal{S} := [H_2O, CH_4, CO, CO_2, H_2]$$

are constructed from the atoms

$$\mathcal{A} := [H, O, C]\,.$$

The objective is to find a matrix with a basis for the null space computing the reduced row-echolon form of the homogeneous equation set. First we construct a table with the rows being the number of atoms in the molecules in the columns.

	H_2O	CH_4	CO	CO_2	H_2
H	2	4	0	0	2
O	1	0	1	2	0
C	0	1	1	1	0

The coefficient matrix $\underline{\underline{\mathbf{C}}}$ is then the values in the table, thus:

$$\underline{\underline{\mathbf{C}}} := \begin{bmatrix} 2 & 4 & 0 & 0 & 2 \\ 1 & 0 & 1 & 2 & 0 \\ 0 & 1 & 1 & 1 & 0 \end{bmatrix}$$

In a first step the first row is divided by 2, which is achieved by multiplying with a corresponding matrix:

$$\underline{\underline{C}}_1 := \begin{bmatrix} 0.5 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 4 & 0 & 0 & 2 \\ 1 & 0 & 1 & 2 & 0 \\ 0 & 1 & 1 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 0 & 0 & 1 \\ 1 & 0 & 1 & 2 & 0 \\ 0 & 1 & 1 & 1 & 0 \end{bmatrix}$$

Next we eliminate the first 1 in the second row by replacing the second row by the difference of the second row minus the first row:

$$\underline{\underline{C}}_2 := \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 0 & 0 & 1 \\ 1 & 0 & 1 & 2 & 0 \\ 0 & 1 & 1 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 0 & 0 & 1 \\ 0 & -2 & 1 & 2 & -1 \\ 0 & 1 & 1 & 1 & 0 \end{bmatrix}$$

The second row is divided by -2 yielding a 1 in the first non-zero position:

$$\underline{\underline{C}}_3 := \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1/2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 0 & 0 & 1 \\ 0 & -2 & 1 & 2 & -1 \\ 0 & 1 & 1 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 0 & 0 & 1 \\ 0 & 1 & -1/2 & -1 & 1/2 \\ 0 & 1 & 1 & 1 & 0 \end{bmatrix}$$

The 1 in the third row is the next target to eliminate:

$$\underline{\underline{C}}_4 := \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 0 & 0 & 1 \\ 0 & 1 & -1/2 & -1 & 1/2 \\ 0 & 1 & 1 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 0 & 0 & 1 \\ 0 & 1 & -1/2 & -1 & 1/2 \\ 0 & 0 & 1.5 & 2 & -0.5 \end{bmatrix}$$

Now we scale the third equation:

$$\underline{\underline{C}}_5 := \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2/3 \end{bmatrix} \begin{bmatrix} 1 & 2 & 0 & 0 & 1 \\ 0 & 1 & -1/2 & -1 & 1/2 \\ 0 & 0 & 1.5 & 2 & -0.5 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 0 & 0 & 1 \\ 0 & 1 & -1/2 & -1 & 1/2 \\ 0 & 0 & 1 & 4/3 & -1/3 \end{bmatrix}$$

The result is the upper triagonal matrix. In the next steps the non-zero elements above the pivot 1 are eliminated.

$$\underline{\underline{C}}_6 := \begin{bmatrix} 1 & -2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 0 & 0 & 1 \\ 0 & 1 & -1/2 & -1 & 1/2 \\ 0 & 0 & 1 & 4/3 & -1/3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 & 2 & 0 \\ 0 & 1 & -1/2 & -1 & 1/2 \\ 0 & 0 & 1 & 4/3 & -1/3 \end{bmatrix}$$

and:

$$\underline{\underline{C}}_7 := \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 & 2 & 0 \\ 0 & 1 & -1/2 & -1 & 1/2 \\ 0 & 0 & 1 & 4/3 & -1/3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 2/3 & 1/3 \\ 0 & 1 & -1/2 & -1 & 1/2 \\ 0 & 0 & 1 & 4/3 & -1/3 \end{bmatrix}$$

and:

$$\underline{\underline{\mathbf{C}}}_8 := \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0.5 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 2/3 & 1/3 \\ 0 & 1 & -1/2 & -1 & 1/2 \\ 0 & 0 & 1 & 4/3 & -1/3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 2/3 & 1/3 \\ 0 & 1 & 0 & -1/3 & 1/3 \\ 0 & 0 & 1 & 4/3 & -1/3 \end{bmatrix}$$

Writing the result as a block matrix with the index of the identity matrix indicating its dimension. The matrix is in the desired reduced row echolon form:

$$\underline{\underline{\mathbf{R}}} := \underline{\underline{\mathbf{C}}}_8 := \begin{bmatrix} \underline{\underline{\mathbf{I}}}_3 & \underline{\underline{\mathbf{F}}} \end{bmatrix}$$

The null space is found quite easily by noticing that

$$\begin{bmatrix} \underline{\underline{\mathbf{I}}}_3 & \underline{\underline{\mathbf{F}}} \end{bmatrix} \begin{bmatrix} -\underline{\underline{\mathbf{F}}} \\ \underline{\underline{\mathbf{I}}}_2 \end{bmatrix} = -\underline{\underline{\mathbf{F}}} + \underline{\underline{\mathbf{F}}} = \underline{\underline{\mathbf{0}}}$$

The set of vectors spanning the null-space, captured in a matrix, is thus:

$$\text{Null}\left(\underline{\underline{\mathbf{C}}}\right) := \begin{bmatrix} -\underline{\underline{\mathbf{F}}} \\ \underline{\underline{\mathbf{I}}}_2 \end{bmatrix} = \begin{bmatrix} -2/3 & -1/3 \\ 1/3 & -1/3 \\ -4/3 & 1/3 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

The 2 identity matrix is the standard choice for the free variables providing the special solution. The basis for the null space is thus not orthonormal.

The reaction rates are the variables in 8.6. Thus the vectors spanning the computed null space provide a solution, namely the ratios of the reaction rates and thus the stoichiometric coefficients. The number of free variables in this case is 2 whilst the rank is 3. So two relations, being two reactions, exist and the vectors spanning the null space is thus directly the stoichiometric matrix for the two independent reactions. Scaling with 3 gives:

$$\underline{\underline{\mathbf{N}}}^T := \begin{bmatrix} -2 & -1 \\ 1 & -1 \\ -4 & 1 \\ 3 & 0 \\ 0 & 3 \end{bmatrix}$$

The two independent reactions are thus:

$$\begin{aligned} 2\,H_2O + 4\,CO &\leftrightarrows CH_4 + 3\,CO_2 \\ H_2O + CH_4 &\leftrightarrows CO + 3\,H_2 \end{aligned}$$

In the above example, the row-echolon form has no zero subspace. In general, allowing also for column swapping, the reduced row echolon matrix takes the form:

$$\underline{\underline{\mathbf{R}}} := \begin{bmatrix} \underline{\underline{\mathbf{I}}}_{r,n} & \underline{\underline{\mathbf{F}}} \\ \underline{\underline{\mathbf{0}}}_{s,n} & \underline{\underline{\mathbf{0}}}_{s,f} \end{bmatrix}$$

where r is the rank of the coefficient matrix, s the number of zero rows and f the number of free variables, latter defining the number of independent reactions. The matrix spanning the null space is as before:

$$\text{Null}\left(\underline{\underline{\mathbf{C}}}\right) := \begin{bmatrix} -\underline{\underline{\mathbf{F}}} \\ \underline{\underline{\mathbf{I}}}_f \end{bmatrix}$$

The reactions obtained from the atom-species matrix, provides no information about the mechanism that makes the change happening. It only provides a summary reaction, an overall picture on what is happening from a process-global viewpoint. Even on that level, the set of reactions being defined by the null space of the coefficient matrix is not unique. In many cases different sets of reactions are possible. If we define our initial table with the inverse set of species, so:

	H_2	CO_2	CO	CH_4	H_2O
H	2	0	0	4	2
O	0	2	1	0	1
C	0	1	1	1	0

The matrix spanning the null space now is:

$$\underline{\underline{\mathbf{N}}}^T := \begin{bmatrix} -2 & -1 \\ 1 & -1 \\ -2 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

In this case the two independent reactions are :

$$\begin{aligned} 2\,H_2 + 2\,CO &\leftrightarrows CO_2 + CH_4 \\ H_2 + CO_2 &\leftrightarrows CO + H_2O \end{aligned}$$

So choosing a different set of free variables by permuting the species set gives different spanning vectors for the same null space.

8.2.2 Relation between different sum reactions for the same system

How are the two reaction sets related. Both of them represent sum reactions, but are constructed on a different arrangement of the species. Since they both are sum reactions, there must exist a linear transformation between the two stoichiometric matrices. We notice that the atom-species coefficient matrix of the second case is generated by permuting the columns, in this case we reverse the order. So let $\underline{\underline{\mathbf{C}}}_a$ be the coefficient matrix of the first case, thus the using the species set:

$$\{H_2O, CH_4, CO, CO_2, H_2\}$$

and $\underline{\underline{\mathbf{C}}}_b$ the one using the inverted set:

$$\{H_2, CO_2, CO, CH_4, H_2O\}.$$

The second matrix is obtained by permuting the columns $\underline{\underline{\mathbf{C}}}_b := \underline{\underline{\mathbf{C}}}_a \, \underline{\underline{\mathbf{P}}}$. The reduced row-echolon form of the matrix $\underline{\underline{\mathbf{C}}}_i$ is $\underline{\underline{\mathbf{R}}}_i$ this for $i \in \{a, b\}$. The transformation between the two is then done with a matrix $\underline{\underline{\mathbf{L}}}$ remembering also that the columns are permuted:

$$\underline{\underline{\mathbf{L}}} \, \underline{\underline{\mathbf{R}}}_a \, \underline{\underline{\mathbf{P}}} = \underline{\underline{\mathbf{R}}}_b$$

Since the $\underline{\underline{\mathbf{R}}}_i$ are not square, the solution is a little bit more complicated. We multiply the equation with $\left(\underline{\underline{\mathbf{R}}}_a \, \underline{\underline{\mathbf{P}}}\right)^T$ producing on the left-hand side a square matrix of the defined product:

$$\begin{aligned}
\underline{\underline{\mathbf{L}}} \left(\underline{\underline{\mathbf{R}}}_a \, \underline{\underline{\mathbf{P}}}\right) \left(\underline{\underline{\mathbf{R}}}_a \, \underline{\underline{\mathbf{P}}}\right)^T &= \underline{\underline{\mathbf{R}}}_b \left(\underline{\underline{\mathbf{R}}}_a \, \underline{\underline{\mathbf{P}}}\right)^T \\
\underline{\underline{\mathbf{L}}} \left(\underline{\underline{\mathbf{R}}}_a \, \underline{\underline{\mathbf{P}}} \, \underline{\underline{\mathbf{P}}}^T \underline{\underline{\mathbf{R}}}_a^T\right)^T &= \underline{\underline{\mathbf{R}}}_b \left(\underline{\underline{\mathbf{R}}}_a \, \underline{\underline{\mathbf{P}}}\right)^T \\
\underline{\underline{\mathbf{L}}} \left(\underline{\underline{\mathbf{R}}}_a \, \underline{\underline{\mathbf{R}}}_a^T\right)^T &= \underline{\underline{\mathbf{R}}}_b \left(\underline{\underline{\mathbf{R}}}_a \, \underline{\underline{\mathbf{P}}}\right)^T
\end{aligned}$$

Note that the product $\underline{\underline{\mathbf{P}}} \, \underline{\underline{\mathbf{P}}}^T$ is identity. So:

$$\underline{\underline{\mathbf{L}}} = \underline{\underline{\mathbf{R}}}_b \left(\underline{\underline{\mathbf{R}}}_a \, \underline{\underline{\mathbf{P}}}\right)^T \left(\underline{\underline{\mathbf{R}}}_a \, \underline{\underline{\mathbf{R}}}_a^T\right)^{-1}$$

In our case the transformation matrix becomes:

$$\underline{\underline{\mathbf{L}}} := \begin{bmatrix} 1 & 2 & 0 \\ 1 & -1 & 0 \\ -1 & 2 & 1 \end{bmatrix}$$

Yet another approach: Partitioning the matrix $\underline{\underline{\mathbf{R}}}_b := [\underline{\underline{\mathbf{I}}}, \underline{\underline{\mathbf{F}}}_b]$ the left-hand-side expression can also be partitioned accordingly:

$$\underline{\underline{\mathbf{L}}}\,\underline{\underline{\mathbf{R}}}_a\,\underline{\underline{\mathbf{P}}} = \underline{\underline{\mathbf{R}}}_b := [\underline{\underline{\mathbf{I}}}, \underline{\underline{\mathbf{F}}}_b]$$

$$\left[\underline{\underline{\mathbf{L}}}\,\underline{\underline{\mathbf{R}}}_a\,\underline{\underline{\mathbf{P}}}_1, \underline{\underline{\mathbf{L}}}\,\underline{\underline{\mathbf{R}}}_a\,\underline{\underline{\mathbf{P}}}_2\right] = \underline{\underline{\mathbf{R}}}_b := [\underline{\underline{\mathbf{I}}}, \underline{\underline{\mathbf{F}}}_b]$$

making the transformation for the F-part:

$$\underline{\underline{\mathbf{L}}}\,\underline{\underline{\mathbf{R}}}_a\,\underline{\underline{\mathbf{P}}}_2 = \underline{\underline{\mathbf{F}}}_b$$

There is an alternative approach to the above, by utilising more the structure of the matrix equation:

$$\underline{\underline{\mathbf{L}}}\,\left[\underline{\underline{\mathbf{I}}}, \underline{\underline{\mathbf{F_a}}}\right]\begin{bmatrix}\underline{\underline{\mathbf{P}}}_{11} & \underline{\underline{\mathbf{P}}}_{12}\\ \underline{\underline{\mathbf{P}}}_{21} & \underline{\underline{\mathbf{P}}}_{22}\end{bmatrix} = \left[\underline{\underline{\mathbf{I}}}, \underline{\underline{\mathbf{F_b}}}\right]$$

$$\underline{\underline{\mathbf{L}}}\,\left[\underline{\underline{\mathbf{P}}}_{11} + \underline{\underline{\mathbf{F_a}}}\,\underline{\underline{\mathbf{P}}}_{21}, \underline{\underline{\mathbf{P}}}_{12} + \underline{\underline{\mathbf{F_a}}}\,\underline{\underline{\mathbf{P}}}_{22}\right] = \left[\underline{\underline{\mathbf{I}}}, \underline{\underline{\mathbf{F_b}}}\right]$$

So this makes

$$\underline{\underline{\mathbf{L}}}\,\left(\underline{\underline{\mathbf{P}}}_{11} + \underline{\underline{\mathbf{F_a}}}\,\underline{\underline{\mathbf{P}}}_{21}\right) = \underline{\underline{\mathbf{I}}}$$

and

$$\underline{\underline{\mathbf{L}}} = \left(\underline{\underline{\mathbf{P}}}_{11} + \underline{\underline{\mathbf{F_a}}}\,\underline{\underline{\mathbf{P}}}_{21}\right)^{-1}$$

The second equation gives essentially the transformation between the two null spaces:

$$\underline{\underline{\mathbf{L}}}\,\left(\underline{\underline{\mathbf{P}}}_{12} + \underline{\underline{\mathbf{F_a}}}\,\underline{\underline{\mathbf{P}}}_{22}\right) = \underline{\underline{\mathbf{F}}}_b$$

Here $\underline{\underline{\mathbf{P}}}_{22}$ is of special interest. If the permutation moves the rows out of the free variable space into the basis space, the matrix is zero, with which expression simplifies to:

$$\underline{\underline{\mathbf{L}}}\,\underline{\underline{\mathbf{P}}}_{12} = \underline{\underline{\mathbf{F}}}_b$$

8.2.3 Formal stoichiometry

Finaly we introduce a notation for the reactions by defining a formal equation-like object for the reaction:

$$\sum_i \nu_i\,\mathsf{S}_i = 0 \qquad (8.7)$$

whereby the index i runs over the species set. The stoichiometric coefficients are denoted by ν_i, which are negative for the reactants, positive for the products and zero for those species that do not participate in the reaction. The species are appearing as symbols $\mathtt{S}_i$, thus the above is not a "real" equation, but a formal representation of a reaction. The stoichiometric matrix is formed by spanning a table with the reactions as the row indicators and the species as the column headings.

8.2.4 Direction

In chemistry it is customary to indicate in which direction the reaction is going using a arrow as an indicator. In the above reaction representation the arrow is replaced by the left-right arrows thus avoiding the indication of the direction. Reason being that the procedure of finding a stoichiometric matrix does not provide an indication of the direction. If one looks at the two reaction sets, then what changed was the order of the species and consequently which of the variables are chosen to be free and later taken as the basis for the null-space representation. The direction is determined by the difference in the chemical potential of the reactants and the products at the conditions the reaction is taking place.

The chemical potential is the partial derivative of the internal energy with respect to the component molar masses, thus

$$\mu_i = \left(\frac{\partial U}{\partial n_i} \right)_{S,V,n_{j \neq i}}$$

. Base values can be obtained from tables usually listing them at some normal conditions for example 298.15 K and 101.325 Pa. For the species in the two reactions we find:

	H_2O	CH_4	CO	CO_2	H_2	units
μ	-228,59	-50,75	-137,15	-394,36	0	kJ/mol

The reaction goes into the direction of negative Gibbs energy. Thus for the first set we get:

$$\begin{array}{rcll} 2\,H_2O + 4\,CO & \Rightarrow & CH_4 + 3\,CO_2 & -228.05 \quad kJ/mol \\ H_2O + CO & \Leftarrow & CH_4 + 3\,H_2 & -315 \quad kJ/mol \end{array}$$

and for the second:

$$\begin{array}{rcll} 2\,H_2 + 2\,CO & \Rightarrow & CO_2 + CH_4 & -170.81 \quad kJ/mol \\ H_2 + CO_2 & \Leftarrow & CO + H_2O & -28.62 \quad kJ/mol \end{array}$$

8.3 Reaction rates

The subject is chemical kinetics. In order for a reaction to take place the reactants, being the species that participate in the reaction, need to physically meet. But this is usually not enough, they must be in the right geometrical position to each other and they may require a certain energy level in order to be able to interact sufficiently intensively so as to undergo the reaction. It seems logical that the rate with which the reaction is taking place is a function of the number of collisions of the different ”reaction ingredients” per unit time. The collision frequency rises as the concentration of the ingredients rises and it also rises as the velocity of the individual particles increases. The probability of a collision of the required ingredients decreases with the number of ingredients required. The latter relation suggests a power function of the ingredients scaled by the volume, which is what one usually takes as the first suggestion for the reaction law. This sounds quite OK, if one knows the mechanism of the reaction, meaning that one actually knows what ingredients do react. A reaction mechanism is often much more complex than it appears from a summary reaction equation as it comes from the previous section’s discussion. Very often an overall reaction splits into several steps each of which has its own dynamics. For example, literature suggest a great number of different mechanism for the water gas shift reaction. The simple overall reaction:

$$CO + H_2O \leftrightarrows CO_2 + H_2$$

may be split into different steps[1]:

$$\begin{aligned} H_2O + S &\leftrightarrows O \cdot S + H_2 \\ O \cdot S + CO &\leftrightarrows CO_2 + S \end{aligned}$$

[1] Temkin, M. I. The kinetics of some industrial heterogeneous catalytic reactions. In Advances in Catalysis; Eley, D. D., Pines, H., Weisz, P. B., Eds.; Academic Press: New York, 1979; Vol. 28; pp 173.

with the S being an active site on a given catalyst. Or another suggested mechanis is much more complex[2]

$$\begin{array}{ccc}
CO+S & \leftrightarrows & CO\cdot S \\
H_2O+S & \leftrightarrows & H_2O\cdot S \\
H_2O\cdot S+S & \leftrightarrows & OH\cdot S+H\cdot S \\
CO\cdot S+OH\cdot S & \leftrightarrows & HCOO\cdot S+S \\
HCOO\cdot S & \leftrightarrows & CO2+H\cdot S \\
2\,H\cdot S & \leftrightarrows & H_2+2\cdot S
\end{array}$$

Knowing the steps, the reaction rates are often formulated as a power function of the reactants' composition, whereby the exponent is the absolute value of the stoichiometric coefficient of the respective species. So if we only consider the overall reaction we would be tempted to write the forward reaction as a second order reaction in the species on the left-hand-side and the backward reaction as a second order in the species on the right-hand-side. This power function is scaled with a ”reaction constant”, which is not constant, but in contrary a very strong function of the temperature. All of which is given per unit volume, for homogeneous reactions that take place in a volume. For non-homogeneous systems where the reaction takes place on a surface of a catalyst, one may norm with the surface and provide the additional information on how much surface is available in a volume given a certain geometry or granularity of the catalyst.

The latter suggests that norming with volume or surface is a good idea, but also, from the null space computation we notice that there is only one variable per reaction necessary to define the rate at which it progresses. The rate for each species is then obtained by multiplying with the stoichiometric coefficient of the respective species denoted by $\nu_{i,r}$, where the i is the species index and the r the reaction index.

So the typical reaction law takes the form:

$$\tilde{\xi}_r := {k^r}_r(T)\, g_r(\underline{\mathbf{c}}) \tag{8.8}$$

$$g_r(\underline{\mathbf{c}}) := \prod_{i\in\{\text{reactants}\}} c_i^{|\nu_{i,r}|} \tag{8.9}$$

where $\tilde{\xi}$ is the rate of change of the extent of reaction. The extent of reaction itself ξ takes values between 0 and 1. The decorator indicates that

[2] Campbell, C. T.; Daube, K. A. "A surface science investigation of the water-gas shift reaction on copper(111)", J. Catal. 1987, 104, 109.

it is a production rate[3]. Since one extent of reaction is being defined for each reaction, the association with the reaction is typically shown as an index. So we would write $\tilde{\xi}_r$ for the r-th reaction. The $k(T)$ is the "reaction constant", which really is anything else than constant, but a strong function of the temperature, for which typically the Arrhenius relation applies:

$$k^r{}_r(T) := k^0{}_r \, e^{-E_{A_r}/RT} \tag{8.10}$$

where k^0 is the pre-exponential factor, E_A is the activation energy, R is the gas constant and T is the absolute temperature.

Whilst there is a lot of logic in the formulation of the kinetic laws, there exists no basic theory for their derivation. The kinetic gas theory provides insight into the basic mechanisms in gases and the basic structure but not more, besides that one very often really does not know much about the reaction steps or on a sufficient level of detail to get an appropriate expression for the reaction rates. The kinetic gas theory provides insight into the form of the Arrhenius equation, for example, confirming the need for an exponential dependency on the temperature. But then in some cases Arrhenius is not sufficiently precisely reflecting the behaviour and more consequently complex relations have been formulated. The probably best known one is Rice-Ramsperger-Kassel-Marcus relation. The consequence of all of this being that the reaction rate expressions or laws, as they are also called, are to be considered as empirical relations, so-called black-box models, though here with some whiteness added, making them a little grey.

8.4 Reactions in an isolated containment

If we take the above discussed view of separating the reaction from the physical containment viewing it to occur separately in an imaginary volume, then we can discuss the nature of the reaction system in separation. To get there a little easier, we assume the reaction takes place in an ideally stirred tank reactor, that is, in a completely mixed volume, where the mixing dynamics is infinitely fast. No feed and no product stream shall be connected, thus the system's model is:

$$\underline{\dot{\mathbf{n}}} = \underline{\tilde{\mathbf{n}}} \tag{8.11}$$

The vector of production rate is in the volume V and the stoichiometry $\underline{\underline{\mathbf{N}}}$:

$$\underline{\tilde{\mathbf{n}}} := V \underline{\underline{\mathbf{N}}}^T \, \underline{\tilde{\xi}} \tag{8.12}$$

[3] We have chosen to make a difference between accumulation, transport and production rate, by defining separate decorators. Apparently the idea is to keep them separate even though they all are measured in unit mass per unit time.

So what we achieve is that the accumulation term in the separate containment is identical with the production rate enabling us to study the nature of the reaction in isolation.

8.4.1 Case: Highly reactive intermediates

When studying reaction mechanisms it often occurs that the overall transposition of one set of species into another one is conceived as a sequence of reaction steps that form a reaction network. In this reaction network, some intermediate species appear. In many cases these intermediates are very reactive and have a short live time. This leads to the assumption that such species are only present in a (very) low concentration and quite constant during the reaction process. Based on this a simplification is commonly applied, which in its mathematical nature is similar to what is being discussed for fast heat transfer in 10.2.1. Here the assumption is that the amount of intermediate species is more or less constant. Thus rate equation is split into two parts, a part of constant intermediates and the other species. Mathematically this can be done by multiplying the reaction rate equation with a permutation matrix that changes the order of the species in the vectors. Making this permutation matrix be a block of two matrices, the split can be made visible at the same time. Let the permutation/splitting matrix (see A.1.1) be:

$$\underline{\underline{\mathbf{P}}} := \begin{bmatrix} \underline{\underline{\mathbf{P}}}_d & \underline{\underline{\mathbf{P}}}_i \end{bmatrix} \tag{8.13}$$

then applying to the rate definition equation we get:

$$\begin{bmatrix} \underline{\dot{\mathbf{n}}}_d \\ \underline{\dot{\mathbf{n}}}_i \end{bmatrix} := V \begin{bmatrix} \underline{\underline{\mathbf{N}}}_d^T \\ \underline{\underline{\mathbf{N}}}_i^T \end{bmatrix} \underline{\tilde{\xi}}(\underline{\mathbf{c}}) \tag{8.14}$$

With the intermediates assumed to not change with time, this simplifies to:

$$\begin{bmatrix} \underline{\dot{\mathbf{n}}}_d \\ \underline{\mathbf{0}} \end{bmatrix} := V \begin{bmatrix} \underline{\underline{\mathbf{N}}}_d^T \\ \underline{\underline{\mathbf{N}}}_i^T \end{bmatrix} \underline{\tilde{\xi}}(\underline{\mathbf{c}}) \tag{8.15}$$

Literature often refers to this assumption as a ”pseudo-steady state assumption”. These are two sets of equations in which the top set are differential equations and the bottom ones are algebraic equations. The second equation set can be used to solve for some of the compositions, usually the composition of the intermediate and substitute the result into the first equation thereby reducing the number of variables. Formally this could be written as:

$$\underline{\mathbf{c}}_i(\underline{\mathbf{c}}_d) := \mathrm{root}\left(\underline{\underline{\mathbf{N}}}_i^T \, \underline{\tilde{\xi}}(\underline{\mathbf{c}}_d, \underline{\mathbf{c}}_i)\right) \tag{8.16}$$

which then is used to simplify the differential equations:

$$\underline{\dot{\mathbf{n}}}_d = V \, \underline{\underline{\mathbf{N}}}_d^T \, \underline{\tilde{\xi}}(\underline{\mathbf{c}}_d, \underline{\mathbf{c}}_i(\underline{\mathbf{c}}_d)) = V \, \underline{\underline{\mathbf{N}}}_d^T \, \underline{\tilde{\xi}}'(\underline{\mathbf{c}}_d) \tag{8.17}$$

Let us illustrate the idea on a simple but common example:

$$\texttt{A} \xrightarrow{\tilde{\xi}_1} \texttt{B} \xrightarrow{\tilde{\xi}_2} \texttt{C} \xrightarrow{\tilde{\xi}_3} \texttt{D}$$

The three reactions are of first order:

$$\begin{aligned}
\tilde{\xi}_1 &:= {k^r}_1 \, c_A \\
\tilde{\xi}_2 &:= {k^r}_2 \, c_B \\
\tilde{\xi}_3 &:= {k^r}_3 \, c_C
\end{aligned}$$

which in more compactly reads in vector form:

$$\underline{\tilde{\xi}} := \underline{\underline{\mathbf{K}}} \, g(\underline{\mathbf{c}})$$

whereby the vector of functions $g\,(\underline{\mathbf{c}})$ is simply

$$g(\underline{\mathbf{c}}) := \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \underline{\mathbf{c}} \qquad\qquad := \underline{\underline{\mathbf{S}}} \, \underline{\mathbf{c}}$$

with $\underline{\mathbf{c}}$

$$\underline{\mathbf{c}}^T := \begin{bmatrix} c_A & c_B & c_C & c_D \end{bmatrix}$$

The stoichiometric matrix is easily constructed:

Reac	A	B	C	D
1	-1	1	0	0
2	0	-1	1	0
3	0	0	-1	1

It is assumed that both intermediate are very reactive, motivating a pseudo-steady-state assumption for these two species. We permute and split the vector of production rates so that the production rate for species `A` and `D` are the first two forming the dynamic block and the intermediates are in the second block for which the pseudo-steady state assumption is being made. The permutation/splitting matrix is then:

$$P := \left[\begin{array}{cccc} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ \hline 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right]$$

Thus the expression for the reaction rates is first permuted and split:

$$\begin{aligned}
\underline{\underline{\mathbf{P}}}\,\underline{\dot{\mathbf{n}}} &:= V\,\underline{\underline{\mathbf{P}}}\,\underline{\underline{\mathbf{N}}}^T\,\underline{\tilde{\xi}} \\
&:= V\,\underline{\underline{\mathbf{P}}}\,\underline{\underline{\mathbf{N}}}^T\,\underline{\underline{\mathbf{K}}}\,\underline{\underline{\mathbf{S}}}\,\underline{\mathbf{c}}
\end{aligned}$$

Making the assumption and substituting $\underline{\underline{\mathbf{S}}}\,\underline{\mathbf{c}}$:

$$\begin{aligned}
\left[\begin{array}{c} \dot{n}_A \\ \dot{n}_D \\ \hline 0 \\ 0 \end{array}\right] &:= V \left[\begin{array}{cccc} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ \hline 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right] \begin{bmatrix} -1 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} {k^r}_1 & 0 & 0 \\ 0 & {k^r}_2 & 0 \\ 0 & 0 & {k^r}_3 \end{bmatrix} \begin{bmatrix} c_A \\ c_B \\ c_C \end{bmatrix} \\
&:= V \begin{bmatrix} -1 & 0 & 0 \\ 0 & 0 & 1 \\ 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} {k^r}_1 & 0 & 0 \\ 0 & {k^r}_2 & 0 \\ 0 & 0 & {k^r}_3 \end{bmatrix} \begin{bmatrix} c_A \\ c_B \\ c_C \end{bmatrix} \\
&:= V \left[\begin{array}{c} -{k^r}_1\,c_A \\ {k^r}_3\,c_C \\ \hline {k^r}_1\,c_A - {k^r}_2\,c_B \\ {k^r}_2\,c_B - {k^r}_3\,c_C \end{array}\right]
\end{aligned}$$

The two bottom equations, the algebraic equations, we extract the composition of the two intermediates, for which we find:

$$c_B := \frac{{k^r}_1}{{k^r}_2}\,c_A \qquad\qquad c_C := \frac{{k^r}_2}{{k^r}_3}\,c_B := \frac{{k^r}_1}{{k^r}_3}\,c_A$$

which when substituted simplify the differential equations:

$$\begin{aligned}
\dot{n}_A = \tilde{n}_A &:= -{k^r}_1\,c_A \\
\dot{n}_D = \tilde{n}_D &:= +{k^r}_1\,c_A
\end{aligned}$$

which is what we expected.

8.4.2 Case: equilibrium reactions

Reactions are by definition reversible, because matter has internal energy, which in parts is in the form of internal oscillations, stretching and contracting of distances between atoms, rotations etc. Also molecules move around, have kinetic energy and if they collide the energy is dissipated in one or the other form. This may lead to a reversion of the reaction, or another reaction. If we formulate a reaction as irreversible, we assume that the back reaction is negligibly small. This is saying that in the stochastic framework of the particles, the likelihood of a reaction to reverse, that go over the potential hump in the energetically unfavourable direction is very, very small. If however the barrier is not very large and the two local equilibria of reactants and products are close, the reaction may go both ways forming a mixture of all species.

For the purpose of the derivation we assume that we have only one such reversible reaction, then the stoichiometric matrix takes a special form in that the respective vectors of stoichiometric coefficients for the forward and the backward reaction are the same except they have inverse signs. So let the forward reaction, shown as the reaction labelled with f be given by:

$$\sum_{i,f} \nu_{i,f}\, \mathtt{S}_i = 0$$

and for the backward reaction, shown as the reaction labelled with b:

$$\sum_{i,b} \nu_{i,b}\, \mathtt{S}_i = 0$$

Then the respective vectors are:

$$\underline{\boldsymbol{\nu}}_r := \Big[\nu_{i,r}\Big]_{\forall i} \qquad\qquad r \in \{f, b\}$$

and

$$\underline{\boldsymbol{\nu}}_f = -\underline{\boldsymbol{\nu}}_b$$

So the stoichiometric matrix is:

$$\underline{\underline{\mathbf{N}}} := \begin{bmatrix} \underline{\boldsymbol{\nu}}_f^T \\ \underline{\boldsymbol{\nu}}_b^T \end{bmatrix} := \begin{bmatrix} +\underline{\boldsymbol{\nu}}_f^T \\ -\underline{\boldsymbol{\nu}}_f^T \end{bmatrix}$$

The expression for the production rates then takes the form:

$$\begin{aligned} \underline{\tilde{\mathbf{n}}} &:= V \underline{\underline{\mathbf{N}}}^T \underline{\tilde{\xi}} \\ &:= V\, \underline{\underline{\mathbf{N}}}^T\, \underline{\underline{\mathbf{K}}}\, g(\underline{\mathbf{c}}) \\ &:= V\, \underline{\underline{\mathbf{N}}}^T \begin{bmatrix} k^r{}_1 & 0 \\ 0 & k^r{}_2 \end{bmatrix} g(\underline{\mathbf{c}}) \end{aligned}$$

Both the reaction constants are large. We scale with one of them, say $k^r{}_1$:

$$\frac{1}{k^r{}_1} \underline{\tilde{\mathbf{n}}} := V \begin{bmatrix} \underline{\boldsymbol{\nu}}_f & -\underline{\boldsymbol{\nu}}_f \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \frac{k^r{}_2}{k^r{}_1} \end{bmatrix} g(\underline{\mathbf{c}})$$

With ${k^r}_1$ being large, the left-hand-side is close to zero and we get:

$$\underline{\mathbf{0}} \approx V \begin{bmatrix} \underline{\boldsymbol{\nu}}_f & -\underline{\boldsymbol{\nu}}_f \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \frac{{k^r}_2}{{k^r}_1} \end{bmatrix} g(\underline{\mathbf{c}})$$

For each species we thus get an equation. For example for species i we get:

$$0 = \nu_{i,f} \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \frac{{k^r}_2}{{k^r}_1} \end{bmatrix} g(\underline{\mathbf{c}})$$

and for the species that appear in the reaction, the stoichiometric coefficient is non-zero. Thus for each of those we get the same equation, namely:

$$\begin{aligned} 0 &= \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \frac{{k^r}_2}{{k^r}_1} \end{bmatrix} g(\underline{\mathbf{c}}) \\ &= g_1(\underline{\mathbf{c}}) - \frac{{k^r}_2}{{k^r}_1} g_2(\underline{\mathbf{c}}) \end{aligned}$$

which reshapes into:

$$\frac{{k^r}_2}{{k^r}_1} = \frac{g_1(\underline{\mathbf{c}})}{g_2(\underline{\mathbf{c}})}$$

being the equilibrium relation.

8.5 Behaviour of reactive systems

The reaction is going on inside the system, thus if we now include this into our description as done in Integral Balance (4.2) expanding 4.2 with a reaction system we get:

$$\underline{\dot{\mathbf{\Phi}}} = \underline{\hat{\mathbf{\Phi}}} + \underline{\tilde{\mathbf{\Phi}}}$$

The derivation of 4.3 is identical as it only affects the behaviour in terms of the boundary. The production term just ”hangs on” as the integral of the local turn over of the species measured by the extent of reaction for each of the ongoing reaction mapped onto the species space by the stoichiometric matrix. The derived integral balance then is, when including the transposition term:

$$\int_V \frac{\partial \underline{\varphi}}{\partial t} dV = -\int_V \frac{\partial}{\partial \underline{\mathbf{r}}} \underline{\hat{\varphi}} \, dV + \int_V \underline{\tilde{\varphi}} \, dV$$

which implies that:

$$\frac{\partial \underline{\varphi}}{\partial t} = -\frac{\partial}{\partial \underline{\mathbf{r}}} \underline{\hat{\varphi}} + \underline{\tilde{\varphi}}$$

Again substituting a simple transfer law 5.2

$$\hat{\varphi} := -c\,\frac{\partial\,\pi}{\partial\,\underline{\mathbf{r}}}\,,$$

that yields

$$\int_V \frac{\partial\,\underline{\varphi}}{\partial\,t}\,dV = \int_V c\,\frac{\partial}{\partial\,\underline{\mathbf{r}}}\,\frac{\partial\,\pi}{\partial\,\underline{\mathbf{r}}}\,dV + \int_V \underline{\tilde{\varphi}}\,dV$$

and:

$$\frac{\partial\,\underline{\varphi}}{\partial\,t} = c\,\frac{\partial^2\,\pi}{\partial\,\underline{\mathbf{r}}^2} + \underline{\tilde{\varphi}}$$

This now describes a finite spacial domain in which diffusion and reaction takes place.

8.5.1 Lumped systems

The species mass balance equation for the reactive lumped system expands the 4.6 by a homogeneous reaction term, thereby implementing the uniformity condition for the intensive properties. Thus for the integral of the reaction term we simplify to:

$$\int_V \underline{\tilde{\varphi}}\,dV := \underline{\underline{\mathbf{N}}}s^T\,\underline{\tilde{\mathbf{\Phi}}}_s$$

Thus the balance for the reactive lumped systems becomes:

$$\underline{\dot{\mathbf{\Phi}}}_s := \underline{\underline{\mathbf{F}}}_s\,\underline{\hat{\mathbf{\Phi}}}_s + \underline{\underline{\mathbf{N}}}s^T\,\underline{\tilde{\mathbf{\Phi}}}_s \tag{8.18}$$

9

Energy and mass

Synopsis *The energy balance provides the dynamic information about the energy household of a plant. Noticeably, mass carries energy, and thus mass induces energy besides that energy is being transferred across the system's boundary in different forms, giving rise to define morphologies for energy as it is being transported.*

Energy is a curious quantity as Feynman (Feynman et al. (1966)) notes. *Its nature we really do not know, but we have observed that it is conserved, that is if we compute a numerical value for the energy before a process starts and then account for all bits and pieces of energy again after the process has finished, we end up with the same numerical value, that is if we have really included everything.*

Mass carries energy merely by its pure existence and its internal microscopic mechanical energies and the macroscopic mechanical energies. Mechanical energy comes in different forms, and includes kinetic energy and potential energy. On the microscopic scale this can be in the form of rotation, vibration, translation but also potential due to structural changes. On the macroscopic scale mass translates in space and gravitational or other fields that effect mass. Energy reflects into mass, as a compressed spring is slightly heavier than a relaxed spring. Equipped with this fact, it is though then no surprise that the balancing of energy involves mass. If we assume no conversion of mass into energy and vice versa, then the mass balance can be drawn up independent of the energy balance, whilst the opposite is not the case.

9.1 Energy balance - somewhat simplified

Since energy comes in different forms, one first has to ask the question of what form of energies need to be considered when modelling a particular system. For example one requires answers to the questions if the process itself is moving like a rocket, a car, a bicycle or the like. If so, then one needs to decide if one includes the movement viewed from a stationary co-ordinate system or if one decides to just simply ”sit” on it and implement a moving co-ordinate system that does not bother about the movement of the system itself, which, by the way, is what we do all the time, as the earth moves relative to any observer outside. Or does one have to consider heat stream, radiation, elasticity, or any other fields than the earth's gravitational field. And, how about other, very common fields such as magnetic fields, pressure fields (sound) and electrical fields. Chemical and biological systems mostly get away without the latter.

For the time being we shall limit ourself to gravitational fields and kinetic energy, besides the obvious need for the internal energy. With this limitation in mind, we define the total energy as the sum of the internal energy U, the kinetic energy and the potential energy:

$$E := U + K + P \tag{9.1}$$

So a system's energy is affected by mass flows and different forms of energy flows all of which on a closer look are driven by potential fields of one or the other kind. The consequence is that when drawing up the energy balance of the system, one has to consider a variety of possible interactions between the system and its environment. In a chemical engineering context this is often, but not always, limited to mass transfer, having internal energy, kinetic energy, potential energy and also has volume work to perform when crossing a systems boundary, simply because it is of finite volume. Besides mass the most frequent form of energy being transferred that is not bound to mass is heat, which is the effect of momentum transfer of mechanically interacting species particles, but also radiation and mechanical work that is done on the system. The latter could for example be the energy input through the mixing device, such as a stirrer. Electrical energy is also frequently encountered. So for chemical-biological systems the energy balance for an arbitrary system takes usually the form ():

$$\dot{E}_S = \sum_{\forall m} \alpha_{s,m}\, \hat{E}_m + \sum_{\forall q} \alpha_{s,q}\, \hat{q}_q + \sum_{\forall w} \alpha_{s,w}\, \hat{w}_w$$

or putting it into a more compact form:

$$\dot{E}_S = \underline{\underline{\mathbf{F}}}^{m}{}_{S}\, \underline{\hat{\mathbf{E}}} + \underline{\underline{\mathbf{F}}}^{q}{}_{S}\, \underline{\hat{\mathbf{q}}} + \underline{\underline{\mathbf{F}}}^{w}{}_{S}\, \underline{\hat{\mathbf{w}}}$$

whereby the $\underline{\underline{\mathbf{F}}}$ is again the incidence matrix (see). The index s indicates the system and the superscripts m, q, w indicate the nature of the network, being mass, heat and work.

This formulation is to be seen in the light of history. It includes three main terms each being a sum. The first sum runs over all mass streams using the index m. The second term is the sum over all heat streams, which would typically also include radiation. The index used is q. Note that we use here the same symbol for the index as for the quantity itself with the idea to enhance readability. The last term includes more or less all other energy transfer forms, foremost the volume work term associated with the mass transfer, but also the volume work associated with the system itself expanding or contracting. Yet another term associated with mass flow is due to friction, with *surface friction* being the most common relevant part. *Internal friction* in a stream converts mechanical energy into internal energy by "heating up" the fluid. It also includes the mentioned mechanical and electrical work term. If we lift out the volume work associated with the mass streams by decorating the work flow symbol with a "v", and the friction work with the decorator "if" and "sf", for the internal friction and the surface friction respectively, we may collect all the mass-flow related terms into one summation:

$$
\begin{aligned}
\dot{E}_S &= \sum_{\forall m} \alpha_{s,m} \, (\hat{E}_m + \hat{w}^v_m + \hat{w}^{if}_m + \hat{w}^{sf}_m) + \sum_{\forall q} \alpha_{s,q} \, \hat{q}_q + \sum_{\forall r} \alpha_{s,r} \, \hat{w}_r \\
&= \underline{\underline{\mathbf{F}}}^{m}{}_{S} \, \left(\underline{\hat{\mathbf{E}}} + \underline{\hat{\mathbf{w}}}^{v} + \underline{\hat{\mathbf{w}}}^{if} + \underline{\hat{\mathbf{w}}}^{sf})\right) + \underline{\underline{\mathbf{F}}}^{q} \, \underline{\hat{\mathbf{q}}} + \underline{\underline{\mathbf{F}}}^{w}{}^{r} \, \underline{\hat{\mathbf{w}}}^{r}
\end{aligned}
$$

Consequently we also need to change the summation of the remaining work streams.

Having made a couple of initial assumption, we continue by branching into different classes of processes characterised by different order-of-magnitude assumptions:

9.2 Assumption of non-moving

Adding the assumption that the process is not moving, the accumulation term is simplified, as the kinetic and the potential energy remain constant and thus what remains is the internal energy:

$$
\begin{aligned}
\dot{U}_S &= \sum_{\forall m} \alpha_{s,m} \, (\hat{E}_m + \hat{w}^v_m + \hat{w}^{if}_m + \hat{w}^{sf}_m) + \sum_{\forall q} \alpha_{s,q} \, \hat{q}_q + \sum_{\forall r} \alpha_{s,r} \, \hat{w}_r \\
&= \underline{\underline{\mathbf{F}}}^{m}{}_{S} \, \left(\underline{\hat{E}} + \underline{\hat{\mathbf{w}}}^{v} + \underline{\hat{\mathbf{w}}}^{if} + \underline{\hat{\mathbf{w}}}^{sf})\right) + \underline{\underline{\mathbf{F}}}^{q} \, \underline{\hat{\mathbf{q}}} + \underline{\underline{\mathbf{F}}}^{w}{}^{r} \, \underline{\hat{\mathbf{w}}}^{r}
\end{aligned}
$$

9.3 Assumption: Stationary and dominating mechanical processes

Quite frequently one models flow systems that show little side effects in terms of thermal effects, that exhibit no reactions, and are in addition operating in a stationary mode. In these cases all those terms that have to do with the flow are dominating and the heat flow terms are negligible. Since the process is stationary, the accumulation term is also negligible:

$$\begin{aligned} 0 &= \sum_{\forall m} \alpha_{s,m} \, (\hat{K}_m + \hat{P}_m + \hat{w}^v_m + \hat{w}^{if}_m + \hat{w}^{sf}_m) + \sum_{\forall r} \alpha_{s,r} \, \hat{w}_r \\ &= \underline{\underline{\mathbf{F}}}^m{}_s \, \left(\underline{\hat{\mathbf{K}}} + \underline{\hat{\mathbf{P}}} + \underline{\hat{\mathbf{w}}}^v + \underline{\hat{\mathbf{w}}}^{if} + \underline{\hat{\mathbf{w}}}^{sf}) \right) + \underline{\underline{\mathbf{F}}}^{w}{}^{r} \, \underline{\hat{\mathbf{w}}}^r \end{aligned}$$

The remaining work term is often reduced to a mechanical work flow only.

So what are examples of such systems? Good examples are pipe networks with bends, all type of valves, measurement facilities, contractions and expansions and not at least any type of pumping device, fans, compressors or the like. All of which are so common that this will need a bit more discussion, which we will do in .

9.4 Common reactive systems

Many processing units operate at nearly constant pressure. This includes most biological processes, but also many industrial processes. A standard example is a batch reactor that is open to the outside, in which we feed a couple of reactants, possibly in a solvent, which then react in the reactor to produce the product. The reactor is usually equipped with a type of heat exchanger, either a jacket or a set of internal coils, as many reactions are exothermic and thus energy in the form of heat must be removed in order to keep the temperature under control.

In fact most chemical reactions are characterised by a quite large difference of energy content between reactants and products, for which reason the temperature rises in the reaction mixture quite quickly, whilst this is to significantly lesser degree the case for biological systems. In the latter case one usually is below 10 W/liter, which is in the same order of magnitude as the mechanical energy input through stirring devices. Even smaller is usually the effect of the kinetic energy associated with the inflows and the outflows of material and the energy due to the gravitation and the same applies to the friction terms.

Focusing on the most common case, thus negligible potential, kinetic energy and friction, the energy balance reduces to:

$$\dot{U}_S = \underline{\underline{\mathbf{F}}}^m{}_S\,(\underline{\hat{\mathbf{U}}} + \underline{\hat{\mathbf{w}}}^v) + \underline{\underline{\mathbf{F}}}^q{}_S\,\underline{\hat{\mathbf{q}}} + \underline{\underline{\mathbf{F}}}^w{}_S\,\underline{\hat{\mathbf{w}}}$$

which for the system shown in 9.1 gives the matrices $\underline{\underline{\mathbf{F}}}^m | \underline{\underline{\mathbf{F}}}^q | \underline{\underline{\mathbf{F}}}^w$:

	$\underline{\underline{\mathbf{F}}}^m$			$\underline{\underline{\mathbf{F}}}^q$			$\underline{\underline{\mathbf{F}}}^w$
	$A\vert T$	$B\vert T$	$T\vert P$	$C\vert J$	$J\vert H$	$T\vert J$	$T\vert R$
A	−1						
B		−1					
T	1	1	−1			−1	−1
P			1				
R							1
J				1	−1	1	
C				−1			
H					1		

The system may change volume, but the other remaining work terms are assumed negligible, in particular we neglect the energy carried in by mixing units. The only term then remaining is the volume-work term associated with the change of the size of the system, which is $\hat{w}_{s|e} := -p\,\dot{V}$. The volume work terms of the flows are $\hat{w}^v_m := p_m\,\hat{V}_m$ and we only have one single heat flow that is relevant, thereby assuming that evapouration and condensation on the lid or heat losses to the environment are not essential in this application, a subject we have discussed in . The abstraction of this system is shown in 9.1

Thus for this simplified system we get:

$$\begin{aligned}\dot{U}_T &= \sum_{\forall m} \alpha_{T,m}\,(\hat{U}_m + p_m\,\hat{V}_m) + (-1)\,\hat{q}_{T|J} + (-p)\,\dot{V}_T \\ &= \underline{\underline{\mathbf{F}}}^m{}_T\,\left(\left[\hat{U}_m + p_m\,\hat{V}_m\right]\right) + \underline{\underline{\mathbf{F}}}^q{}_T\underline{\hat{\mathbf{q}}} + (-p)\,\dot{V}_T\end{aligned}$$

Above the heat flow has not been substituted with a heat flow model, whilst the volume work term associated with the change of the system's volume has been substituted. Reason being that we are after a reformulation of the energy balance: For each mass flow we now have two terms namely an internal energy and a volume work flow. Defining a new energy function, called enthalpy $H := U + pV$ we simplify the term in the sum over the mass streams. We can also look at the accumulation term. The differential enthalpy we get $dH = dU + dp\,V + p\,dV$ or as change of time $\dot{H} = \dot{U} + \dot{p}\,V +$

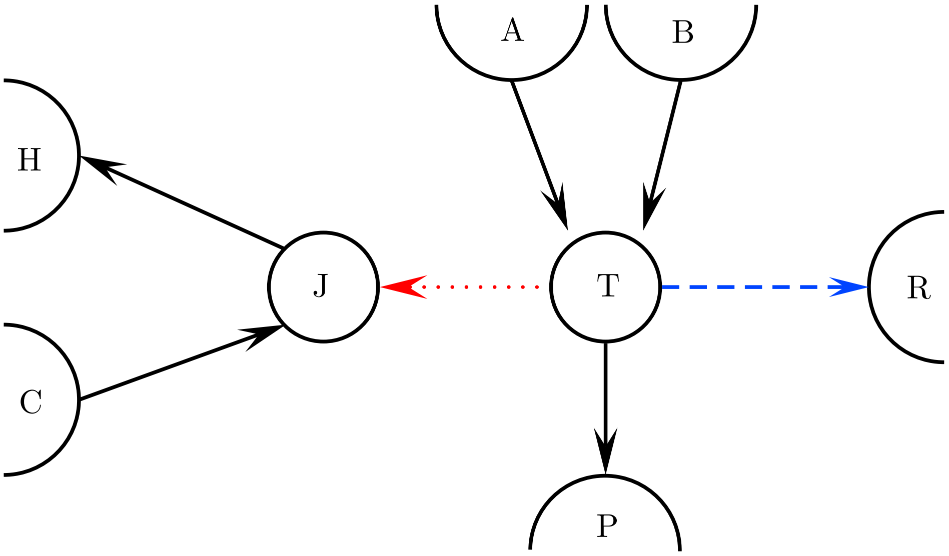


Figure 9.1: An abstraction of a jacketed tank reactor with two feeds. Assuming no heat losses, no stirrer energy input, jacket as a uniform capacity exchanging fluid with a cold and a hot reservoir and a product stream leaving the tank.

$p\,\dot{V}$. Which simplifies considering that the pressure is assumed constant. If we take the volume work term of the system to the left-hand-side we get:

$$\dot{U}_T + p\,\dot{V}_T = \underline{\underline{\mathbf{F}}}^{m}{}_{T}\,\underline{\hat{\mathbf{H}}} + \underline{\underline{\mathbf{F}}}^{q}{}_{T}\,\underline{\hat{\mathbf{q}}}$$

So for constant pressure systems the left-hand-side is the change of enthalpy of the system because:

$$dH := d(U + p\,V) = dU + dp\,V + p\,dV = dU + p\,dV$$

and we get:

$$\dot{H}_T = \underline{\underline{\mathbf{F}}}^{m}{}_{T}\,\underline{\hat{\mathbf{H}}} + \underline{\underline{\mathbf{F}}}^{q}{}_{T}\,\underline{\hat{\mathbf{q}}} \tag{9.2}$$

This equation evolved from the initial assumption of a lumped system and the assumption of negligible kinetic, potential energy effects and no friction associated with the mass flows. In addition we now also introduced the assumption of constant pressure.

So far the system is represented in the space defined by the component mass and the energy latter either total energy, internal energy or enthalpy. Thus the change of view is the change of the "type of energy" in which we represent and thus also view the system. When discussing systems from the thermodynamic view, one often talks about "measurable" quantities such as temperature, pressure and composition. Textbook derivations aim at representing the model in a set of these variables, mostly in those that can be measured or directly observed or are accessible. From a mathematical/system's point of view this implies that one transforms the behaviour description into another state space, whereby the information content is not changing, but the viewpoint is.

The objective is thus to change the state space from now being $\underline{\mathbf{n}}$ and H into $\underline{\mathbf{n}}$ and T. So we are asked to do a variable transformation:

$$\begin{aligned} dH = & \left(\frac{\partial H(T,p,\underline{\mathbf{n}})}{\partial T}\right)_{p,\underline{\mathbf{n}}} dT + \left(\frac{\partial H(T,p,\underline{\mathbf{n}})}{\partial p}\right)_{T,\underline{\mathbf{n}}} dp \\ & + \left(\frac{\partial H(T,p,\underline{\mathbf{n}})}{\partial \underline{\mathbf{n}}^T}\right)_{T,p} d\underline{\mathbf{n}} \end{aligned}$$

With the pressure being constant, the second term drops out and we get:

$$\begin{aligned} \left(\frac{\partial H_T(T_T,p_T,\underline{\mathbf{n}}_T)}{\partial T_T}\right)_{p_T,\underline{\mathbf{n}}_T} \frac{dT_T}{dt} + \left(\frac{\partial H_T(T_T,p_T,\underline{\mathbf{n}}_T)}{\partial \underline{\mathbf{n}}_T^T}\right)_{T_T,p_T} \frac{d\underline{\mathbf{n}}_T}{dt} \\ = \underline{\underline{\mathbf{F}}}^m{}_T \,\hat{\underline{\mathbf{H}}} + \underline{\underline{\mathbf{F}}}^q{}_T \,\hat{\underline{\mathbf{q}}} \end{aligned} \tag{9.3}$$

To keep things under control, we now carry along all the variables for which the enthalpy is a function of. Whilst this makes writing and reading somewhat cumbersome, it is necessary to keep track of the representation space.

At this point it is handy to remember that the enthalpy is a thermodynamic state function which with the given assumptions is being conserved and thus is an Euler-one-type of function with respect to mass. So the superposition applies and we can write:

$$H(T,p,\underline{\mathbf{n}}) = \left(\frac{\partial H(T,p,\underline{\mathbf{n}})}{\partial \underline{\mathbf{n}}^T}\right)_{T,p} \underline{\mathbf{n}}$$

The partial derivatives are called the partial molar enthalpies for which we introduce a short notation:

$$\underline{h}(T,p,\underline{\mathbf{n}}) = \left(\frac{\partial H(T,p,\underline{\mathbf{n}})}{\partial \underline{\mathbf{n}}^T}\right)_{T,p}$$

The other partial derivative can also be normed by the mass:

$$\left(\frac{\partial H(T,p,\underline{\mathbf{n}})}{\partial T}\right)_{p,\underline{\mathbf{n}}} = m\, c_p(T,p,\underline{\mathbf{n}})$$

whereby the such normed partial derivative of the enthalpy with respect to the temperature is called the specific heat capacity at constant pressure for which we use the symbol c_p. The total mass is to be computed from the vector of molar masses and the mole-masses $\underline{\boldsymbol{\lambda}}_m$:

$$m := \underline{\boldsymbol{\lambda}}_m^T \,\underline{\mathbf{n}}$$

It should be noted that the specific heat capacity is a function of the same variables as the enthalpy, thus this is the specific heat capacity of the mixture.

Looking at this , we have now two terms on the left-hand side. But inspection shows immediately that we have an equation for the change of mass,

$$\frac{d\,\underline{\mathbf{n}}}{d\,t} = \underline{\underline{\mathbf{F}}}^{m}{}_{T}\,\underline{\mathbf{n}} + \underline{\underline{\mathbf{N}}}^{T}\,V_T\,\underline{\tilde{\xi}}_T \tag{9.4}$$

for which we also used the definition of the production term implied in . This above equation we can substitute:

$$m_T\,c_p(T_T,p_T,\underline{\mathbf{n}}_T)\,\frac{d\,T_T}{d\,t} + \underline{\mathbf{h}}_T^T(T_T,p_T,\underline{\mathbf{n}}_T)\,\left(\underline{\underline{\mathbf{F}}}^{m}{}_{T}\,\underline{\mathbf{n}} + \underline{\underline{\mathbf{N}}}^{T}\,V\,\underline{\tilde{\xi}}\right)$$
$$= \underline{\underline{\mathbf{F}}}^{m}{}_{T}\,\underline{\hat{\mathbf{H}}} + \underline{\underline{\mathbf{F}}}^{q}{}_{T}\,\underline{\hat{\mathbf{q}}}$$

Next step is quite obvious: one combines the sums on the left with the one on the right:

$$m_T\,c_p(T_T,p_T,\underline{\mathbf{n}}_T)\,\frac{d\,T_T}{d\,t} = \underline{\underline{\mathbf{F}}}^{m}{}_{T}\,\left[\left(\underline{\mathbf{h}}_m^T(T_m,p_m,\underline{\mathbf{n}}_m) - \underline{\mathbf{h}}_T^T(T_T,p_T,\underline{\mathbf{n}}_T)\right)\,\underline{\hat{\mathbf{n}}}_m\right]_{\forall m}$$
$$- \underline{\mathbf{h}}_T^T(T_T,p_T,\underline{\mathbf{n}}_T)\,\underline{\underline{\mathbf{N}}}^{T}\,V_T\,\underline{\tilde{\xi}}_T + \underline{\underline{\mathbf{F}}}^{q}{}_{T}\,\underline{\hat{\mathbf{q}}}$$

The product $\underline{\mathbf{h}}^T(T,p,\underline{\mathbf{n}})\,\underline{\underline{\mathbf{N}}}^T = \left(\underline{\underline{\mathbf{N}}}\,\underline{\mathbf{h}}(T,p,\underline{\mathbf{n}})\right)^T$ is a linear combination of the partial molar enthalpies, whereby the weights are the stoichiometric coefficients of the individual reactions. These are vectors appearing as rows in the stoichiometric matrix. Defining vectors as column vectors, the vector of stoichiometric coefficients for reaction r is denoted by $\underline{\boldsymbol{\nu}}_r^T$, and multiplying it with the vector of partial molar enthalpies gives the reaction enthalpy:

$$\Delta h_r := \underline{\boldsymbol{\nu}}_r^T\,\underline{\mathbf{h}}(T,p,\underline{\mathbf{n}})$$

thus for the product $\underline{\underline{\mathbf{N}}}\,\underline{\mathbf{h}}(T,p,\underline{\mathbf{n}})$:

$$\underline{\boldsymbol{\Delta} h} := \underline{\underline{\mathbf{N}}}\,\underline{\mathbf{h}}(T,p,\underline{\mathbf{n}})$$

Substitution gives us the text-book equation:

$$m_T\,c_p(T_T,p_T,\underline{\mathbf{n}}_T)\,\frac{d\,T_T}{d\,t} = \underline{\underline{\mathbf{F}}}^{m}{}_{T}\,\left[\left(\underline{\mathbf{h}}_m^T(T_m,p_m,\underline{\mathbf{n}}_m) - \underline{\mathbf{h}}_T^T(T_T,p_T,\underline{\mathbf{n}}_T)\right)\,\underline{\hat{\mathbf{n}}}_m\right]_{\forall m}$$
$$- \underline{\boldsymbol{\Delta} h}_T^T\,V_T\,\underline{\tilde{\xi}}_T + \underline{\underline{\mathbf{F}}}^{q}{}_{T}\,\underline{\hat{\mathbf{q}}}$$

The reaction-related term is often called an ”energy production” term, which ought to cause some excitation in the reader’s mind, as energy is conserved and cannot be generated! So what is fishy here? Well we simply have moved terms from the accumulation to the other side. So this is not really any more an enthalpy balance, but it describes the process in a state space of a secondary, but intensive variable ”temperature”, which is thought to be often measurable instead of in the enthalpy state space. Along these lines it should be noted that the mass flow now is weighted with the difference of the enthalpy in the respective stream minus the enthalpy of the container's content. Thus for the out-flowing flow, this difference is zero, which makes sense, as a mass outflow does not affect the fluid body it comes from as long as it can be seen as uniform in the intensive properties. Having more of a look into the logics of the equation, it says:

- As there is more content, things go slower
- As the content has a higher specific heat capacity, things go slower
- As the molar partial enthalpies of the incoming streams are bigger, the effect of the stream on the temperature increases correspondingly
- As the inflow increases, the effect on the temperature increases
- If the reaction is exothermic, thus $\Delta h < 0$ the temperature increases
- Heat loss decreases the temperature

all of which sounds rather OK if we compare it with our common experience.

At this point text books go often a step further by giving an expression for the partial molar enthalpy:

$$h(T,p,\underline{\mathbf{n}}) := \int_{T_{ref}}^{T} c_p(T,p,\underline{\mathbf{n}})\, dT$$

The used c_p is a function of the composition, which is the composition of the mixture. Often one assumes that the mixing effects are negligible and one estimates the specific heat capacity as a linear combination of the component molar mass and the specific heat capacities of the pure species. Indicating pure species properties with a black bullet decorator:

$$\underline{\mathbf{h}}(T,p,\underline{\mathbf{n}}) := \underline{\mathbf{h}}(T,p)^{\bullet}$$

So if we now further assume that the heat capacity is not a function of the temperature, or that the temperature dependency in the given interval can be neglected:

$$\underline{\mathbf{c}}_p(T,p,\underline{\mathbf{n}}) := \underline{\mathbf{c}}_p{}^{\bullet}(p)\,,$$

then we get

$h(T, p, \underline{\mathbf{n}}) := c_p{}^{\bullet}(p)\,(T - T_{ref})$ and:

$$\begin{aligned}\underline{\mathbf{c}}_p{}^{\bullet}(p)_T{}^T\ \underline{\mathbf{n}}_T\,\frac{d\,T_T}{d\,t} = &\underline{\underline{\mathbf{F}}}^m\left[\left(c_p{}^{\bullet}(p_m)\,(T_m - T_{ref}) - c_p{}^{\bullet}(p_T)\,(T_T - T_{ref})\right)\,\hat{\underline{\mathbf{n}}}_m\right]_{\forall m}\\ &- \underline{\boldsymbol{\Delta} h}_T^T\,V_T\,\tilde{\underline{\xi}}_T + \underline{\underline{\mathbf{F}}}^q{}_T\hat{\underline{\mathbf{q}}}\end{aligned}$$

which for the pressure in the flows being equal to the pressure in the tank simplifies to:

$$\begin{aligned}\underline{\mathbf{c}}_p{}^{\bullet}(p)_T{}^T\ \underline{\mathbf{n}}_T\,\frac{d\,T_T}{d\,t} = &\underline{\underline{\mathbf{F}}}^m\left[\left(c_p{}^{\bullet}(p_T)\,(T_m - T_T)\right)\,\hat{\underline{\mathbf{n}}}_m\right]_{\forall m}\\ &- \underline{\boldsymbol{\Delta} h}_T^T\,V_T\,\tilde{\underline{\xi}}_T + \underline{\underline{\mathbf{F}}}^q{}_T\hat{\underline{\mathbf{q}}}\end{aligned}$$

which spans together the new state space with the component mass balance , as desired.

9.5 More on spaces

So far we have used component mass and energy as the state space in which we constructed the mathematical model. We also have talked about spacial co-ordinates and implicitly time. At one point we promised to carry all the variables along, but then if we are checking very carefully we often did not include either the spacial co-ordinates or the time; so for example equation . Since we model a lumped system it is by definition not a function of the position, but it is certainly a function of time:

$$\dot{H}_T(T, p, \underline{\mathbf{n}}; t) = \sum_{\forall m} \alpha_{s,m}\,\hat{H}_m(T_m, p_m, \underline{\mathbf{n}}_m; t) - \hat{q}_{T|J}(T_T, T_J; t)\,.$$

Classical "thermodynamics" is not dynamic, but static. Thus the introduction of time should make us enter the domain of non-equilibrium thermodynamics.

The subject enters also in the material description, thus in the modelling of the involved materials. These elements include most of the material properties, such as conductivities of heat, diffusivity of mass, density, chemical potential, heat capacity etc. etc. Most of the modelling we have done is on the macroscopic scale, whilst the material properties reflect the small-scale behaviour. If one dives into the smaller scale, one will again meet similar representation that often are based on control volumes and their respective content. If we zoom in time scale fitting the smaller length scale, also these systems are described by the dynamic conservation principles. As one zooms in scale by scale, one eventually leaves the domain of continuous mass and one enters the domain of particles, which may be molecules and further atoms. The ab initio computation methods allow us to compute models on the nano-scale and below. Schrödinger equations can be solved at least approximately for atoms and their components, combine them to molecules and density function theory enables us to compute for example the ternary structure of molecules and mixtures. Based on these models it is now possible to estimate the energy surface and the main properties of mixtures.

The models for the scales form layers, or like an onion a layer of shells. Each layer differ in around 2-3 orders of magnitude both in the time scale and the length scale. On each level, one assumes the smaller system to reach an equilibrium very fast, so that it can be considered event-dynamic. The behaviour of the underlying system is then approximated by a surrogate model, replacing the small-scale model with a simpler integral behaviour. This forms thus a hierarchical modelling system.

The surrogate models have often a mechanistic background that is based on an ideal behaviour of the modelled system, such as the Van der Waals model for the pressure-temperature-volume relation.

One of the main keys of computational thermodynamics is the increasing computer power. As the number of particles that we can compute increases, and the number of considered interactions increases, the accuracy of the prediction will also increase. So it appears to be only a function of time that laboratory experiments are substituted by computational experiments. This has a huge advantage of not being limited by physical limitations and it is also significantly cheaper and saver to perform.

The space of the event-dynamic systems "sits" orthogonal on the trajectory of the slower system, so-to-speak: in every point in time and space, the faster system is assumed to be in equilibrium. Thus equilibrium thermodynamics applies.

Trying to depict this, we define a state-space using the notation $\underline{\mathbf{x}} : \underline{\mathbf{x}} := \left[\underline{\mathbf{n}}^T, H\right]^T$ whereby the enthaply is spanned in the "thermodynamic" state space, namely temperature, pressure and composition. So if we compute the enthalpy at any point in time from the heat capacity, for example, we integrate in the thermodynamic state space over temperature, what we have referred to as the "orthogonal state space" above. If we try to picture this, we have the obvious problem of extending into more than just 3 dimensions, which makes it impossible to draw. Thus simplifying the picture showing just two thermodynamic state variables, say x_1, x_2 the trajectory may look like shown in 9.2 The underlying assumption is that the process is locally in equilibrium, this we always move on a equilibrium surface formed by the

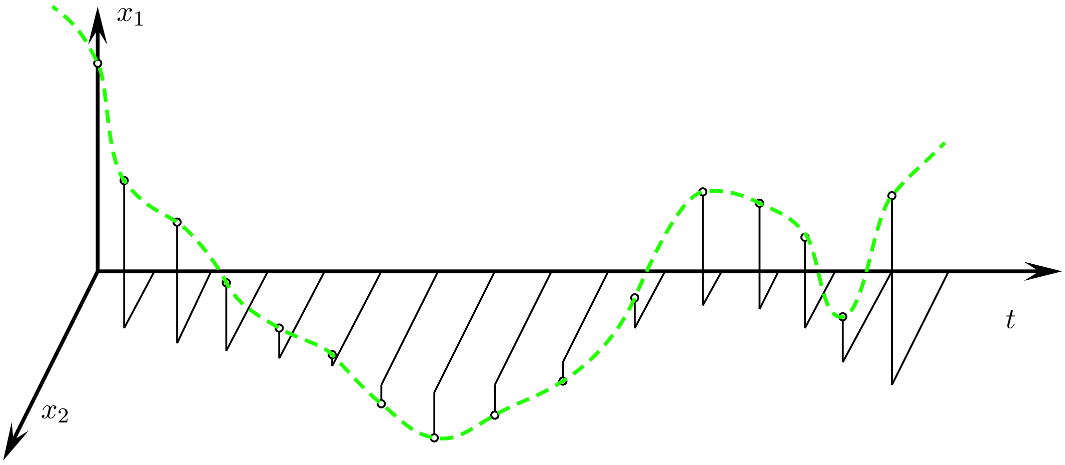


Figure 9.2: A trajectory in the dynamic thermodynamic state space

bundle of local equilibria. So if we compute the enthalpy by integrating the heat capacity over the temperature, for example then we integrate in one of the x-directions in a subspace forming a plane perpendicular to the respective time.

10

From distributed systems to transfer laws

Synopsis *The splitting into control volumes is not only motivated by discontinuities of some intensive properties, thus phase boundaries, but also by the relative dynamics of the sub-processes. This really introduces the notion of "time scales" and relative dynamics. Since equipment is build to transfer heat, or mass, for example it is no surprise that these transfer systems are usually very fast, they really are designed to have this property. This motivates the idea to use the difference of the dynamics in the different parts to make dynamic order-of-magnitude assumptions. Besides of a lot of other results, this yield, when applied to distributed models of such transfer systems, to quite simple transfer laws, which are in many of the engineers backpack.*

10.1 Nature of transfer systems

Transfer systems are sandwiched between two subsystems, providing the physical carrier for the extensive quantity being transferred between the two connected subsystems. In order to transfer extensive quantity, it must be conductive for the extensive quantity. Nature is such that it is often not just the one quantity that is being transferred, but often it is more than just one extensive quantities that is transferred. For example, a pipe transferring fluid due to a pressure difference from one end to the other, will invariably also conduct heat, if there is a temperature difference. Thought this conductive heat transfer is often not important, it nevertheless is there.

Transfer systems are distributed systems thus the state is a function of the position and time. The transfer is driven by a difference in the effort variables at the two opposite boundaries, which it shares with its environment. Apparently this may apply to any system. So we use the term "transfer systems" for system that are in the said "sandwich" position where they are relative to the environment very fast. In this case we model an idealised transport, which just "happens".

Such sandwiched systems may also be exhibit quite the opposite behaviour, namely that they block a transfer of an extensive quantity thereby isolating one part from the other. Mathematically this decouples systems in terms of the blocked extensive quantity transfer. In the case of fast transfer things look different, in that the capacity effect of the transfer relative to the transfer rate is disappearing. This leads to a mathematical description in which the state of the fast transfer system is a function of the boundaries only.

10.2 Heat transfer

Our daily experience is the main source of our understanding of temperature. But then do we really have an understanding of temperature? A simple example makes us quickly a little more cautious: Our hand is one of our body parts that we consider very sensitive to temperature. Touching an object gives us a sensation, which we usually interpret as temperature, but is it temperature? If we touch the metal leg of our chair or table or what ever else of a metal object in our room and then change over to touch a wooden object in the same room, such as the top of the table or the seat of the chair, does it feel like having the same temperature? Well no, the metal will ”feel” colder in comparison to the wooden object, but then are they of different temperature? If something is warmer than the other and they are directly in touch with each other, common experience indicates that they will exchange ”energy” until they are of the same temperature. With the two objects standing in the same room for an extended period of time, one would expect that a local equilibrium has been reached, at least close enough, so that the two objects are of the same temperature. So if we really would feel temperature, the two objects should feel the same, but they do not. Since both objects are typically colder than our body temperature, what we seem to feel is the heat flux and not the temperature per say, from which we conclude that what we feel is energy flow, thus ”heat” and not temperature[1].

In the early days of science after the invention of the steam engine, heat was seen as a massless material, which comes with the matter. It took quite some time for the scientists to come to terms with heat (Truesdell (1980)) and identify it as a measure closely related to entropy and internal energy. Today heat is defined as an energy in transition. Consequently heat cannot

[1] New research indicates that this picture needs to be further detailed as heat-detecting ion-channels in the skin are temperature dependent and seem to show significant different behaviours in different temperature ranges.This can then give indirect information about the temperature range **?**

be stored as heat, but as internal energy. Temperature reflects the different mechanical energies associated with matter, such as internal vibration, rotation and the molecules moving about. Heat is flowing spontaneously between two bodies of different temperature. The discussion on this subject has not subdued. In fact it is ongoing, though with a different flavour. It is argued that one should better introduce entropy in place of temperature, as entropy fits the every-days conception better (Hermann (2004)).

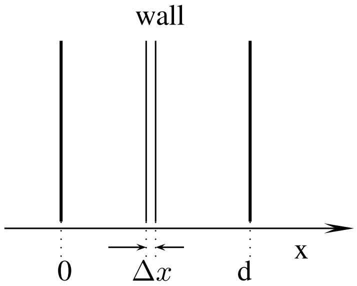


Figure 10.1: A one-dimensional heat diffusion problem

A mathematical description on how energy in the form of heat is being dissipated in a material can be derived from a the energy balance drawn up around a small, representative volume element in a problem-fitting co-ordinate system. In the most simple case this is a one-dimensional system such as a wall, where the temperature only changes across the wall. So the physical picture is shown in 10.1. The unit cell of the description is the small slice of the wall (see 4.4). With no material flowing in and out, the energy balance only includes the energy flows: one into the slice and one out, the difference of which is the accumulation in the slice:

$$\dot{H} = \hat{q}|_{x} - \hat{q}|_{x+\Delta x}$$

The heat flow is given proportional to the area (A), the thermal conductivity (k^q) and with the temperature being the effort variable, the negative temperature gradient $\frac{\partial T}{\partial x}$:

$$\hat{q}|_{x} := -\, A\, k^q \,\frac{\partial T}{\partial x}$$

The flow at the position $x + \Delta x$ is given by the first variation:

$$\begin{aligned} \hat{q}|_{x+\Delta x} &:= \hat{q}|_{x} + \left(\frac{\partial}{\partial x}\hat{q}\right)_{x} \Delta x \\ &:= \hat{q}|_{x} - \Delta V\, k^q \,\frac{\partial}{\partial x}\frac{\partial T}{\partial x} \end{aligned}$$

So the enthalpy balance then simplifies to:

$$\dot{H} = \Delta V\, k^q \,\frac{\partial^2 T}{\partial x^2}$$

and with the enthalpy being $H := \rho \Delta V \, c_p T$ the c_p, the specific heat capacity at constant pressure being $\frac{\partial H}{\partial T}$ we can easily take the limit to a differential change in the x-co-ordinate. The result is the familiar heat diffusion equation of Fourier:

$$\frac{\partial T}{\partial t} = \alpha \frac{\partial^2 T}{\partial x^2} \qquad , \alpha := \frac{k^q}{\rho \, c_p}$$

In contrast to what had been shown before, namely the integral method (see 4.2) this derivation is known as shell balance method and is widely celebrated in chemical engineering (see 4.4).

10.2.1 A fast heat transfer system

Insulation is the uninteresting case: nothing is flowing, at least ideally. Fast internal heat transfer it more interesting. In fact it is a very common model component because plants are usually designed to enable fast heat transfer such as in heat exchange equipment. The assumption is that when changing the temperature at one of the boundaries, the internal equilibrium is reached very quickly, much quicker than the changes occur at the boundaries. 10.2 shows the fast behaviour of a uniform wall after the temperature has changed on the right boundary from 0 to 1 in an instance. The profile, starting with a step, adjust to a line connecting the two boundary temperatures.

The result can be derived formally using the singular perturbation argument (D) using $\frac{1}{\alpha}$ as the singular perturbation parameter, which shows that the system reaches the equilibrium quickly in a large time scale. Thus this then gives:

$$\begin{aligned} \lim_{\alpha \to \infty} \frac{1}{\alpha} \frac{\partial T}{\partial t} &= 0 \\ &= \tfrac{d^2 T}{d x^2} \end{aligned}$$

Which of course is a straight line connecting the conditions at the boundaries. The two parameters, namely the slope a and the interception b is found readily by means of substitution:

$$\begin{aligned} &\text{for } x = 0 \; : \quad T(0) := a\,0 + b \\ &\text{for } x = d \; : \quad T(d) := a\,d + b \end{aligned}$$

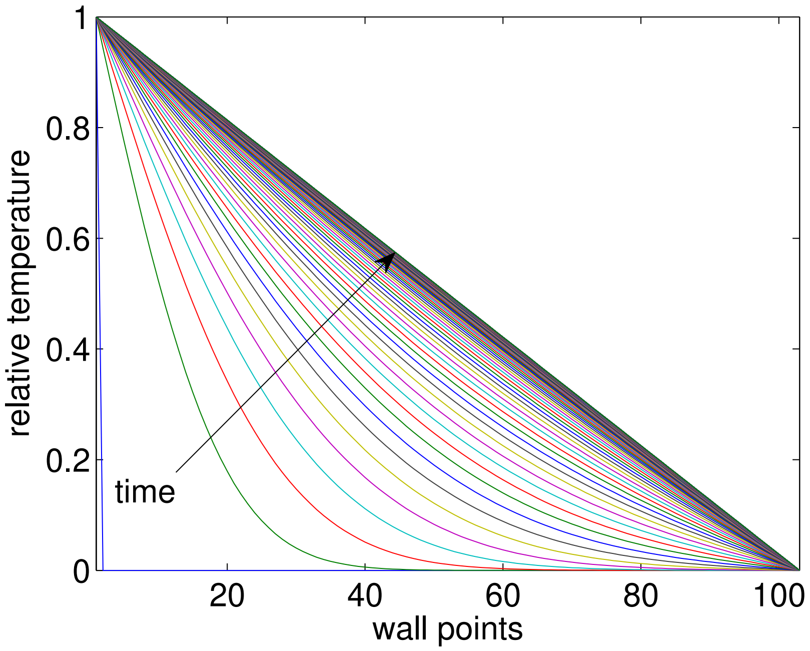


Figure 10.2: Keeping the temperatures constant on each end after changing the left boundary from 0 to 1 at time zero, the temperature approach the equilibrium, which is a straight line between the two boundary temperatures

which gives for:

$$b := T(0)$$
$$a := \frac{T(d) - T(0)}{d}$$

This is a straight line, but more so, that the state of the transfer system is solely given by the conditions at the two boundaries! So the state of the transfer system becomes obsolete, in the sense that it can be completely constructed from the boundary conditions.

10.3 Fluid flow

Besides the mass balance the core of modelling fluid flow is the momentum balance, which in differential form is known as the Navier-Stokes equation (Bird et al. (2001)). In contrast to the energy conservation and the mass conservation, the momentum conservation has four independent variables, namely the three spatial co-ordinates and time. This makes this equation one of the most difficult ones to handle. For incompressible fluids, though, the situation is not really so terribly difficult. Let us have a look:

$$\rho \left(\frac{\partial \underline{\mathbf{v}}}{\partial t} + \underline{\mathbf{v}} \bigtriangledown \underline{\mathbf{v}} \right) = - \bigtriangledown p + \bigtriangledown T + \underline{\mathbf{f}}$$

is the generic Navier-Stokes equation with ρ the density, $\underline{\mathbf{v}}$ the velocity, $\bigtriangledown$ the gradient operator, p the pressure, T the stress tensor, $\underline{\mathbf{f}}$ the external body forces per unit volume. One often finds that the left-hand-side is written as $\rho \frac{D \underline{\mathbf{v}}}{D t}$ which is called the material derivative.

The assumption of incompressible fluid leads to a significant simplification:

$$\overbrace{\rho \left(\underbrace{\frac{\partial \underline{\mathbf{v}}}{\partial t}}_{\substack{\text{mass}\\ \text{accelleration}}} + \underbrace{\underline{\mathbf{v}} \bigtriangledown \underline{\mathbf{v}}}_{\substack{\text{convective}\\ \text{accelleration}}} \right)}^{\text{inertia per volume}} = \overbrace{- \underbrace{\bigtriangledown p}_{\substack{\text{pressure}\\ \text{gradient}}} + \underbrace{\mu \bigtriangledown^2 \underline{\mathbf{v}}}_{\substack{\text{effect of}\\ \text{viscosity}}}}^{\text{divergence of stress}} + \underbrace{\underline{\mathbf{f}}}_{\substack{\text{other body}\\ \text{forces}}} \tag{10.1}$$

whereby the viscosity term is a diffusion of momentum. The other body force term may include mechanical action on the fluid as well as the effect of the gravitation field.

Let us step aside for a moment and think what happens in a straight piece of pipe: We feed the fluid in there at an elevated pressure compared to the outlet. In between, energy is dissipated through friction on the wall and internal between the molecules, as they move past each other at different relative velocity. These losses are entropy losses and make the wall and the fluid to change temperature. In many cases this is not very significant and thus it is usually quite simply ignored. If we do so, then the equations become even more simple. Formulating the conservation of mass (4.4), for example, the balance reduces to:

$$\frac{\partial \rho}{\partial t} + \bigtriangledown \cdot (\rho \underline{\mathbf{v}}) = 0 \quad \text{simplifies to} \quad \bigtriangledown \underline{\mathbf{v}} = 0$$

10.3.1 The steady-state

So let us do the same analysis as we did for the heat transfer: At steady state, the mass is not accelerated, thus the velocity is not changing with time, which makes the first term in 10.1 zero. For simplicity, we take a straight horizontal piece of pipe, which makes the convective acceleration term zero. The viscosity diffusion term is usually quite small, at least in a thin liquid, so let us assume it is negligible. We also assume not much is happening inside the pipe with respect to turbulences, velocity profiles or the like. For the straight pipe we have to consider only one co-ordinate. We choose to call it the x-co-ordinate and friction occurs also only in the x-direction and is a loss so the sign is negative. The position we measure from the high pressure end. Then:

$$0 = -\nabla p - \underline{\mathbf{f}} \quad \text{simplifies to} \quad 0 = -\frac{\partial\, p}{\partial\, x} - f_x$$

This equation is to be integrated over the length of the pipe:

$$\int_0^x \frac{\partial\, p}{\partial\, x}\, dx = -\int_0^x f_x\, dx$$

With the friction not being a function of the position, the result is a linear relation: the pressure drops linearly with length:

$$p(x) = p(0) - f_x\, x$$

Literature also offers another approach for the analysis: the energy balance. We draw up an overall energy balance enclosing the starting end to any position along the pipe. The green envelope signifies the domain on which the balance is drawn up in 10.3. On the left-hand-side, the flow is coming into the control volume and on the right it is leaving at the position x. If we assume an incompressible fluid, then the two flows will be identical. Further, mass carries energy and it does volume work on the system as it comes in and on the environment when it leaves. In form of an equation this is:

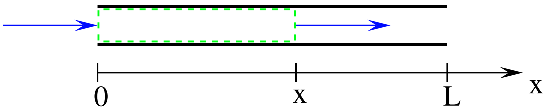


Figure 10.3: A one-dimensional flow in a straight pipe

$$\frac{d\, E}{d\, t} = \hat{E}(0) - \hat{E}(x) + \hat{w}^V(0) - \hat{w}^V(x) - \hat{w}^f$$

The energy consists of three parts, the internal energy, the kinetic energy and the potential energy. The first contribution is associated with the material, whilst the second is the effect of the mass moving. The third is the effect of the gravitational field. If other field effects are present then those have to be considered as well, but for the purpose of this exposition we leave it at the gravitational field. The last term is an integral friction loss term. Algebraically the energies are:

$$E := U + K + P \quad ; \quad K := \frac{m\,v^2}{2} \quad ; \quad P(x) := m\,g\,h(x)$$

Since internal energy is a fundamental extensive quantity, thus is of Euler degree one, we can represent the internal energy as a product of the internal energy per unit mass $\frac{\partial U}{\partial m}$ and the mass. To complete the picture, we need to provide an equation for the volume work, which is the pressure times the volumetric flow:

$$\hat{w}^V(x) := p(x)\,\hat{V}$$

The volumetric flow is constant as the fluid is incompressible and the density is assumed to be constant.

First we expand:

$$\begin{aligned}\dot{U} + \dot{K} + \dot{P} = {} & \hat{U}(0) + \hat{K}(0) + \hat{P}(0) \\ & - \left(\hat{U}(x) + \hat{K}(x) + \hat{P}(x)\right) \\ & + p(0)\,\hat{V} - p(x)\,\hat{V} - \hat{w}^f\end{aligned}$$

and then substitute the respective definitions:

$$\begin{aligned}\dot{U} + \frac{d\,\frac{m\,v^2}{2}}{d\,t} + \frac{d\,m}{d\,t}\,g\,h(x) = {} & \hat{U}(0) + \frac{\hat{m}\,v^2(0)}{2} + \hat{m}\,g\,h(0) \\ & - \hat{U}(x) - \frac{\hat{m}\,v^2(x)}{2} - \hat{m}(x)\,g\,h(x) \\ & + p(0)\,\hat{V} - p(x)\,\hat{V} - \hat{w}^f\end{aligned}$$

At steady state we have no dynamics and the left-hand-side is zero.

$$\begin{aligned}0 = {} & \hat{U}(0) - \hat{U}(x) + \frac{\hat{m}\,v^2(0)}{2} - \frac{\hat{m}\,v^2(x)}{2} + \hat{m}\,g\,h(0) - \hat{m}\,g\,h(x) \\ & + p(0)\,\hat{V} - p(x)\,\hat{V} - \hat{w}^f\end{aligned}$$

Assuming that the temperature effect is negligible, the internal energy does not change. If we divide by the volumetric flow, which by assumption is constant over the system (incompressible, no thermal effects and no change in composition), we get :

$$0 = \frac{\rho}{2}\,\left(v^2(0) - v^2(x)\right) + \rho\,g\,\left(h(0) - h(x)\right) + \left(p(0) - p(x)\right) - f_x\,x$$

If in addition we assumed that the diameter of the pipe does not change, which eliminates the kinetic energy term, and since the pipe is horizontal, the two potential terms are the same. So what is left is the volume work terms and the friction term:

$$\begin{aligned} 0 &= p(0)\,\hat{V} - p(x)\,\hat{V} - \hat{w}^f \\ &= (p(0) - p(x)) - f_x\,x \end{aligned}$$

which again gives us the same linear relation. This approach is largely due to Bernoulli, though no friction was included at the time. The isothermal energy balance is also called a mechanical energy balance.

10.3.2 Controlling and measuring flow

The facts reflected in Bernoulli's equation can be used in for measuring flows. We all are supposed to know that if we take two sheets of paper, preferably slightly curved and blow in between that the two pieces will move towards each other and not apart as one would first assume.The flow between the two papers generates a lower pressure between the papers, so the air pressure outside pushes the two papers together.

Putting constraint into in the form of a nozzle into a pipe, a plate with a hole or a smooth constriction generates the same effect, where the pressure being measured in the constriction or shortly after gives the information about the flow.

10.5 shows the situation for a simple nozzle, a smooth nozzle and below a smooth constriction. The contraction itself does not cause much friction. In the case of the nozzles, the loss is mainly due to the eddies that form at the outlet of the stream, which also cause the minimal flow cross-section to be somewhat down-stream of the nozzle. This stream contraction is called the vena contracta.

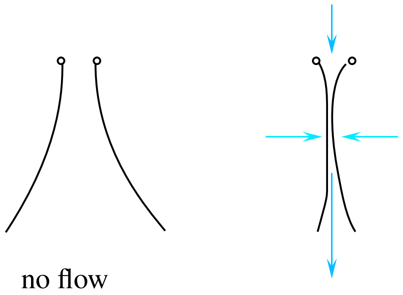


Figure 10.4: A Bernoulli experiment: two slightly bent papers hanged up on a pivot. Blowing in between makes them move together

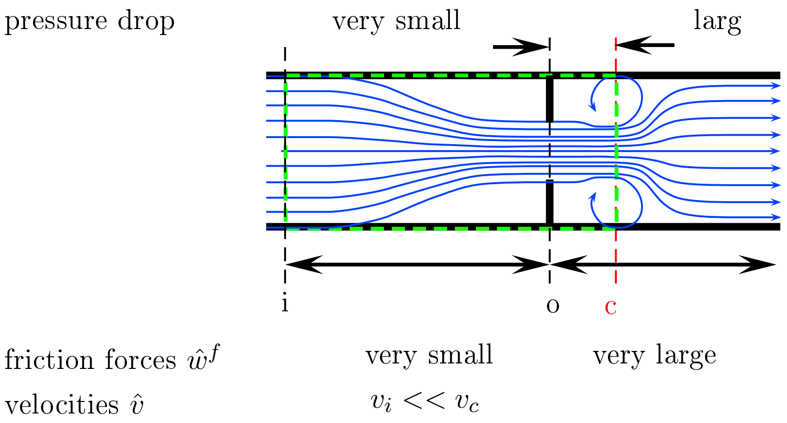


Figure 10.5: A nozzle in a pipe

The friction otherwise is comparable with the normal pipe friction. Since the pressure measurement is usually done in the domain of the vena contracta, but the diameter of the nozzle being used in the equation, the equations are not quite reflecting the physical situation. This issue is significantly lessened when constructing a smooth contraction, which in most cases is simply more expensive. Drawing up the energy balance around the green area again, we find

$$0 = \frac{\hat{m}\, v_i^2}{2} - \frac{\hat{m}\, v_c^2}{2} + \hat{m}\, g\, h_i - \hat{m}\, g\, h_c + p_i\, \hat{V} - p_c\, \hat{V} - \hat{w}^f$$

with the friction causing very little of the pressure drop from i to c, and the pipe being horizontal the flow rate is:

$$0 = \frac{\hat{m}\, v_i^2}{2} - \frac{\hat{m}\, v_c^2}{2} + p_i\, \hat{V} - p_c\, \hat{V}$$

Experiments show that the pressure drop on the expansion is very small and that the pressure in the vena contracta is close to the output pressure. Also the velocity at the inflow is much smaller than the one in the vena contracta. Thus:

$$\begin{aligned} 0 &= -\frac{\hat{m}\, v_c^2}{2} + p_i\, \hat{V} - p_o\, \hat{V} \\ &= -\frac{\rho\, v_c^2}{2} + (p_i - p_o) \end{aligned}$$

So if we know the diameter of the nozzle, we can compute the flow as the product of the velocity, computed from the pressure drop, and the cross sectional area even though in many applications the contraction is not done smoothly. At this point a correction factor is introduced that compensates for some of the made assumptions:

$$1/c\, v_c^2 = p_i \; - p_o \quad \text{and the volume flow is} \quad \hat{V} := c_V \sqrt{p_i - p_o}$$

where c_V is a characteristic for the device generating the stream contraction and it also includes the density of the material. These "constants" are usually obtained through calibration. Whilst one of the applications is a cheap flow-rate measurement, the same equation applies to a valve. The c_V is then called the valve constant though it is certainly not constant as the valve changes the cross section when manipulating the flow. The c_V is then often given for the completely open valve and another function is introduced which provides the information on how the flow changes with the valve position relative to the rest of the equation, meaning the valve constant and the square root of the pressure drop.

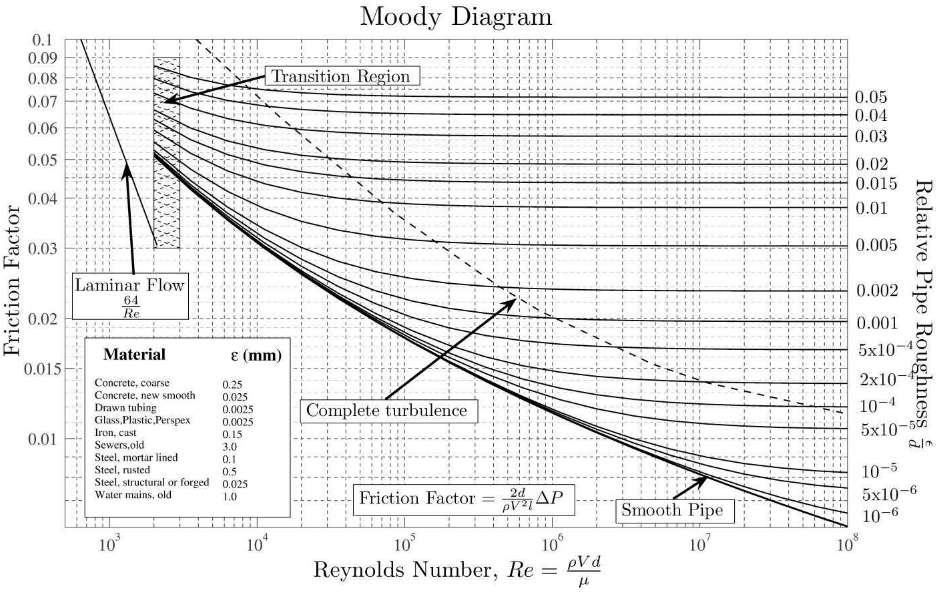
10.4 More on pipes

Pipes are a common means of transferring material from one part of a plant to another. It is quite common practice that pipes are essentially not modelled, implying that they simply appear in models as ideal flow units that have no effect what-so-ever on the plant other than moving material from one end to the other in zero time. Assuming an ideal behaviour thus also includes: no thermal effects, no time delay, and no mixing. Thus if one thinks of a modelling as a hierarchy in the sense of having more crude models on the top and finer models as one moves down in the hierarchy, then these models are towards the top. If one though requires pressure drop information, then one has to move a little down and consider the steady-state behaviour of flow in pipes a bit more carefully. Apparently this is a common engineering problem and correspondingly a lot of work has been done on this subject. The fact that this work was done over a very long period indicates that it is difficult whilst also important for the application. The subject is not closed as recent literature clearly indicates. Thus we cannot cover it comprehensively but just scratch the surface.

Flow in pipes is a complex. In the case of low velocities, the fluid behaves like moving in layers. The stream lines follow the pipe and behave nicely, to which we refer to as laminar flow. With friction acting on the interface between the wall and the fluid and between the fluid layers, the flow in the centre is fastest and decreases as one approaches the wall. The shape of the profile depends on the nature of the fluid. Ideally it will be parabolic, but with non-ideal fluid the shape changes mostly to flatter profiles. As one increases the flow rate, things become unstable and the flow becomes turbulent: the stream lines do not any more follow the shape of the pipe but form eddies of various sizes. The nature of the switching from laminar to turbulent flow is not completely understood, but commonly associated with small local disturbances. It is though speculated that the Navier-Stokes equation is covering the scales from above the particle level, i.e. every scale that can be described by macroscopic field theory. Such has been suggested as early as 1926 by L F Richardson and 1941 by A Kolmogorov, latter being considered a very influential piece of work.

On the simplistic level, a scaling law is being suggested, which also is in the spirit of Reynold and others that sought scale-independent representations. Scaling of the Navier-Stokes equation can be achieved by multiplying it with a factor $D/\rho\,\bar{v}^2$. D is a characteristic dimension. In pipes this is usually the hydraulic diameter. ρ is the density and $\bar{v}$ the mean velocity. Defining the scaled variables:

$$v' := \frac{v}{\bar{v}}\,, \quad p' := \frac{p}{\rho\,\bar{v}^2}\,, \quad f' := f\,\frac{D}{\bar{v}^2}\,, \quad \frac{\partial}{\partial\,t} := \frac{D}{\bar{v}}\frac{\partial}{\partial\,t}\,, \quad \nabla' := D\,\nabla$$



Material	**ε (mm)**
Concrete, coarse	0.25
Concrete, new smooth	0.025
Drawn tubing	0.0025
Glass,Plastic,Perspex	0.0025
Iron, cast	0.15
Sewers,old	3.0
Steel, mortar lined	0.1
Steel, rusted	0.5
Steel, structural or forged	0.025
Water mains, old	1.0

Figure 10.6: Moody diagram showing the change of the Darcy-Weisbach friction factor as a function of the Reynolds Number and the pipe roughness as a family parameter. (source: Wikipedia)

the Navier-Stokes 10.1 then takes the form:

$$\frac{\partial \underline{\mathbf{v}}'}{\partial t} + \underline{\mathbf{v}}' \bigtriangledown \underline{\mathbf{v}}' = -\bigtriangledown' p' + \frac{1}{\mathrm{Re}} \bigtriangledown'^2 \underline{\mathbf{v}}' + \underline{\mathbf{f}}'$$

So the scaling indicates that the flows behave similar if they have the same Reynolds number latter being defined as:

$$\mathrm{Re} := \frac{\rho \, D \, \bar{v}}{\mu}$$

In 1856 Henry Darcy, a French engineer suggested a model that links the volumetric flow to the pressure drop, simply as the volume flow being driven by the pressure gradient slowed down by friction in the pipe:

$$\hat{V} := \frac{k^M \, A}{\mu} \, (p_o - p_i)$$

where k^M is the "permeability" of momentum, A the cross sectional area, μ the viscosity and p_i, p_o the pressure in, and the pressure out of the pipe, respectively. The Darcy-Weisbach equation is a phenomenological equation where the pressure drop Δp is given by:

$$\Delta p = f \, \frac{L}{D} \, \frac{\rho \, \bar{v}^2}{2}$$

with L the length of the pipe. This provides a practical mean for estimating the pressure drop in a pipe. The factor f is the Darcy friction factor, which is, according to the similarity statement, a function of the Reynolds number.

The Moody diagram shows the relation between Reynolds number and the **Darcy-Weisbach friction factor**. The Darcy-Weisbach friction factor f is **four times larger than the Fanning friction factor**, latter being more commonly used in Chemical Engineering, whilst the former is more used in civil and mechanical engineering.

So for a horizontal pipe of length L the volumetric flow and the pressure we choose to relate by the Darcy-Weisbach relation:

$$p_i - p_o := f \, \frac{L}{\mathcal{D}} \, \frac{\rho \, v^2}{2}$$

For laminar flow, one gets the friction factor:

$$f := \frac{64}{\mathrm{Re}} := \frac{64 \, \mu}{\rho \, D \, v}$$

which for the pressure drop gives:

$$\begin{aligned} p_i - p_o &:= \frac{64\,\mu}{\rho\,D\,v}\,\frac{L}{\mathcal{D}}\,\frac{\rho\,v^2}{2} \\ &:= \frac{32\,\mu\,L}{D^2}\,v \\ &:= \frac{32\,\mu\,L}{D^2\,A}\,\hat{V} \\ &:= \frac{\mu}{k^M\,A}\,\hat{V} \end{aligned}$$

which is Darcy's equation.

For high Reynold numbers, where the friction factor becomes constant, one finds that the volumetric flow rate is a proportional to the square root of the pressure drop, thus

$$\begin{aligned} \hat{V} &:= -k^M\,A\,\sqrt{p_i - p_o} \\ &:= \sqrt{\frac{2\,D}{f\,L\,\rho}}\,A\,\sqrt{p_i - p_o} \end{aligned}$$

whereby f is constant with respect to the Reynolds number but gets bigger as the ratio surface roughness / hydraulic diameter increases.

10.4.0.1 Correlations

Apparently, the critical piece of information is the friction factor. The published Moody diagrams are the result of the collection of data from experiments over a long period, and have found wide acceptance and consequent applications. Since diagrams are not usable for computing, many empirical models are used for fitting the different behaviours, namely laminar, transition and turbulent flows. A quite large number of such empirical models exist and are readily found in the literature. The laminar one can be derived analytically and is mentioned above. For smooth turbulent flow, often the Blasius relation is used:

$$f := (100\,\mathrm{Re})^{-1/4}$$

For rough pipes one may use relations listed in appendix E.

10.4.1 A turbulent excursion

Turbulence is a fascinating subject. The public wiki has a nice exposure of the subject:

”In a turbulent flow, there is a range of scales of the time-varying fluid motion. The size of the largest scales of fluid motion (sometimes called eddies) are set by the overall geometry of the flow. For instance, in an industrial smoke stack, the largest scales of fluid motion are as big as the diameter of the stack itself. The size of the smallest scales is set by the Reynolds number. As the Reynolds number increases, smaller and smaller scales of the flow are visible. In a smoke stack, the smoke may appear to have many very small velocity perturbations or eddies, in addition to large bulky eddies. In this sense, the Reynolds number is an indicator of the range of scales in the flow. The higher the Reynolds number, the greater the range of scales. The largest eddies will always be the same size; the smallest eddies are determined by the Reynolds number.

What is the explanation for this phenomenon? A large Reynolds number indicates that viscous forces are not important at large scales of the flow. With a strong predominance of inertial forces over viscous forces, the largest scales of fluid motion are undamped – there is not enough viscosity to dissipate their motions. The kinetic energy must "cascade" from these large scales to progressively smaller scales until a level is reached for which the scale is small enough for viscosity to become important (that is, viscous forces become of the order of inertial ones). It is at these small scales where the dissipation of energy by viscous action finally takes place. The Reynolds number indicates at what scale this viscous dissipation occurs. Therefore, since the largest eddies are dictated by the flow geometry and the smallest scales are dictated by the viscosity, the Reynolds number can be understood as the ratio of the largest scales of the turbulent motion to the smallest scales.”

If visualised turbulences produce beautiful patterns.

10.7 [2] is a Landsat 7 image of clouds off the Chilean coast near the Juan Fernandez Islands (also known as the Robinson Crusoe Islands) on September 15, 1999, shows a unique pattern called a ”von Karman vortex street”. This pattern has long been studied in the laboratory, where the vortices are created by oil flowing past a cylindrical obstacle, making a string of vortices only several tens of centimetres long. Study of this classic ”flow past a circular cylinder” has been very important in the understanding of laminar and turbulent fluid flow that controls a wide variety of phenomena, from the lift under an aircraft wing to Earth’s weather.

Here, the cylinder is replaced by Alejandro Selkirk Island (named after the true ”Robinson Crusoe”, who was stranded here for many months in the early 1700s). The island is about 1.5 km in diameter, and rises 1.6 km into a

[2] Image and caption courtesy Bob Cahalan, NASA GSFC

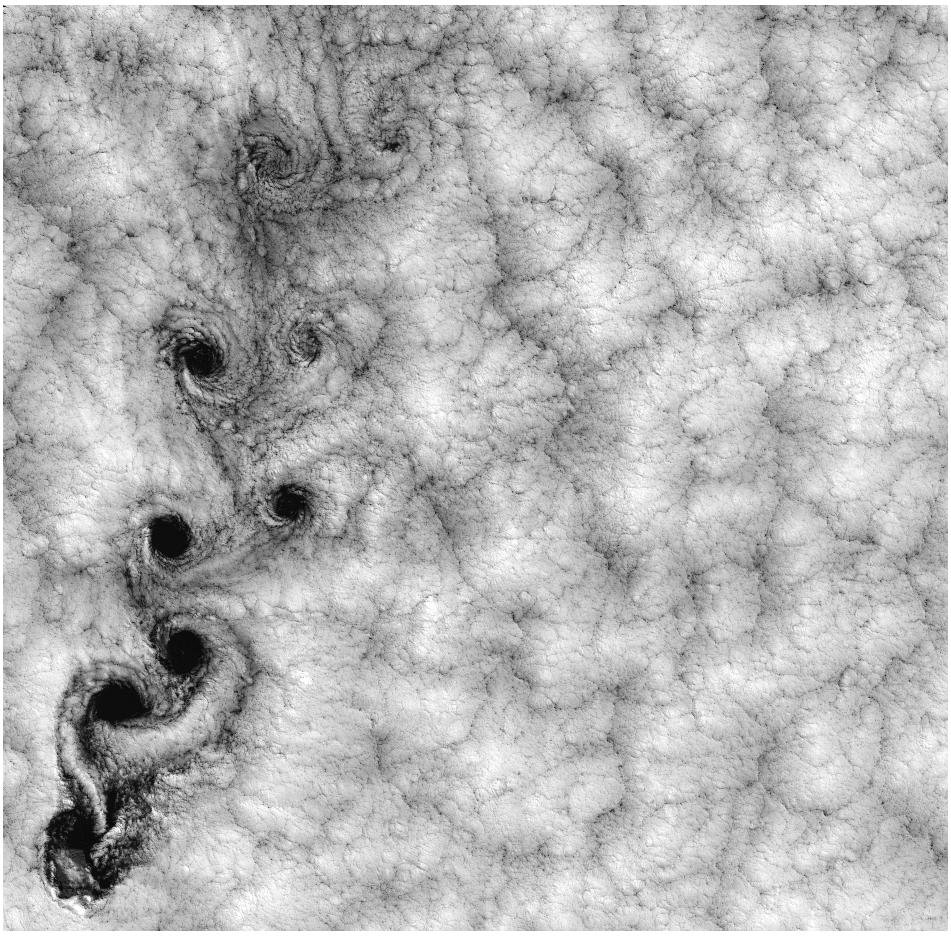


Figure 10.7: LandSat 7 image of clouds off the Chilean coast

layer of marine strato-cumulus clouds. This type of cloud is important for its strong cooling of the Earth's surface, partially counteracting the Greenhouse warming. An extended, steady equatorward wind creates vortices with clockwise flow off the eastern edge and counter clockwise flow off the western edge of the island. The vortices grow as they advect hundreds of kilometres downwind, making a street 10,000 times longer than those made in the laboratory. Observing the same phenomenon extended over such a wide range of sizes dramatizes the "fractal" nature of atmospheric convection and clouds. Fractals are characteristic of fluid flow and other dynamic systems that exhibit "chaotic" motions.

Both clockwise and counter-clockwise vortices are generated by flow around the island. As the flow separates from the island's leeward (away from the source of the wind) side, the vortices ”swallow” some of the clear air over the island. (Much of the island air is cloudless due to a local ”land breeze” circulation set up by the larger heat capacity of the waters surrounding the island.) The ”swallowed” gulps of clear island air get carried along within the vortices, but these are soon mixed into the surrounding clouds.

Landsat is unique in its ability to image both the small-scale eddies that mix clear and cloudy air, down to the 30 meter pixel size of Landsat, but also having a wide enough field-of-view, 180 km, to reveal the connection of the turbulence to large-scale flows such as the subtropical oceanic gyres. Landsat 7, with its new on-board digital recorder, has extended this capability away from the few Landsat ground stations to remote areas such as Alejandro Island, and thus is gradually providing a global dynamic picture of evolving human-scale phenomena.

10.5 Mass diffusion

Mass diffusion is the effect of small-scale anisotropic movement of species relative to others. In a fixed phase this may be a chemical species that moves relative to the material making up the fixed phase. In fluids it is the movement of one or several species relative to each other. Diffusion movement is thus defined relative to a co-ordinate system. Apparently this can get quite involved as the diffusing material is moving and thus contributes to the definition of the overall movement of the mixture one describes. So for a binary mixture, for example one may define a mole-average velocity:

$$v := \frac{c_1 \, v_1 + c_2 \, v_2}{c_1 + c_2}$$

which for multicomponent systems extents to:

$$v := \frac{\sum_i c_i \, v_i}{\sum_i c_i}$$

The velocity v_i for species i is thereby measured relative to a equipment-relative stationary co-ordinate system. The molar concentrations c_i are the weights of the averaging process. The sum of the molar diffusion fluxes relative to the molar average velocity is zero in any mixture. Thus in a binary mixture the two diffusion fluxes are equal (Bird et al. (2001)): one talks about equimolar counter diffusion, as one species moves in the opposite direction relative to the other. The flow of the individual species relative to the average velocity is then of the form:

$$j_1 := -\mathcal{D}_i \, \bigtriangledown_1 \varkappa_1$$

where x is here the mole fraction. Quite commonly, instead of using the mole fraction, one uses the molar composition, thus the number of moles normed by the volume. From our earlier discussion in 5.1 the chemical potential is a good candidate, but it is rarely used because it leads to a complex structure of the diffusion equation, which is easy to recognise when substituting a model for the chemical potential into the generic diffusion equation: 4.5. The extensive quantity is then the molar concentration or molar density and the chemical potential is the effort variable:

$$\frac{\partial c_1}{\partial t} = \mathcal{D}_1 \frac{\partial}{\partial \underline{\mathbf{r}}} \frac{\partial \mu_1}{\partial \underline{\mathbf{r}}}$$

and the model for the chemical potential, which links the chemical potential to the composition, is typically a logarithmic relation of the kind:

$$\mu_i := \mu_i^o + R\,T \,\ln \varkappa_i$$

with the ^{o}-decorator indicating the standard potential and the μ the gas constant, T the temperature and $\varkappa_i$ the mole fraction, which is the fraction of the molar concentration and the total molar concentration.

One also talks about diffusion when one has intense mixing, where the mechanics of the fluid flow generates eddies of very small scale as this is the case in (highly) turbulent flow. With these eddies shifting fluid particles passed each other, local diffusion can take place, whilst the eddy flows keep on changing the local environment. Molecular diffusion is often referred to as pure diffusion. It is largely driven by the species thermal movement, which has a preferred direction, determined by the gradient of the chemical potential, the temperature (thermo diffusion) or the pressure (Dufour diffusion). The latter two are in most technical application of negligible dimension.

10.5.1 Common diffusion model

As mentioned, the most common form being used is to make the composition the driving force, because it leads to seemingly simple equation, namely Fick's second law:

$$\frac{\partial c}{\partial t} = \mathcal{D} \frac{\partial}{\partial \underline{\mathbf{r}}} \frac{\partial c}{\partial \underline{\mathbf{r}}}$$

We have now dropped the index on the composition, assuming diffusion of one species in a seemingly stationary phase, which simplifies the question of relative velocity. This description thus fits diffusion in for example membranes but it is also used to describe phase-transfer behaviours, where it leads to an interesting problem.

10.5.1.1 Inside a single phase

If we stay within a phase, that is we do not cross a phase boundary, the mathematics of mass diffusion is the same as the one of heat diffusion. So making order-of-magnitude assumptions is handled in the same way and yields the same results: A fast diffusion system can be approximated by a linear stationary solution for the transfer system (see 10.2.1).

11

Going back and forth the phase boundary

Synopsis *One of the main arguments for splitting a spacial domain is the discontinuities of some intensive properties at the phase boundaries. Transport across the phase boundaries is thus a very common model component.*

11.1 A physical picture

On the microscopic scale molecules move or are moved about by external forces exerted by the immediate neighbours in the form of impulse transfer and external fields. As the condition changes in the environment, the movement changes, so since the conditions changes towards the boundary, the movement of the molecule also changes. The movement of the molecules is not quite random but it is skewed towards one or the other side, which one observes on the macroscopic scale as the conditions change along the spatial co-ordinate. Consequently the molecules are either attracted to the boundary or pushed away. The interface itself is thereby a non-sharp domain within which the conditions change more rapidly over the spatial co-ordinates as compared to either side of the coupled spatial domains. The gradient of all properties tend to change, in particular the effort variables.

On the macroscopic scale, then, the gradient changes discretely, has a jump at the boundary, which on this scale has no volume, but reduces to a surface. The immediate attached domain serve as the transport system in which the conditions change more gradually. Apparently, the description of these changes depends on the nature of the transport system. One of the main issues being if the material is a fluid or a solid and if it is a fluid, if it is stagnant or moving. Whilst the result is in its nature often the same, the path to the result varies.

The result is either a macroscopic model in the form of a film in which the conditions change or it is a stochastic model in which volume elements interact with each other across the interface. The existence of such a micro-volume transport can be derived from fluid models, primarily when one deals with turbulent flow. Then the small eddies can be seen as separate volume elements, that are whilst spinning, also moving and hitting the interface with a certain probability. Once they are at the interface they have a stochastic time interval during which they stay attached to the interface and interact with the neighbour across the boundary. This latter exchange is usually modelled very much the same as the continuous macroscopic model, namely with local diffusion.

These considerations have lead to the formation of different models describing the exchange across phase boundaries. One of the earliest documented one is due to Nernst.

11.2 Boundary layer theory

11.2.1 Nernst diffusion layer

The idea of a diffusion layer in fluids is attributed to Nernst (1888). It is based on the thought that the fluid is stagnant on the surface and that the effort variable changes gradually from the bulk of the fluid as one approaches the interface. Geometrically the stagnant film is relatively thin compared to the bulk dimension. Since the intensities change gradually, the definition of the film thickness is not readily done. One can think of different approaches, for example the thickness of the film is where the driving force has dropped down to 1 % or the like. Nernst suggested to use the gradient at the interface for the definition of the film thickness. Specifically, the film thickness is defined as the intersection point of a straight line with the slope of the the gradient of the effort variable and the axis interaction being the effort variable π at the interface.

The thickness of the diffusion film is defined by the straight line given by:

$$\left(\frac{d\,\pi}{d\,x}\right) := \frac{\pi_{\text{bulk}} - \pi_i}{d}$$

Whilst Nernst used it to define diffusion of ions, the model can be applied to various problems. One of the common celebrated problems are falling film adsorbers in which a gas is adsorbed in a fluid that runs down a tube in form of a fluid film. In this case the flux changes with time, as the bulk composition changes as one moves down the tube. The consequence being that the diffusion film gets thicker.

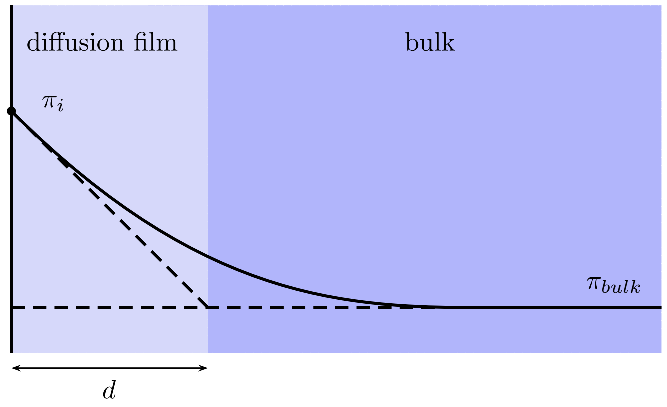


Figure 11.1: Nernst diffusion boundary layer

11.2.2 Two-film theory

The transfer of mass across a phase boundary is often described by two coupled diffusion processes. If the two phases are both solids, this view is natural. In the case where one or both phases are liquids, this view of the process' behaviour is based on Schlichting's boundary-layer school (Schlichting et al. (2004)): The model assumes the formation of two films on either side of the phase. Again in the case of stationary fluids, this is a rather natural assumption. Not so much though for a flowing fluid the arguments of which we leave the reader to extract from the transport-phenomena literature (Bird et al. (2001)).

For interfaces between fluids, the original idea was developed early by Lewis (1916) and Whitman (1923). This approach assumes that the turbulence dies out towards the interface and that the transfer is determined by two fictitious diffusion films one on each side of the interface. It is postulated that at the interface an equilibrium establishes as soon as the two faces come into contact. This implies that on the macroscopic scale, the effort variables are continuous. Thus on the small scale we assume the equilibrium to be established instantaneously, here at the interface.

The continuity condition for the effort variables generally applies to macroscopic system with the exception of shocks. With the interface itself having no capacity, its behaviour is of a event-dynamic system and the flux in is equal to the flux out for all the conserved quantities. The theory simply uses Nernst's diffusion layer picture twice, coupling the two films through the continuity condition of the effort variables and the flux condition on the interface assuming the interface does not have capacity.

11.2.2.1 Heat diffusion

The pure heat diffusion is the simplest process. This assumes that no mass is being transferred between the two phases, but purely energy in the form of heat. The driving force for the conductive heat transfer is the gradient in the temperature, the temperature being the effort variable. The energy balance equation for the boundary, being again assumed to be of negligible length scale and thus of event-dynamic nature thus exhibiting no capacity effects, provides the analytical statement. The energy balance for the interface I is then simply:

$$\dot{E}_I := \hat{q}_{-\epsilon} - \hat{q}_{+\epsilon} := 0$$

With no capacity of the interface, the left-hand-side becomes zero. And the flow of heat in is identical to the flow out. This is the flux condition at the surface. The second condition is the continuity of the effort variable, that is, the temperature on either side of the interface is the same.

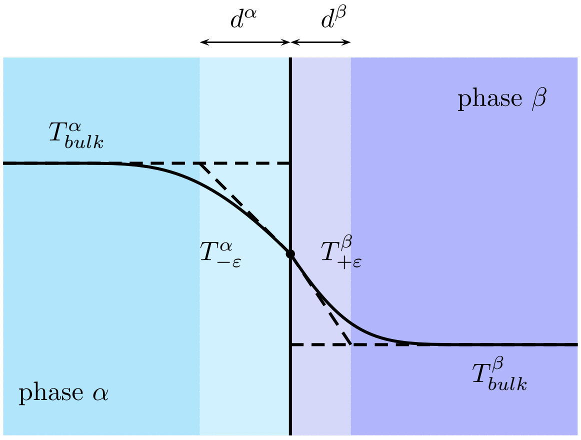


Figure 11.2: Pure heat transfer across a phase boundary

11.2.2.2 Mass diffusion

The situation is somewhat more challenging in the case of mass diffusion. The main reason is that whilst the effort variable is the chemical potential, historically it is the gradient in the composition that is responsible for the mass-diffusion transfer. The relation is know as Fick's first law and can be seen as a linearised version of the transfer law that is based on the gradient of the chemical potential. Let the relation between chemical potential and the concentration be:

$$\mu_i := \mu_i^0 + RT \log \frac{c_i}{c_i^0}$$

and

$$\begin{aligned}\frac{\partial \mu_i}{\partial x} &:= RT \frac{\partial}{\partial x} \log \frac{c_i}{c_i^0} \\ &:= RT \frac{\partial}{\partial c_i} \left(\log \frac{c_i}{c_i^0} \right) \frac{\partial c_i}{\partial x} \\ &:= RT \frac{1}{c_i} \frac{\partial c_i}{\partial x}\end{aligned}$$

Fick's law is the most commonly used transport equation. Thus if one uses Fick's first and second law in their common form, it is the concentration that is seen as the driving force, even though in reality it is the chemical potential that drives the transfer. The fact that it is the chemical potential that is continuous across the boundary, makes the composition to change discretely. There is a jump (see 11.3) and the gradient changes discretely at the interface.

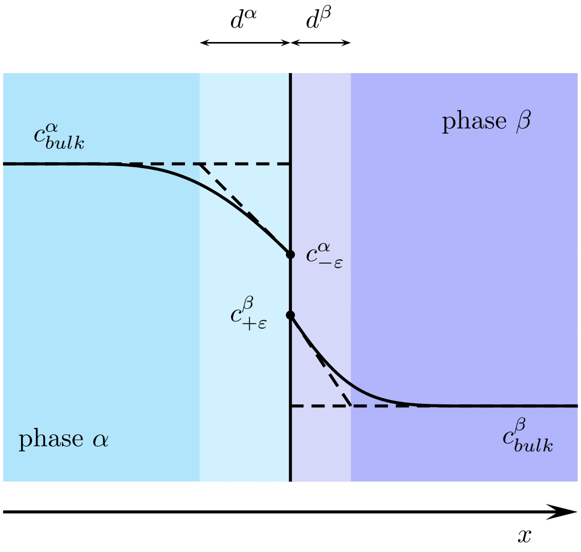


Figure 11.3: Concentration jump at the boundary and Nernst approximations

The jump in the composition for each species can be computed from the model for the chemical potential and the equilibrium condition at the boundary:

$$\begin{aligned}\mu_i^\alpha &= \mu_i^\beta \\ \mu_i^{o,\alpha} + RT\,\ln \varkappa_i^\alpha &= \mu_i^{o,\beta} + RT\,\ln \varkappa_i^\beta\end{aligned}$$

which leads to:

$$\begin{aligned}\mu_i^{o,\alpha} - \mu_i^{o,\beta} &= RT\left(\ln \varkappa_i^\beta - \ln \varkappa_i^\alpha\right) \\ \frac{\mu_i^{o,\alpha} - \mu_i^{o,\beta}}{RT} &= \ln \frac{\varkappa_i^\beta}{\varkappa_i^\alpha}\end{aligned}$$

In the common case where the composition is dominated by the stationary species on each side, the above expression simplifies to:

$$\frac{\mu_i^{o,\alpha} - \mu_i^{o,\beta}}{RT} = \ln \frac{c_i^\beta}{c_i^\alpha} := \ln K_i^{\alpha,\beta}$$

The expression on the left-hand side is constant and its exponential is often called Nernst's distribution constant. Giving the equation for the jump:

$$c_i^\beta := c_i^\alpha\,K_i^{\alpha,\beta}$$

The diffusion flow model uses the gradient of the composition, thus it is natural to define a diffusion layer thickness based on the gradient. The thickness is then defined by the intersection of the bulk composition and the gradient at the boundary (see 11.3):

$$-\mathcal{D}_i\,\triangledown c_i = -\frac{\mathcal{D}_i}{d^\alpha}\left(c_{i,-\varepsilon}^\alpha - c_{i,\text{bulk}}^\alpha\right)$$

and

$$d^\alpha = \frac{c_{i,-\varepsilon}^\alpha - c_{i,\text{bulk}}^\alpha}{\triangledown c_i}$$

which similarly applies to the β-phase.

11.2.2.3 Momentum diffusion

The same concept can be applied to momentum transport being described by the Navier-Stokes equation. Again a film-layer thickness can be defined along the same lines as done for the heat diffusion and the mass diffusion. One observes, that in general the layer thickness for the heat transfer is the smallest, whilst the one for the momentum transfer is the largest, thus the mass transfer is somewhere in between.

11.3 Penetration theory

The underlying thought is that under turbulent conditions, fluid bodies as a whole penetrate the boundary layer and attach for some time to the interface 11.4. During this contact time, the fluid body exchanges extensive quantities across the boundary to the other side, which again can be a fluid body or a film model. The idea is attributed to Higbie (1935) in which the absorption of gas into a still fluid is being described. This model has the "feel" of small eddies from the bulk to penetrate the film layer formed on the surface due to friction, which then "stick" to the interface for some time before they return to the bulk again. The contact time would be stochastically distributed, thus in general not be constant.

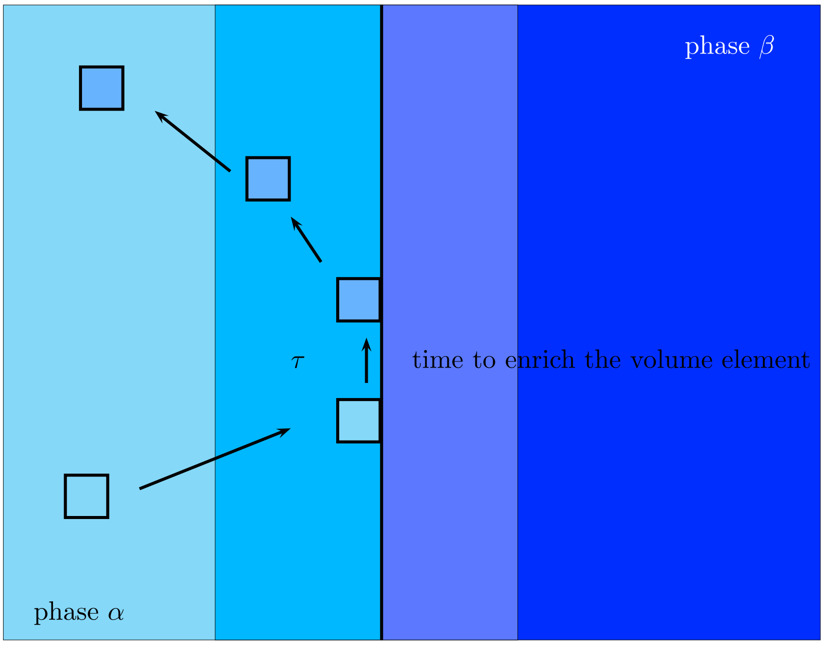


Figure 11.4: A volume element that travels to the surface, exchanges material over time τ and returns again to the bulk

The standard approaches use diffusion into the half plane to describe the transfer during the contact time.

12

Approximating distributed systems

Synopsis *Whilst essentially everything is distributed and we can usually write the equations all right, solving them analytically is only possible in very simple cases, where the mathematics are very simple and regular. With the computing devices to become available and fast, using approximations and numerical solutions has become the norm.*

12.1 From Network to Continuum

Distributed systems are mathematically described as partial differential equations. Most of them have no analytical solutions. Only in very special cases can we solve them, thus we do depend on numerical methods for finding solutions. Given numerical solutions are so central it should not surprise that the literature body is very extensive and also that the different disciplines have spawned many specialised activities in this field. Classic books in this field are Hildebrand (1956); Schwarz (1989); Atkinson (1989), but as said, there are many more.

If we have a look at our derivation using what we called a shell balance 4.4, then we defined a unit cell, formulated a model for the flow and then used the first variation of the flow to formulate the balance about the unit cell. Letting the cell dimension going to infinitesimally small, we obtained the desired partial differential equation. For each cell, the mathematical model was an ordinary differential equation. Representing the overall space by a set of unit cells gives a communicating network, representing the system's behaviour as a set of ordinary differential equations. Since all cells are communicating in series with each other, the number of cells in each direction gives us a definition for an order of the involved differential equations. This gives rise to the use of the term infinite-order systems for distributed systems. This process of transmogrifying a network of lumped system into a partial differential equations can also be reversed, at least in spirits: One can approximate partial differential equations as networks of ordinary differential equations. We have used the term ”network” of ordinary differential equations as an alternative to a set of coupled ordinary differential

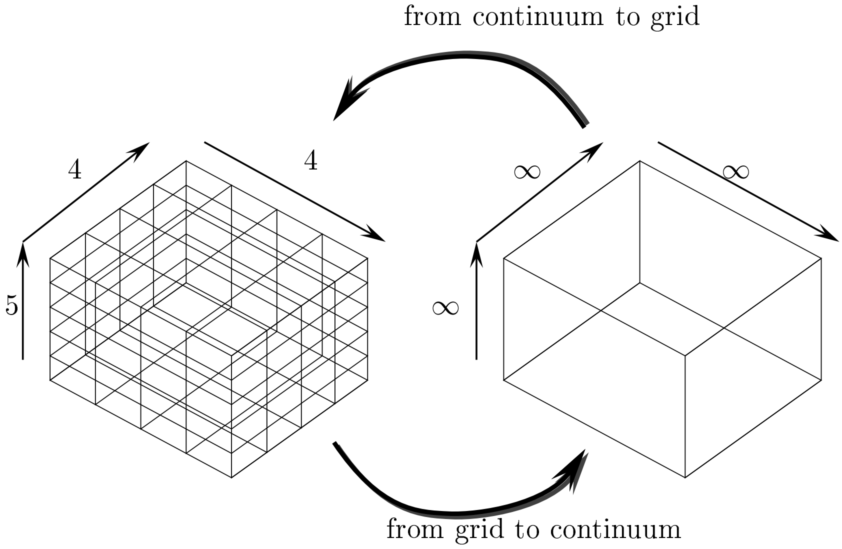


Figure 12.1: Derivation of PDE: from grid to continuum an back

equations with the objective to provide this link back to the mentioned derivation. Many different approximation methods are known, which on the background of having to find a numerical solution, is not difficult to rationalise. Most methods are purely mathematically motivated and have little to do with our picture of a network of control volumes. In fact many of the methods deviate very much from this picture. Nevertheless, this exposition should reflect the underlying nature of approximating distributed systems descriptions.

12.2 Approximating Derivatives

The basic idea is to introduce a grid in the direction the distribution effect is present. This can be done in all independent variables, thus spatial co-ordinates as well as time. Gridding only in the spatial co-ordinate leads to the networks of ordinary differential equations. If we also grid the time, then the result is a set of difference equations.

12.2.1 Equally spaced grids

Let us do the former, thus only grid in the spatial co-ordinate and for the time being we also assume only a one-dimensional distributed system. This is the simplest of the cases which serves our purpose of introducing the core idea and background of the method. So, let x be a state and r a scalar independent spatial variable, thus $\frac{d\,x}{d\,r}$ is a scalar first derivative of x with respect to r and $\frac{d^2\,x}{d\,r^2}$ be the corresponding 2^{nd} derivative. Let further r_k denote k^{th} point in the one-dimensional equally-spaced grid with a grid width of h. Having the objective to approximate 2-nd order derivatives, the minimal number of approximation points is three. A generic set of points is defined labelling the three points with the subscript -1,0,1 with -1 indicating the point k-1, 0 the point k, and 1 the point k+1. In each point the state function can be extended in a Taylor series:

$$x(r_k + h) := \sum_{i:=0}^{n} \frac{1}{i!} \left.\frac{\partial^i\, x}{\partial\, r^i}\right|_{r_k} h^i + \frac{1}{(n+1)!} \left.\frac{\partial^{n+1}\, x}{\partial\, r^{n+1}}\right|_{\xi} h^{n+1}\,.$$

Making two approximations for each point provides six equations enabling to solve for the first and the second derivatives. The solutions are obtained easily by taking the difference and the sum of the two equations that contain the desired approximate derivative. In the case of taking the sum, the zero-th and the even derivatives remain, whilst in the case of taking the difference, the odd derivatives are eliminated. This reflects into the error estimates for the approximations.

The six equations are, not showing the error terms:

$$x_{-1} := x_0 + \left.\frac{\partial\, x}{\partial\, r}\right|_{r_0} (-h) + \frac{1}{2} \left.\frac{\partial^2\, x}{\partial\, r^2}\right|_{r_0} (-h)^2 + \dots\,, \tag{12.1}$$

$$x_{-1} := x_1 + \left.\frac{\partial\, x}{\partial\, r}\right|_{r_1} (-2\,h) + \frac{1}{2} \left.\frac{\partial^2\, x}{\partial\, r^2}\right|_{r_1} (-2\,h)^2 + \dots\,, \tag{12.2}$$

$$x_0 := x_{-1} + \left.\frac{\partial\, x}{\partial\, r}\right|_{r_{-1}} h + \frac{1}{2} \left.\frac{\partial^2\, x}{\partial\, r^2}\right|_{r_{-1}} h^2 + \dots\,, \tag{12.3}$$

$$x_0 := x_1 + \left.\frac{\partial\, x}{\partial\, r}\right|_{r_1} (-h) + \frac{1}{2} \left.\frac{\partial^2\, x}{\partial\, r^2}\right|_{r_1} (-h)^2 + \dots\,, \tag{12.4}$$

$$x_1 := x_{-1} + \left.\frac{\partial\, x}{\partial\, r}\right|_{r_{-1}} 2\,h + \frac{1}{2} \left.\frac{\partial^2\, x}{\partial\, r^2}\right|_{r_{-1}} (2h)^2 + \dots\,, \tag{12.5}$$

$$x_1 := x_0 + \left.\frac{\partial\, x}{\partial\, r}\right|_{r_0} h + \frac{1}{2} \left.\frac{\partial^2\, x}{\partial\, r^2}\right|_{r_0} h^2 + \dots\,, \tag{12.6}$$

Assuming a constant grid, the grid constant is denoted by h, which further simplifies the writing. Choosing the appropriate pairs, one extracts the first and second derivative at one of the three points. Taking the pair 12.1 and

12.6 and ignoring the error terms for the time being :

$$\begin{aligned} x_{-1} - x_1 &:= -2 \left.\frac{\partial x}{\partial r}\right|_{r_0} h \,, \\ \left.\frac{\partial x}{\partial r}\right|_{r_0} &:= -\frac{x_1 - x_{-1}}{2\,h} \,. \end{aligned}$$

For the second derivative one finds:

$$\left.\frac{\partial^2 x}{\partial r^2}\right|_{r_0} := \frac{x_{-1} - 2x_0 + x_1}{h^2} \,.$$

For the error terms of the first derivative one finds:

$$\begin{aligned} O(h^2) &:= -\frac{h^3}{3!\,2\,h}\left(\left.\frac{\partial^3 x}{\partial r^3}\right|_{\xi_{-1,0}} + \left.\frac{\partial^3 x}{\partial r^3}\right|_{\xi_{0,1}}\right) \,, \\ &\approx -\frac{h^3}{3!\,2\,h}\, 2 \left.\frac{\partial^3 x}{\partial r^3}\right|_{\xi_{-1,0}} \,, \\ &\approx -\frac{h^2}{3!} \left.\frac{\partial^3 x}{\partial r^3}\right|_{\xi_{-1,1}} \,. \end{aligned}$$

where ξ_{ab} is the value of $r \in [r_a, r_b]$ with $\left| \left.\frac{\partial^3 x}{\partial r^3}\right|_{\xi_{-1,1}} \right|$ is maximal.

For the second derivative truncation only occurs at the 4-th order term:

$$\begin{aligned} O(h^2) &:= -\frac{h^4}{4!\,h^2}\left(\left.\frac{\partial^4 x}{\partial r^4}\right|_{\xi_{-1,0}} + \left.\frac{\partial^4 x}{\partial r^4}\right|_{\xi_{0,1}}\right) \,, \\ &\approx -\frac{h^2}{12} \left.\frac{\partial^4 x}{\partial r^4}\right|_{\xi_{-1,1}} \,. \end{aligned}$$

The pair 12.3 and 12.5 yields the two approximations for the derivatives at r_{-1} and finally the pair 12.2 and 12.4 gives the two at r_1.

Derivative	Approximation	Error Estimate
$\left.\frac{\partial x}{\partial r}\right\|_{r_{-1}}$	$\frac{1}{2h}\left(-3x_{-1}+4x_0-x_1\right)$	$\frac{h^2}{3}\left.\frac{\partial^3 x}{\partial r^3}\right\|_{\xi}$
$\left.\frac{\partial x}{\partial r}\right\|_{r_0}$	$\frac{1}{2h}\left(-x_{-1}+x_1\right)$	$-\frac{h^2}{6}\left.\frac{\partial^3 x}{\partial r^3}\right\|_{\xi}$
$\left.\frac{\partial x}{\partial r}\right\|_{r_1}$	$\frac{1}{2h}\left(x_{-1}-4x_0+3x_1\right)$	$-\frac{h^2}{3}\left.\frac{\partial^3 x}{\partial r^3}\right\|_{\xi}$
$\left.\frac{\partial^2 x}{\partial r^2}\right\|_{r_{-1}}$	$\frac{1}{h^2}(1x_{-1}-2x_0+x_1)$	$-h\left.\frac{\partial^3 x}{\partial r^3}\right\|_{\xi_1}+\frac{h^2}{6}\left.\frac{\partial^4 x}{\partial r^4}\right\|_{\xi_2}$
$\left.\frac{\partial^2 x}{\partial r^2}\right\|_{r_0}$	$\frac{1}{h^2}(1x_{-1}-2x_0+x_1)$	$-\frac{h^2}{12}\left.\frac{\partial^4 x}{\partial r^4}\right\|_{\xi}$
$\left.\frac{\partial^2 x}{\partial r^2}\right\|_{r_1}$	$\frac{1}{h^2}(1x_{-1}-2x_0+x_1)$	$h\left.\frac{\partial^3 x}{\partial r^3}\right\|_{\xi_1}+\frac{h^2}{6}\left.\frac{\partial^4 x}{\partial r^4}\right\|_{\xi_2}$

Table 12.1: Three point approximations

The 12.1 lists all the three point approximations ($\xi := \xi_{-1,1}$)
The analysis can be extended to more points thereby increasing the accuracy with which the derivation is approximated though with the cost of increasing complexity of the expressions. Most commonly the 3 point approximations are being used and significantly less often the 5-point approximations (see tables 12.2 and 12.3)

12.2.2 Non-equally spaced grids

In many cases things happen fast in one part of the considered system, whilst in other parts things are a little less hectic. So we really want to look a little closer at the part where things happen fast whilst in the slow parts things do not change so much with distance. We could certainly introduce a grid that is simply fine enough for the fast part and use it across the whole distributed domain, but then that is not very efficient indeed. This defines a request for defining a variable grid, so that we can adjust it to the dynamic of the process locally. To derive the equations, we repeat the derivation above, but now with a grid that changes. Since the error terms do not change, we give the equation without them. Again if we look at the 3-point approximation, we get the six equations of above, but now with changing grid width. Defining $h_1 := r_k - r_{k-1}$ and $h_2 := r_{k+1} - r_k$ we get:

Derivative	Approximation
$\frac{\partial x}{\partial r}\Big\|_{r_{-2}}$	$\frac{1}{12\,h}\left(-25x_{-2} - 48x_{-1} + 36x_0 - 16x_1 - 3x_2\right)$
$\frac{\partial x}{\partial r}\Big\|_{r_{-1}}$	$\frac{1}{12\,h}\left(-3x_{-2} - 10x_{-1} + 18x_0 - 6x_1 + x_3\right)$
$\frac{\partial x}{\partial r}\Big\|_{r_0}$	$\frac{1}{12\,h}\left(x_{-2} - 8x_{-1} + 8x_1 - x_2\right)$
$\frac{\partial x}{\partial r}\Big\|_{r_1}$	$\frac{1}{12\,h}\left(-x_{-2} + 6x_{-1} - 18x_0 + 10x_1 + 3x_2\right)$
$\frac{\partial x}{\partial r}\Big\|_{r_2}$	$\frac{1}{12\,h}\left(3x_{-2} - 16x_{-1} + 36x_0 - 48x_1 + 25x_2\right)$
$\frac{\partial^2 x}{\partial r^2}\Big\|_{r_{-2}}$	$\frac{1}{24h^2}(70x_{-2} - 208x_{-1} + 228x_0 - 112x_1 + 224x_2)$
$\frac{\partial^2 x}{\partial r^2}\Big\|_{r_{-1}}$	$\frac{1}{24h^2}(22x_{-2} - 40x_{-1} + 12x_0 + 8x_1 - 2x_2)$
$\frac{\partial^2 x}{\partial r^2}\Big\|_{r_0}$	$\frac{1}{24h^2}(-2x_{-2} + 32x_{-1} - 60x_0 + 32x_1 - 2x_2)$
$\frac{\partial^2 x}{\partial r^2}\Big\|_{r_1}$	$\frac{1}{24h^2}(-2x_{-2} + 8x_{-1} + 12x_0 - 40x_1 + 22x_2)$
$\frac{\partial^2 x}{\partial r^2}\Big\|_{r_2}$	$\frac{1}{24h^2}(22x_{-2} - 112x_{-1} + 228x_0 - 20x_1 + 70x_2)$

Table 12.2: Five-point approximations

Derivative	Error Estimate
$\frac{\partial x}{\partial r}\Big\|_{r_{-2}}$	$\frac{h^4}{5}\,f^{(5)}(\xi)$
$\frac{\partial x}{\partial r}\Big\|_{r_{-1}}$	$-\frac{h^4}{20}\,f^{(5)}(\xi)$
$\frac{\partial x}{\partial r}\Big\|_{r_0}$	$\frac{h^4}{30}\,f^{(5)}(\xi)$
$\frac{\partial x}{\partial r}\Big\|_{r_1}$	$-\frac{h^4}{20}\,f^{(5)}(\xi)$
$\frac{\partial x}{\partial r}\Big\|_{r_1}$	$-\frac{h^2}{3}\,f^{(5)}(\xi)$
$\frac{\partial^2 x}{\partial r^2}\Big\|_{r_{-2}}$	$-\frac{5}{6}h^3\,f^{(5)}(\xi_1) + \frac{h^4}{15}\,f^{(6)}(\xi)$
$\frac{\partial^2 x}{\partial r^2}\Big\|_{r_{-1}}$	$\frac{h^3}{12}\,f^{(5)}(\xi_1) - \frac{h^4}{60}\,f^{(6)}(\xi)$
$\frac{\partial^2 x}{\partial r^2}\Big\|_{r_0}$	$\frac{h^4}{90}\,f^{(6)}(\xi)$
$\frac{\partial^2 x}{\partial r^2}\Big\|_{r_1}$	$-\frac{h^3}{12}\,f^{(5)}(\xi_1) - \frac{h^4}{60}\,f^{(6)}(\xi)$
$\frac{\partial^2 x}{\partial r^2}\Big\|_{r_2}$	$\frac{5}{6}h^3\,f^{(5)}(\xi_1) + \frac{h^4}{15}\,f^{(6)}(\xi)$

Table 12.3: Errors for the five-point approximations

Derivative	Approximation
$\frac{\partial x}{\partial r}\Big\vert_{r_{-1}}$	$\frac{\left(-2\,h_1\,h_2-{h_2}^2\right)x_{-1}}{h_1\,h_2\,(h_1+h_2)}+\frac{\left({h_1}^2+2\,h_1\,h_2+{h_2}^2\right)x_0}{h_1\,h_2\,(h_1+h_2)}-\frac{h_1\,x_1}{h_2\,(h_1+h_2)}$
$\frac{\partial x}{\partial r}\Big\vert_{r_0}$	$-\frac{h_2\,x_{-1}}{h_1\,(h_1+h_2)}-\frac{\left(-{h_2}^2+{h_1}^2\right)x_0}{h_1\,h_2\,(h_1+h_2)}+\frac{h_1\,x_1}{h_2\,(h_1+h_2)}$
$\frac{\partial x}{\partial r}\Big\vert_{r_1}$	$\frac{h_2\,x_{-1}}{h_1\,(h_1+h_2)}-\frac{\left({h_1}^2+2\,h_1\,h_2+{h_2}^2\right)x_0}{h_1\,h_2\,(h_1+h_2)}-\frac{\left(-{h_1}^2-2\,h_1\,h_2\right)x_1}{h_1\,h_2\,(h_1+h_2)}$
$\frac{\partial^2 x}{\partial r^2}\Big\vert_{r_{-1}}$	$\frac{x_{-1}}{h_1\,(h_1+h_2)}-\frac{x_0}{h_1\,h_2}+\frac{x_1}{h_2\,(h_1+h_2)}$
$\frac{\partial^2 x}{\partial r^2}\Big\vert_{r_0}$	$\frac{x_{-1}}{h_1\,(h_1+h_2)}-\frac{x_0}{h_1\,h_2}+\frac{x_1}{h_2\,(h_1+h_2)}$
$\frac{\partial^2 x}{\partial r^2}\Big\vert_{r_1}$	$\frac{x_{-1}}{h_1\,(h_1+h_2)}-\frac{x_0}{h_1\,h_2}+\frac{x_1}{h_2\,(h_1+h_2)}$

Table 12.4: Three approximations for variable grid

The six equations are, not showing the error terms:

$$\begin{aligned}
x_{-1} &:= x_0 && +\left.\frac{\partial x}{\partial r}\right|_{r_0}(-h_1) && +\frac{1}{2}\left.\frac{\partial^2 x}{\partial r^2}\right|_{r_0}(-h_1)^2\\
x_{-1} &:= x_1 && +\left.\frac{\partial x}{\partial r}\right|_{r_1}(-h_1-h_2) && +\frac{1}{2}\left.\frac{\partial^2 x}{\partial r^2}\right|_{r_1}(-h_1-h_2)^2\\
x_0 &:= x_{-1} && +\left.\frac{\partial x}{\partial r}\right|_{r_{-1}}h_1 && +\frac{1}{2}\left.\frac{\partial^2 x}{\partial r^2}\right|_{r_{-1}}h_1^2\\
x_0 &:= x_1 && +\left.\frac{\partial x}{\partial r}\right|_{r_1}(-h_2) && +\frac{1}{2}\left.\frac{\partial^2 x}{\partial r^2}\right|_{r_1}(-h_2)^2\\
x_1 &:= x_{-1} && +\left.\frac{\partial x}{\partial r}\right|_{r_{-1}}(h_1+h_2) && +\frac{1}{2}\left.\frac{\partial^2 x}{\partial r^2}\right|_{r_{-1}}(h_1+h_2)^2\\
x_2 &:= x_0 && +\left.\frac{\partial x}{\partial r}\right|_{r_0}h_2 && +\frac{1}{2}\left.\frac{\partial^2 x}{\partial r^2}\right|_{r_0}h_2^2
\end{aligned}$$

An equation system we can solve for the 6 derivatives:

The 12.4 lists all the three point approximations ($\xi := \xi_{-1,1}$):

Whilst these expressions look rather complicated, there is a very generic structure there and their derivation is really not difficult, though it is a lot of writing if one does not use a symbolic manipulation program.

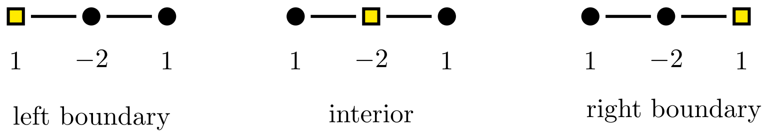


Figure 12.2: Stencils for 3 point approximations: left boundary, interior, right boundary

12.3 Approximation of the diffusion equation

We take the example of finding numerical solutions to the diffusion problems – most simple: heat diffusion. The energy balance for a body exhibiting only heat conduction, also called heat diffusion, is known as Fourier's heat diffusion equation. Using the symbol T for temperature and x for the spacial co-ordinate in which direction the system is distributed, defining $\mathcal{D}$ as the heat diffusivity and as usual t for time, Fourier's law takes the familiar form:

$$\frac{\partial T}{\partial t} = \mathcal{D} \frac{\partial^2 T}{\partial r^2}$$

Having a second-order derivative to approximate, the minimal number of points in the approximation required is three. This three point approximation introduces a regular pattern for the internal points and a separate one for each side at the boundary. This can be readily visualised using three stencils (12.2)

Using the stencils the partial differential equation becomes a network of ordinary differential equations. Defining n interior points, labelling them with the indices $i := 1, \dots n$, and a regular grid width of Δr the equations for the interior points are:

$$\frac{dT_i}{dt} = \frac{\mathcal{D}}{\Delta r^2} \left(T_{i-1} - 2\,T_i + T_{i+1} \right) \quad ; \quad i := 1, \dots, n \tag{12.7}$$

and for the left and the right boundary the respective equations are:

$$\begin{aligned} \frac{dT_0}{dt} &= \frac{\mathcal{D}}{\Delta r^2} \left(-2\,T_0 + T_1 + T_2 \right) \\ \frac{dT_{n+1}}{dt} &= \frac{\mathcal{D}}{\Delta r^2} \left(T_{n-1} + T_n - 2\,T_{n+1} \right) \end{aligned}$$

When formulating distributed problems one requires additional information to specify a proper mathematical problem. In most cases this will be the boundaries and the initial conditions. In the case where the state at the boundary is specified, the last two equations are replaced by the boundary condition. For example in our case this may be that the temperatures at the two boundaries are specified as a function of time. So the problem would read like:

$$\begin{aligned}
\text{dynamic:} \quad & \frac{dT_i}{dt} = \frac{\mathcal{D}}{\Delta r^2}\left(T_{i-1} - 2T_i + T_{i+1}\right) \quad ; \quad i := 1, \ldots, n \\
\text{left boundary:} \quad & T_0 = f_l(t) \\
\text{right boundary:} \quad & T_{n+1} = f_r(t)
\end{aligned}$$

The state are those quantities that are on the left-hand-side, the time derivatives, here the temperatures in the internal grid points. The dynamics are driven by the temperatures at the boundary, thus those become the input to the dynamic system, being the heat diffusion system. With the equations being linear, we must put it into a matrix form by defining:

$$\begin{aligned}
\text{state:} \quad & \underline{x} := \left[T_1, T_2, \ldots, T_n\right]^T := \left[T_i\right]_{i:=1,\ldots,n} \\
\text{left input:} \quad & u_l := T_0 \\
\text{right input:} \quad & u_r := T_{n+1}
\end{aligned}$$

we get the nicely patterned dynamic matrix equation:

$$\dot{\underline{x}} = \frac{\mathcal{D}}{\Delta r^2} \begin{bmatrix} -2 & 1 & 0 & \ldots & \ldots & \ldots & 0 \\ 1 & -2 & 1 & 0 & \ldots & \ldots & \vdots \\ 0 & 1 & -2 & 1 & 0 & \ldots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & 0 & 1 & -2 & 1 & 0 \\ \vdots & \ldots & \ldots & 0 & 1 & -2 & 1 \\ 0 & \ldots & \ldots & \ldots & 0 & 1 & -2 \end{bmatrix} \underline{x} + \frac{\mathcal{D}}{\Delta r^2} \begin{bmatrix} 1 & 0 \\ 0 & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & 0 \\ 0 & 1 \end{bmatrix} \underline{\mathbf{u}}\,.$$

This equation can be readily solved because it is a linear set of ordinary differential equation, naturally given the initial conditions and the boundary temperature as a function of time.

12.3.1 Extension to Higher-Dimensional Problems

This extension is straightforward as the approximation is done in each direction separately. As long as the grid is equally spaced and equally scaled in each direction, the stencils look quite simple and the equations are equally regular in terms of the patterns. So the equations can be written very concisely using the fact that the central grid points are in the diagonal. 12.3

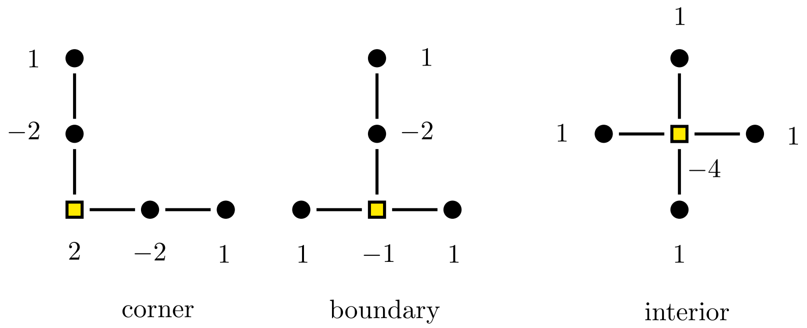


Figure 12.3: 2-D, 3 point stencil for equally spaced

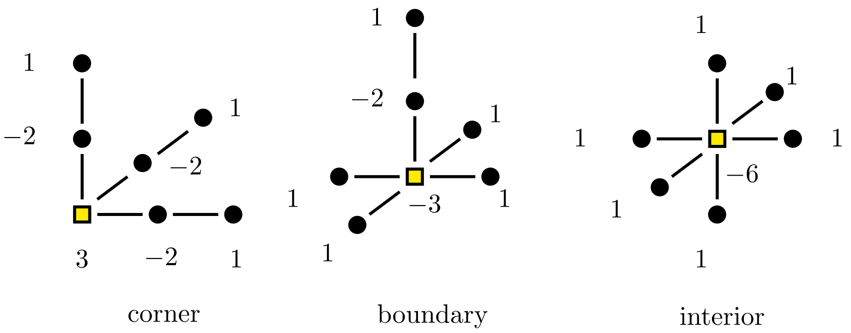


Figure 12.4: 3-D, 3 point stencil for equally spaced grid

and 12.4 show the stencils and the respective weights for the two and the three dimensional equally-spaced grids.

Constructing the matrix equations requires, as in the 1-D case, a stacking up procedure for the state variables in the different locations. In any case the indices for the different directions are ”looped” in one or the other sequence. In any case the dimensionality grows with the power of the dimension. The resulting matrices are all sparse, meaning that there are very few elements in the matrices that are different from zero. Most numerical packages have special procedures to handle sparse matrices, storing only the non-zero elements and a map of where they are located in the matrices and vectors.

These schemes have been further developed mainly because a rectangular grid is not always a very good idea but it is much better to adapt the grid to the dynamics of the process, for example to stream lines. These methods have been developed and computational tools exist that perform adaptive gridding very effectively.

13

Material models are everywhere

Synopsis *Materials have particular properties with regard to the intensive properties, like density. The equations of state represent the core knowledge two of which are required to generate an energy function, which in turn provides access to the other energy functions through Legendre transformations.*

13.1 Fundamentals

Earlier we already elaborated on mass carrying energy or being energy, for that matter. Not too surprising then, that the description of material is also tightly interlinked with the energy concept. The material is seen as a system and its behaviour, which we observe, is being described in terms of energy due to changes in its internal structure or effects imposed by the embedding environment. The fundamental measure for the behaviour of a material is the internal energy, a concept that is completely abstract, which has been developed based on observations in the 17th and 18th century being continuously refined, but not fundamentally changed. On the macroscopic level, the internal energy is seen to be a function of a number of key extensive quantities, which together with the energy form the fundamental thermodynamic state space. In fact the term ”state” is most likely being derived from thermodynamics so it is not clear on who used it first. Certainly Caratheodory (1909) did use the term ”Zustand” (state) already. Not too surprisingly, the term is very closely related to the abstract definition of the behaviour of a system. In essence it leads to a circular definition of state and behaviour. The concept of state and behaviour description is thus to be seen as a purely abstract concept and as such cannot be proven to be correct but is only supported by our continuous observations of the correctness of the postulates forming the foundation of this abstraction.

Several versions of postulates exist. The most commonly cited ones are due to Callen (1985):

1. *There exist particular states (called equilibrium states) of simple sys-*

tems that, macroscopically, are characterized completely by the internal energy U, the volume V, and the mole numbers n_1, n_2 , ..., n_n of the chemical components. (Callen p 13)

2. *There exists a function (called the entropy S) of the extensive parameters of any composite system, defined for all equilibrium states and having the following property: The values assumed by the extensive parameters in the absence of an internal constraint are those that maximize the entropy over the manifold of constrained equilibrium states. (Callen p 27)*

3. *The entropy of a composite system is additive over the constituent subsystems. The entropy is continuous and differentiable and is a monotonically increasing function of the energy. (Callen p 28)*

4. *The entropy of any system [is non-negative and] vanishes in the state for which $(\frac{\partial U}{\partial S} = 0$ (that is, at the zero of temperature). (Callen p 30)*

The building of thermodynamic functions can be constructed on these postulates. We shall not do this here but we shall have a close look at the mathematical structure of this building.

For this purpose we pick up the concept of internal energy and the thermodynamic space:

Statement 1: There exists a function called internal energy, which is a fundamental function of some extensive variables. $U\left(S, \underline{\mathbf{X}}\right)$, where the $\underline{\mathbf{X}}$ is a vector of extensive quantities consistent of volume V, components mass measured in moles $\underline{\mathbf{n}}$.

Statement 2: There exists a function called entropy, which is a fundamental function of some extensive variables $S\left(U, \underline{\mathbf{X}}\right)$

Both are functions of Euler degree 1. Thus:

$$U\left(\lambda\, S, \lambda\, \underline{\mathbf{X}}\right) = \lambda U\left(S, \underline{\mathbf{X}}\right)$$

and

$$S\left(\lambda\, U, \lambda\, \underline{\mathbf{X}}\right) = \lambda S\left(U, \underline{\mathbf{X}}\right)$$

None of the two fundamental functions S, U are known, or at least only very simply idealised systems can be computed from quantum theory. But some of the structures of these functions are known, which give at least a template of the equations representing the material properties. These "templates" are then replaced or filled in with empirical models that have the requested structure. This replacing of unknown relations by intelligent guesses is quite common throughout science and engineering.

The scientist of the 18th century came eventually to the conclusion that the thermodynamic phase space is spanned by the a set of extensive quantities and their derivatives. Even more, they also found that they always appear in pairs, called conjugate pairs. In a first round it was the pairs (T, S) and, (p, V) that characterised the internal energy, where the conjugate to the extensive quantities are the intensive quantities being the partial derivative of the internal energy with respect to the paired extensive quantity. Even more specific they found that it is the product of the quantities in the pairs that make up the internal energy. It took some time to add the remaining pairs, namely (μ_i, n_i), the chemical potential and the component mass, which is considered to be due to Gibbs, published in 1873 in a paper entitled "A Method of Geometrical Representation of the Thermodynamic Properties of Substances by Means of Surfaces". The structural statements then being that:

$$U = T\,S - p\,V + \sum_{\forall i} \mu_i n_i$$

and

$$dU = T\,dS - p\,dV + \sum_{\forall i} \mu_i\,dn_i \tag{13.1}$$

with:

$$T := \frac{\partial U}{\partial S} \quad -p := \frac{\partial U}{\partial V} \quad \mu_i := \frac{\partial U}{\partial n_i}$$

The thermodynamic phase space is then:

$$F\left(U, S, V, \underline{\mathbf{n}}, \frac{\partial U}{\partial S}, \frac{\partial U}{\partial V}, \frac{\partial U}{\partial \underline{\mathbf{n}}}\right)$$

which is a first-order partial differential equation.

The role of U and S can thereby be switched. Thus if we count s species we have $n := s + 2$ related variables, then the dimension of the configuration space is $2n + 1$ in dimension, namely U or S as the "+1" standing out quantity, the extensive quantities S or U, namely the not chosen one, and V together with the s component masses and the derivatives of the selected quantity with respect to the remaining extensive quantities. In this configuration space we have n dependencies, namely the derivatives and we have one to be satisfied, namely the first law of thermodynamics (13.1). Thus there exists 2n+1 equations that describe a thermodynamic system. The theory of first-order partial differential equations provides these equations Duff (1956).

In Gibbs terms, there are the two alternatives for the fundamental equation:

$$\begin{aligned} \text{internal energy} \quad & U := U(S, V, \underline{\mathbf{n}}) \\ \text{entropy} \quad & S := S(U, V, \underline{\mathbf{n}}) \end{aligned}$$

and the above-defined derivatives provide the effort variables, which are the equations of state:

$$
\begin{aligned}
T &:= \frac{\partial U}{\partial S} := T(S, V, \underline{\mathbf{n}}) \\
-p &:= \frac{\partial U}{\partial V} := -p(S, V, \underline{\mathbf{n}}) \\
\boldsymbol{\mu} &:= \frac{\partial U}{\partial \underline{\mathbf{n}}} := \boldsymbol{\mu}(S, V, \underline{\mathbf{n}})
\end{aligned}
$$

So all these are a function of the n dimensional space or the $2n + 1$ dimensional thermodynamic phase space. There are n equations of state, from which we can eliminate n variables. The result is a function of $n + 1$ variables.

For example if we have U, S, V, n, thus a single substance, the dimension of the configuration space can be reduced by looking at the densities only:

$$
F(u, s, v, \frac{\partial u}{\partial s}, -\frac{\partial u}{\partial v}) := F(u, s, v, T, p)\,,
$$

with u, s, v being $U/(nR), S/(nR), V/(nR)$ the respective densities formed by norming with the molar mass and the gas constant. For a monoatomic gas and choosing s as the base variable and given two of the linking equations[1] are:

$$
\begin{aligned}
T &:= p\,v \\
u &:= \frac{3}{2}\,T
\end{aligned}
$$

The remaining 3 equations are readily constructed. First the base variable s:

$$
\begin{aligned}
ds - \frac{p}{T}\,dv - \frac{1}{T} &= 0 \\
ds - \frac{1}{v}\,dv - \frac{3}{2}\,\frac{1}{u}\,du &
\end{aligned}
$$

and integrating:

$$
\begin{aligned}
s &= \ln v + \frac{3}{2}\,\ln u + \text{const} \\
s &= \ln u^{2/3}\,v
\end{aligned}
$$

[1] current physics literature terms all of these 5 equations equations of state, which is in some contrast with the use in engineering

defining the specific entropy up to a constant. The remaining two equations are simply:

$$\frac{1}{T} := (\frac{\partial s}{\partial u})_v \quad ; \quad \frac{p}{T} := (\frac{\partial s}{\partial v})_u$$

Given the appropriate information, namely n+1 equations, the system is completely defined. In chemical engineering it common to use the internal energy as the base variable and not as above the entropy. In this case the internal energy is only a function of the temperature, so for purpose of demonstrating the concept entropy was chosen as the base variable and thus its function was constructed.

Choosing the appropriate set of n equations, and eliminating n variables the equation of states are being constructed; the equations of state as are commonly used in engineering. For ideal gas, this is the celebrated equation:

$$p\,V = n\,R\,T$$

with R being the ideal gas constant.

14

Mixing patterns

Synopsis *Two extremes are commonly analysed: the ideal mixing, where mixing is really fast compared to anything else in the system, and, no mixing. The first is termed ideally mixed tank whilst the latter is the celebrated plug flow.*

14.1 Two Extremes

The abstraction of the process-relevant universe into systems separated by idealised walls from its environment can be further abstracted. Two important cases of flow systems are generated by analysing two extreme, limiting cases: one in which the flow system is flown through with minimal internal re-circulation and one in which the in-flow and the out-flow is minimal compared to the internal re-circulation. The limit is in both cases a reduction of *minimal* flow to *no* flow (14.1).

In the first case, one assumes a zero internal re-circulation whilst in the second case one sets the flow across the boundary zero. The argument is not only an order of magnitude assumption in the flow, but also in the time scale: it is assumed that the dynamic window for the internal process is clearly in the short time scale, compared to the dynamics of the flows across the system's boundary.

14.2 Small Internal Re-Circulation, No Reactions and Slow Changes at the Boundary

In this case, the stationary and constant control volume is placed in a flow field. The modelling is done in a range of the time-scale, where the changes at the boundary are very slow, thus one may assume a stationary flow field, which has no internal re-circulation, that is, the curl of the flow field is zero (Deen (1998); Bird et al. (2001)). This in turn implies that the accumulation terms in both the basic balances, the integral balance (4.3) and the differential balance (4.4) approach zero, which is often referred to

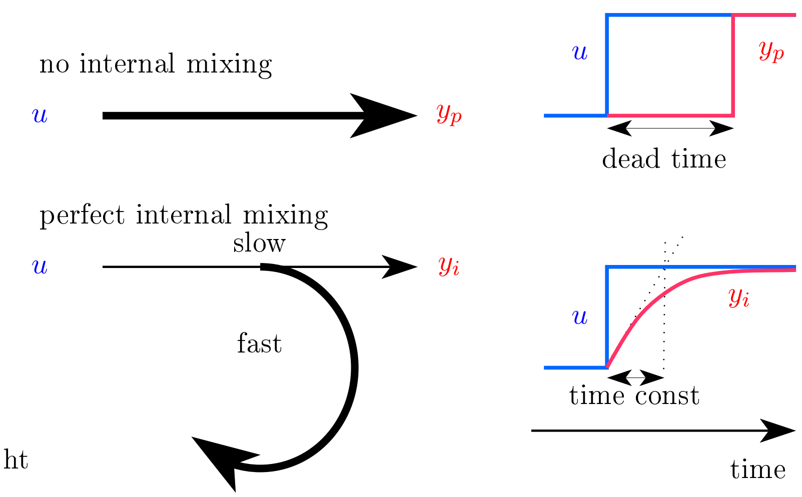


Figure 14.1: Two extremes: on the top no mixing at all resulting in a plug flow, whilst below there is nearly only mixing, at least in the short time scale resulting in the ideally mixed volume.

as *pseudo-steady state.*

Thus the integral balance (4.3) reduces to:

$$\underline{\mathbf{0}} := -\int_S \underline{\underline{\hat{\phi}}}^T \, \underline{\mathbf{n}} \, dS \, , \tag{14.1}$$

To simplify one step further, we assume that the boundary of the system is split into pieces in which each piece has uniform conditions. This simplifies the handling of the flows across the boundary and we have a directionality defining a local reference co-ordinate, pointing inwards or pointing outwards for each flow. Thus the behaviour of this system with "lumpy" boundaries writes:

$$\underline{\mathbf{0}} := \sum_c \alpha_c \, \underline{\hat{\mathbf{\Phi}}}_c \, , \tag{14.2}$$

$$:= \underline{\underline{\mathbf{F}}}_S \, \underline{\hat{\mathbf{\Phi}}} \, , \tag{14.3}$$

where in the second line we have wrapped the direction indicators into a matrix and the vector of flows into a stack of vectors. In terms of the behaviour, for the steady-state behaviour, the inflows balance the outflows, which matches our expectations.

The differential balance (4.4) simplifies to:

$$\underline{\mathbf{0}} := -\frac{\partial}{\partial \, \underline{\mathbf{r}}} \, \underline{\underline{\hat{\phi}}} \tag{14.4}$$

14.2.1 Pure transport systems

Equation 14.4 describes an idealised fast transfer system, in which the internal transport is fast compared to the changes at the boundary. The transport is a function of the state of the system and the state of the connected system. With the accumulation term ”disappearing”, the resulting set of equations become algebraic from which the stationary distribution of the state can be computed as a function of the conditions at the boundaries as we have discussed before.

Substituting the simple isotropic gradient transport law 5.4, one gets:

$$\underline{\mathbf{0}} := -\frac{\partial}{\partial \underline{\mathbf{r}}}\left(\underline{\underline{\mathbf{C}}}\,\frac{\partial}{\partial \underline{\mathbf{r}}}\pi\right) . \tag{14.5}$$

So this is a second-order differential equation in π. For the transfer to be computable, the solution to the second-order differential equation must exist (Lin and Segel (1988), p121). The existence of a solution is discussed early in the literature Courant et al. (1928). Lin and Segel, though, expressed the fact Lin and Segel (1988), p418) that *most scientists on most occasions do not concern themselves with the thorny philosophical questions that emerge from a searching examination of what lies at the foundation of their endeavours. ...*

The solution forms a hyper-surface with the boundary condition defining the position of this surface. Integrating above equation once states that the flux tensor $\underline{\underline{\hat{\phi}}}$ is constant:

$$\underline{\underline{\hat{\phi}}} := -\underline{\underline{\mathbf{C}}}\,\frac{\partial}{\partial \underline{\mathbf{r}}}\pi := \text{const} . \tag{14.6}$$

Two important lessons are to be drawn from this, namely the facts that

- the state is eliminated and
- there is no time effect associated with the transfer

For simple two-active boundary systems such as discussed in 10.2.1 the time-scale assumption leads to a simplification of the transfer system to a simple resistance, which is what the arrows in the first picture of the decomposition represent.

14.2.2 Plug flow reactors

Flow systems of this type that exhibit reactions are called plug-flow reactors. They are idealised reactors which represent one of the two extreme reactor systems. The shell balance equation for a plug-flow with lumpy boundaries is:

$$\underline{\mathbf{0}} := \underline{\hat{\mathbf{n}}}_r - \underline{\hat{\mathbf{n}}}_{r+dr} + \underline{\tilde{\mathbf{n}}}(\underline{\mathbf{c}}_r)\, dV \ ,$$

The outflow of the differential volume is approximated by the first variation on the inflow:

$$0 = \underline{\hat{\mathbf{n}}}_r - (\underline{\hat{\mathbf{n}}}_r + d\underline{\hat{\mathbf{n}}}_r) + \underline{\tilde{\mathbf{n}}}(\underline{\mathbf{c}}_r)\, dV \ ,$$

We divide by the differential volume:

$$0 = -d\underline{\hat{\mathbf{n}}}_r + \underline{\tilde{\mathbf{n}}}(\underline{\mathbf{c}}_r)\, dV$$

Expanding the production term and assuming constant density and constant volumetric flow then yields:

$$0 = -\hat{V}\, d\underline{\mathbf{c}}_r + \underline{\underline{\mathbf{N}}}^T\, \underline{\tilde{\xi}}(\underline{\mathbf{c}}_r)\,.$$

So with

$$\hat{V} := A\, v \quad , \quad V := r\, A\,,$$

Substitution results in:

$$\begin{aligned} 0 &= -\, A\, v\, d\underline{\mathbf{c}}_r + \underline{\underline{\mathbf{N}}}^T\, \underline{\tilde{\xi}}(\underline{\mathbf{c}}_r)\, A\, dr \\ 0 &= -\, v\, \frac{d\, \underline{\mathbf{c}}_r}{d\, r} + \underline{\underline{\mathbf{N}}}^T\, \underline{\tilde{\xi}}(\underline{\mathbf{c}}_r) \end{aligned}$$

Since $\frac{dr}{v} =: dt$ we get the result

$$\frac{d\, \underline{\mathbf{c}}_r}{d\, t} = \underline{\underline{\mathbf{N}}}^T\, \underline{\tilde{\xi}}(\underline{\mathbf{c}}_r)$$

If we now substitute a kinetic model we get a first-order differential equation with the concentration vector being the state. This is also the model for a batch reactor, as we shall see shortly. Thus one of the interpretations of the result is to look at the tubular reactor to transport batches of reactive mixture. The residence time of these batches is determined by the length of the tube and the flow rate. With the flow being constant, each batch is the same.

14.3 Maximal Internal Flow, Slow Reactions and Small, Slow Flows Across the Boundaries

In this case one assumes strictly no flow across the boundary and maximal internal flow. Placing the dynamic window into the small time scale, where the reactions are slow and thus the turnover very small compared to the internal flows, the differential balance 8.5 reduces to

$$\frac{\partial \underline{\varphi}_S}{\partial t} := -\frac{\partial}{\partial \underline{\mathbf{r}}} \underline{\underline{\hat{\phi}}},$$

Further assuming that the inflow and the outflow from the control volume are small compared to the internal flows, the equilibrium is reached quickly. Thus on the larger time scale, the internal fast dynamics are in equilibrium and no change with time is observed:

$$\underline{\mathbf{0}} := -\frac{\partial}{\partial \underline{\mathbf{r}}} \underline{\underline{\hat{\phi}}},$$

Since the inflow is negligible in this time scale, the system is closed and the solution is a constant. So the intensive quantity $\underline{\varphi}_S$ is constant everywhere in the region.

With the conditions in the contents being uniform, we shift time scale to a longer one. Now 8.5 simplifies significantly: the densities are constant everywhere in the volume, thus the volume integrals involving the densities change to the volume times the densities, which is simply the corresponding extensive quantity:

$$\frac{d\,\underline{\mathbf{\Phi}}_S}{d\,t} = -\int_S \underline{\underline{\hat{\phi}}}^T \underline{\mathbf{n}}\, dS + \underline{\tilde{\mathbf{n}}}_S\,.$$

Lumping the boundary (4.5) and assigning the global co-ordinate, the equation for the reactive, ideally-mixed domain emerges:

$$\frac{d\,\underline{\mathbf{\Phi}}_S}{d\,t} = \underline{\underline{\mathbf{F}}}_S\, \underline{\hat{\mathbf{\Phi}}} + \underline{\tilde{\mathbf{n}}}_S\,.$$

which for the component mass reads:

$$\frac{d\,\underline{\mathbf{n}}_S}{d\,t} = \underline{\underline{\mathbf{F}}}_S\, \underline{\hat{\mathbf{n}}} + \underline{\tilde{\mathbf{n}}}_S\,.$$

This equation describes an idealised capacity, namely a lumped system. The above-made fast mixing assumption yields that the intensive quantities are uniform within the control volume at a time scale that is large relative to the internal mixing. In chemical engineering this type of reactor is referred to as ideally stirred tank reactor.

In the case where there is no inflow and no outflow, this becomes the model of a batch reactor, where the reactor is an ideally stirred volume. In the case of constant volume, one often applies a state variable transformation:

$$\underline{\mathbf{n}} := V\,\underline{\mathbf{c}}$$

which can readily be differentiated with respect to time:

$$\frac{d\,\underline{\mathbf{n}}}{d\,t} := V\,\frac{d\,\underline{\mathbf{c}}}{d\,t}$$

and expanding the production term:

$$\frac{d\,\underline{\mathbf{c}}_S}{d\,t} = \underline{\underline{\mathbf{N}}}^T\,\underline{\tilde{\xi}}(\underline{\mathbf{c}}_S)\,,$$

which is identical to what we got for the plug-flow reactor.

15

Surfing on the state concept

Synopsis *The term state has been coined implying that knowing the state of a system one knows everything about the system at the point in time the state is reported. Defining the state as a identifiable unique quantity associated with the process provides the stage for an abstraction, which is suitable for analysis and computational engineering.*

15.1 The mechanism of ”things”

The behaviour of a system is described using a set of suitable variables that are called ”the state” living in the state space. The state contains all the information one requires to predict the next state given knowledge of the current state and the input. Or as Kalman formulated it 1963 (Kalman, 1963): *”The state is to be regarded always as an abstract quantity. Intuitively speaking, the state is the minimal amount of information about the past history of the system which suffices to predict the effect of the past upon the future.”* The state of a system can change due to internal changes or interactions with the environment in which the system is embedded.

For physical macroscopic, classical systems the behaviour of a system is described by the dynamic conservation laws that state: the change of the state, which is the vector of conserved quantities, is the consequence of exchanging conserved quantities with the environment in one or the other form. For example, placing a hot body into a cold room will induce a heat exchange between the hot body and the cold room - as we all experienced, the body will get colder and the room may get slightly warmer, probably not noticeably, if the body is not large compared to the room. Similarly, pulling the plug in the full bath tub will make the water to run out and if you leave a coloured object like a newspaper on your nicely lacquered desk you may experience that the dye diffuses into the lacquer permanently.

For physical systems the state is extended with quantities that collectively satisfy a conservation law, but may undergo conversion between them. The application of the later extension is species. They may undergo reactions or phase changes. Whilst the species itself is not conserved, the total mass is. For an example, we all know that if we have a container full of oxygen and hydrogen and initiate the reaction with a spark, the mixture will exhibit a very fast violent reaction where the product is water.

15.2 The grand scheme

Having discussed all the major components, the "grand scheme" can be assembled. It has four major components: The balance equations providing the description of the dynamics, the transport of the balanced extensive quantity across the systems' boundaries, the internal transposition and reactions, and the state-variable transformations that are required for the description of the transport and transposition that includes material properties, thus thermodynamics and geometry.

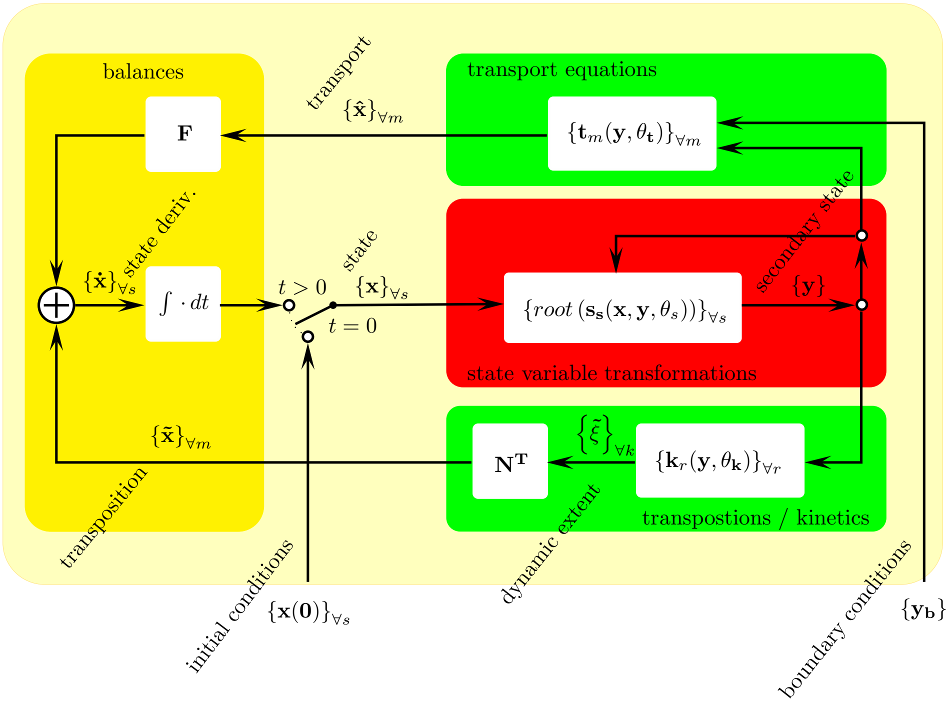

balances
transport
transport equations
F
$\{\hat{\mathbf{x}}\}_{\forall m}$
$\{\mathbf{t}_m(\mathbf{y},\theta_\mathbf{t})\}_{\forall m}$
state deriv.
$\{\dot{\mathbf{x}}\}_{\forall s}$
$\int \cdot dt$
state
$t > 0$
$\{\mathbf{x}\}_{\forall s}$
$t = 0$
$\{root\,(\mathbf{s_s}(\mathbf{x},\mathbf{y},\theta_s))\}_{\forall s}$
secondary state
$\{\mathbf{y}\}$
state variable transformations
$\{\tilde{\mathbf{x}}\}_{\forall m}$
$\mathbf{N^T}$
$\{\tilde{\xi}\}_{\forall k}$
$\{\mathbf{k}_r(\mathbf{y},\theta_\mathbf{k})\}_{\forall r}$
transpostions / kinetics
transposition
dynamic extent
initial conditions
$\{\mathbf{x}(\mathbf{0})\}_{\forall s}$
boundary conditions
$\{\mathbf{y_b}\}$

Figure 15.1: The grand scheme as a block diagram

15.1 shows the four main components and indicates the principle sequence in which they are established. It should be noted though, that before one can construct this equation system, which is a differential algebraic system of index one, there are two major steps to be taken before: (i) the process must be identified usually in the form of a technical drawing or a corresponding sketch, which makes it clear to the user of the model as to which process is being modelled and what embedding into the environment is. This is followed by (ii) an abstraction of the process into a physical topology so as to discuss and settle the dynamic contents of the plant, answering the question of what is precisely modelled with what type of dynamic system component. This is to be supplemented by filling in the "chemistry" and energy that is essential for the process description. Only then can the above scheme be established.

The topology maps into the incidence matrix of the directed graph, which in general will have a block structure, as each stream is possibly transporting several quantities such as different species. The incidence matrix provides the directionality coefficients that indicate the reference co-ordinate for each flow, namely the direction that is shown in the graph. The actual flow may well go in the opposite direction namely when it is negative.

The scheme also shows a matrix $\underline{\underline{\mathbf{N}}}$, which in the case where reactions are taking place, is the stoichiometric matrix for all systems This stoichiometric matrix is a block matrix with one block for each system.

All the arrows in the block diagram represent flow of stacks of vectors indicated by the two wiggly brackets. So the matrices are also block matrices. Whilst the latter are technical details that naturally fall into place on use, it is essential to capture the overall structure of these models. This structure only assumes lumped models.

15.2.1 Dynamics: balances

The conservation principle are strict in the sense that they are always satisfied and thus form a strong foundation for building the models on. They also clearly identify the state, which is the quantities that are changing with time, and thus are in the accumulation term:

$$\begin{gathered}\textbf{accumulation of conserved extensive quantity}\\ =\\ \textbf{transport of conserved extensive quantity across the boundary}\end{gathered}$$

On a global base, the internal transposition that may take place in the form of reactions, for example, is not affecting the total conserved quantity. Thus conversion of species in reactions is mass neutral. If however, what we nearly always do, formulate the component balances:

$$\begin{gathered}\textbf{accumulation of extensive quantity}\\ =\\ \textbf{transport of extensive quantity across the boundary}\end{gathered}$$

+

internal conversion

Important so, is that the sum over the component masses follows the first version of the conservation law, thus the sum of the internal conversion terms *must* be annihilated.

For the sake of simplicity, and the fact that we can always approximate distributed systems as a set of lumped systems, we shall limit our discussion to lumped systems, which here is represented by the member labelled with s. In the state space notation, in which the state is denoted by $\underline{\mathbf{x}}$ and the understanding that the above applies:

$$\underline{\dot{\mathbf{x}}}_s = \sum_{\forall m} \alpha_{s,m} \underline{\hat{\mathbf{x}}}_m + \underline{\tilde{\mathbf{x}}}_s \tag{15.1}$$

15.2.2 Exchange: transport

Transport can only take place between adjacent systems being driven by the difference in the effort variables. The two system being couple have a common boundary, which is one way in which the geometry comes into the formulation. So the transfer is

$$\underline{\hat{\mathbf{x}}}_{a|b} := \underline{\hat{\mathbf{x}}}_{a|b} \left(\underline{\mathbf{y}}_a, \underline{\mathbf{y}}_b, \underline{\mathbf{p}}_{a|b} \right) \tag{15.2}$$

The flows introduce the effort variables and some parameters. Some of these parameters may be manipulated from the outside making them the manipulating elements controlling the flow, such as valves.

15.2.3 Internals: transposition / reaction

Internal conversion of one type of extensive quantity into another one is characteristic for the manipulation of species as they occur in reactive systems both on human-made processes as well as natural processes. Every conversion goes in ratios. For reactive systems this is the stoichiometric coefficients. The reaction dynamics itself is described by the change in the extent of reaction, which we denote by $\underline{\tilde{\xi}}$. Thus:

$$\underline{\tilde{\mathbf{x}}}_s := \underline{\underline{\mathbf{N}}}_s^T \, \underline{\tilde{\xi}}_r \tag{15.3}$$

with the rate being a empirical relation that reflects the fact that species have to physically meet to undergo reaction and that the reaction itself usually must be promoted in terms of energy:

$$\underline{\tilde{\xi}}_r := \underline{\tilde{\xi}}_r (\underline{\mathbf{y}}_s, \underline{\mathbf{p}}_r) \tag{15.4}$$

The most noticeable part of this relation is that again a vector of quantities is introduced. The stoichiometric matrix is here defined in the space of species x reactions, whilst also the transposed definition is being used. Whilst the reaction occurs inside the system, at least for homogeneous systems, the rate expression is for the reaction. It is the secondary state, usually a material density such as composition and the intensive secondary state temperature, that drives the reaction.

15.2.4 Link: fundamental state and secondary state

The ”new variables” being introduced by the transport and transposition relations are all a function of the state. So the last block must provide these links, namely the mapping of the fundamental or primary state into the “new variables”, which we term secondary states. This block is the most difficult part to fill in. It includes some simple definitions like composition as a function of the molar masses; but also more complex relations like the equation of states or derivatives of the fundamental energy function with respect to the extensive quantities like entropy, volume, component mass, the latter being the effort variables. Very commonly the relation between mass and volume also comes into this block, thus the density, for which we hardly have any analytical expressions beyond empirical models in the form of equations of state. Geometry comes into the picture more often through the fact that the exchange between two adjacent systems is going through the common piece of interface. Thus the interface relative to the two systems comes into the definition of the process. Another reason geometry comes into the description is the need for ”measurable” quantities such as levels or pressures. This ”linking box” can be build by starting with the secondary state variables, recursively extending the set of equations thereby constructing a tree of equations, the leaves being variables that are known, usually parameters of one or the other correlation, or components of the fundamental state for the system. In both cases the extension stops, thus forms the bottom-out rule in the recursion. Often these equations form a nearly triagonal equation/variable structure, which is easy to handle, except that it is also quite common that some variables are defined by implicit relations. One of these is often temperature. The tree is usually not very deep. Three to four layers is common and correspondingly a handful of secondary variables. Rarely it increases to the order of 10.

The definition of the model requires that the relation between the two set of quantities must exist, thus at least numerically it must be possible to solve the equation for the secondary state variables as a function of the primary state variables.

Lumping these relative versatile set of equations together in one we could write:

$$\underline{\mathbf{0}} := \underline{\mathbf{s}}(\underline{\mathbf{y}}_s, \underline{\mathbf{p}}_i, \underline{\mathbf{x}}_s, \underline{\mathbf{\Theta}}_s) \tag{15.5}$$

or considering the fact that this set of at least partially implicit equations must be solvable:

$$\underline{\mathbf{y}}_s, \underline{\mathbf{p}}_i := \text{root}\left(\underline{\mathbf{s}}(\underline{\mathbf{y}}_s, \underline{\mathbf{p}}_i, \underline{\mathbf{x}}_s, \underline{\mathbf{\Theta}}_s)\right) \tag{15.6}$$

15.2.4.1 Complications

All of the above "constitutive" equations: transport, transposition/reaction, and state variable transformations are often a function of the state, in most cases the secondary state. Probably. the most outstanding one is the reaction "constant", which is anything but constant: it is modelled by an exponential function of the temperature, thereby indicating a first-order dynamic dependency of the probability of reactants to meet. The consequence is that another set of "constituent" equations is to be added, which is the dependency on the (secondary) state of the above-defined parameters.

$$\underline{\mathbf{p}}_i := \underline{\mathbf{p}}_i(\underline{\mathbf{x}}_s, \underline{\mathbf{y}}_s, \underline{\mathbf{p}}_i) \tag{15.7}$$

This set of equations is material dependent, such as transport parameters reflect the property of the physical transport system such as conductivity, capacity and density, which not too surprisingly are a function of the state of the transport system. Similar is the case with the reaction parameters. As has been mentioned they are a function of the state.

The "devil" is often in the "link" between the fundamental state and the secondary variables and in the continuation the "parameters". The insight actually questions the use of the term "parameter" as, whilst a characteristic of the system/material/reaction etc, it is state dependent and not really constant. If one follows this line of through, "parameters" in the sense of constant characteristic quantities do hardly exist. It is only the quantities appearing in empirical relations that are fitted that are kind of "parameters" and physical constants, some being universal, that, strictly speaking, fit reasonably well into this description. In this context the often used term process parameters for utility temperature, pressures or the like are better termed as "conditions".

15.2.5 Structures

Some textbooks consider it "good practise" to substitute equations as early as possible in the modelling process with the main argument being that one is interested in the measured quantities. The person writing the model is thus asked to go through the ordeal of executing a set of substitutions and variable transformations on the above differential algebraic equation system. This is not only work, but it also is a great source of errors both in algebra and simply copying from one line to the next or whatever goes into transcribing material.

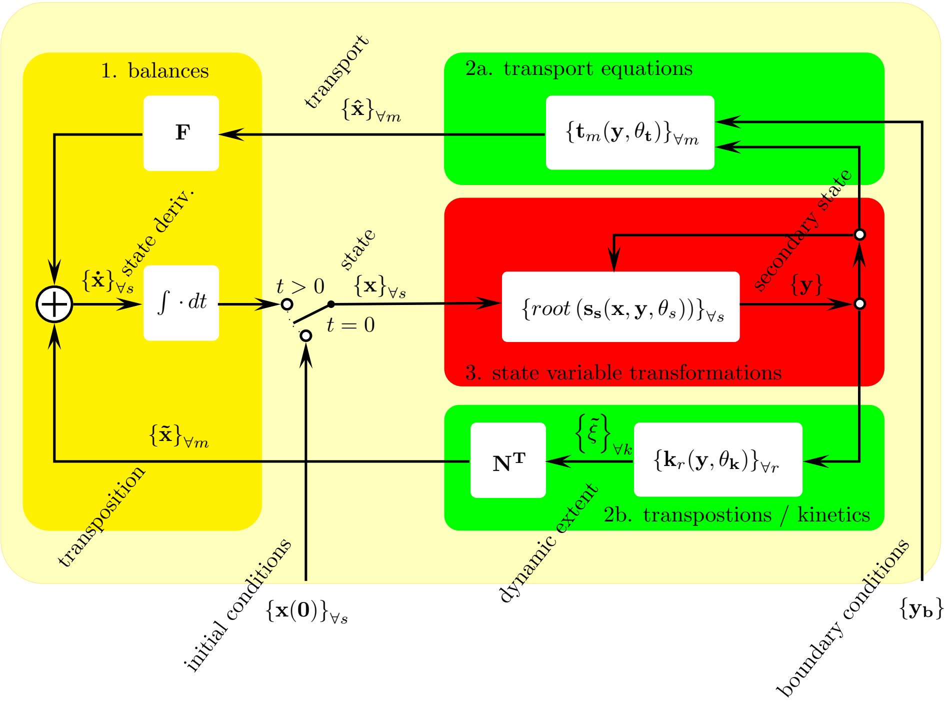

1. balances
transport
$\{\hat{\mathbf{x}}\}_{\forall m}$
2a. transport equations
$\mathbf{F}$
$\{\mathbf{t}_m(\mathbf{y}, \theta_\mathbf{t})\}_{\forall m}$
state deriv.
secondary state
state
$\{\dot{\mathbf{x}}\}_{\forall s}$
$\int \cdot dt$
$t > 0$
$\{\mathbf{x}\}_{\forall s}$
$t = 0$
$\{root\,(\mathbf{s_s}(\mathbf{x}, \mathbf{y}, \theta_s))\}_{\forall s}$
$\{\mathbf{y}\}$
3. state variable transformations
$\{\tilde{\mathbf{x}}\}_{\forall m}$
$\mathbf{N^T}$
$\left\{\tilde{\xi}\right\}_{\forall k}$
$\{\mathbf{k}_r(\mathbf{y}, \theta_\mathbf{k})\}_{\forall r}$
transposition
dynamic extent
2b. transpostions / kinetics
initial conditions
$\{\mathbf{x(0)}\}_{\forall s}$
boundary conditions
$\{\mathbf{y_b}\}$

Figure 15.2: The computation sequence: start with state (initial state at time zero), 1. compute secondary state, 2. compute transport and transposition and 3. the conservations

If we have a good look at the above equation set and define the problem of putting it into a simulation, thus integrating the equations given initial conditions and the "parameters", the computation sequence is precisely the inverse of the definition sequence as it was chosen above: one works its way around the block diagram (15.2) starting with the fundamental state $\underline{\mathbf{x}}$.

First one computes the secondary state, which is the most complex task. It potentially requires the solution of the simultaneous set of equation:

$$\underline{\mathbf{y}}_s, \underline{\mathbf{p}}_i := \text{root}\left(\underline{\mathbf{s}}(\underline{\mathbf{y}}_s, \underline{\mathbf{p}}_i, \underline{\mathbf{x}}_s, \underline{\mathbf{\Theta}}_s)\right) \quad , \forall s,\ i \in [\{a|b\}, r, s]$$

Once the secondary state vector for all systems s are known, the transport and the kinetics can be calculated:

$$\begin{aligned}
\underline{\hat{\mathbf{x}}}_{a|b} &:= \underline{\hat{\mathbf{x}}}_{a|b}\left(\underline{\mathbf{y}}_a, \underline{\mathbf{y}}_b, \underline{\mathbf{p}}_{a|b}\right) \\
\underline{\tilde{\xi}}_r &:= \underline{\tilde{\xi}}_r(\underline{\mathbf{y}}_s, \underline{\mathbf{p}}_r) \\
\underline{\tilde{\mathbf{x}}}_s &:= \underline{\underline{\mathbf{N}}}_s^T\, \underline{\tilde{\xi}}_r
\end{aligned}$$

Which then makes it possible to compute the right-hand sides of the differential equations:

$$\underline{\dot{\mathbf{x}}}_s = \sum_{\forall m} \alpha_{s,m}\underline{\hat{\mathbf{x}}}_m + \underline{\tilde{\mathbf{x}}}_s$$

15.2.5.1 What is to be accurate

When we simulate a plant, we want to have an accurate response from the simulator, whatever it is. But what is accurate? If we think about putting energy into a small system and having limited amount of information about the local geometry, one has to resort to approximations, which often have the feature that the volume/surface ratio is incorrect. Thus putting energy into this system and having a state space which includes the temperature as one of its variables, it is quite obvious that the resulting temperature will be arbitrary and may even approach infinity as the volume goes to zero. This kind of phenomena, which are typical for intensive properties, do not occur when using conserved/extensive quantities as the state, and thus is the fundamental state.

To solve differential equations, we readily use computing components that we take from libraries. Numerical mathematicians have spent a lot of time and effort to make these algorithms robust, which often includes a control of step lengths based on the estimated accuracy of the solution. The common pieces of information the integration facility requires are thus not only the model in the form of a function that computes the state derivatives with time as a function of the inputs and the current state, the initial conditions, the final time, but also a measure for the accuracy besides other things like output intervals. The accuracy measure goes into the part of the integration algorithm that controls the solution progress. Taking the above case into consideration and comparing the representation of the system in the space of energy in contrast to temperature has a obvious advantage as the criterion applies to the energy and not the temperature, latter being extremely sensitive to the modelling of the geometry and physical properties.

In composite systems, where the software will compose the overall model from sub-models, it is desirable that one maintains certain global properties, which in the author's opinion are the closure of the balances, having top priority. One cannot have any confidence in a solution that does not satisfy this basic condition. Unfortunately this is the point to raise a big warning flag indicating that generations of chemical engineers have been educated to represent systems in the state space of intensive properties.

15.2.5.2 Solvability

Numerically one cannot make much of a statement, because at this point no numbers are given. However, in terms of structure one can do a simple analysis by substituting the algebraic parts into the differential equation to show that indeed it reduces to a set of ordinary differential equations, thus the problem is of differential index 1:x

$$\begin{aligned}
\underline{\mathbf{p}}_i &:= \underline{\mathbf{p}}_i(\underline{\mathbf{x}}_s, \underline{\mathbf{y}}_s, \underline{\mathbf{p}}_i) := \underline{\mathbf{s}}^{-1}(\underline{\mathbf{x}}_s, \underline{\boldsymbol{\Theta}}_s) \\
\underline{\hat{\mathbf{x}}}_{a|b} &:= \underline{\hat{\mathbf{x}}}_{a|b}\left(\underline{\mathbf{y}}_a, \underline{\mathbf{y}}_b, \underline{\mathbf{p}}_{a|b}\right) := \underline{\hat{\mathbf{x}}}(\underline{\mathbf{s}}^{-1}(\underline{\mathbf{x}}_s, \underline{\boldsymbol{\Theta}}_s)) := \underline{\hat{\mathbf{x}}}'_m(\underline{\mathbf{x}}_s, \underline{\boldsymbol{\Theta}}_s) \\
\underline{\tilde{\xi}}_r &:= \underline{\tilde{\xi}}_r(\underline{\mathbf{y}}_s, \underline{\mathbf{p}}_r) := \underline{\tilde{\xi}}'_r(\underline{\mathbf{s}}^{-1}(\underline{\mathbf{x}}_s, \underline{\boldsymbol{\Theta}}_s)) \\
\underline{\tilde{\mathbf{x}}}_s &:= \underline{\underline{\mathbf{N}}}_s^T\, \underline{\tilde{\xi}}_r := \underline{\tilde{\mathbf{x}}}(\underline{\mathbf{s}}^{-1}(\underline{\mathbf{x}}_s, \underline{\boldsymbol{\Theta}}_s)) := \underline{\tilde{\mathbf{x}}}'_s(\underline{\mathbf{x}}_s, \underline{\boldsymbol{\Theta}}_s) \\
\underline{\dot{\mathbf{x}}}_s &= \sum_{\forall m} \alpha_{s,m} \underline{\hat{\mathbf{x}}}_m + \underline{\tilde{\mathbf{x}}}_s \\
&= \sum_{\forall m} \alpha_{s,m} \underline{\hat{\mathbf{x}}}'_m(\underline{\mathbf{x}}_s, \underline{\boldsymbol{\Theta}}_s) + \underline{\tilde{\mathbf{x}}}'(\underline{\mathbf{x}}_s, \underline{\boldsymbol{\Theta}}_s)
\end{aligned}$$

which is only a function of the fundamental state and the parameters $\underline{\boldsymbol{\Theta}}_s$

15.2.6 The Text Book Representation

The above representation 15.1- 15.5 represents the plant in the form of a set of ordinary differential equations augmented with a set of differential equations. Traditionally one teaches to represent the model in the space of the observed quantities, where *observed* is to be interpreted as *measurable*. Thus text books will usually aim at a representation in the space of a set of intensive variables such as concentration, temperature and pressure accompanied with an extensive quantity, often volume. Such a representation is usually not minimal but includes obsolete information. For example choosing the concentration vector, temperature, pressure and volume contains at least one obsolete variable as one of the concentrations is a function of the others and the volume. The mapping into the space of *observables* is motivated by the fact that one is most often interested in these quantities and not in the conserved ones. Also, it results in a set of ordinary differential equations that can be solved using standard techniques. No need for a DAE solver. It is thus no surprise that one can find statements such as: *early substitution is a good practice* in standard textbooks.

The two representations are formally linked by a variable transformation. Let the model 15.1- 15.5 be captured in the form

$$\begin{align} \dot{\underline{\mathbf{x}}} &:= \underline{\mathbf{f}}(\underline{\mathbf{v}})\,, \tag{15.8}\\ \underline{\mathbf{v}} &:= \underline{\mathbf{g}}(\underline{\mathbf{y}})\,, \tag{15.9}\\ \underline{\mathbf{y}} &:= \underline{\mathbf{h}}(\underline{\mathbf{x}})\,. \tag{15.10} \end{align}$$

where $\underline{\mathbf{x}}$ is the vector of balanced extensive quantities and $\underline{\mathbf{v}}$ the vector of transports and transpositions and $\underline{\mathbf{y}}$ again the vector of secondary states.

Seeking a representation in the secondary state, for example, one differentiates the secondary state with respect to time:

$$\begin{aligned} \dot{\underline{\mathbf{y}}} &:= \frac{\partial\,\underline{\mathbf{h}}(\underline{\mathbf{x}})}{\partial\,\underline{\mathbf{x}}}\,\dot{\underline{\mathbf{x}}}\,,\\ &:= \frac{\partial\,\underline{\mathbf{h}}(\underline{\mathbf{x}})}{\partial\,\underline{\mathbf{x}}}\,\underline{\mathbf{f}}(\underline{\mathbf{g}}(\underline{\mathbf{y}}))\,. \end{aligned}$$

If the expression $\frac{\partial\,\underline{\mathbf{h}}(\underline{\mathbf{x}})}{\partial\,\underline{\mathbf{x}}}$ is not function of $\underline{\mathbf{x}}$ the result is, as desired, a function of $\underline{\mathbf{y}}$ only. Otherwise it must be possible to find an explicit relation between $\underline{\mathbf{x}}$ and $\underline{\mathbf{y}}$. Whilst the latter mapping must exist, it may only be possible to do so numerically.

In some cases 15.10 is given explicit in the primary state:

$$\underline{\mathbf{x}} := \underline{\mathbf{d}}(\underline{\mathbf{y}})\,,$$

then

$$\begin{aligned} \dot{\underline{\mathbf{x}}} &:= \frac{\partial\,\underline{\mathbf{d}}(\underline{\mathbf{y}})}{\partial\,\underline{\mathbf{y}}}\,\dot{\underline{\mathbf{y}}}\,,\\ \dot{\underline{\mathbf{y}}} &:= \left(\frac{\partial\,\underline{\mathbf{d}}(\underline{\mathbf{y}})}{\partial\,\underline{\mathbf{y}}}\right)^{-1}\,\underline{\mathbf{f}}(\underline{\mathbf{g}}(\underline{\mathbf{y}}))\,, \end{aligned}$$

in which case $\left(\frac{\partial\,\underline{\mathbf{d}}(\underline{\mathbf{y}})}{\partial\,\underline{\mathbf{y}}}\right)$ must be invertible.

15.3 A B C

The $\{\underline{\underline{\mathbf{A}}}, \underline{\underline{\mathbf{B}}}, \underline{\underline{\mathbf{C}}}, \underline{\underline{\mathbf{D}}}\}$ representation of a process model is obtained by linearisation of the above non-linear model (see C.2.3). The four matrices are an abbreviation of the linear-time invariant system (LTI system):

$$\begin{aligned}\underline{\dot{\mathbf{x}}} &= \underline{\underline{\mathbf{A}}}\,\underline{\mathbf{x}} + \underline{\underline{\mathbf{B}}}\,\underline{\mathbf{u}} \\ \underline{\mathbf{y}} &:= \underline{\underline{\mathbf{C}}}\,\underline{\mathbf{x}} + \underline{\underline{\mathbf{D}}}\,\underline{\mathbf{u}}\end{aligned}$$

The four matrices are not a function of time and not a function of either the state $\underline{\mathbf{x}}$ or the input $\underline{\mathbf{u}}$ and are thus constant or *time-invariant*. Interpreting the LTI systems in the physical context is relatively straight-forward, if we have not transformed the original mechanistic description. In this case, the meaning of the state has not really changed except that it represents the linear approximated counter part of the balanced extensive quantities. The fact that it is linearised calls for some cautiousness in the interpretation. The second static equations essentially contain the other three ”boxes” in our representation as shown in 15.1, though with the restriction that the $\underline{\mathbf{y}}$ is typically only a election of the $\underline{\mathbf{y}}$ as shown in 15.1.

The linear time invariant systems are the pet systems of control because they have all the nice things: they are linear in the state, the time derivative of the state and the input. The equations can be readily integrated (C.1.1).

16

Time-scale based model reduction

Synopsis *Models need only be as good as required by the application. Thus target is to be good enough. If one has an unnecessary complex model, why then not simplify it !*

16.1 Nature of models

So models only need to be good enough; they must fit the purpose, the application. There is no question of right or wrong: Models are always wrong simply because a process model never is the process it is modelling. It is though imperative that one must have the purpose, the application of the model in mind, when designing it, particularly when one chooses the structure and the sub-models making up the mimicked behaviour of the modelled process. Having already constructed a model and thus gone through a modelling process makes it desirable to also have the facility to modify existing models such that it fits the new purpose. If the requirement for a modification are such that one requires more detail, one needs to get back to the very beginning and reconsider the topology, thus the basic structure of the model. In the case the structure is generated in a programmed environment, as this is the case in computational fluid dynamics (CFD), then it is the meshing that needs to be repeated generating a new, improved spatial discretisation, thus a new network. In case this has been done manually, one has to go back to the drawing board, which makes it interesting to have a well-designed graphical interface. However in the case where one requires less accuracy, one can consider to simplify the existing model, which is the subject of this chapter.

We construct proper model, which have four blocks: the balances that model the dynamics of the process, the transport providing the information of what is crossing over the boundaries between the adjacent communicating control volumes, the capacity-internal dynamics, namely the transposition of tokens and the state-variable transformations that close the gap between the conserved quantities and the secondary state variables required for the transport and the transposition sub-models. Simplifications are possible for all these parts and all cases these are order-of-magnitude assumptions. Here we discuss the first two, the order-of-magnitude assumptions for reactive systems, namely fast vs slow reactions have been discussed in Section 8.4.2. The technique used to reduce the model is singular perturbation Appendix D.

16.2 Conservation

The dynamics is representation as a network of capacities interacting by exchanging extensive quantities such as mass, energy and momentum.

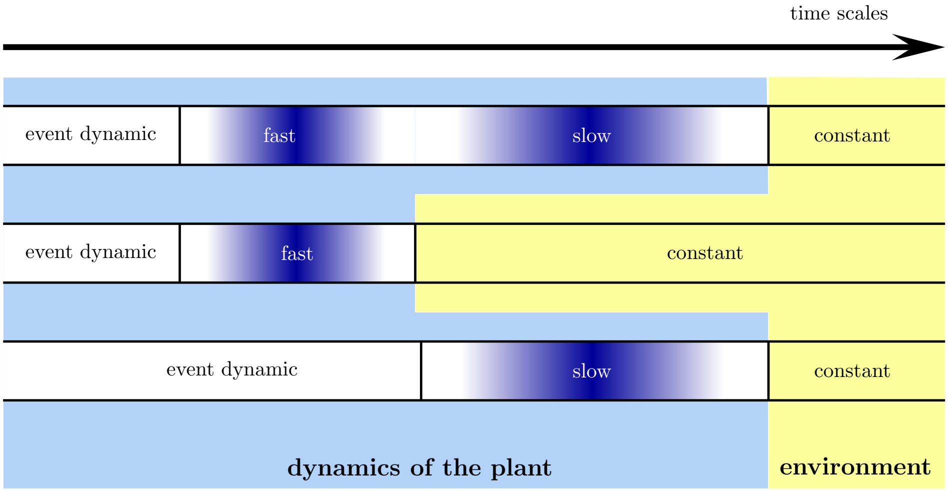


Figure 16.1: Shrinking the dynamic range.

Any representation in the form of a physical topology, the network representing the physical containment, is the result of making time-scale assumptions as it was discussed earlier 2.1.1.2. Simplification of the representation is achieved by narrowing the middle part of the time scales, which is the dynamic range considered by the model. The simplification is only meaningful, if the dynamic range can be split into two time scales, namely a slow one and a fast one. There must be a significant gap between them, which, as a rule of thumb, should be at least of two order-of-magnitudes. If they are closer or even overlap, any split is not appropriate and will result in serious deficiencies of the model's ability to represent the plant's dynamics.

Figure 16.1 shows how the dynamic range can be shrank. In the case where the constant part of the network is made bigger, one shrinks the plant in favour of the environment that is constant by *moving* the slow part into the constant domain, whilst if one extends the event-dynamic part by *moving* the fast parts into the event-dynamic range.

16.2.1 A minimal example

Three system represent a minimal configuration to demonstrate the shifting of the two time scale boundaries, namely three serially connected lumps.

The setup is such that the three lumps have a significant different capacity. The left one is very large, whilst the left one is very small.

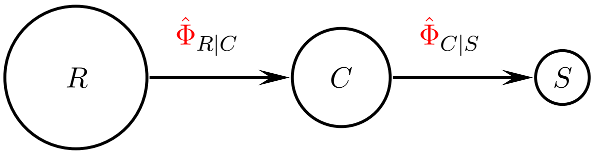


Figure 16.2: A minimal example consisting of three serially connected lumps with decreasing capacity for the token.

The full model would take the form:

$$\begin{aligned}\dot{\Phi}_R &= -\hat{\Phi}_{R|C}\\ \dot{\Phi}_C &= +\hat{\Phi}_{R|C} - \hat{\Phi}_{C|S}\\ \dot{\Phi}_S &= +\hat{\Phi}_{C|S}\end{aligned}$$

16.2.1.1 Expanding the constant domain

The objective of the simplification is to shift the large system R into the constant time-scale domain, thus make it a reservoir. This means that we shift the boundary between the dynamic system and enlarge the time scale characterised as constant.

The simplification involves the ”size” of the system. We need to norm the system to size and then allow it to approach the limit to infinity. This we can achieve by scaling the extensive property Φ in the order of another extensive property, usually the volume. We define the scaling factor ε:

$$\varphi := \frac{\Phi}{\varepsilon}$$

As desired, the volume now shows up as a scaling factor.

The modified model then reads:

$$\begin{aligned}\varepsilon_R\,\dot{\varphi}_R &= -\hat{\Phi}_{R|C}\\ \therefore \quad \dot{\varphi}_R &= -\frac{\hat{\Phi}_{R|C}}{\varepsilon_R}\end{aligned}$$

Now we let the volume approach infinity:

$$\begin{aligned}\dot{\varphi}_R &= -\lim_{\varepsilon_R \to \infty} \left(\frac{\hat{\Phi}_{R|C}}{\varepsilon_R} \right) \\ \dot{\varphi}_R &= 0\end{aligned}$$

which yields that the intensive property is not changing with time, thus is constant. The second consequence is that we cannot balance the system R, but we have to get access to the flow rate $\hat{\Phi}_{R|C}$ either in that the flow can be measured and thus appears as an input or that there exists a model with that is driven be a difference in the effort variables, whereby the effort variable in the reservoir again is measurable. The model thus reduces to the two balances for C and S.

16.2.1.2 Introducing an event-dynamic domain

The original model had no event-dynamic parts. So we introduce one by assuming the capacity of the lump S is small compared to the other two. We choose a scaling factor that is in the order of the volume. The scaling factor is the singular perturbation parameter:

$$\varepsilon_S \, \dot{\varphi}_S^j = \hat{\Phi}_{C|S}$$

Taking the limit of the volume to go to zero:

$$\lim_{\varepsilon_S \to 0} \varepsilon_S \, \dot{\varphi}_S^j = \hat{\Phi}_{C|S}$$

which says that the token exchange is negligible. So if start with the model with the extended constant domain, we end up with a single balance, namely the one for the middle lump C.

$$\dot{\Phi}_C = +\hat{\Phi}_{R|C}$$

with the $\hat{\Phi}_{R|C}$ being known. This represents the outer solution for the $R-S$ system.

16.2.1.3 Analysing the fast system

The modeller may not only be interested on the inner solution, in which the dynamic of the fast part dominates. This represents an extension of the constant domain, but now seen from the small system S. The approach is identical to the one taken above, just now we find the result that the lump C is being shifted into the constant domain and we are left with the dynamics of S.

16.3 Transport

In contrast to the capacity discussion, here order-of-magnitude assumptions are made about transport. One side is simple, namely nothing is flowing between two capacities, which simply nullifies the corresponding connection, it removes it. The interesting assumption is a transfer to be fast, faster than the other relevant transfers. Our sample system serves as an exercise ground. If we take model we obtained after expanding the constant domain, we have two balances and we now need in addition the two models for the transfers. We use the simple models where the transfer is driven by the discrete gradient of the effort variables on either side of the connection:

$$\begin{aligned} \hat{\Phi}_{R|C} &:= -\Theta_{R|C}\ (\pi_C - \pi_R) \\ \hat{\Phi}_{C|S} &:= -\Theta_{C|S}\ (\pi_S - \pi_C) \end{aligned}$$

Since we apply the order-of-magnitude assumption to the transfer parameters and the said parameter will define the singular perturbation parameter, we need to substituting the two transfer laws into the balance equations first:

$$\begin{aligned} \dot{\Phi}_C &= -\Theta_{R|C}\ (\pi_C - \pi_R) + \Theta_{C|S}\ (\pi_S - \pi_C) \\ \dot{\Phi}_S &= -\Theta_{C|S}\ (\pi_S - \pi_C) \end{aligned}$$

We now introduce the order-of-magnitude assumption that the $\Theta_{C|S}$ is much larger than the $\Theta_{R|C}$.

16.3.1 Small time scale

We zoom into the time and stretch the time axis such that τ is in the order of 1 in the boundary layer:

$$\tau := \varepsilon^{-1}\, t$$

The dynamic equations in the ”new time” introduce the scaling factor:

$$\begin{aligned} \varepsilon^{-1}\frac{d\,\Phi_C}{d\,\tau} &= -\Theta_{R|C}\ (\pi_C - \pi_R) + \Theta_{C|S}\ (\pi_S - \pi_C) \\ \varepsilon^{-1}\frac{d\,\Phi_S}{d\,\tau} &= -\Theta_{C|S}\ (\pi_S - \pi_C) \end{aligned}$$

Multiplication with the scale (ε) yields

$$\begin{aligned} \frac{d\,\Phi_C}{d\,\tau} &= -\varepsilon\,\Theta_{R|C}\ (\pi_C - \pi_R) + \varepsilon\,\Theta_{C|S}\ (\pi_S - \pi_C) \\ \frac{d\,\Phi_S}{d\,\tau} &= -\varepsilon\,\Theta_{C|S}\ (\pi_S - \pi_C) \end{aligned}$$

In order to see boundary layer behaviour, the scale is chosen in the same order of magnitude as the large transport parameter. Thus the product $\varepsilon\,\Theta_{C|S}$ is much larger than the product $\varepsilon\,\Theta_{R|C}$. Which leads to the simplification:

$$\begin{aligned}\frac{d\,\Phi_C}{d\,\tau} &= +\varepsilon\,\Theta_{C|S}\;(\pi_S - \pi_C)\\ \frac{d\,\Phi_S}{d\,\tau} &= -\varepsilon\,\Theta_{C|S}\;(\pi_S - \pi_C)\end{aligned}$$

which is the same as:

$$\begin{aligned}\dot{\Phi}_C &= +\Theta_{C|S}\;(\pi_S - \pi_C)\\ \dot{\Phi}_S &= -\Theta_{C|S}\;(\pi_S - \pi_C)\end{aligned}$$

Since we have in addition assumed that the capacity of the lump C is much larger than the capacity of lump S, we can in addition apply the simplification of the large/small capacity for the short time scale and get:

$$\dot{\Phi}_S = -\Theta_{C|S}\;(\pi_S - \pi_C)$$

with π_C being constant.

16.3.2 Large time scale

This is the more commonly used approximation, as one is more often interested in the long time scale. For the singular perturbation to apply, we divide both equations by the large parameter $\Theta_{C|S}$:

$$\begin{aligned}\Theta_{C|S}^{-1}\,\dot{\Phi}_C &= -\frac{\Theta_{R|C}}{\Theta_{C|S}}\;(\pi_C - \pi_R) + (\pi_S - \pi_C)\\ \Theta_{C|S}^{-1}\,\dot{\Phi}_S &= -\,(\pi_S - \pi_C)\end{aligned}$$

and take the limit of $\Theta_{C|S}^{-1} \to 0$. The result is that

$$0 = (\pi_S - \pi_C)$$

which implies that the two lumps C and S are in equilibrium with respect to the transferred token. But, having made the order-of-magnitude assumption, we do not know on how much of the token is being transferred between the two lumps.

16.4 Fazit

16.4.1 Small time scale

The focus is on the fast system shifting the slow system into the constant domain. Thus the constant domain is extended and one zooms into the time scale where things happen quickly

Small/large capacities: The fast system is analysed in isolation, the large capacities are seen as constant.

Slow/fast transfer: The slow transport is ignored and focus is on the fast transport

16.4.2 Large time scale

Small/large capacities: Reduction based on small capacity simply leads to neglecting the small capacity.

Slow/fast transfer: The result is that the two strongly-connected systems will, for all practical purposes, be at equilibrium with regard to the transferred token. But we do not know how much of the token is being transferred. It provides though an argument to simplify the topology for this particular token by simply eliminating the unknown transfer by combining the two connected primitive system.

16.5 Fast networks and slow networks

Many systems have the properties of being represented as multiple networks some of which have quite different dynamic properties than the other. In processes any of the units may be represented as a sub-network, for example a distillation, a reactor, a heat exchanger. The fact that the model components are now networks, in contrast to primitive system, requires additional considerations.

The reduction process aims at lumping relative small capacities between which extensive quantities are transferred relatively quickly. If the model has been transformed into the space of intensive variables, one can argue on the basis of time constants, that is, on the ratio of the stream parameter and the system capacity. This ratio expresses the inverse of the time constant for the process and a particular flow. If the time constant is significantly shorter in one case, then the two systems coupled by the mentioned stream have nearly the same intensive properties in the longer time scale. The reduction is thus the implementation of an order-of-magnitude assumption with regards to the capacities and flows in the frame of a longer time scale than what was used to establish the original model.

Figure 16.4 shows the generic structure of a fast network coupled with a slow network.

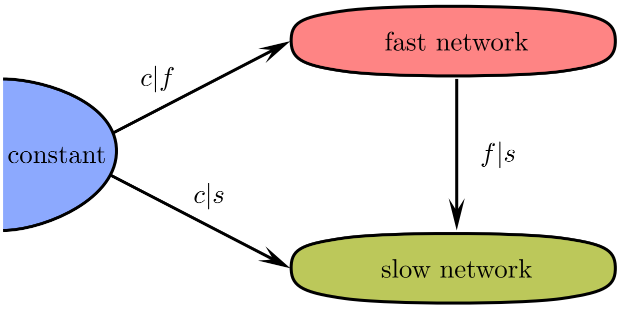


Figure 16.3: A slow network connected to a fast network

The splitting of the dynamic part of the network into a fast and a slow network is formally done using selection matrices that grab the respective parts. We start with the network repesentation:

$$\dot{\underline{\mathbf{\Phi}}} := \underline{\underline{\mathbf{F}}}\,\hat{\underline{\mathbf{\Phi}}}$$

Multiplying the set of equations with the respective row selection matrix $\underline{\underline{\mathbf{S}}}^r_f$ and $\underline{\underline{\mathbf{S}}}^r_s$, we get the desired split indicated by the two indices s, f:

$$\begin{aligned} \dot{\underline{\mathbf{\Phi}}}_f &:= \underline{\underline{\mathbf{R}}}_f\,\underline{\underline{\mathbf{F}}}\,\hat{\underline{\mathbf{\Phi}}} \\ \dot{\underline{\mathbf{\Phi}}}_s &:= \underline{\underline{\mathbf{R}}}_s\,\underline{\underline{\mathbf{F}}}\,\hat{\underline{\mathbf{\Phi}}} \end{aligned}$$

Next we split the flow vector into the 5 parts, one for the internal flows in the slow and the fast sub-networks, and one for each flow as indicated in Figure 16.4:

$$\begin{aligned} \dot{\underline{\mathbf{\Phi}}}_f &:= \underline{\underline{\mathbf{R}}}_f\,\underline{\underline{\mathbf{F}}}\,\underline{\underline{\mathbf{C}}}_{f|f}\,\underline{\underline{\mathbf{C}}}^T_{f|f}\hat{\underline{\mathbf{\Phi}}} + \underline{\underline{\mathbf{R}}}_f\,\underline{\underline{\mathbf{F}}}\,\underline{\underline{\mathbf{C}}}_{f|s}\,\underline{\underline{\mathbf{C}}}^T_{f|s}\hat{\underline{\mathbf{\Phi}}} + \underline{\underline{\mathbf{R}}}_f\,\underline{\underline{\mathbf{F}}}\,\underline{\underline{\mathbf{C}}}_{c|f}\,\underline{\underline{\mathbf{C}}}^T_{c|f}\hat{\underline{\mathbf{\Phi}}} \\ \dot{\underline{\mathbf{\Phi}}}_s &:= \underline{\underline{\mathbf{R}}}_s\,\underline{\underline{\mathbf{F}}}\,\underline{\underline{\mathbf{C}}}_{s|s}\,\underline{\underline{\mathbf{C}}}^T_{s|s}\hat{\underline{\mathbf{\Phi}}} + \underline{\underline{\mathbf{R}}}_s\,\underline{\underline{\mathbf{F}}}\,\underline{\underline{\mathbf{C}}}_{f|s}\,\underline{\underline{\mathbf{C}}}^T_{f|s}\hat{\underline{\mathbf{\Phi}}} + \underline{\underline{\mathbf{R}}}_s\,\underline{\underline{\mathbf{F}}}\,\underline{\underline{\mathbf{C}}}_{c|s}\,\underline{\underline{\mathbf{C}}}^T_{c|s}\hat{\underline{\mathbf{\Phi}}} \end{aligned}$$

This can be compacted by defining:

$$\underline{\underline{\mathbf{F}}}_{\bullet|\circ} := \underline{\underline{\mathbf{R}}}_{\bullet}\,\underline{\underline{\mathbf{F}}}\,\underline{\underline{\mathbf{C}}}_{\bullet|\circ}$$
$$\underline{\hat{\mathbf{\Phi}}}_{\bullet|\circ} := \underline{\underline{\mathbf{C}}}^{T}_{\bullet|\circ}\,\underline{\hat{\mathbf{\Phi}}}$$

yielding:

$$\underline{\dot{\mathbf{\Phi}}}_f := \underline{\underline{\mathbf{F}}}_{f|f}\,\underline{\hat{\mathbf{\Phi}}}_{f|f} + \underline{\underline{\mathbf{F}}}_{f|s}\,\underline{\hat{\mathbf{\Phi}}}_{f|s} + \underline{\underline{\mathbf{F}}}_{c|f}\,\underline{\hat{\mathbf{\Phi}}}_{c|f}$$
$$\underline{\dot{\mathbf{\Phi}}}_s := \underline{\underline{\mathbf{F}}}_{s|s}\,\underline{\hat{\mathbf{\Phi}}}_{s|s} + \underline{\underline{\mathbf{F}}}_{s|f}\,\underline{\hat{\mathbf{\Phi}}}_{f|s} + \underline{\underline{\mathbf{F}}}_{c|s}\,\underline{\hat{\mathbf{\Phi}}}_{c|s}$$

Focusing onto the short time scale does not include any surprises; it simply shifts the slow network into the constant domain. More interesting is the opposite, namely a focus on the slow time scale. With "fast network" we imply that for the fast network, the capacities are comparable in "size" and the flows between them are fast compared to the capacities. The simplification is then to eliminate all internal streams in the fast network by combining all nodes together into one. Formally this is the null space of the matrix $\underline{\underline{\mathbf{F}}}_{f|f}$. We define the left null space matrix $\underline{\underline{\mathbf{\Omega}}}$ and get:

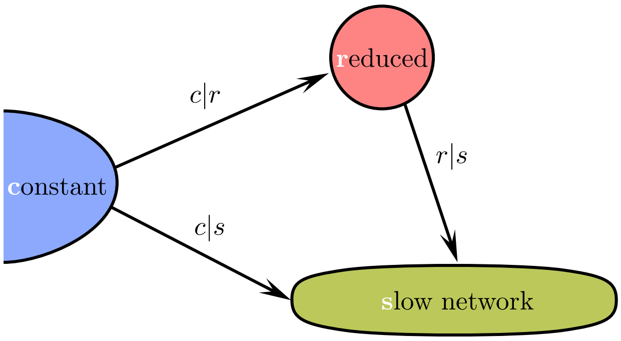


Figure 16.4: A slow network connected to a fast network

$$\underline{\underline{\mathbf{\Omega}}}\,\underline{\dot{\mathbf{\Phi}}}_f := \underline{\underline{\mathbf{\Omega}}}\,\underline{\underline{\mathbf{F}}}_{f|f}\,\underline{\hat{\mathbf{\Phi}}}_{f|f} + \underline{\underline{\mathbf{\Omega}}}\,\underline{\underline{\mathbf{F}}}_{f|s}\,\underline{\hat{\mathbf{\Phi}}}_{f|s} + \underline{\underline{\mathbf{\Omega}}}\,\underline{\underline{\mathbf{F}}}_{c|f}\,\underline{\hat{\mathbf{\Phi}}}_{c|f}$$

with

$$\underline{\underline{\mathbf{0}}} := \underline{\underline{\mathbf{\Omega}}}\,\underline{\underline{\mathbf{F}}}_{f|f}$$

The null space is easy to find as each column in the $\underline{\underline{\mathbf{F}}}$ has only two elements, namely a $+1$ and a -1. Thus the null space matrix is simply a row vector of ones. The topology has changed, the reduction replaces the fast network with one single node. It represents the cumulative capacity of all the nodes in the fast network. This node is shown as ”reduced” in Figure 16.4. The new incidence list are formally obtained by replacing all node identifiers now lumped into ”reduced” are replaced by the identifier of ”reduced”. So, the flows between the fast network and the other two networks simplify to be all linked to this single node that replaces the fast network.

$$\dot{\Phi}_r := \underline{\underline{\mathbf{F}}}_{r|s}\,\underline{\hat{\mathbf{\Phi}}}_{r|s} + \underline{\underline{\mathbf{F}}}_{c|r}\,\underline{\hat{\mathbf{\Phi}}}_{c|r}$$

with $\underline{\underline{\mathbf{F}}}_{r|s} := \underline{\underline{\mathbf{\Omega}}}\,\underline{\underline{\mathbf{F}}}_{f|s}$ and $\underline{\underline{\mathbf{F}}}_{c|r} := \underline{\underline{\mathbf{\Omega}}}\,\underline{\underline{\mathbf{F}}}_{c|f}$.

16.5.1 Assumptions on Assemblies

It is not uncommon that one has knowledge about a state-dependent quantity of an assembly of primitive system and thus stimulates implementing an assumption. A well-known example is the assumption of constant, known volume of a multiphase system that is enclosed in a common confinement. A flash tank or a tank with an overflow are a well-known examples.

Given the standard network model

$$\underline{\dot{\mathbf{x}}} := \underline{\underline{\mathbf{F}}}\,\underline{\hat{\mathbf{x}}} + \underline{\tilde{\mathbf{x}}}\ ,$$

one can split the network into two subsection thereby isolating the part for which the assumption shall be made. Let matrix $\underline{\underline{\mathbf{S}}}_a$ be a selection matrix that is non-square and isolates the part for which the assumption shall be made. Further let $\underline{\underline{\mathbf{\Omega}}}$ be a matrix of the dimension k x n , then it a typical assembly assumption is

$$\underline{\underline{\mathbf{\Omega}}}\,\underline{\underline{\mathbf{S}}}_a\,\underline{\dot{\mathbf{x}}} := \underline{\underline{\mathbf{\Omega}}}\,\underline{\underline{\mathbf{S}}}_a\,\underline{\underline{\mathbf{F}}}\,\underline{\hat{\mathbf{x}}} + \underline{\underline{\mathbf{\Omega}}}\,\underline{\underline{\mathbf{S}}}_a\,\underline{\tilde{\mathbf{x}}} := \underline{\mathbf{0}}\,.$$

that is a linear combination of the states is constant. This then defines k algebraic constrains providing equations for k dependent algebraic variables. The above equations may be used to determine a set of dependent quantities. Bi-partite graph analysis can here help to determine the set of possible quantities that can be determined in a specific case. Further, the above equations can be added to the other part thereby eliminating the connecting streams, but providing the opportunity to possibly compute quantities that depend on the algebraic constraints.

16.5.2 Assumptions in the Space of the Secondary States

The network models, being formulated in the space of the conserved quantities, which we term the primary state space, can be transformed into secondary state space by means of state variable transformations. In fact in chemical engineering models in the secondary states are more common than in the primary, because substituting as early as possible is considered a good mathematical praxis. Thus one usually does not use the models in the primary state space, which is also a minimal space. The approach discussed here represents therefore a deviation from the standard chemical engineering practice.

The transformation can be formalized readily yielding

$$\underline{\underline{\mathbf{J}}}_x^y \, \dot{\underline{\mathbf{x}}} := \underline{\underline{\mathbf{J}}}_x^y \, \underline{\underline{\mathbf{F}}} \, \hat{\underline{\mathbf{x}}} + \underline{\underline{\mathbf{J}}}_x^y \, \underline{\underline{\mathbf{S}}} \, \tilde{\underline{\mathbf{x}}} \, .$$

with:

$$\underline{\underline{\mathbf{J}}}_x^y := \frac{\partial \, \underline{\mathbf{y}}}{\partial \, \underline{\mathbf{x}}^T} \, .$$

resulting in a transformed model:

$$\dot{\underline{\mathbf{y}}} := \underline{\underline{\mathbf{F}}}^y \, \hat{\underline{\mathbf{x}}}(\underline{\mathbf{y}}) + \underline{\underline{\mathbf{S}}}^y \, \tilde{\underline{\mathbf{x}}}(\underline{\mathbf{y}}) \, .$$

From this point on one can implement the same assumptions as they were discussed above. Very common is the assumption of constant volume for single systems or assemblies.

16.5.3 Unmodelled Components

When modelling a plant it is quite common that one does not have much of a clue on what precisely happens in a plant, but may have a thought on what the effect is. Two very common problems are that one does not know on how to model a flow precisely or what reaction is taking place. However, one knows that certain things are being controlled, for example the temperature, or that the volume is approximately constant in an overflow situation etc., or the capacity effects can be neglected.

Thus the before mentioned simplifications can be used to ”determine” the missing model components by means of making order-of-magnitude assumptions. One may assume the capacity effect is zero, or a secondary respective-dependent state variable is constant, or event dynamics on the flow in question yields an algebraic condition for the missing stream information.

Part II

Model & experiment

17

System Identification

17.1 Matching the Model to the Plant

There is really only one purpose for system identification, and that is to find an appropriate model for the modelled plant limited to the range of operating conditions in which the identification experiments can be performed.

So why do these models have to be matched with, and what about these operating conditions? *Matching* is necessary because the model is not a precise image of the plant and it is not always such that all the information about the plant's behaviour is known in all details thought some of them may be necessary to be included in the model in order to meet the specification one has defined for the use of the model. Thus system identification is done to find a model describing the process on the level of details required for the application of the model. Application can be anything from just trying to understand the behaviour of the system, to using it for design and operational tasks such as control.

What about the operating conditions? Plants, or any system for that matter, must be disturbed, excited, as the specialist calls it, in order for the process to reveal his behaviour. For example, in order to find out on how heavy something is, one has to accelerate it or expose it to a gravitational field. The same for any other process: it must be moved about in order to test out its behaviour. For the purpose of identification one thus injects a well-controlled disturbance, an excitation signal, that *moves* the process, that is, changes its state. The model is then fed with the same excitation signal and the behaviour of the plant and the simulated process is compared on the basis of which the model is changed. The model is changed until its behaviour fits *satisfactorily* within the plant's range of operation, whereby *satisfactorily* is determined by introducing a measure for the difference between the plant and the model.

Process identification has been subject of research for as long as models are defined. The recent literature body includes the review paper of Åström and Eykhoff, the book by Eykhoff and the book on the subject by Ljung (Ljung (1987); Eykhoff (1974); Astroem and Eykhoff (1971)). The subject has also been of interest in the statistics community in particular associated with parameter identification and signal processing.

17.2 Defining System Identification

We shall define system identification as follows:

Given a set of models $\mathcal{M} := \{M_i | \forall i\}$ where each of the models may belong to a specific class of models, the system identification task is to find the best model in the defined set given records of input-output data $\mathcal{D}$ from the plant obtained under operating conditions $\mathcal{C}$ where *best* is measured by the criterion $\mathbf{J}$.

In the case where the model set $\mathcal{M}$ consists of structurally different models, one talks about *system identification*. In the case where the set consists of one parameterised model with the varying parameters being the set generator, one talks about *parameter identification*[1].

Fitting *best* implies that system and parameter identification is an optimisation problem. The measure must be suitable to be used in an optimisation and convexity is a desired property. The sum of squares of the deviation, where deviation needs to be defined, is the most commonly used criterion, though also other norms are suitable for the purpose, the 2-norm being mathematically easy to handle.

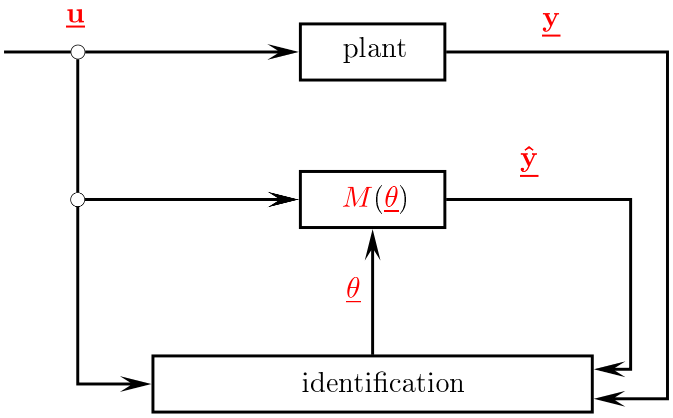


Figure 17.1: The grand scheme of parameter identification: The plant is excited with a sufficiently rich signal to stimulate the interesting modes of the plant. The same input is used to simulate the model's (M) behaviour. All three generated signals, namely the excitation signal $\underline{\mathbf{u}}$, the plant's response $\underline{\mathbf{y}}$, and the model's response $\underline{\hat{\mathbf{y}}}$ is used to compute an estimate of the model's parameters, which then are used to update the model.

[1] The two things may overlap in that a parameter appearing as a factor in an expression may eliminate the associated term from the model as this parameter assumes the value zero. The *zero* takes thus a somewhat special position when interpreting model structures. This fact is extensively used in network representations such as neural nets

The fact that the necessary experiments can usually only be done in a limited range of operating conditions is often not sufficiently appreciated, because the identified model is strictly speaking only valid for the range the model has been validated, which usually coincides with the range in which the model has been identified. There is no guarantee on the extrapolation ability of the model, even if the model is a mechanistic model. For dynamic models the operating conditions may be best characterised by the frequency range and the amplitude range in which the identification experiments were performed. They define some kind of spectral conditions, which for example in robust control become very handy to have available.

In any case, the accuracy of the identified model should ultimately be judged in the framework of the application of the model. Thus for example if the model is used for controller design, the performance of the controlled process should be taken as the ultimate measure. This underlines the statement that the model is being constructed for a particular purpose, a fact that should be kept in mind at all times.

17.2.1 Consequences

Having defined the task *system identification*, it is apparent that the identified models are a function of all the elements entering the procedure: the data, the set of models and the criterion: The **criterion** provides the measure, thus the result is obviously dependent on what yard stick is being used. The most common choice is the sum of squares, mostly because of its nice mathematical properties. It usually serves its purpose very well indeed to the extent that most people would not spend a thought on the choice. The **model** set has a rather obvious effect on the result as the parameters are strictly speaking defined in the context of the model. Not having a model in the set straightforwardly means that it is not being considered – obvious indeed, but not in a hidden context. The **input**, namely the **excitation signal** being used for the identification period, has a huge impact on the result. This fact is much too often ignored and ”standard” excitation signals are being applied without being aware what effect they have on the plant and consequently on the identified model parameters. If one views the problem in the frequency domain one gets quite quickly a good insight. Figure 17.2 shows the frequency behaviour of two models for the same process. The one with the steeper asymptote and higher phase shift is the more complex one. Assuming that the more complex model indeed describes the plant better, one observes that the simpler model does very well up to a frequency of about 1 Hz. Above the phase changes to the double quite quickly. If one thinks about identification, then one observes two major

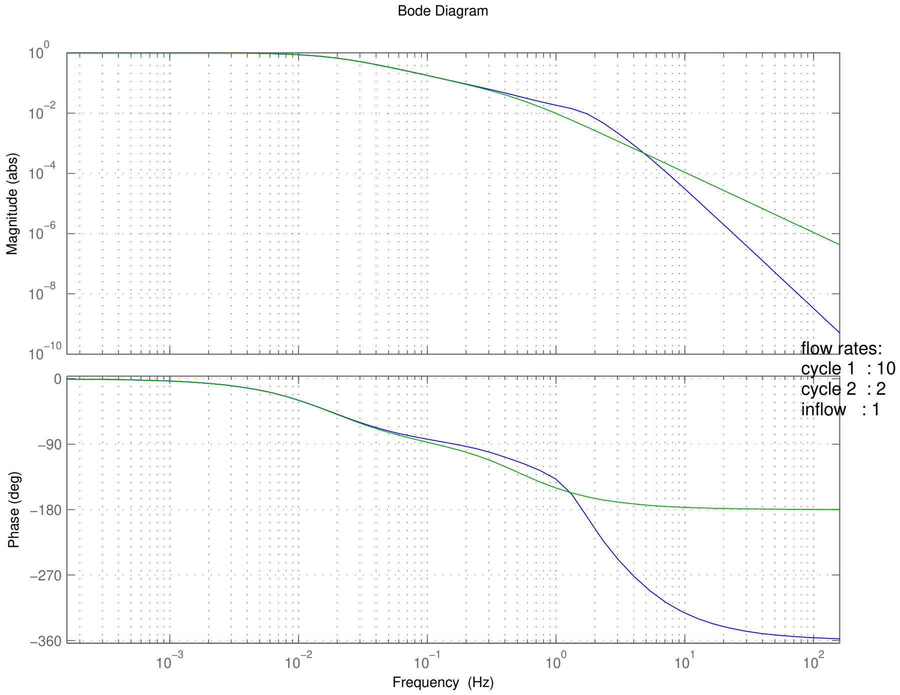


Figure 17.2: Bode plot of two models, a complicated and a simplified one.

parameters, each represented by a corner or a bend in the amplitude plot: One 3 10^{-2} Hz and the other around 2 Hz for the complex case and about 1 Hz for the simplified model. If one uses an input signal that is on the high end frequency limited around 0.01 Hz, none of the corners can be extracted as the output signal will be essentially the same as the input signal. Thus only the steady state value can be obtained from this experiment. If one increases the frequency contents to 0.1 or 1 Hz, one will see the effect of the first corner: The amplitude drops and the output is shifted by about 90 degrees. However, the output signal will have no information about the second corner and the beyond. In order to find this second corner, one has to experiment in the domain of 1 to 10 Hz.. This example makes it apparent that the frequency contents of the excitation signal is essential for the procedure and it is recommendable that one spend some time on designing the experiment so as to tickle the process at the right point, so-to-speak.

17.3 Models

With models being the main objective, it is put into the centre, whilst the methodologies associated with identification is put into the second place as it is extensively treated in the literature: for example Ljung (1987); Eykhoff (1974); Astroem and Eykhoff (1971).

Models are typically classified using attributes such as linear, nonlinear, stochastic, parameterised, discrete and continuous.

But for example what does linearity mean? Most commonly the term *linearity* is used in connection with the state, so more precisely: linear-in-the-state systems, which reflects that on e is primarily interested in the evolution of the state, thus process simulation. In identification, one is mostly interested in linear-in-the-parameters, as it is the parameters that one is solving for. Nonlinearities in the inputs are usually quite manageable, whilst if one is interested in using the model for control, nonlinearities represent a major obstacle.

Literature uses often the term *parameterised* and the opposite *un-parameterised* for model classification. This attribute is not seen as a very descriptive and we rather use data-driven instead of un-parameterised.

Discrete and continuous models: On the macroscopic scale nature is well approximated by continuous systems, that is, the state is a continuous function of time and spatial co-ordinates.

17.3.1 Data-driven Models

Data models come as input-output data or time series. As such system responses belong into this class such as impulse response and step response being the two main ones. The impulse response is the chemical engineer's residence time distribution. Numerically convoluting the impulse response with an input series results the response of the system.

Assuming discretely changing inputs one can apply the step response for each time step to obtain the response of the system. The step response as a model is extensively used in model predictive control applications. The background is linear time-invariant systems.

Using fast Fourier transform techniques, one can use tabled information about the transfer function to obtain input/output data.

17.3.2 Special Forms

17.3.2.1 Hammerstein Model

Nonlinear input transformation followed by a linear dynamic system. Hammerstein models appear for example when making event-dynamic assumptions for parts of the systems which convert these parts into transfer systems.

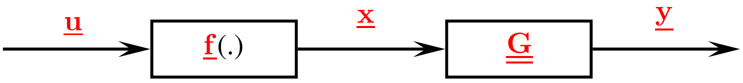


Figure 17.3: Hammerstein model

17.3.2.2 Wiener Model

Linear dynamic system followed by a nonlinear output transformation. Wiener models appear for example when making event-dynamic assumptions about reactions occurring in a capacity. The dynamics part then represent the hydraulic, whilst the static part represents the event-dynamic reaction system.

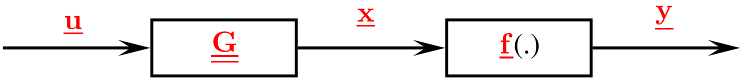


Figure 17.4: Wiener model

17.3.2.3 Static L-i-P (Linear-in-Parameters) Models

This type of model is used to describe the stationary behaviour of processing systems. With chemical engineering being traditionally taught based on stationary behaviour of continuous plants, this type of model is of particular interest to this group.

A rather generic formulation of the lip model is multi-input, single output:

$$y := \underline{\mathbf{f}}^T(\underline{\mathbf{u}})\,\underline{\theta}$$

where the vector of functions $\underline{\mathbf{f}}$ may be nonlinear in the input vector $\underline{\mathbf{u}}$. The model is clearly linear in the parameters $\underline{\theta} \in \mathbb{R}^k$.

The nonlinearity of the function $\underline{\mathbf{f}}$ of $\underline{\mathbf{u}}$ is virtually arbitrary. The most common structures being used are polynomials and exponentials. For example:

$$\underline{\mathbf{f}}^T(\underline{\mathbf{u}}) := [u^r]_{r:=1,2,\ldots,1/2,1/3\ldots},$$

but also mixed nonlinearities are permitted.

17.4 The Analytical Framework

The following sections are devoted to process identification both for stationary systems and dynamic systems. In the previous section we defined *identification* as an operation that fits a model to the process. With the process usually being a physical object, one has no means to quantify the mismatch between the model and the process except than comparing responses to excitations. So quality can only be assessed on the basis of comparative descriptive power measured on the deviation of the observed process quantities and the corresponding signals obtained from the model simulation using the same excitation signal. This is what we reflected in Figure 17.5.

If one though works on the methods, one is interested to learn what effects certain model-mismatches have one the results, being the parameter estimates and their stochastic properties. So the logical approach is to replace the physical model by an abstract mathematical object, which we call *nominal model*. This nominal model can then be tailored to exactly deviate from the identified model as desired. We also added two noise signals: one that is added to the "plants" input and one that adds to the observation. This again is a simplification of the "reality" in that this assumes that the noise is additive to the respective signals.

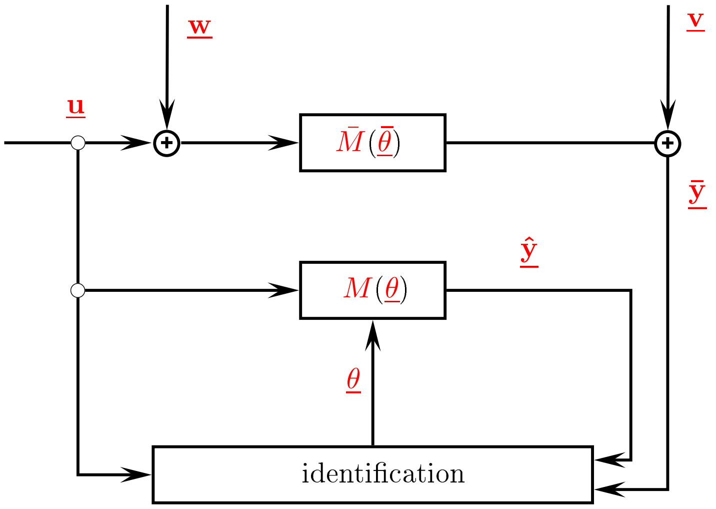


Figure 17.5: Analysing grand scheme of parameter identification: The plant is replace by a nominal model of which one has complete knowledge, a mathematical model. Otherwise the scheme remains the same: All three generated signals, namely the excitation signal $\underline{\mathbf{u}}$, the plant's response $\underline{\bar{\mathbf{y}}}$, and the model's response $\underline{\hat{\mathbf{y}}}$ is used to compute an estimate of the model's parameters, which then are used to update the model.

17.5 Point Estimators

The fitting of a model to an experiment uses the input/ouput data from the process to which the model is being fitted. Estimators Ψ are "rules" on how to compute the parameters of a model[2]. The process is usually a physical process, which we can observe. If we are interested in discussing the properties of the estimation procedure, then comparing what was observed on the process and compare it with the reconstructed observation using the model is not providing us with enough information to discuss the properties. For this purpose we have to introduce an abstract substitute of which we have complete knowledge: a mathematical model. This model we term *nominal model*, which in general will be more complex than the model being fitted.

Let our nominal model be:

$$\underline{\bar{\mathbf{y}}} := \bar{M}(\underline{\mathbf{u}}, \underline{\bar{\theta}}) + \underline{\mathbf{v}} \tag{17.1}$$

The $\underline{\mathbf{v}}$ is a vector of random variables that satisfy the Gauss-Markov as-

[2] The following follows closely the book of Goodwin and Payne (1977)

sumptions (see Section G):

$$\mathbf{E}\left[\underline{\mathbf{v}}\right] := 0$$
$$\mathbf{E}\left[\left(\underline{\mathbf{v}} - \mathbf{E}[\underline{\mathbf{v}}]\right)^2\right] := \sigma^2 < \infty$$

The analysis of estimators aims at defining its quality, so how *good* the estimator is. The first property we going to test a estimator for is if it gives back the information we expect, namely the parameters of the model if applied to the data generated by the same model. So we apply the estimator $\underline{\bar{\boldsymbol{\Psi}}} := \Psi_i\left(\underline{\mathbf{u}}, \underline{\bar{\mathbf{y}}}\right)$ to our nominal model $\bar{M}$ and we expect the nominal parameters $\underline{\bar{\theta}}$ back:

Property - Unbiased : An estimator is called unbiased if

$$\mathbf{E}\left[\underline{\bar{\boldsymbol{\Psi}}}\right] = \underline{\bar{\theta}}$$

Property - Uniformly minimal mean square error : An estimator $\underline{\bar{\boldsymbol{\Psi}}}$ for a parameter $\underline{\bar{\theta}}$ is said to be *uniformly minimal mean square error* if

$$\mathbf{E}\left[\left(\underline{\bar{\boldsymbol{\Psi}}}_i - \underline{\bar{\theta}}\right)\,\left(\underline{\bar{\boldsymbol{\Psi}}}_i - \underline{\bar{\theta}})^T\right)\right] \leq \mathbf{E}\left[\left(\underline{\bar{\boldsymbol{\Psi}}}_j\left(\underline{\mathbf{u}}, \underline{\bar{\mathbf{y}}}\right) - \underline{\bar{\theta}}\right)\,\left(\underline{\bar{\boldsymbol{\Psi}}}_j\left(\underline{\mathbf{u}}, \underline{\bar{\mathbf{y}}}\right) - \underline{\bar{\theta}})^T\right)\right]$$

for all estimators in the set $\left\{\underline{\bar{\boldsymbol{\Psi}}}_j | \forall j\right\}$

Definition - Minimum variance unbiased estimator MVUE : Estimator that is unbiased and has the property uniformly minimal means square error.

Definition - Best linear unbiased estimator BLUE : MVU Estimator that is a linear function of the data.

It is often not feasible to find a MVUE or a BLUE estimator, but usually suffices to use an estimator that approaches the lower variance bound defined by the Cramer-Rao inequality Goodwin and Payne (1977):

Theorem 17.5.1 (Cramer-Rao inequality). *Let $P_{\underline{\bar{\theta}}}$ be a family of distributions on a sample space Ω with the density $p_{\underline{\bar{\mathbf{y}}}|\underline{\bar{\theta}}}$, then, subject to some regularity conditions, the covariance $\underline{\underline{\mathbf{V}}}(\underline{\bar{\boldsymbol{\Psi}}})$ of any unbiased estimator $\underline{\bar{\boldsymbol{\Psi}}}$ of $\underline{\bar{\theta}}$ satisfies the inequality*

$$\underline{\underline{\mathbf{V}}}(\underline{\bar{\boldsymbol{\Psi}}}) \geq \underline{\underline{\mathbf{M}}}_{\underline{\bar{\theta}}}^{-1}$$

with $\underline{\underline{\mathbf{V}}}(\underline{\bar{\boldsymbol{\Psi}}}) = \mathbf{E}\left[\left((\underline{\bar{\boldsymbol{\Psi}}}_i) - \underline{\bar{\theta}}\right)\,\left((\underline{\bar{\boldsymbol{\Psi}}}_i) - \underline{\bar{\theta}}\right)^T\right]$ and were the matrix $\underline{\underline{\mathbf{M}}}_{\underline{\bar{\theta}}}^{-1}$, called the Fisher information matrix, is defined by

$$\underline{\underline{\mathbf{M}}}_{\underline{\bar{\theta}}} := \mathbf{E}\left[\left(\frac{\partial \log p_{(}\underline{\bar{\mathbf{y}}}|\underline{\bar{\theta}})}{\partial\,\underline{\bar{\theta}}}\right)^T\,\left(\frac{\partial \log p_{(}\underline{\bar{\mathbf{y}}}|\underline{\bar{\theta}})}{\partial\,\underline{\bar{\theta}}}\right)\right]$$

Proof. $\underline{\bar{\boldsymbol{\Psi}}}$is an unbiased estimator of $\underline{\bar{\theta}}$, thus:

$$\mathbf{E}\left[\underline{\bar{\boldsymbol{\Psi}}}\right] = \underline{\bar{\theta}}$$

i.e.

$$\begin{aligned}
\int_{\Omega} \underline{\bar{\boldsymbol{\Psi}}}\, p\left(\underline{\bar{\mathbf{y}}}|\underline{\bar{\theta}}\right)\, d\underline{\bar{\mathbf{y}}} &= \underline{\bar{\theta}} \\
\frac{\partial}{\partial \underline{\bar{\theta}}} \int_{\Omega} \underline{\bar{\boldsymbol{\Psi}}}\, p\left(\underline{\bar{\mathbf{y}}}|\underline{\bar{\theta}}\right)\, d\underline{\bar{\mathbf{y}}} &= \underline{\underline{\mathbf{I}}}
\end{aligned}$$

Assuming regularity under the integral

$$\begin{aligned}
\int_{\Omega} \underline{\bar{\boldsymbol{\Psi}}}\, \frac{\partial\, p\left(\underline{\bar{\mathbf{y}}}|\underline{\bar{\theta}}\right)}{\partial \underline{\bar{\theta}}}\, d\underline{\bar{\mathbf{y}}} &= \underline{\underline{\mathbf{I}}} \\
\int_{\Omega} \underline{\bar{\boldsymbol{\Psi}}}\, \frac{\partial \log p\left(\underline{\bar{\mathbf{y}}}|\underline{\bar{\theta}}\right)}{\partial \underline{\bar{\theta}}}\, p\left(\underline{\bar{\mathbf{y}}}|\underline{\bar{\theta}}\right)\, d\underline{\bar{\mathbf{y}}} &= \underline{\underline{\mathbf{I}}} \\
\mathbf{E}\left[\underline{\bar{\boldsymbol{\Psi}}}\, \frac{\partial \log p\left(\underline{\bar{\mathbf{y}}}||\underline{\bar{\theta}}\right)}{\partial \underline{\bar{\theta}}}\right] &= \underline{\underline{\mathbf{I}}}
\end{aligned}$$

Also we have:

$$\begin{aligned}
\mathbf{E}\left[\frac{\partial \log p\left(\underline{\bar{\mathbf{y}}}|\underline{\bar{\theta}}\right)}{\partial \underline{\bar{\theta}}}\right] &= \int_{\Omega} \frac{\partial \log p\left(\underline{\bar{\mathbf{y}}}|\underline{\bar{\theta}}\right)}{\partial \underline{\bar{\theta}}}\, p\left(\underline{\bar{\mathbf{y}}}|\underline{\bar{\theta}}\right)\, dy = \int_{\Omega} \frac{\partial\, p\left(\underline{\bar{\mathbf{y}}}|\underline{\bar{\theta}}\right)}{\partial \underline{\bar{\theta}}}\, d\underline{\bar{\mathbf{y}}} \\
&= \frac{\partial}{\partial \underline{\bar{\theta}}} \int_{\Omega} p\left(\underline{\bar{\mathbf{y}}}|\underline{\bar{\theta}}\right)\, d\underline{\bar{\mathbf{y}}} = \frac{\partial}{\partial \underline{\bar{\theta}}}(1) = \underline{\mathbf{0}}^T \qquad (17.2)
\end{aligned}$$

With the 17.5 and 17.2, the covariance of $\frac{\partial \log p\left(\underline{\bar{\mathbf{y}}}|\underline{\bar{\theta}}\right)}{\partial \underline{\bar{\theta}}}$ and $\underline{\bar{\boldsymbol{\Psi}}}\left(\underline{\bar{\mathbf{y}}}\right)$ is

$$\mathbf{E}\left[\begin{pmatrix} \left(\underline{\bar{\boldsymbol{\Psi}}}\left(\underline{\bar{\mathbf{y}}}\right) - \underline{\bar{\theta}}\right) \\ \frac{\partial \log p\left(\underline{\bar{\mathbf{y}}}|\underline{\bar{\theta}}\right)}{\partial \underline{\bar{\theta}}} \end{pmatrix} \left[\left(\underline{\bar{\boldsymbol{\Psi}}}\left(\underline{\bar{\mathbf{y}}}\right) - \underline{\bar{\theta}}\right)^T \quad \frac{\partial \log p\left(\underline{\bar{\mathbf{y}}}|\underline{\bar{\theta}}\right)}{\partial \underline{\bar{\theta}}}\right]\right] = \begin{bmatrix} \underline{\underline{\mathbf{V}}}(\underline{\bar{\boldsymbol{\Psi}}}) & \underline{\underline{\mathbf{I}}} \\ \underline{\underline{\mathbf{I}}} & \underline{\underline{\mathbf{M}}}_{\underline{\bar{\theta}}} \end{bmatrix} \qquad (17.3)$$

which is clearly non negative since it is a covariance matrix. Thus

$$\left[\underline{\underline{\mathbf{I}}}, -\underline{\underline{\mathbf{M}}}_{\underline{\bar{\theta}}}^{-1}\right] \begin{bmatrix} \underline{\underline{\mathbf{V}}}(\underline{\bar{\boldsymbol{\Psi}}}) & \underline{\underline{\mathbf{I}}} \\ \underline{\underline{\mathbf{I}}} & \underline{\underline{\mathbf{M}}}_{\underline{\bar{\theta}}} \end{bmatrix} \begin{bmatrix} \underline{\underline{\mathbf{I}}} \\ -\underline{\underline{\mathbf{M}}}_{\underline{\bar{\theta}}}^{-1} \end{bmatrix} \geq 0$$

yielding

$$\underline{\underline{\mathbf{V}}}(\underline{\bar{\boldsymbol{\Psi}}}) - \underline{\underline{\mathbf{M}}}_{\underline{\bar{\theta}}}^{-1} \geq 0$$

□

Property - Efficiency : The unbiased estimator is efficient if its covariance is equal the Cramer-Rao bound, i.e. the inverse of the Fisher information matrix.

Theorem 17.5.2. *Subject to regularity conditions, there exists an efficient unbiased estimator for* $\underline{\bar{\theta}}$ *if and only if we can express* $\frac{\partial \log p\left(\underline{\bar{\mathbf{y}}}|\underline{\bar{\theta}}\right)}{\partial \underline{\bar{\theta}}}$ *in the form*

$$\left[\frac{\partial \log p\left(\underline{\bar{\mathbf{y}}}|\underline{\bar{\theta}}\right)}{\partial \underline{\bar{\theta}}}\right]^T = \underline{\underline{\mathbf{A}}}(\underline{\bar{\theta}})\left[\underline{\bar{\mathbf{\Psi}}} - \underline{\bar{\theta}}\right]$$

where $\underline{\underline{\mathbf{A}}}(\underline{\bar{\theta}})$ *is a matrix not depending upon* y

Proof. Sufficiency: Assume the theorem holds then Equation 17.3 becomes:

$$\mathbf{E}\left[\begin{bmatrix} \left(\underline{\bar{\mathbf{\Psi}}}\left(\underline{\bar{\mathbf{y}}}\right) - \underline{\bar{\theta}}\right) \\ \underline{\underline{\mathbf{A}}}(\underline{\bar{\theta}})\left[\underline{\bar{\mathbf{\Psi}}} - \underline{\bar{\theta}}\right] \end{bmatrix} \begin{bmatrix} \left(\underline{\bar{\mathbf{\Psi}}}\left(\underline{\bar{\mathbf{y}}}\right) - \underline{\bar{\theta}}\right) & \underline{\underline{\mathbf{A}}}(\underline{\bar{\theta}})\left[\underline{\bar{\mathbf{\Psi}}} - \underline{\bar{\theta}}\right] \end{bmatrix}\right]$$
$$= \begin{bmatrix} \underline{\underline{\mathbf{V}}}(\underline{\bar{\mathbf{\Psi}}}) & \underline{\underline{\mathbf{V}}}(\underline{\bar{\mathbf{\Psi}}})\,\underline{\underline{\mathbf{A}}}^T(\underline{\bar{\theta}}) \\ \underline{\underline{\mathbf{A}}}(\underline{\bar{\theta}})\,\underline{\underline{\mathbf{V}}}(\underline{\bar{\mathbf{\Psi}}}) & \underline{\underline{\mathbf{M}}}_{\underline{\bar{\theta}}} \end{bmatrix}$$

which from Equation 17.3 is:

$$= \begin{bmatrix} \underline{\underline{\mathbf{V}}}(\underline{\bar{\mathbf{\Psi}}}) & \underline{\underline{\mathbf{I}}} \\ \underline{\underline{\mathbf{I}}} & \underline{\underline{\mathbf{M}}}_{\underline{\bar{\theta}}} \end{bmatrix}$$

which gives

$$\underline{\underline{\mathbf{A}}}(\underline{\bar{\theta}})\,\underline{\underline{\mathbf{V}}}(\underline{\bar{\mathbf{\Psi}}}) = \underline{\underline{\mathbf{I}}}$$

and

$$\underline{\underline{\mathbf{A}}}(\underline{\bar{\theta}})\,\underline{\underline{\mathbf{V}}}(\underline{\bar{\mathbf{\Psi}}})\,\underline{\underline{\mathbf{A}}}^T(\underline{\bar{\theta}}) = \underline{\underline{\mathbf{M}}}_{\underline{\bar{\theta}}}$$

hence

$$\underline{\underline{\mathbf{V}}}(\underline{\bar{\mathbf{\Psi}}}) = \underline{\underline{\mathbf{M}}}_{\underline{\bar{\theta}}}^{-1} \tag{17.4}$$

Necessity: Assume Equation 17.4 then from Equation 17.3

$$\mathbf{E}\left[\begin{bmatrix} \left[\underline{\bar{\mathbf{\Psi}}}\left(\underline{\bar{\mathbf{y}}}\right) - \underline{\bar{\theta}}\right] \\ \frac{\partial \log p\left(\underline{\bar{\mathbf{y}}}|\underline{\bar{\theta}}\right)}{\partial \underline{\bar{\theta}}} \end{bmatrix} \begin{bmatrix} \left[\underline{\bar{\mathbf{\Psi}}}\left(\underline{\bar{\mathbf{y}}}\right) - \underline{\bar{\theta}}\right]^T & \frac{\partial \log p\left(\underline{\bar{\mathbf{y}}}|\underline{\bar{\theta}}\right)}{\partial \underline{\bar{\theta}}} \end{bmatrix}\right] = \begin{bmatrix} \underline{\underline{\mathbf{M}}}_{\underline{\bar{\theta}}}^{-1} & \underline{\underline{\mathbf{I}}} \\ \underline{\underline{\mathbf{I}}} & \underline{\underline{\mathbf{M}}}_{\underline{\bar{\theta}}} \end{bmatrix}$$

Premultiplying with $\left[\underline{\underline{\mathbf{M}}}_{\underline{\bar{\theta}}}, -\underline{\underline{\mathbf{I}}}\right]$ and postmultiplying with $\left[\underline{\underline{\mathbf{M}}}_{\underline{\bar{\theta}}}, -\underline{\underline{\mathbf{I}}}\right]^T$ gives:

$$\mathbf{E}\left[\left[\underline{\underline{\mathbf{M}}}_{\underline{\bar{\theta}}}\left[\underline{\bar{\boldsymbol{\Psi}}}\left(\underline{\bar{\mathbf{y}}}\right) - \underline{\bar{\theta}}\right] - \left(\frac{\partial \log p\left(\underline{\bar{\mathbf{y}}}|\underline{\bar{\theta}}\right)}{\partial\, \underline{\bar{\theta}}}\right)^T\right] \times \right.$$
$$\left. \times \left[\underline{\underline{\mathbf{M}}}_{\underline{\bar{\theta}}}\left[\underline{\bar{\boldsymbol{\Psi}}}\left(\underline{\bar{\mathbf{y}}}\right) - \underline{\bar{\theta}}\right] - \left(\frac{\partial \log p\left(\underline{\bar{\mathbf{y}}}|\underline{\bar{\theta}}\right)}{\partial\, \underline{\bar{\theta}}}\right)^T\right]^T\right] = 0$$

Consequently

$$\underline{\underline{\mathbf{M}}}_{\underline{\bar{\theta}}}\left[\underline{\bar{\boldsymbol{\Psi}}}\left(\underline{\bar{\mathbf{y}}}\right) - \underline{\bar{\theta}}\right] = \left(\frac{\partial \log p\left(\underline{\bar{\mathbf{y}}}|\underline{\bar{\theta}}\right)}{\partial\, \underline{\bar{\theta}}}\right)$$

which proves the theorem. □

Corollary (17.5.2.1)**.** The proof also reveals that if the theorem applies then $\underline{\underline{\mathbf{A}}}(\underline{\bar{\theta}}) = \underline{\underline{\mathbf{M}}}_{\underline{\bar{\theta}}}$, the Fisher information matrix.

17.5.1 Least-Squares Estimator and L-i-P Models

17.5.1.1 Getting the Best Parameters

Let the instance of the multiple-input, single-output, l-i-p model Equation 17.3.2.3 be:

$$\hat{y} := \underline{\mathbf{f}}^T(\underline{\mathbf{u}})\,\underline{\theta} \quad \in \mathbb{R}^1\,,$$

with $\underline{\theta} \in \mathbb{R}^k$. We assume having n instances of input-output experimental data available.

To condense the equations, we stack the n input-output instances up:

$$\begin{aligned}\underline{\hat{\mathbf{y}}} &:= \left[\hat{y}_i\right]_{\forall i}\,,\\ \underline{\underline{\mathbf{F}}} &:= \left[\underline{\mathbf{f}}^T(\underline{\mathbf{u}}_i)\right]_{\forall i} \quad \in \mathbb{R}^{n\times k}\,,\end{aligned}$$

in order to get:

$$\underline{\hat{\mathbf{y}}} := \underline{\underline{\mathbf{F}}}\,\underline{\theta} \quad \in \mathbb{R}^n\,.$$

Let in addition the observation corresponding to the input $\underline{\mathbf{u}}_i$ be y_i, which we also stack up:

$$\underline{\mathbf{y}} := \left[y_i\right]_{\forall i}\,.$$

In order to define the cost function, we first define the error as the difference between the response of the plant and the response of the model to the excitation signal applied to both identically:

$$\underline{\mathbf{e}}(\underline{\theta}) := \underline{\mathbf{y}} - \underline{\hat{\mathbf{y}}}(\underline{\theta})\,,$$

with the model this gets:

$$\underline{\mathbf{e}}(\underline{\theta}) := \underline{\mathbf{y}} - \underline{\underline{\mathbf{F}}}\,\underline{\theta}\,,$$

and the cost function being the $\underline{\underline{\mathbf{Q}}}$-weighted sum of squares:

$$\mathbf{J}(\underline{\theta}) := \underline{\mathbf{e}}(\underline{\theta})^T\,\underline{\underline{\mathbf{Q}}}\,\underline{\mathbf{e}}(\underline{\theta}) \tag{17.5}$$

with $\underline{\underline{\mathbf{Q}}}$ being a positive semi-definite weighting matrix.

The regression problem is then an optimisation problem by defining the optimal parameter being the one that minimises the cost function, thus leads to a minimal sum of square error. Let the optimal solution be marked with a *. Then:

$$\begin{align}
\left.\frac{\partial\,\mathbf{J}(\underline{\theta})}{\partial\,\underline{\theta}}\right|_{\underline{\theta}^\star} &:= 0 \notag\\
0 &:= 2\left(\left(\frac{\partial\,\underline{\mathbf{e}}(\underline{\theta})}{\partial\,\underline{\theta}}\right)^T\,\underline{\underline{\mathbf{Q}}}\,\underline{\mathbf{e}}(\underline{\theta})\right)_{\underline{\theta}^\star} \notag\\
0 &:= -2\left(\left(\frac{\partial\,\underline{\hat{\mathbf{y}}}(\underline{\theta})}{\partial\,\underline{\theta}}\right)^T\,\underline{\underline{\mathbf{Q}}}\,\underline{\mathbf{e}}(\underline{\theta})\right)_{\underline{\theta}^\star} \notag\\
0 &:= \underline{\underline{\mathbf{F}}}^T\,\underline{\underline{\mathbf{Q}}}\,\underline{\mathbf{e}}(\underline{\theta}) \tag{17.6}\\
0 &:= \underline{\underline{\mathbf{F}}}^T\,\underline{\underline{\mathbf{Q}}}\,\left(\underline{\mathbf{y}} - \underline{\underline{\mathbf{F}}}\,\underline{\theta}^\star\right) \notag\\
0 &:= \underline{\underline{\mathbf{F}}}^T\,\underline{\underline{\mathbf{Q}}}\,\underline{\mathbf{y}} - \underline{\underline{\mathbf{F}}}^T\,\underline{\underline{\mathbf{Q}}}\,\underline{\underline{\mathbf{F}}}\,\underline{\theta}^\star\,. \notag
\end{align}$$

The equation Equation 17.6 is also called the normal equation. It is also called the stating that the error is orthogonal to the function of the input, thus no more information can be extracted from the input.

Re-arranging to solve for the parameter vector gives:

$$\underline{\theta}^\star \;:=\; \left(\underline{\underline{\mathbf{F}}}^T\,\underline{\underline{\mathbf{Q}}}\,\underline{\underline{\mathbf{F}}}\right)^{-1}\,\underline{\underline{\mathbf{F}}}^T\,\underline{\underline{\mathbf{Q}}}\,\underline{\mathbf{y}}\,. \tag{17.7}$$

17.5.1.2 Effect of Measurement Noise

Measurement noise is one of the most common problems with measured data. Making a couple of assumptions, it is straightforward to estimate the effect of the measurement noise on the estimated parameters. Again assuming a single-output system, let the additive measurement error be v, then the key assumptions are:

1. inputs are uncorrelated,
2. $E(v) := 0$ :: mean error is zero,
3. $\text{var}(v) := \sigma^2$ being the variance σ^2 being the standard deviation of the error distribution.

Simplifying the writing of the unity-weighted estimator:

$$\begin{aligned}
\underline{\theta}^\star &:= \left(\underline{\underline{\mathbf{F}}}^T \, \underline{\underline{\mathbf{F}}}\right)^{-1} \underline{\underline{\mathbf{F}}}^T \, \underline{\mathbf{y}} \, , \\
\underline{\theta}^\star &:= \underline{\underline{\mathbf{S}}} \, \underline{\mathbf{y}} \, .
\end{aligned}$$

The estimated variance of the parameter vector is

$$\begin{aligned}
\text{var}\left(\underline{\theta}^\star\right) &:= \text{var}\left(\underline{\underline{\mathbf{S}}} \, \underline{\mathbf{y}}\right) \\
&:= \underline{\underline{\mathbf{S}}} \, \text{var}\left(\underline{\mathbf{y}}\right) \, \underline{\underline{\mathbf{S}}}^T \\
&:= \underline{\underline{\mathbf{S}}} \, \sigma^2 \, \underline{\underline{\mathbf{S}}}^T \\
&:= \underline{\underline{\mathbf{S}}} \, \underline{\underline{\mathbf{S}}}^T \, \sigma^2 \\
&:= \left(\underline{\underline{\mathbf{F}}}^T \, \underline{\underline{\mathbf{F}}}\right)^{-1} \underline{\underline{\mathbf{F}}}^T \, \underline{\underline{\mathbf{F}}} \, \left(\underline{\underline{\mathbf{F}}}^T \, \underline{\underline{\mathbf{F}}}\right)^{-1} \sigma^2 \\
&:= \left(\underline{\underline{\mathbf{F}}}^T \, \underline{\underline{\mathbf{F}}}\right)^{-1} \sigma^2 \, . \qquad (17.8)
\end{aligned}$$

The result is a symmetric matrix called the *variance-covariance matrix*, the diagonal being the variances and the co-diagonal the respective covariances. The covariance implies that a change in the expectation (average) of one parameter will also change the correlated parameter in the direction and magnitude indicated by the respective covariance. As a normed measure one uses the correlation.

Correlation The correlation matrix is the variance-covariance matrix normed by the variances:

$$\begin{aligned}
\underline{\underline{\mathbf{R}}} &:= \left[\frac{\text{cov}(\underline{\theta}^\star{}_i \, \underline{\theta}^\star{}_j)}{\left(\text{var}\underline{\theta}^\star{}_i \; \text{var}\underline{\theta}^\star{}_j\right)^{1/2}} \right]_{\forall i, \forall j} , \\
&:= \left[\frac{\text{cov}(\underline{\theta}^\star{}_i \, \underline{\theta}^\star{}_j)}{\sigma^2{}_i \, \sigma^2{}_j} \right]_{\forall i, \forall j} , \\
&:= [r_{i,j}]_{\forall i, \forall j} \, .
\end{aligned}$$

The correlation varies between -1 and 1 being completely negative or completely positively correlated. In most applications correlation is an undesired property and large correlation can be an indication of pure experimental design.

17.5.1.3 Expected Accuracy

One can use the estimated parameters to predict the behaviour of the plant for a particular instance. Let the instance be

$$\hat{y}_i := \underline{\mathbf{f}}^T(\underline{\mathbf{u}}_i)\,\underline{\theta}^\star + v\,.$$

where v is a measurement error that is normally distributed and has a zero mean. Under these conditions, the variance is:

$$\begin{aligned}
\mathrm{var}\left(\hat{y}_i\right) &:= \mathrm{var}\left(\underline{\mathbf{f}}^T(\underline{\mathbf{u}}_i)\,\underline{\theta}^\star\right) + \mathrm{var}\left(v\right) \\
&:= \underline{\mathbf{f}}^T(\underline{\mathbf{u}}_i)\,\mathrm{var}\left(\underline{\theta}^\star\,\underline{\mathbf{f}}(\underline{\mathbf{u}}_i)\right) + \mathrm{var}\left(v\right) \\
&:= \left(\underline{\mathbf{f}}^T(\underline{\mathbf{u}}_i)\,\left(\underline{\underline{\mathbf{F}}}^T\,\underline{\underline{\mathbf{F}}}\right)^{-1}\,\underline{\mathbf{f}}(\underline{\mathbf{u}}_i) + 1\right)\,\sigma^2\,.
\end{aligned}$$

If we repeat the experiment m times, we can improve the estimate of the variance:

$$\mathrm{var}\left(\hat{y}_{i,m}\right) \quad := \quad \left(\underline{\mathbf{f}}^T(\underline{\mathbf{u}}_i)\,\left(\underline{\underline{\mathbf{F}}}^T\,\underline{\underline{\mathbf{F}}}\right)^{-1}\,\underline{\mathbf{f}}(\underline{\mathbf{u}}_i) + \frac{1}{m}\right)\,\sigma^2\,.$$

Given the variance, the confidence limits (upper :: u, lower :: l) are:

$$\left[\hat{y}_{i,l}, \hat{y}_{i,u}\right] \quad := \quad \underline{\mathbf{f}}^T(\underline{\mathbf{u}}_i)\,\underline{\theta}^\star \pm 2\,\sigma^2\sqrt{\left(\underline{\mathbf{f}}^T(\underline{\mathbf{u}}_i)\,\left(\underline{\underline{\mathbf{F}}}^T\,\underline{\underline{\mathbf{F}}}\right)^{-1}\,\underline{\mathbf{f}}(\underline{\mathbf{u}}_i) + 1\right)}\,.$$

If one estimates the variance, the ”2” is replaced by the respective value from the student t-distribution with the appropriate degrees of freedom and the chosen confidence limit.

17.5.1.4 Confidence Limits for Parameters

Having found the *best* parameters poses the question on how confident one can be in them. So how does the cost function change with the parameters? Let the cost function be the identity-weighted version as given in equation

17.5, then its change with the parameters is:

$$\begin{aligned}
\mathbf{J}(\underline{\theta}) &:= \underline{\mathbf{e}}^T(\underline{\theta})\,\underline{\mathbf{e}}(\underline{\theta})\,,\\
&:= \left(\underline{\mathbf{y}} - \underline{\underline{\mathbf{F}}}\,\underline{\theta}\right)^T \left(\underline{\mathbf{y}} - \underline{\underline{\mathbf{F}}}\,\underline{\theta}\right)\,,\\
&:= \left(\left(\underline{\mathbf{y}} - \underline{\underline{\mathbf{F}}}\,\underline{\theta}^\star\right) - \underline{\underline{\mathbf{F}}}\,\left(\underline{\theta} - \underline{\theta}^\star\right)\right)^T \left(\left(\underline{\mathbf{y}} - \underline{\underline{\mathbf{F}}}\,\underline{\theta}^\star\right) - \underline{\underline{\mathbf{F}}}\,\left(\underline{\theta} - \underline{\theta}^\star\right)\right)\\
&:= \left(\underline{\mathbf{y}} - \underline{\underline{\mathbf{F}}}\,\underline{\theta}^\star\right)^T \left(\underline{\mathbf{y}} - \underline{\underline{\mathbf{F}}}\,\underline{\theta}^\star\right)\\
&\quad - \left(\underline{\mathbf{y}} - \underline{\underline{\mathbf{F}}}\,\underline{\theta}^\star\right)^T \underline{\underline{\mathbf{F}}}\,\left(\underline{\theta} - \underline{\theta}^\star\right) - \left(\underline{\theta} - \underline{\theta}^\star\right)^T \underline{\underline{\mathbf{F}}}^T\,\left(\underline{\mathbf{y}} - \underline{\underline{\mathbf{F}}}\,\underline{\theta}^\star\right)\\
&\quad + \left(\underline{\underline{\mathbf{F}}}\,\left(\underline{\theta} - \underline{\theta}^\star\right)\right)^T \underline{\underline{\mathbf{F}}}\,\left(\underline{\theta} - \underline{\theta}^\star\right)\,,\\
&:= \mathbf{J}(\underline{\theta}^\star) + \left(\underline{\theta} - \underline{\theta}^\star\right)^T \underline{\underline{\mathbf{F}}}^T\,\underline{\underline{\mathbf{F}}}\,\left(\underline{\theta} - \underline{\theta}^\star\right)\,,
\end{aligned}$$

where we used the fact of equation Equation 17.6 twice for the middle terms. Thus

$$\mathbf{J}(\underline{\theta}) - \mathbf{J}(\underline{\theta}^\star) \quad := \quad \left(\underline{\theta} - \underline{\theta}^\star\right)^T \underline{\underline{\mathbf{F}}}^T\,\underline{\underline{\mathbf{F}}}\,\left(\underline{\theta} - \underline{\theta}^\star\right)\,.$$

This is an ellipsoid in the parameter space. The length of the axis is related to the eigenvalues of the matrix $\underline{\underline{\mathbf{C}}} := \underline{\underline{\mathbf{F}}}^T\,\underline{\underline{\mathbf{F}}}$ and the eigenvectors to the orientation: The equatorial radii are the inverse of the square roots of the eigenvalues whilst the eigenvectors, which, due to the spectral theorem, are orthogonal, determine the direction. [3]

The α confidence limits of the parameters are given by the corresponding value of the F-distribution:

$$\left(\underline{\theta} - \underline{\theta}^\star\right)^T \underline{\underline{\mathbf{F}}}^T\,\underline{\underline{\mathbf{F}}}\,\left(\underline{\theta} - \underline{\theta}^\star\right) \quad \leq \quad k\,s^2\,F^{\alpha}_{k,n-k}(\alpha)\,.$$

The confidence ellipsoid is thus defined by the matrix C normed by the right-hand-side value.

The variance can be estimated from the cost function:

$$s^2 \quad := \quad \frac{\underline{\mathbf{e}}^T(\underline{\theta}^\star)\,\underline{\mathbf{e}}(\underline{\theta}^\star)}{n-k}\,,$$

where n is the number of observations and k the number of estimated parameters. In the case of a BLUE estimator, and assuming the statistics to be normal, then one can prove that the BLUE estimator approaches the Cramer-Rao minimal variance bound (Goodwin and Payne, 1977).

[3] Since $\underline{\underline{\mathbf{C}}}$ is symmetric, $\underline{\underline{\mathbf{C}}} = \underline{\underline{\mathbf{V}}}\,\underline{\underline{\mathbf{\Lambda}}}\,\underline{\underline{\mathbf{V}}}^{-1} = \underline{\underline{\mathbf{C}}}^T = \left(\underline{\underline{\mathbf{V}}}\,\underline{\underline{\mathbf{\Lambda}}}\,\underline{\underline{\mathbf{V}}}^{-1}\right)^T$ Thus $\underline{\underline{\mathbf{V}}}^T = \underline{\underline{\mathbf{V}}}^{-1}$ the quadratic form $\underline{\mathbf{x}}^T\,\underline{\underline{\mathbf{F}}}^T\,\underline{\underline{\mathbf{F}}}\,\underline{\mathbf{x}}$ can be rewritten as $\underline{\mathbf{x}}^T\,\underline{\underline{\mathbf{V}}}\,\underline{\underline{\mathbf{\Lambda}}}\,\underline{\underline{\mathbf{V}}}^T\,\underline{\mathbf{x}} := \underline{\mathbf{z}}^T\,\underline{\underline{\mathbf{\Lambda}}}\,\underline{\mathbf{z}}$ with $\underline{\mathbf{z}} := \underline{\underline{\mathbf{V}}}^T\,\underline{\mathbf{x}}$.

17.5.1.5 How Good is the Identified Model: Variance Analysis

Because of the experimental errors one will not get the same response from the plant when repeating an experiment using the same input. If the responses are within the limits of the expected output error, one has no reason to be suspicious about the model appropriately describing the process. If one tries to fit the same data a more complex model, one will find no improvement. Naturally if one performs more experiments, it may show that the model is indeed not the best one can find. Latter aspect is used to design experiment focusing on the weak parts of the model.

The means to check on the model is to analyse the variance for the various contributions. Again we start with the sum of squares of the error, the cost function Equation 17.5 which we expand:

$$\begin{aligned}
\underline{\mathbf{e}}^T\,\underline{\mathbf{e}} &:= \left(\underline{\mathbf{y}} - \underline{\underline{\mathbf{F}}}\,\underline{\theta}^\star\right)^T\,\left(\underline{\mathbf{y}} - \underline{\underline{\mathbf{F}}}\,\underline{\theta}^\star\right) \\
&:= \underline{\mathbf{y}}^T\,\underline{\mathbf{y}} - \underline{\theta}^{\star T}\,\underline{\underline{\mathbf{F}}}^T\,\underline{\mathbf{y}} - \underline{\mathbf{y}}^T\,\underline{\underline{\mathbf{F}}}\,\underline{\theta}^\star + \underline{\theta}^{\star T}\,\underline{\underline{\mathbf{F}}}^T\,\underline{\underline{\mathbf{F}}}\,\underline{\theta}^\star \\
&:= \underline{\mathbf{y}}^T\,\underline{\mathbf{y}} - \underline{\theta}^{\star T}\,\underline{\underline{\mathbf{F}}}^T\,\underline{\underline{\mathbf{F}}}\,\underline{\theta}^\star - \underline{\theta}^{\star T}\,\underline{\underline{\mathbf{F}}}^T\,\underline{\underline{\mathbf{F}}}\,\underline{\theta}^\star + \underline{\theta}^{\star T}\,\underline{\underline{\mathbf{F}}}^T\,\underline{\underline{\mathbf{F}}}\,\underline{\theta}^\star \\
&:= \underline{\mathbf{y}}^T\,\underline{\mathbf{y}} - \underline{\theta}^{\star T}\,\underline{\underline{\mathbf{F}}}^T\,\underline{\underline{\mathbf{F}}}\,\underline{\theta}^\star
\end{aligned}$$

Isolate the total sum of squares over the outputs:

$$\underline{\mathbf{y}}^T\,\underline{\mathbf{y}} \quad := \quad \underline{\theta}^{\star T}\,\underline{\underline{\mathbf{F}}}^T\,\underline{\underline{\mathbf{F}}}\,\underline{\theta}^\star + \underline{\mathbf{e}}^T\,\underline{\mathbf{e}}\,.$$

The total sum of squares is thus the sum of the regression sum of squares plus the rest sum of squares. Each of these terms is connected to a degree of freedom being used to compute the respective term. The sum of squares of the outputs uses n :: number of observations. The regression sum of squares is computed from k :: number of parameters normal equations. Thus the difference $n - k$ is are the left degrees of freedom for the rest sum of squares.

It is customary to show this in a table:

	SSQ	DOF
total SSQ	$\underline{\mathbf{y}}^T\,\underline{\mathbf{y}}$	n
regression SSQ	$\underline{\theta}^{\star T}\,\underline{\underline{\mathbf{F}}}^T\,\underline{\underline{\mathbf{F}}}\,\underline{\theta}^\star$	k
rest SSQ	$\underline{\mathbf{e}}^T\,\underline{\mathbf{e}}$	$n - k$

One can show that if $\frac{\underline{\mathbf{e}}^T\,\underline{\mathbf{e}}}{n-k}$ estimates the variance of the experimental error, then the model is describing the process appropriately. If we take the rest

SSQ divided by the respective degrees of freedom as an estimate for the variance, thus

$$s_e^2 := \frac{\underline{\mathbf{e}}^T \, \underline{\mathbf{e}}}{n-k} \, ,$$

and knowing the actual variance of the experimental error to be σ^2 then the ratio of the estimated variance and the $n-k$ scaled variance is χ^2 distributed thus algebraically:

$$s^2 \, \frac{n-k}{\sigma^2} \sim {\chi^2}_{n-k} \, .$$

One has good reasons do declare the model as not fitting well and thus reconsider its structure if :

$$\frac{s^2}{\sigma^2} > \frac{\chi^2(\alpha)}{n-k} \, ,$$

α being the confidence limit.

Not knowing the variance The variance of the experimental error is usually not known and must be estimated from the data. Assuming that we make n_i experiments for the input $\underline{\mathbf{u}}_i$ and obtain a corresponding set of responses $\underline{\mathbf{y}}_i$ and repeat the experiments by varying $i := 1, \ldots, q$, then the estimate for the variance is computed by :

$$\begin{aligned} s_e^2 &:= \frac{\sum_{i=1}^q \left(\underline{\mathbf{y}} - \mathbf{E}\left[\underline{\mathbf{y}}\right]\right)^2}{\sum_{i=1}^q \left(n_i - 1\right)} \\ &:= \frac{\sum_{i=1}^q \left(\underline{\mathbf{y}} - \mathbf{E}\left[\underline{\mathbf{y}}\right]\right)^2}{\sum_{i=1}^q n_i \ - q} \, . \end{aligned}$$

The -1 thus the reduction of the degree of freedom by one is due to the mean being $\mathbf{E}\left[\underline{\mathbf{y}}\right]$, which is calculated from the same data. So for the s_e^2 total degree of freedom is:

$$n_e := \sum_{i=1}^q n_i \ - q$$

If the model fits well, then the experimental error is also estimated by the rest sum of squares. The two variance estimates can be compared with each other as one can show that their ratio is F-distributed with the respective two degrees of freedom.

If the ratio gets too large the model does not fit well and one may consider the model to be a bad fit:

$$\frac{1}{n-k}\,\frac{\underline{\mathbf{e}}^T\,\underline{\mathbf{e}}}{s_e^2} > F^{\alpha}_{n-k,n_e}\,.$$

The above test assumes that the variance is estimated with one set and the parameters with another. It is, though, meaningful to use all experiments for the regression and split the variance accordingly:

source deviation	SSQ	DOF	average SSQ
regression	$\underline{\theta}^{\star T}\,\underline{\underline{\mathbf{F}}}^T\,\underline{\underline{\mathbf{F}}}\,\underline{\theta}^{\star}$	k	$\frac{\underline{\theta}^{\star T}\,\underline{\underline{\mathbf{F}}}^T\,\underline{\underline{\mathbf{F}}}\,\underline{\theta}^{\star}}{k}$
lack of fit	$\underline{\mathbf{e}}^T\,\underline{\mathbf{e}} - \sum_{i=1}^{q}\left(\underline{\mathbf{y}} - \mathbf{E}\left[\underline{\mathbf{y}}\right]\right)^2$	$n-k-n_e$	$\frac{\underline{\mathbf{e}}^T\,\underline{\mathbf{e}} - \sum_{i=1}^{q}\left(\underline{\mathbf{y}} - \mathbf{E}\left[\underline{\mathbf{y}}\right]\right)^2}{n-k-n_e}$
pure error	$\sum_{i=1}^{q}\left(\underline{\mathbf{y}} - \mathbf{E}\left[\underline{\mathbf{y}}\right]\right)^2$	n_e	$\frac{\sum_{i=1}^{q}\left(\underline{\mathbf{y}} - \mathbf{E}\left[\underline{\mathbf{y}}\right]\right)^2}{n_e}$
total SSQ	$\underline{\mathbf{y}}^T\,\underline{\mathbf{y}}$	n	

Defining the variances:

$$s_{ef}^2 := \frac{\underline{\mathbf{e}}^T\,\underline{\mathbf{e}} - \sum_{i=1}^{q}\left(\underline{\mathbf{y}} - \mathbf{E}\left[\underline{\mathbf{y}}\right]\right)^2}{n-k-n_e}\,,$$

$$s_e^2 := \frac{\sum_{i=1}^{q}\left(\underline{\mathbf{y}} - \mathbf{E}\left[\underline{\mathbf{y}}\right]\right)^2}{n_e}\,,$$

the lack of fit test is then:

$$\frac{s_{ef}^2}{s_e^2} \le F^{\alpha}_{n-k-n_e,n_e}$$

to accept the model.

How to proceed Identification is an iterative process. One fits a model, checks if it fits well and if not, modifies the model until one is satisfied. The lack-of-fit measure is thus used as the decision criterion if or if not a modified model should be adopted: The lack-of-fit for the new model compared to the old model must be statistically significant better. An appropriate F-test provides the information.

17.5.1.6 Bias

Under certain circumstances the estimator will not deliver the desired result, but an estimate that is contaminated with a bias. With $\mathbf{E}\left[\underline{\bar{\mathbf{\Psi}}}\right]$ being the estimated parameters and $\underline{\bar{\theta}}$ the true parameter values, a biased estimator is defined as:

$$\mathbf{E}\left[\underline{\bar{\mathbf{\Psi}}}\right] := \underline{\bar{\theta}} + \underline{\mathbf{b}}\,.$$

If $\underline{\mathbf{b}}$ is not equal zero, the estimator is called biased, otherwise the estimator is unbiased.

Bias due to omitted variables This is unfortunately a very common case, as one often does not know what variables do affect the output of the plant. The linear model that one identifies thus may not include all those variables and the effect is a bias in the estimate. To show the effect, let the plant be represented by a mathematical object, namely:

$$\bar{y} := \underline{\mathbf{f}}^T(u)\,\underline{\theta} + \underline{\mathbf{g}}^T(u)\,\underline{\theta}\,.$$

The model to be fitted shall be identical to the first term of the plant, thus the second term is the omitted one, the output of which we abbreviate as $\underline{\bar{\mathbf{z}}}$.

$$\hat{y} := \underline{\mathbf{f}}^T(u)\,\underline{\theta}\,.$$

Consequently one can write the plant output as:

$$\bar{y} := \hat{y} + \underline{\bar{\mathbf{z}}}\,,$$

Using the model as a basis and the analogue stacking of the individual experiment instances, the estimator Equation 17.7 is

$$\begin{aligned}
\underline{\theta}^{\star} &:= \left(\underline{\underline{\mathbf{F}}}^T\,\underline{\underline{\mathbf{F}}}\right)^{-1}\,\underline{\underline{\mathbf{F}}}^T\,\underline{\bar{\mathbf{y}}} \\
&:= \left(\underline{\underline{\mathbf{F}}}^T\,\underline{\underline{\mathbf{F}}}\right)^{-1}\,\underline{\underline{\mathbf{F}}}^T\,\left(\underline{\hat{\mathbf{y}}} + \underline{\bar{\mathbf{z}}}\right) \\
&:= \underline{\bar{\theta}} + \left(\underline{\underline{\mathbf{F}}}^T\,\underline{\underline{\mathbf{F}}}\right)^{-1}\,\underline{\underline{\mathbf{F}}}^T\,\underline{\bar{\mathbf{z}}}\,.
\end{aligned}$$

Clearly the second term is now the bias of the estimate.

Bias due to correlation in output noise Asymptotically the bias is given by Astroem and Eykhoff (1971)

$$\mathbf{E}\left[\underline{\theta}^{\star} - \underline{\theta}\right] := \left(\mathbf{E}\left[\underline{\underline{\mathbf{F}}}^T\,\underline{\underline{\mathbf{F}}}\right]\right)^{-1}\mathbf{E}\left[\underline{\underline{\mathbf{F}}}^T\,\underline{\mathbf{e}}\right]\,.$$

where $\underline{\mathbf{e}}$ is the output error.

Bias due to input noise Bias is also introduced into the parameter estimation if the input has a stochastic component. The mathematical treatment of this case is rather involved and closely linked to the derivation of the Kalman filter.

17.5.1.7 Instrumental Variables

The least squares estimator can be obtained from the model

$$\underline{\hat{\mathbf{y}}} := \underline{\underline{\mathbf{F}}}(\underline{\mathbf{u}})\,\underline{\theta} + \underline{\mathbf{e}}$$

by multiplying both sides of the error-free model with $\underline{\underline{\mathbf{F}}}^T$:

$$\underline{\underline{\mathbf{F}}}^T\,\underline{\hat{\mathbf{y}}} := \underline{\underline{\mathbf{F}}}^T\,\underline{\underline{\mathbf{F}}}\,\underline{\theta}^\star\,.$$

The estimate will be unbiased if the term $\underline{\underline{\mathbf{F}}}^T\,\underline{\mathbf{e}}$ has zero mean, which is not the case when the error is correlated. The instrumental variable method replaces the $\underline{\underline{\mathbf{F}}}^T$ matrix by an instrumental variable matrix $\underline{\underline{\mathbf{W}}}^T$ in above's manipulation. It is a matrix which is a function of the data with the properties

$$\begin{aligned}\mathbf{E}\left[\underline{\underline{\mathbf{W}}}^T\,\underline{\underline{\mathbf{F}}}\right] &:: \text{not singular}\\ \mathbf{E}\left[\underline{\underline{\mathbf{W}}}^T\,\underline{\mathbf{e}}\right] &:= 0\,.\end{aligned}$$

The corresponding estimator is

$$\begin{aligned}\underline{\underline{\mathbf{W}}}^T\,\underline{\hat{\mathbf{y}}} &:= \underline{\underline{\mathbf{W}}}^T\,\underline{\underline{\mathbf{F}}}\,\underline{\theta}^\star\\ \underline{\theta}^\star &:= \left(\underline{\underline{\mathbf{W}}}^T\,\underline{\underline{\mathbf{F}}}\right)^{-1}\,\underline{\underline{\mathbf{W}}}^T\,\underline{\hat{\mathbf{y}}}\,,\end{aligned}$$

which is unbiased.

Choice of instruments For dynamic systems (17.8), most commonly a filtered input is used as instrument where the filter's discrete transfer function may be

$$g(q) := \frac{D(q)}{C(q)}\,.$$

An attractive alternative is the modulating function filters introduced by Maletinsky (1978); Preisig (1984); Preisig and Rippin (1993a,b).

17.5.2 Maximum Likelihood Estimator

The maximum likelihood estimator selects the most likely parameter Box and Tiao (1973); Ljung (1987). The approach is based on Bayes' theorem Equation G.1. Given the vector of observations $\underline{\mathbf{y}} := \{y_i\}$ the joint density function is $p(\underline{\mathbf{y}}\,,\,\underline{\theta})$ that depends on a vector of parameters $\underline{\theta}$. This density may be interpreted in two ways:

$$\begin{aligned} p(\underline{\mathbf{y}}\,\underline{\theta}) &:= p(\underline{\mathbf{y}}|\underline{\theta})\,p(\underline{\theta})\,, \\ &:= p(\underline{\theta}|\underline{\mathbf{y}})\,p(\underline{\mathbf{y}})\,. \end{aligned}$$

The conditional distribution of $\underline{\theta}$ is:

$$p(\underline{\theta}|\underline{\mathbf{y}}) := \frac{p(\underline{\mathbf{y}}|\underline{\theta})\,p(\underline{\theta})}{p(\underline{\mathbf{y}})}\,.$$

The denominator can be rewritten as:

$$p(\underline{\mathbf{y}}) := \begin{cases} \int\, p(\underline{\mathbf{y}}|\underline{\theta})\,p(\underline{\theta}), & \underline{\theta}\ \text{continuous} \\ \sum p(\underline{\mathbf{y}}|\underline{\theta})\,p(\underline{\theta}), & \underline{\theta}\ \text{discrete} \end{cases}$$

$p(\underline{\theta})$ is denoted as *prior probability*, $p(\underline{\theta}|\underline{\mathbf{y}})$ as *posterior probability* and $p(\underline{\mathbf{y}}|\underline{\theta})$ as *likelihood*.

In contrast to the least squares method, the maximum likelihood method assumes the parameter to be distributed and not the measurement depending on the parameter. This assumption is exactly inverted for the least squares method Johnston and DiNardo (1997); Koch (2007); Box and Tiao (1973).

17.6 Non-linear Regression

17.6.1 Finding the best parameters

In the previous section we assumed the model is linear in the parameters, which in many cases is a stringent assumption requiring the linearisation of the process model with respect to the parameters. Often this is not feasible and one has to resort to solve the non-linear problem. The model is then of the form:

$$\hat{y} := \mathrm{f}(\underline{\mathbf{u}}, \underline{\theta})$$

The least-square estimator then minimises the objective function:

$$\mathbf{J}(e) := \underline{\mathbf{e}}^T\,\underline{\mathbf{e}}$$

with the output error being:

$$e_i := \hat{y} - y_i := \mathrm{f}(\underline{\mathbf{u}}, \underline{\theta}) - y_i$$

If the non-linearities are mild, then one can consider to linearize the model and revert to linear regression. With more severe non-linearities, one may still try the approach, but may well have to submit to solve the optimisation as a non-linear problem. The difference is that in the linear case, the optimisation has a close solution, whilst in the non-linear case, one may be faced with multiple minima and a non-convex surface of the objective function in the parameter space.

The solution is usually to use a surface search method starting with a initial guess, one "walks" the response surface, here being the objective function in the parameter space. Finding the global optimum can, as usual not be guaranteed in cases where one has multiple minima. The methods are being discussed in more details in the corresponding section <ref>.

17.6.2 Confidence limits

As in all regression problems, one is interested in the confidence regions. If the problem was linearised, then the linear confidence regions can be used as was discussed before. They are ellipsoids. For the non-linear case, one has to compute the confidence limits based on the observation that the ratio:

$$\frac{\mathbf{J}(\Theta) - \mathbf{J}(\Theta^\star)}{\mathbf{J}(\Theta^\star)} \leq \frac{k}{n-k} F^{\alpha}_{k,n-k}$$

follows an F-distribution.

Re-arranging the above equation gives the objective function for a point in the parameter space

17.7 Robust Data Analysis

17.7.1 The Issue

If we look at a data set that was generated as the result of some process we often find data points that seem to be contaminated with an extraordinary amount of error. These data are called outliers, implying that indeed something extraordinary has happened when they were taken. As such this may be just a simple error, such as in hand-taken data a comma error, a writing error or when taken automatically the mistake can be in any of the involved components, the sensor, the measurement instrument, the data converesion, the transmission line, the transmission elecronic, the transmission software, the program picking it up, a writing error in the memory, the disc etc etc. Sources are manifuld even though we have usually a high confidence in the involved systems. Certainly the error may also come from the process itself. And again dependent on the process a multitude of sources may be suggested in each case. The outlier thus may very well contain interesting and relevant information, and people indeed learnt to pay attention to this issue and analyse outliers. But for the model identification, these outliers represent an unnecessary or avoidable increase in the uncertainty of the identified model. Thus one is interested in taking them out of the data set for the purpose of identification, and deal with them separately for the mentioned reasons. For small data set, the elimination may be done manually, but for larger sets or more complex underlying structures, the exclusion of data is not a trivial matter which triggered new research in the 1960/70ties. The key references are Hampel et al. (2005); Huber (2009); Maronna et al. (2006).

The problem of outliers can be readily demonstrated when computing the average and the standard deviation which are common estimates of the distribution function underlying the data set. For the discussion we assume that we know a ”true” data set including the underlying distribution. So let us assume we have an ordered list of random variables $Y := [y_i]$ of length n for which we know the distribution to be symmetrical. The average of the data then is an estimate for the centre[4] of the distribution $\hat{\mu}$:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} y_i$$

And for the variance the standard estimator is given by:

$$s := \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (y_i - \hat{\mu})^2}$$

Adding an outlier, say $x = \infty$ it is apparent that the estimate is now also ∞. The estimator of the mean value of the distribution is thus extremely sensitive to outliers. The same applies to any of the central moments thus also to the estimator of the variance and skewness, etc. Hampel defined a the breakdown point of an estimator theoretically (Hampel et al. (2005)). In loose terms, the breakdown point is the maximal fraction of outlying objects in the data, that the estimator can handle yielding acceptable estimates.

[4] Robust statistics uses the term location estimator

For instance, the breakdown point of the mean estimator equals 0% being the smallest possible.

The median is commonly used as a robust estimator for the mean location estimator. For an odd number of samples, the median is the one in the middle of the ordered set of data points, whilst if the cardinality is even, one takes the average of the two middle points. The median has the highest possible breakdown point, namely 50 %. In a multi-dimensional one can use an Eucledian-based measure:

$$\min_{\hat{\mu}} \sum_{i=1}^{n} ||y_i - \hat{\mu}||_1$$

with $||\cdot||_1$ being the L1-norm. It can be shown that this estimator is less sensitive to outliers than the L2-based estimator (min sum of squares estimator). Again this estimator has a breakdown point of 50 %.

Different measures of the quantiles are known to be robust estimators. Quantiles are the points when taking regular intervals from the cumulative distribution function or a random variable. Applying it to ordered lists of data taken from a random process, the size-ordered data are divided into q equal-sized sub-lists with the quantiles being the data points marking the boundary between the sub-lists. Some of the quantiles have special names:

number	name	symbol
2	median	M
3	teriles or terciles	T
4	quartiles	Q
5	quintiles	QU
6	sixtiles	S
10	deciles	D
100	percentiles	P
1000	permilles	Pr

The median is, as mentioned a measure for the location of the mean. The standard deviation may be estimated using the quartils or sixtiles differences as follows:

$$s \approx \frac{3}{4}(Q_3 - Q_1)$$
$$s \approx \frac{1}{2}(S_5 - S_1)$$
$$s \approx c\, m_i(|Y_i - m_j(Y_j)|)$$

The third estimator is called the MAD estimator with MAD standing for median absolute deviation. It is the median of the absolute deviations form the data's median. The constant c depends on the underlying distribution function. For normally distributed data $c = 1.4826$.

For the median estimate of the average one can in addition also provide an estimate for the 95 % confidence limits, which is $\sqrt{m}$.

17.7.2 Visualisation and Data Transformations

It is quite common practice to first visualise the data primarily with the objectives to learn about the properties of the data such as underlying distribution function, outliers and trends, ergodicity etc. To learn about the distribution function one has developed several plots. Commonly used are PP plots and QQ plots. The PP plot is plots the estimated cumulative distribution of the data versus a proposed cumulative probability test function. The QQ plot does essentially the same, but uses quantile differences on both axis and is considered to be more robust than the PP plot.

Box plots are used to explore if different data set have the same underlying distribution. For the construction of the plot each data set is ordered and the quartiles are determined. Next a graph is connected in which either horizontally or vertically the experiment number is shown, whilst on the other axis the data scale is shown. For the quartile difference a box or thick beam is shown indicating also the centre (median). From the boxes outwards two "whiskers" are added, which may be of the length: (i) min and max of the data (ii) 1.5 the interquartile range (iii) one standard deviation (iv) 9th and 91st percentile (v) 2nd and 98th percentile. Any data not included in the range of the two extremes of the whiskers are considered outliers.

Often data are treated by empirical transformations, with logarithmic transformations being a popular one. The background of this transformation is the the observation that data are often the result of a first-order process, which has an exponential as a solution. The transformation brings it often back to at least a symmetrical distribution.

Centring is a common operation done mainly for the reason of improving the numerical nature of the estimator. It is routinely done for minimal sum-of-square estimators. So the data are, after having removed the outliers transformed by subtracting the average or a robust estimate of the average:

$$z_i = y_i - \hat{\mu}$$

Auto-scaling is another popular transformation technique, which is applied after centring. The data are scaled by the (robust) estimate of the variance:

$$v_i = \frac{z_i}{s}$$

17.7.3 Robust regression

As for the estimate of basic distribution parameters of stochastic variables, also parameter identification is strongly affected by outliers. The detection of outliers may not be a trivial matter. Graphical representation of the data is certainly one of the first actions one can take, though if the input/output interaction does not result in a very regular pattern, then identifying outliers is not possible based on a visual pattern and one must resort to take the model into the outlier detection process, which lead to the introduction of what is labelled as ”robust” regression.

The idea is to identify those observations that give a very large deviation from the expected observation, which gives a large output error. Robust regression modifies the optimisation criterion by giving large errors less weight than the smaller ones. This implies that one adds additional empirical information to the regression problem definition. Technically the method can be implemented by directly defining a objective function that has the desired property or use a ”standard” objective function combined with a chosen weighting function. Both methods are used in implementations.

To facilitate the definition we introduce the objective function as a function of the output error:

$$\mathbf{J}(e) := \sum_{i:=1}^{n} \rho(e_i)$$

17.8 Selected Dynamic Systems

In this section two commonly used dynamic models are being introduced, which then are extended to a generic transfer-function model, which captures a large family of models. In terms of overall structure, one distinguishes between three structures computing the error in the three different ways: 1) equation error, 2) output error, 3) input error.

17.8.1 Auto-Regressive-eXtra-input (ARX) Model

The ARX model is an equation error model and is given by the following discrete polynomial representation (Ljung (1987)) using the shift operator

Type	$\rho(e)$
L_1	$\lvert e \rvert$
L_2	$1/2\, e^2$
Huber	$\begin{cases} 1/2\, e^2 & ; \lvert e \rvert \le k \\ k\,\lvert e \rvert - 1/2\, k^2 & ; \lvert e \rvert > k \end{cases}$
Welsch	$e^2/2\left[1 - \exp\left(-(e/c)^2\right)\right]$
Bisquare	$\begin{cases} k^2/6\left[1 - \left[1 - (e/k)^2\right]\right] & ; \lvert e \rvert \le k \\ k^2/6 & ; \lvert e \rvert > k \end{cases}$

Table 17.1: A selection of "robustifying" objective functions

q **??**:

$$A(q)\, y(k) := B(q)\, u(k) + e(k)\,,$$

or

$$y(k) := \frac{B(q)}{A(q)}\, u(k) + \frac{1}{A(q)}\, e(k)$$

with:

$$\begin{aligned} A(q) &:= 1 + \sum_{\mathcal{A}} a_i\, q^{-i}\,, \\ B(q) &:= \sum_{\mathcal{B}} b_i\, q^{-i}\,, \\ \mathcal{A} &:= \{i := 1, \dots, n\}\,, \\ \mathcal{B} &:= \{i := 0, \dots, m\}\,, \end{aligned}$$

and e denoting the error signal. [5]

The ARX acronym derives from the statistics literature labelling the different terms with:

AR	$A(q)\, y(k)$	**A**uto-**R**egressive
X	$B(q)\, u(k)$	e**X**tra input [6]

This model can be cast is linear in the parameters and results a standard

[5] The notation used here is compacted in that only the index sets are shown implying the operation to run over the index set for the implied index. For example the summation denoted by $\sum_{\mathcal{A}-,} a_i$ stands short for $\sum_{i=1}^{n}$ given the definition of $\mathcal{A}$ above.

[6] in economics this term is also referred to as exogenous input

linear regression problem. Let:[7]

$$\underline{\theta} := [[a_i]_{\mathcal{A}}; [b_i]_{\mathcal{B}}]$$
$$\underline{\mathbf{z}}(k) := \left[[-q^{-i}]_{\mathcal{A}}\, y(k); [q^{-i}]_{\mathcal{B}}\, u(k)\right] ,$$

then the model can be written in the form:

$$\hat{y}(k|\underline{\theta}) := \underline{\mathbf{z}}^T(k)\,\underline{\theta}\,.$$

17.8.2 Auto-Regressive-Moving-Average-eXtra-input (ARMAX) Model

The ARX model has a very simple structure with respect to the error. The ARMAX model extends this by defining also dynamics for the error. For this purpose an additional polynomial is introduced (Ljung (1987)):

$$C(q) := \sum_{\mathcal{C}} c_i\, q^{-i} ,$$
$$\mathcal{C} := \qquad \{i := 1, \ldots, o\} ,$$

which is used to define the ARMAX model:

$$y(k) := \frac{B(q)}{A(q)}\, u(k) + \frac{C(q)}{A(q)}\, e(k)\,.$$

The parameter vector is correspondingly expanded too:

$$\underline{\theta} := [[a_i]_{\mathcal{A}}; [b_i]_{\mathcal{B}}; [c_i]_{\mathcal{C}}] \,.$$

In the ARX case we could cast the parameter estimation problem into a simple linear regression form. In order to find a similar form, we first have to construct an estimate for the output for the ARMAX process. For this derivation we compact the notation:

$$y(k) := G(q)\, u(k) + H(q)\, e(k) ,$$

with

$$\begin{aligned} G(q) &:= \frac{B(q)}{A(q)} \\ H(q) &:= \frac{C(q)}{A(q)} \\ &:= 1 + \sum_{i:=1}^{\infty} h_i\, q^{-i}\,. \end{aligned}$$

[7] The notation used here is compacted following the same idea: $[q^{-i}]_{\mathcal{A}}$ stands for $[q^{-i}]_{\forall i}$ or $[q^{-1}, q^{-2}, \ldots, q^{-n}]$

The variance of the error is thus scaled such that the $H(q)$ polynomial is monic, d.h. the leading coefficient is 1. We further define:

$$v(k) := H(q)\, e(k)\,.$$

Thus:

$$\begin{aligned} v(k) &:= e(k) + \left(\sum_{i:=1}^{\infty} h_i\, q^{-i}\right)\, e(k)\,, \\ &:= e(k) + (H(q) - 1)\; e(k)\,. \end{aligned}$$

The expectation of the $v(k)$ given the data at $k-1$ is then:

$$\begin{aligned} \hat{v}(k|k-1) &:= \mathbf{E}\left[v(k)|k-1\right]\,, \\ &:= \mathbf{E}\left[e(k)\right] + \mathbf{E}\left[(H(q)-1)\; e(k)\right]\,. \end{aligned}$$

With the data being known at the time $k-1$, the second term is actually know and the expectation of the error is zero, thus

$$\begin{aligned} \hat{v}(k|k-1) &:= (H(q)-1)\; e(k)\,, \\ &:= (H(q)-1)\; H^{-1}(q)\, v(k)\,, \\ &:= \left(1 - H^{-1}(q)\right)\; v(k)\,. \end{aligned}$$

Now we can assemble the expression for the expected output:

$$\begin{aligned} \hat{y}(k|\Theta) &:= G(q)\, u(k) + \hat{v}(k|k-1)\,, \\ &:= G(q)\, u(k) + \left(1 - H^{-1}(q)\right)\; v(k)\,, \\ &:= G(q)\, u(k) + \left(1 - H^{-1}(q)\right)\; \left(y(k) - G(q)\, u(k)\right)\,, \\ &:= G(q)\, u(k) + \left(1 - H^{-1}(q)\right)\; y(k) - \left(1 - H^{-1}(q)\right)\; G(q)\, u(k)\,, \\ &:= \left(1 - H^{-1}(q)\right)\; y(k) + H^{-1}(q)\, G(q)\, u(k)\,. \end{aligned}$$

Substituting the two polynomials

$$\begin{aligned} \hat{y}(k|\Theta) &:= \left(1 - \frac{A(q)}{C(q)}\right)\; y(k) + \frac{A(q)}{C(q)}\,\frac{B(q)}{A(q)}\, u(k) \\ &:= \left(1 - \frac{A(q)}{C(q)}\right)\; y(k) + \frac{B(q)}{C(q)}\, u(k)\,. \end{aligned} \tag{17.9}$$

we get the one-step predictor for the ARMAX model.
Some more manipulations: First multiply with the $C(q)$ polynomial:

$$C(q)\,\hat{y}(k|\Theta) := (C(q) - A(q))\; y(k) + B(q)\, u(k)\,.$$

Extend on both sides:

$$\begin{aligned}C(q)\,\hat{y}(k|\Theta) + (1 - C(q))\ \hat{y}(k|\Theta) :=&\\ (C(q) - A(q))\ y(k) +&\quad B(q)\,u(k) + (1 - C(q))\ \hat{y}(k|\Theta)\,.\end{aligned}$$

Simplify the left-hand side first and expand the right-hand side aiming at an expression of the prediction error

$$\epsilon(k,\Theta) := (y(k) - \hat{y}(k|\Theta))$$

$$\begin{aligned}\hat{y}(k|\Theta) :=\ & (C(q) - A(q))\ y(k) + B(q)\,u(k) +\\ & (1 - C(q))\ \hat{y}(k|\Theta) - y(k) + y(k)\\ :=\ & B(q)\,u(k) + (1 - A(q))\,y(k) + (C(q) - 1)\ (y(k) - \hat{y}(k|\Theta))\ ,\\ :=\ & B(q)\,u(k) + (1 - A(q))\,y(k) + (C(q) - 1)\,\epsilon(k,\Theta)\,.\end{aligned}$$

The estimated output can thus be written in the form:

$$\hat{y}(k|\Theta) := \underline{\mathbf{z}}^T(t,\Theta)\,\underline{\theta}\,,$$

with :

$$\begin{aligned}\underline{\theta} &:= [[a_i]_{\mathcal{A}};[b_i]_{\mathcal{B}};[c_i]_{\mathcal{C}}]\\ \underline{\mathbf{z}}(k,\Theta) &:= \left[[-q^{-i}]_{\mathcal{A}}\,y(k);[q^{-i}]_{\mathcal{B}}\,u(k);[q^{-i}]_{\mathcal{C}}\,\epsilon(k,\Theta)\right]\ ,\\ \underline{\hat{\mathbf{y}}}(k|\underline{\theta}) &:= \underline{\mathbf{z}}^T(k,\Theta)\,\underline{\theta}\,.\end{aligned} \tag{17.10}$$

which is a nonlinear relation, though looking very much like the linear regression model we had for the ARX model. This form is called pseudo-linear regression.

17.8.3 General Transfer Function Model Structures

Along this line a generic transfer model can be suggested (Ljung (1987)) :

$$A(q)\,y(k) := \frac{B(q)}{F(q)}\,u(k) + \frac{C(q)}{D(q))}\,e(k)\,.$$

The one step predictor for this generic model, analogue to Equation 17.9, is:

$$\hat{y}(k|\Theta) := \qquad \left(1 - \frac{D(q)\,A(q)}{C(q)}\right)\,y(k) + \frac{D(q)\,B(q)}{C(q)\,F(q)}\,u(k)\,.$$

The following table, also taken from Ljung (Ljung (1987) p77) shows the models and their names depending on what polynomials are used in the general model:

polynomial	model	name
B	FIR	finite impulse response
A, B	ARX	auto regressive with extra input
A, B, C	ARMAX	auto regressive moving average with extra input
A, C	ARMA	auto regressive moving average
A, B, D	ARARX	2 (auto regressive) with extra input
A, B, C, D	ARARMAX	2 (auto regressive) moving average with extra input
B, F	OE	output error
B, F, C, D	BJ	Box-Jenkins

This model can also be cast into the pseudo-linear regression form. Again defining the error:

$$\epsilon(k,\Theta) := (y(k) - \hat{y}(k|\Theta))$$

one finds:

$$\epsilon(k,\Theta) := \frac{F(q))}{C(q)} \left(A(q)\, y(k) - \frac{B(q)}{F(q)}\, u(k) \right) .$$

Introducing the variables:

$$\begin{aligned} w(k,\Theta) &:= \frac{B(q)}{F(q)}\, u(k) \\ v(k,\Theta) &:= A(q)\, y(k) - w(k,\Theta)\,, \end{aligned}$$

this simplifies to:

$$\epsilon(k,\Theta) := \frac{F(q)}{C(q)}\, v(k,\Theta)\,.$$

With :

$$\begin{aligned} \underline{\theta} &:= [[a_i]_{\mathcal{A}}\,;\,[b_i]_{\mathcal{B}}\,;\,[c_i]_{\mathcal{C}}\,;\,[d_i]_{\mathcal{D}}\,;\,[f_i]_{\mathcal{F}}] \\ \underline{\mathbf{z}}(k,\Theta) &:= \big[[-q^{-i}]_{\mathcal{A}}\, y(k)\,;\,[q^{-i}]_{\mathcal{B}}\, u(k)\,;\,[q^{-i}]_{\mathcal{C}}\, \epsilon(k,\Theta)\,; \\ &\qquad [-q^{-i}]_{\mathcal{D}}\, v(k)\,;\,[-q^{-i}]_{\mathcal{F}}\, w(k)\big]\ , \end{aligned}$$

one has the model again in the pseudo-linear regression Equation 17.10 form.

17.9 Kalman Filter in Identification

The Kalman filter has been giving its name because the technique got most attention after being published by Rudolf E Kalman but the basic idea has been worked on by several people also earlier. This includes mainly Bucy, who is often also included in the name of the filter but rarely people like Ruslan L. Stratonovich and others[8]. For the sake of briefness it shall be called Kalman filter in the continuation.

The filter and its derivation is interesting as it solves an old problem formulated by Wiener, namely the issue of having stochastic components active at the input of a dynamic system. For linear systems the Kalman filter solves this problem for stochastic components exciting the input an the output independently and both distributions being at least symmetrical.

The nominal model being used for a discrete system is:

$$\begin{aligned}\underline{\bar{\mathbf{x}}}(k) &:= \underline{\underline{\mathbf{\Phi}}}\,\underline{\bar{\mathbf{x}}}(k-1) + \underline{\underline{\mathbf{\Gamma}}}\,\underline{\mathbf{u}}(k) + \underline{\mathbf{w}}(k)\,,\\ \underline{\bar{\mathbf{y}}}(k) &:= \underline{\underline{\mathbf{C}}}\,\underline{\bar{\mathbf{x}}}(k) + \underline{\mathbf{v}}(k)\,.\end{aligned}$$

Where the stochastic components are here assumed to be Gaussian:

$$\begin{aligned}\underline{\mathbf{w}} &\sim N\left(\underline{\mathbf{0}}, \underline{\underline{\mathbf{Q}}}\right)\,,\\ \underline{\mathbf{v}} &\sim N\left(\underline{\mathbf{0}}, \underline{\underline{\mathbf{R}}}\right)\,.\end{aligned}$$

The derivation of the filter can be done in many different ways, including the orthogonality principle, Bayes' theorem, sequential minimal sum of squares, gradient search method for sum of squares and others. We shall not derive the filter, but refer the interested reader to the literature for example Jazwinski (1970) which is still one of the books with the most thorough treatment of this subject.

The Kalman filter works in two steps:

Prediction: of the state and the estimates' covariance

$$\begin{aligned}\text{state} \quad \underline{\hat{\mathbf{x}}}(k|k-1) &:= \underline{\underline{\mathbf{\Phi}}}\,\underline{\hat{\mathbf{x}}}(k-1|k-1) + \underline{\underline{\mathbf{\Gamma}}}\,\underline{\mathbf{u}}(k-1)\,,\\ \text{covariance} \quad \underline{\underline{\mathbf{P}}}(k|k-1) &:= \underline{\underline{\mathbf{\Phi}}}\,\underline{\underline{\mathbf{P}}}(k-1|k-1)\,\underline{\underline{\mathbf{\Phi}}}^T + \underline{\underline{\mathbf{Q}}}(k-1)\,.\end{aligned}$$

Update: of measurement residuals, covariance of measurement residuals; computation of the Kalman gain being used to update the state estimate

[8] see for example Wikipedia http://en.wikipedia.org/wiki/Kalman_filter for more information on this subject

and the covariance of the state estimate:

$$\begin{aligned}
\text{residual} \quad \underline{\mathbf{e}}(k|k-1) &:= \underline{\bar{\mathbf{y}}}(k) - \underline{\underline{\mathbf{C}}}\,\underline{\hat{\mathbf{x}}}(k|k-1) \qquad , \\
\text{residual covariance} \quad \underline{\underline{\mathbf{S}}}(k) &:= \underline{\underline{\mathbf{C}}}\,\underline{\underline{\mathbf{P}}}(k|k-1)\,\underline{\underline{\mathbf{C}}}^T + \underline{\underline{\mathbf{R}}}(k)\,, \\
\text{Kalman gain} \quad \underline{\underline{\mathbf{K}}}(k) &:= \underline{\underline{\mathbf{P}}}(k|k-1)\,\underline{\underline{\mathbf{C}}}^T\,\underline{\underline{\mathbf{S}}}^{-1}(k)\,, \\
\text{state estimate} \quad \underline{\hat{\mathbf{x}}}(k|k) &:= \underline{\hat{\mathbf{x}}}(k|k-1) + \underline{\underline{\mathbf{K}}}(k)\,\underline{\mathbf{e}}(k|k-1) \qquad , \\
\text{estimate covariance} \quad \underline{\underline{\mathbf{P}}}(k|k) &:= \left(\underline{\underline{\mathbf{I}}} - \underline{\underline{\mathbf{K}}}(k)\,\underline{\underline{\mathbf{C}}}\right)\,\underline{\underline{\mathbf{P}}}(k|k-1),.
\end{aligned}$$

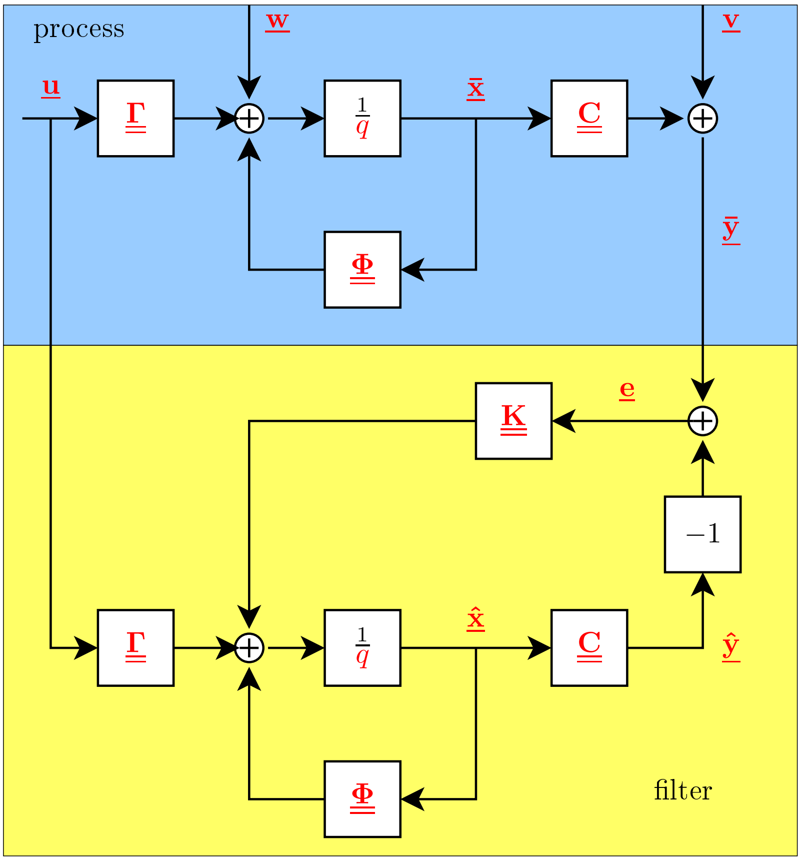

process
$\underline{\mathbf{w}}$
$\underline{\mathbf{v}}$
$\underline{\mathbf{u}}$
$\underline{\underline{\mathbf{\Gamma}}}$
$\frac{1}{q}$
$\underline{\bar{\mathbf{x}}}$
$\underline{\underline{\mathbf{C}}}$
$\underline{\underline{\mathbf{\Phi}}}$
$\underline{\bar{\mathbf{y}}}$
$\underline{\mathbf{e}}$
$\underline{\underline{\mathbf{K}}}$
-1
$\underline{\underline{\mathbf{\Gamma}}}$
$\frac{1}{q}$
$\underline{\hat{\mathbf{x}}}$
$\underline{\underline{\mathbf{C}}}$
$\underline{\hat{\mathbf{y}}}$
$\underline{\underline{\mathbf{\Phi}}}$
filter

Figure 17.6: Nominal model and Kalman filter

The Kalman filter is a state variable filter, meaning that the output of the filter provides an estimate of the state given a vector of observations. The main use of this filter is to reconstruct the state from a set of measurements. It is thus also called an observer. There other observers known in the literature, in particular the Luenberg observer, which differs from the Kalman filter mainly by having a fixed gain. The gain of the Luenberg observer is designed by setting the dynamics of the error propagation setting the poles of the set of linear differential equations that evelove from the derivation of the residual error.

17.9.1 Extetended Kalman Filter

The extended Kalman filter is using a linearised version of the nonlinear model for the computation of the variance propagation, whilst the prediction step is done with the nonlinear model. This idea was extended for the use in parameter identification in that the state of the system is extended by the parameters to be estimated setting the dynamics of these new ”states” to zero. The technique suffers from a number of problems, which are mostly associated convergence and providing a good, or better, workable estimate of the variance-co-variance matrices for the actual state, the parameters and the measurements. The literature is correspondingly rich on modifications to this scheme including constraining the parameters, introduction of forgetting factors, fixed matrices for the Kalman gain etc. etc.

17.10 The Excitation Signal

The behaviour of the plant can only be observed if it is *”moving”*, meaning it has to be disturbed, or *excited* by applying an excitation to the plant. The excitation must be chosen such that the plant is flexed sufficiently so as to make all ”movements” visible. For example, if one wants to know the mass of a physical object, one must to move the object, for example lift it. If one only pulls on it without moving it, one can only state that the mass is bigger than what one applied in force during the un-successful experiment.

This concept applies directly to the plant identification problem: if one wants to obtain information about the plant in a certain time scale, then one needs to excite it in that time scale, which directly translate into applying a certain frequency as an excitation input. This can be nicely demonstrated in the case shown in 17.7. It shows a Bode amplitude plot. The black line represents the behaviour of the true plant. A first-order model is being fit, which has two parameters, the gain and the time constant. One thus needs at least two independent experiments that provide the information necessary to find estimates for the two parameters. In the red case, the two sinusoidals indicated by the two dotted red lines are being used and in the green case it is the corresponding green dotted lines that indicate the input signals. The result is obviously different. In the red case the low-frequency behaviour is captured whilst in the green case more of the high-frequency behaviour is reflected into the model.

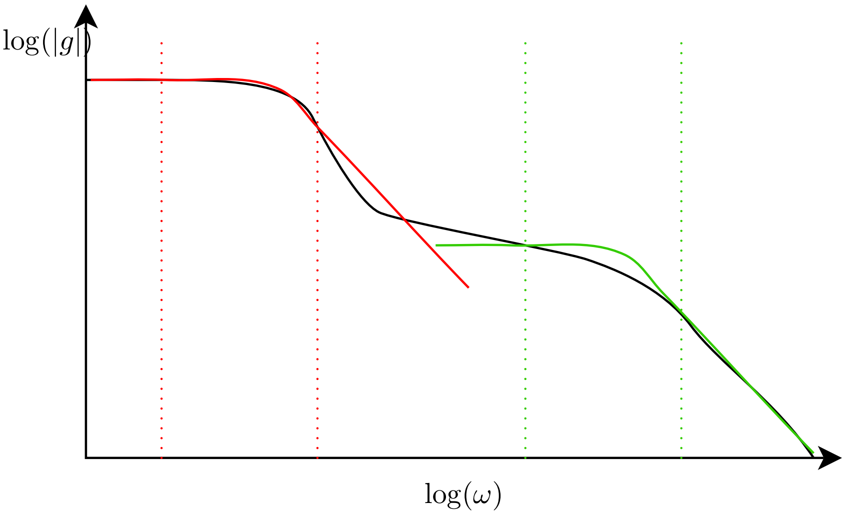


Figure 17.7: The choice of frequencies is essential for the identification experiment

The literature is rich on discussions and suggestions of what type of input signals should or can be used. In many cases people aim at identifying a plant as a kind of "whole", meaning that they do not think in time scales and thus hierarchical models. If one finds a split in the time scale, it is almost always feasible to work with two models, one for the high-dynamic range and one for the low-dynamic range. For example, it is quite thinkable that the red model is used to describe the plant in 17.7 in the low-frequency range whilst in the high-frequency range the green model is used.

If one indeed aims at identifying the whole plant, one must provide a model that is able to capture the behaviour, thus is rich enough. Having such a model, one then must excite the plant persistently, meaning with a signal that is rich enough in frequency contents. For more details on the definition of *persistant* see for example Ljung (1987); Eykhoff (1974), which also include references to work on this subject.

Obviously one of the simple solutions for the latter problem is to use all frequencies, for example a random signal. Since this may not be trivial to apply, one often uses signals that are coming close, such as random binary signals. Adding a variation in the amplitude gives multi-level random signals.

From the practical point of view, one should also keep the signal-noise ratio in mind. If one applies for example a random binary signal, the energy one has usually available in a limited amount, is spread over all the frequencies equally, at least ideally. In any case it is spread making the signal of the individual frequency less "strong" and thus more likely to be "covered" by noise components acting on the equipment, for example the measurement devices. It is thus often better to apply a frequency or a selected set of frequencies. In any case though, one should keep in mind in what time-scale the model will be used and thus to what detail the process should or must be described.

17.11 Sampling

Dynamic systems, which will be discussed in the next section, are usually connected to computer systems that provide the control for the excitation signal and that take the samples. With the computing device being a discrete-time device, both operation occur in a sampled-data environment, meaning there is a sampler on the measurement side and a signal output on the other side. In other processes, this sampling is not so obvious, for example if one deals with a large population, such it occurs frequently in connection with quality control. Examples are inflow of material that comes in discrete units, parcels, groups, lumps of one or the other kind. Consider for example a stream of potatoes, which are to be priced based on their contents of starch. It is not feasible to check every potato but samples must be taken. Similarly if one receives a train full of material of non-uniform composition. Being faced with having to assess the distribution of the population, people have invented a set of sampling techniques each of which has its own characteristics and takes care of one or the other special situation one may occur in this context.

17.11.1 Random sampling

This is one of quite obvious methods to get an idea of the distribution in a population. It requires though, that one really accesses the population uniformly, meaning that each member in the population is selectable with the same probability. This is known as *equal probability selection*. All samples are consequently given the same weight. If the uniform selectable condition is not satisfied, then one talks about exclusion bias.

17.11.2 Systematic sampling

The population is pre-ordered and the sampling is adjusted to this order. Sampling is then done in a systematic way making sure that one samples uniformly over the ordered population. If the order is having a periodic characteristic, then the sampling must not sample periodic, but must randomise within a period as otherwise the wrong distribution is obtained. The sampling must thus be in accordance with the frequency distribution of the population.

17.11.3 Stratified samples

If the population can be ordered or split into classes or layers, the sampling must be done in each of these layers. This approach also provides information about the differences between the different strata. One may also tailor the sampling to the individual stratum. Equally well is it possible to shift focus or to put more or less weight on one or the other stratum.

17.11.4 Proportional-probability sampling

If one has a priori knowledge, such as the size of a population, relative to another one, one may sample in the corresponding proportions.

17.11.5 Cluster sampling

This is commonly implemented as a multi-stage method. On the top level one constructs the clusters whilst on the second level one applies a sampling technique that is appropriate for the cluster.

17.12 Design of Experiments

Models are constructed based on the available knowledge. If the nature of the plant is known in details, one may decide on constructing a mechanistic model such as it was discussed in the earlier chapters. I this is not the case, one has to resort to empirical models, which reflect the plant's behaviour in a functional form that the person modelling the process believes ”fits” the behaviour of the plant best. It is customary to label the first type of models as white-box models, whilst the latter are called black-box models. What was discussed in the earlier chapters is typically a mix of the two. The foundation is usually mechanistic and the more one gets into the internal details, what is often referred to as the constitutive equations, one has less and less information about the basic nature and black-box models are being used selected on experience: The conservation concepts of physics are considered as mechanistic and so are large parts of the macroscopic theory-based description of the hydraulics of a plant. As one gets into the details of transport and in into the description of material properties and reactions, the understanding of the underlying processes becomes thinner and thinner or more and more involved so that one usually has to resort to essentially empirical models. Often some remainders of the underlying concepts are preserved reflecting into the functional form the empirical model takes.

Asking the following three questions leads stepwise to a process model:

- **What affects the plant?** – **screening experiments** aim at providing rudimentary information about the input/output behaviour of the plant.
- **How do Inputs Affects the Plant?** – **Response-surface Method** usually uses simple models, often polynomial models to describe the steady-state input-output behaviour of the plant, searching then for the optimum in the approximate space.
- **Why does the plant do what it is doing?** – **Mechanistic model** are the only models that explain the internal behaviour of the plant.

17.12.1 Single Block Design

Some of the characteristics are not available through deductive studies and must be identified using process identification techniques. The experiments are to be designed to provide most information as possible. Design of experiments has its roots in the statistical literature which refers to the inputs or stimuli signals as factors Box et al. (1978). The most efficient way of arranging experiments is in blocks meaning a set of experiments which modifies the input levels systematically. Potentially to each input a step is being applied. One waits long enough to get sufficiently close to the steady state value of the observation, which implies that one has to wait for at least 5 times the maximal time constant in the plant.

17.12.1.1 Linear in input and linear in input model

The simplest model is linear in the inputs and the parameters:

$$\hat{y} := \left[\underline{\mathbf{u}}^T\right] \underline{\theta} \tag{17.11}$$

with:

$$\underline{\mathbf{u}} := \underline{\underline{\text{diag}}}\left([\Delta \underline{\mathbf{u}}]\right) \underline{\bar{\mathbf{u}}}$$

The elements in the diagonal matrix are the perturbations in the individual inputs around a central point defined as $\underline{\mathbf{u}}^0$. The perturbations are done in both the positive and the negative direction. If one uses all combinations, then this is a complete design.

For example for a 3-input systems, one gets the $\underline{\underline{\mathbf{S}}}$ matrix:

$$\underline{\underline{\mathbf{S}}} \;:=\; \begin{bmatrix} +1 & +1 & +1 \\ -1 & +1 & +1 \\ +1 & -1 & +1 \\ -1 & -1 & +1 \\ +1 & +1 & -1 \\ -1 & +1 & -1 \\ +1 & -1 & -1 \\ -1 & -1 & -1 \end{bmatrix} .$$

This forms a cube with the corners being the Δu away from the central point. The plan generates the inputs :

$$\underline{\underline{\mathbf{F}}}(\underline{\mathbf{u}}) := \underline{\mathbf{F}}^o(\underline{\mathbf{u}}^o) + \underline{\underline{\mathrm{diag}}}\,[\Delta \underline{\mathbf{u}}]\; \underline{\underline{\mathbf{S}}}^T .$$

The response of the model to the plan is:

$$\underline{\hat{\mathbf{y}}} = \underline{\underline{\mathbf{F}}}(\underline{\mathbf{u}})\,\underline{\theta}$$

Simplifying the notation by defining the norming matrix

$$\underline{\underline{\mathbf{D}}} := \underline{\underline{\mathrm{diag}}}\,[\Delta \underline{\mathbf{u}}] \tag{17.12}$$

Then the regression problem can be reformulated. More compactly we write:

$$\underline{\underline{\mathbf{F}}} := \underline{\underline{\mathbf{F}}}^o + \underline{\underline{\mathbf{D}}}\,\underline{\underline{\mathbf{S}}}^T .$$

Substituting the input matrix in Equation 17.3.2.3 one gets

$$\underline{\hat{\mathbf{y}}} := \left(\underline{\underline{\mathbf{F}}}^o + \underline{\underline{\mathbf{D}}}\,\underline{\underline{\mathbf{S}}}^T\right)^T \underline{\theta} .$$

Performing experiments as the centre, averaging them and subtracting them from the measurement obtained when executing the plan, one gets

$$\begin{aligned} \underline{\hat{\mathbf{y}}} - \underline{\hat{\mathbf{y}}}^0 &:= \left(\underline{\underline{\mathbf{D}}}\,\underline{\underline{\mathbf{S}}}^T\right)^T \underline{\theta} , \\ &:= \underline{\underline{\mathbf{S}}}\,\underline{\underline{\mathbf{D}}}\,\underline{\theta} . \end{aligned}$$

The least-squares estimator 17.7 is then:

$$\begin{aligned} \underline{\underline{\mathbf{D}}}\,\underline{\theta} &:= \left(\underline{\underline{\mathbf{S}}}^T\,\underline{\underline{\mathbf{S}}}\right)^{-1} \underline{\underline{\mathbf{S}}}^T \left(\underline{\hat{\mathbf{y}}} - \underline{\hat{\mathbf{y}}}^0\right) \\ \underline{\theta} &:= \underline{\underline{\mathbf{D}}}^{-1} \left(\underline{\underline{\mathbf{S}}}^T\,\underline{\underline{\mathbf{S}}}\right)^{-1} \underline{\underline{\mathbf{S}}}^T \left(\underline{\hat{\mathbf{y}}} - \underline{\hat{\mathbf{y}}}^0\right) \end{aligned}$$

The $\underline{\underline{\mathbf{S}}}^T\,\underline{\underline{\mathbf{S}}}$ is orthogonal and thus makes the regression analysis extremely simple:

$$\underline{\underline{\mathbf{S}}}^T\,\underline{\underline{\mathbf{S}}} := n\underline{\underline{\mathbf{I}}}$$

and thus:

$$\underline{\theta} := \frac{1}{n}\,\underline{\underline{\mathbf{D}}}^{-1}\,\underline{\underline{\mathbf{S}}}^T\,\left(\underline{\hat{\mathbf{y}}} - \underline{\hat{\mathbf{y}}}^0\right)$$

17.12.1.2 Linear in parameter models with interactions of the inputs

If the above linear-in-input and linear-in-parameter model does not perform satisfactorily, one would usually increase the complexity of the model by adding interaction terms and quadratic terms. For example one could extent the above model with interaction terms:

$$\hat{y} := [u_1, u_2, u_3, u_1u_2, u_1u_3, u_2u_3, u_1\,u_2\,u_3]\,\underline{\theta}$$

The plan is as above, and so are the inputs. What changes is the formulation of the $\underline{\underline{\mathbf{F}}}$. For the given case it takes the form:

$$\underline{\underline{\bar{\mathbf{S}}}} \quad := \begin{bmatrix} +1 & +1 & +1 & +1 & +1 & +1 & +1 \\ -1 & +1 & +1 & -1 & -1 & +1 & -1 \\ +1 & -1 & +1 & -1 & +1 & -1 & -1 \\ -1 & -1 & +1 & +1 & -1 & -1 & +1 \\ +1 & +1 & -1 & +1 & -1 & -1 & -1 \\ -1 & +1 & -1 & -1 & +1 & -1 & +1 \\ +1 & -1 & -1 & -1 & -1 & +1 & +1 \\ -1 & -1 & -1 & +1 & +1 & +1 & -1 \end{bmatrix}.$$

This matrix is constructed from $\underline{\underline{\mathbf{S}}}$:

$$\underline{\underline{\bar{\mathbf{S}}}} := \left[\underline{\underline{\mathbf{S}}}, \underline{\mathbf{s}}_1 \circ \underline{\mathbf{s}}_2, \underline{\mathbf{s}}_1 \circ \underline{\mathbf{s}}_3, \underline{\mathbf{s}}_2 \circ \underline{\mathbf{s}}_3, \underline{\mathbf{s}}_1 \circ \underline{\mathbf{s}}_2 \circ \underline{\mathbf{s}}_3\right]$$

with $\circ$ being the Hadamard operator.
The $\underline{\underline{\mathbf{F}}}$ is then readily constructed

$$\underline{\underline{\mathbf{F}}} := \underline{\underline{\mathbf{F}}}^o + \underline{\underline{\bar{\mathbf{D}}}}\,\underline{\underline{\bar{\mathbf{S}}}}^T.$$

with the $\underline{\underline{\bar{\mathbf{D}}}}$ being the respective norming matrix, which for our example is:

$$\underline{\underline{\bar{\mathbf{D}}}} := \underline{\underline{\mathrm{diag}}}([\Delta u_1, \Delta u_2,\ \Delta u_3, \Delta u_1\,\Delta u_2, \Delta u_1\,\Delta u_3, \Delta u_2\,\Delta u_3, \Delta u_1\,\Delta u_2\,\Delta u_3])$$

and again the regression problem is orthogonal and simplifies to:

$$\underline{\theta} := \frac{1}{n} \bar{\underline{\underline{\mathbf{D}}}}^{-1} \bar{\underline{\underline{\mathbf{S}}}}^{T} \left(\underline{\hat{\mathbf{y}}} - \underline{\hat{\mathbf{y}}}^{0}\right)$$

17.12.1.3 Models with constant term

It is straightforward to extend the above models with a constant term, if it is to be estimated as well. This then adds an additional parameter, namely the offset and extends the regression matrix correspondingly.

In the approach taken above, any constant term is removed by taking the difference of the measurements at the corners minus the measurement in the centre.

17.12.1.4 Reduced plans

The linear model (Equation 17.11) has three parameters plus a possible offset. Thus four equations are enough to get an estimate of all the parameters and one could use half of the set of equations generated by the full plan. The four equations are chosen from the extremes, namely the ones opposite along the two space diagonals, thus the two pairs of opposite corners. Apparently two configurations are possible corresponding to the two possible blocked plans.

The reduced plans can be generated systematically using plan generators as they are discussed in Box et al. (2005).

17.12.1.5 Handling Additive Noise

The effect of measurement noise can be reduced by replication of individual experiments. This reduces the variations by the square root of the number of identical experiments assuming the noise is stationary. Assuming the variance is σ^2 and one performs n experiments getting n responses $\{y_i\}_{i:1\dots n}$, then mean is

$$\breve{y} := \frac{1}{n} \sum_{1}^{n} y_i$$

The variance is then

$$\begin{aligned}\mathrm{var}(y) &:= \frac{1}{n^2}\sum_{1}^{n} \mathrm{var}(y_i) \\ &:= \frac{1}{n^2}\, n\, \sigma^2 \\ &:= \frac{\sigma^2}{n}\end{aligned}$$

In praxis such repeated experiments are often difficult to perform as a lot of other ”disturbances” affect the plant. Not at least the person or instrument performing the experiments.

17.12.1.6 Reducing Trends

Randomising Trends caused by correlation of the inputs can be reduced by ”randomsing” the experiments. In this process one introduces a random variable selecting the experiments. For example randomizing the experimental plant Equation 17.12 one would randomly mix the columns each of which represent an individual experiments.

Block Designs Trends can be also be reduced in the case when one deals with similar items / plants, which are to analysed modeled exciting the corresponding inputs. For example if one has two different pair of shoes and wants to know their ability to absorb a dye and grease, the shoes are the items / plants and will be blocked. The experiments are then to apply grease / dye respectively in combinations. An experimental plan is then established for each block, which in turn is randomized as discussed above.

17.12.2 Optimal Designs

The optimal experiment design builds on information theory, because for an unbiased estimator the estimator variance is related to Fisher information matrix: Minimizing the variance corresponds to maximizing the information.

There exist several criteria of optimality. The traditional ones build on the invariants of the Fisher information matrix $\underline{\underline{\mathbf{M}}}$. The inverse of the Fisher information matrix is the lower bound of the variance-covariance: 17.5.1. Thus the optimal design attempts to minimize the variance-covariance matrix or, which is for uniform variances the $\underline{\underline{\mathbf{V}}} := \underline{\underline{\mathbf{F}}}^T\,\underline{\underline{\mathbf{F}}}$ (see Equation 17.8). Commonly used measures are:

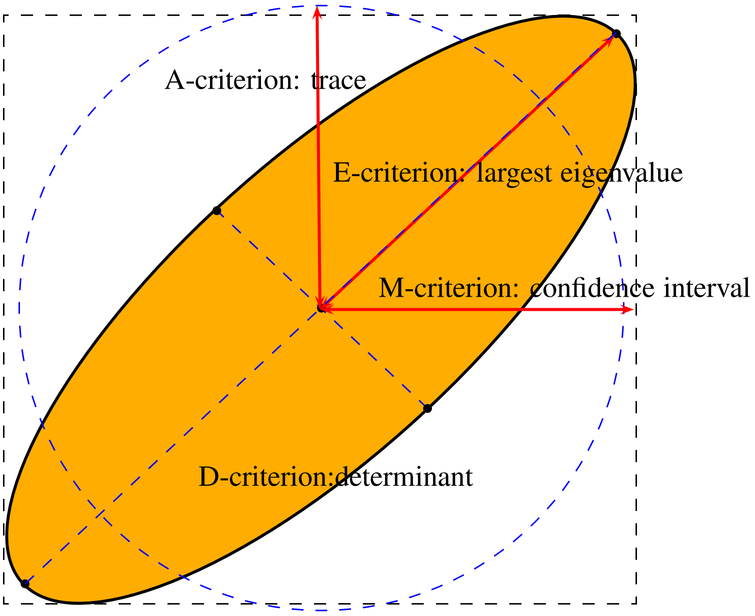

A-criterion: trace
E-criterion: largest eigenvalue
M-criterion: confidence interval
D-criterion:determinant

Figure 17.8: The various criteria visualised

A (average) - optimality: $\min \operatorname{trace}(\underline{\underline{\mathbf{V}}})$
average length of the half axes

D (determinant) - optimality: $\min |\underline{\underline{\mathbf{V}}}|$
volume

E (eigenvalue) - optimality: $\max \lambda \left(\underline{\underline{\mathbf{V}}}\right)$
largest half axis

M (dominant diagonal value) - optimality : $\max \sqrt{m_{ii}}$
largest side of enclosing box

Appendices

A

Linear algebra

A.1 Matrices and vectors

Many engineering and science problems make extensive use of linear algebra. Why chemical engineering? Our problems are almost never one-dimensional, but the dimensionality is given by the number of species plus energy and one to three dimension from the hydraulic. Latter adds 3 dimensions if one solves the momentum balances in the three dimensional space. Whilst overall process models are nonlinear, the fundamental balance equations are linear: the superposition applies, they are of Euler degree one. The nonlinearities enter mainly through the material description and to some extend also through the transport equations. Definitely the reaction add a strongly nonlinear term as the ”reaction constant” is indeed anything else than constant but a strong function of the temperature. Arrhenius being the standard model, which describes the reaction constant as an exponential function of the temperature.

The utilisation of models often utilises linearisation, thereby approximating the nonlinear behaviour locally by a linear model. Overall, it is no surprise that linear algebra forms a core in applied mathematics. There exist many excellent text books and the reader is being recommended to ”consume” one of these text in full. It certainly is useful to have a good background and a reference at hand when the need comes about. Myself I enjoyed particularly the book by Gilbert Strang (Strang (2009)).

This appendix does not cover linear algebra. The idea is to review and collect some of the main results, which are used here and there in this text.

Denoting the real numbers with $\mathbb{R}$, let $x \in \mathbb{R}$, a column vector is a stack of numbers. Let us define a column vector of real numbers as a vertical stack of real numbers:

$$\underline{\mathbf{x}} := \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n \tag{A.1}$$

Or if convinient we may write:

$$\underline{\mathbf{x}} := [x_i]_{i:=1,2,\dots,n} \quad \in \mathbb{R}^n \tag{A.2}$$

The sum of two vectors is calculated by adding component by component:

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix} \tag{A.3}$$

This implies that the two vectors must be of the same length and thus must belong to the same space.
The transposed is a row vector:

$$\underline{\mathbf{x}}^T := \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix} \tag{A.4}$$

A matrix of real numbers is a two-dimensional object, a full table of real numbers:

$$\underline{\underline{\mathbf{A}}} := \begin{bmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m,1} & a_{m,2} & \dots & a_{m,n} \end{bmatrix} \quad \in \mathbb{R}^{m \times n} \tag{A.5}$$

Or again we may write:

$$\underline{\underline{\mathbf{A}}} := [a_{i,j}]_{i:=1,\dots,m,j:=1,\dots,n} \quad \in \mathbb{R}^{m \times n} \tag{A.6}$$

The transposed matrix:

$$\underline{\underline{\mathbf{A}}}^T := \begin{bmatrix} a_{1,1} & a_{2,1} & \dots & a_{n,1} \\ a_{1,2} & a_{2,2} & \dots & a_{n,2} \\ \vdots & \vdots & \vdots & \vdots \\ a_{1,n} & a_{2,n} & \dots & a_{m,n} \end{bmatrix} \quad \in \mathbb{R}^{n \times m} \tag{A.7}$$

Inner product:

$$\underline{\mathbf{x}}^T \, \underline{\mathbf{y}} := \sum_{\forall i} x_i \, y_i \tag{A.8}$$

Outer product:

$$\underline{\mathbf{x}}\,\underline{\mathbf{y}}^T := [x_i\, y_i]_{\forall i,\forall j} \tag{A.9}$$

Matrix - vector product:

$$\underline{\mathbf{b}} := \underline{\underline{\mathbf{A}}}\,\underline{\mathbf{x}} \quad := \left[\sum_{\forall j} a_{i,j}\, y_j\right]_{\forall i} \tag{A.10}$$

Matrix product:

$$\underline{\underline{\mathbf{C}}} := \underline{\underline{\mathbf{A}}}\,\underline{\underline{\mathbf{B}}} := \left[\sum_{\forall j} a_{i,j}\, b_{j,k}\right]_{\forall i,\forall k} \tag{A.11}$$

Inverse:

$$\underline{\underline{\mathbf{A}}}^{-1} := |\underline{\underline{\mathbf{A}}}|^{-1}\, \mathbf{adj}(\underline{\underline{\mathbf{A}}}) \tag{A.12}$$

Example 2 x 2 matrix:

$$\underline{\underline{\mathbf{A}}} := \begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix} \tag{A.13}$$

$$|\underline{\underline{\mathbf{A}}}| := a_{1,1}\, a_{2,2} - a_{1,2}\, a_{2,1} \tag{A.14}$$

$$\mathrm{adj}(\underline{\underline{\mathbf{A}}}) := \begin{bmatrix} a_{2,2} & -a_{1,2} \\ -a_{2,1} & a_{1,1} \end{bmatrix} \tag{A.15}$$

$$\tag{A.16}$$

given matrix $\underline{\underline{\mathbf{A}}} \in \mathbb{R}^{n\times n}$

- Define minor of $\underline{\underline{\mathbf{A}}}$ denoted by $\underline{\underline{\mathbf{M}}}_{i,j} \in \mathbb{R}^{(n-1)\times(n-1)}$ by deleting the i row and the j column.
- Define the i, j co-factor of $\underline{\underline{\mathbf{A}}}$:

$$\underline{\underline{\mathbf{c}}}_{i,j} := (-1)^{i+j}\, |\underline{\underline{\mathbf{M}}}_{i,j}|$$

- Define the adjoint or adjugate of A as the transpose of the co-factor matrix:

$$\mathbf{adj}(\underline{\underline{\mathbf{A}}}) = [c_{i,j}]^T_{\forall i,\forall j} = [c_{j,i}]_{\forall j,\forall i}$$

$$\mathbf{A} = \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$$

$$\mathbf{C} = \begin{pmatrix} +\begin{vmatrix} A_{22} & A_{23} \\ A_{32} & A_{33} \end{vmatrix} & -\begin{vmatrix} A_{21} & A_{23} \\ A_{31} & A_{33} \end{vmatrix} & +\begin{vmatrix} A_{21} & A_{22} \\ A_{31} & A_{32} \end{vmatrix} \\ -\begin{vmatrix} A_{12} & A_{13} \\ A_{32} & A_{33} \end{vmatrix} & +\begin{vmatrix} A_{11} & A_{13} \\ A_{31} & A_{33} \end{vmatrix} & -\begin{vmatrix} A_{11} & A_{12} \\ A_{31} & A_{32} \end{vmatrix} \\ +\begin{vmatrix} A_{12} & A_{13} \\ A_{22} & A_{23} \end{vmatrix} & -\begin{vmatrix} A_{11} & A_{13} \\ A_{21} & A_{23} \end{vmatrix} & +\begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{vmatrix} \end{pmatrix}$$

$$= \begin{pmatrix} +\begin{vmatrix} 5 & 6 \\ 8 & 9 \end{vmatrix} & -\begin{vmatrix} 4 & 6 \\ 7 & 9 \end{vmatrix} & +\begin{vmatrix} 4 & 5 \\ 7 & 8 \end{vmatrix} \\ -\begin{vmatrix} 2 & 3 \\ 8 & 9 \end{vmatrix} & +\begin{vmatrix} 1 & 3 \\ 7 & 9 \end{vmatrix} & -\begin{vmatrix} 1 & 2 \\ 7 & 8 \end{vmatrix} \\ +\begin{vmatrix} 2 & 3 \\ 5 & 6 \end{vmatrix} & -\begin{vmatrix} 1 & 3 \\ 4 & 6 \end{vmatrix} & +\begin{vmatrix} 1 & 2 \\ 4 & 5 \end{vmatrix} \end{pmatrix}$$

$$\begin{aligned}\operatorname{adj}(\mathbf{A}) &= \begin{pmatrix} +\begin{vmatrix} A_{22} & A_{23} \\ A_{32} & A_{33}\end{vmatrix} & -\begin{vmatrix} A_{12} & A_{13} \\ A_{32} & A_{33}\end{vmatrix} & +\begin{vmatrix} A_{12} & A_{13} \\ A_{22} & A_{23}\end{vmatrix} \\ \\ -\begin{vmatrix} A_{21} & A_{23} \\ A_{31} & A_{33}\end{vmatrix} & +\begin{vmatrix} A_{11} & A_{13} \\ A_{31} & A_{33}\end{vmatrix} & -\begin{vmatrix} A_{11} & A_{13} \\ A_{21} & A_{23}\end{vmatrix} \\ \\ +\begin{vmatrix} A_{21} & A_{22} \\ A_{31} & A_{32}\end{vmatrix} & -\begin{vmatrix} A_{11} & A_{12} \\ A_{31} & A_{32}\end{vmatrix} & +\begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22}\end{vmatrix}\end{pmatrix} \\ &= \begin{pmatrix} +\begin{vmatrix} 5 & 6 \\ 8 & 9\end{vmatrix} & -\begin{vmatrix} 2 & 3 \\ 8 & 9\end{vmatrix} & +\begin{vmatrix} 2 & 3 \\ 5 & 6\end{vmatrix} \\ \\ -\begin{vmatrix} 4 & 6 \\ 7 & 9\end{vmatrix} & +\begin{vmatrix} 1 & 3 \\ 7 & 9\end{vmatrix} & -\begin{vmatrix} 1 & 3 \\ 4 & 6\end{vmatrix} \\ \\ +\begin{vmatrix} 4 & 5 \\ 7 & 8\end{vmatrix} & -\begin{vmatrix} 1 & 2 \\ 7 & 8\end{vmatrix} & +\begin{vmatrix} 1 & 2 \\ 4 & 5\end{vmatrix}\end{pmatrix} \\ &= \begin{pmatrix} -3 & 6 & -3 \\ 6 & -12 & 6 \\ -3 & 6 & -3\end{pmatrix}\end{aligned}$$

A.1.1 Special matrices

Permutation matrices: The permutation of rows is done with a matrix that permutes the respective rows of an identity matrix. Premultiplying a $3 \times n$ matrix by the following permutation matrix

$$P := \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \tag{A.17}$$

will exchange the 2rd with the 3th row:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} a & 1 & 0 \\ b & 2 & 0 \\ c & 3 & 0 \end{bmatrix} = \begin{bmatrix} a & 1 & 0 \\ c & 3 & 0 \\ b & 2 & 0 \end{bmatrix} \tag{A.18}$$

A.1.2 Vector spaces

Some facts (Strang (2009))

- The space $\mathbb{R}^n$ contains all column vectors with n components, also the zero vector $\underline{\mathbf{0}}$.
- A subspace $\mathbb{R}^m$ containing $\underline{\mathbf{v}}, \underline{\mathbf{w}} \in \mathbb{R}^m$ must also contain all linear combinations $c\,\underline{\mathbf{v}} + d\,\underline{\mathbf{w}}$ with $\underline{\mathbf{c}}, \underline{\mathbf{w}} \in \mathbb{R}$.
- The combination of all columns of the matrix $\underline{\underline{\mathbf{A}}}$ form the column space of $\underline{\underline{\mathbf{A}}}$ denoted by $C\left(\underline{\underline{\mathbf{A}}}\right)$. One says: the space is spanned by the columns of $\underline{\underline{\mathbf{A}}}$.
- $\underline{\underline{\mathbf{A}}}\,\underline{\mathbf{x}} = \underline{\mathbf{b}}$ has a solution only if $\underline{\mathbf{b}} \in C\left(\underline{\underline{\mathbf{A}}}\right)$

A.1.3 Null space for fat matrices

A.1.3.1 Example:

$$a_x\,x + a_y\,y + a_z\,z = 0$$

is a plane that goes through the origin of the co-ordinate system. This plane is a subspace of $\mathbb{R}^3$. It is the nullspace of of the matrix $\begin{bmatrix} a_x & a_y & a_z \end{bmatrix}$.
The solution to:

$$a_x\,x + a_y\,y + a_z\,z + b = 0$$

also forms a plane but goes not through the origin of the co-ordinate system and is thus not a subspace.
Fat matrices have more columns than rows: $\underline{\underline{\mathbf{A}}} \in \mathbb{R}^{m \times n}, m < n$

- The solutions of $\underline{\underline{\mathbf{A}}}\,\underline{\mathbf{x}} = \underline{\mathbf{0}}$ is the null space of $\underline{\underline{\mathbf{A}}}$ denoted by $\mathrm{Null}\left(\underline{\underline{\mathbf{A}}}\right)$ or kernel of $\underline{\underline{\mathbf{A}}}$ denoted by $\mathrm{kern}\left(\underline{\underline{\mathbf{A}}}\right)$

$$\mathrm{Null}\left(\underline{\underline{\mathbf{A}}}\right) := \left\{\underline{\mathbf{x}} \in \mathbb{R}^n : \underline{\underline{\mathbf{A}}}\,\underline{\mathbf{x}} = \underline{\mathbf{0}}\right\} \tag{A.19}$$

- The null space is invariant to row manipulations, which makes it possible to convert $\underline{\underline{\mathbf{A}}}$ to the row echolon form and by *swapping columns* to the reduced row echelon form or row canonical echelon form, which corresponds to re-ordering the components in the x-vector:

$$\underline{\underline{\mathbf{R}}} := \begin{bmatrix} \underline{\underline{\mathbf{I}}} & \underline{\underline{\mathbf{F}}} \\ \underline{\underline{\mathbf{0}}} & \underline{\underline{\mathbf{0}}} \end{bmatrix}. \tag{A.20}$$

- The non-zero top rows belong to the pivot variables. The lower rows belong to the free variables.

- The rank is the number of non-zero rows, thus the number of pivot variables.

- For fat matrices the matrix $\underline{\underline{\mathbf{A}}}$ has at least one column without a pivot, which thus gives a special solution. Consequently there are non-zero solution vectors in Null $\left(\underline{\underline{\mathbf{A}}}\right)$.

- The result is matrix with the column vectors spanning the null space:

$$\text{Null}\left(\underline{\underline{\mathbf{A}}}\right) := \begin{bmatrix} -\underline{\underline{\mathbf{F}}} \\ \underline{\underline{\mathbf{I}}} \end{bmatrix} . \tag{A.21}$$

A.1.3.2 Procedure

1. $\underline{\underline{\mathbf{A}}}\,\underline{\mathbf{x}} \rightarrow \underline{\underline{\mathbf{U}}}\,\underline{\mathbf{x}}$: Transform equations to upper triangular form through linearly combining rows and swapping rows. The resulting upper triangular matrix $\underline{\underline{\mathbf{U}}}$ has the pivots as the first non-zero element in each row, thus in echolon form.

2. $\underline{\underline{\mathbf{U}}}\,\underline{\mathbf{x}} \rightarrow \underline{\underline{\mathbf{R}}}\,\underline{\mathbf{x}}$: Eliminate the non-zero elements above the pivots through linearly combining rows. This operation decouples the pivot variables. The resulting matrix is called to be in reduced row echolon form. $\underline{\underline{\mathbf{R}}} := \begin{bmatrix} \underline{\underline{\mathbf{I}}} & \underline{\underline{\mathbf{F}}} \\ \underline{\underline{\mathbf{0}}} & \underline{\underline{\mathbf{0}}} \end{bmatrix}$

3. Extract null space: Null $\left(\underline{\underline{\mathbf{A}}}\right) := \begin{bmatrix} -\underline{\underline{\mathbf{F}}} \\ \underline{\underline{\mathbf{I}}} \end{bmatrix}$

A.1.3.3 Example

Let the equation system be defined as:

$$\begin{array}{rcrcrcrcrcc}
2x_1 & & & + & 4x_3 & + & 4x_4 & + & 6x_5 & = & 0 \\
 & & & & x_3 & + & 4x_4 & & 2x_5 & = & 0 \\
4x_1 & & & + & 8x_3 & + & 8x_4 & & 12x_5 & = & 0 \\
6x_1 & + & 4x_2 & + & 2x_3 & + & 12x_4 & & & = & 0
\end{array} \tag{A.22}$$

The coefficient matrix is then:

$$\begin{bmatrix} 2 & 0 & 4 & 4 & 6 \\ 0 & 0 & 1 & 4 & 2 \\ 4 & 0 & 8 & 8 & 12 \\ 6 & 4 & 2 & 12 & 0 \end{bmatrix} \tag{A.23}$$

Eliminate the first element in row 3 by substracting 2 times row 1 and eliminate the first element in row 4 by substracting 3 times row 1:

$$\begin{bmatrix} 2 & 0 & 4 & 4 & 6 \\ 0 & 0 & 1 & 4 & 2 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 4 & -10 & 0 & -18 \end{bmatrix} \tag{A.24}$$

Next: move row 3 one down:

$$\begin{bmatrix} 2 & 0 & 4 & 4 & 6 \\ 0 & 0 & 1 & 4 & 2 \\ 0 & 4 & -10 & 0 & -18 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{A.25}$$

Then swap row 2 and 3:

$$\begin{bmatrix} 2 & 0 & 4 & 4 & 6 \\ 0 & 4 & -10 & 0 & -18 \\ 0 & 0 & 1 & 4 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{A.26}$$

Divide row 1 by 2 and row 2 by 4:

$$\begin{bmatrix} 1 & 0 & 2 & 2 & 3 \\ 0 & 1 & -2.5 & 0 & -4.5 \\ 0 & 0 & 1 & 4 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{A.27}$$

This is now in row echelon form, all the leading elements that are non-equal 0 are one position to the right compared to the one above and zero rows are on the bottom. It looks somewhat like a upper triangular matrix, though it is not square. Next we "make" all element above the leading 1's also zero.

We use row 3 to do so and get:

$$\begin{bmatrix} 1 & 0 & 0 & -6 & -1 \\ 0 & 1 & 0 & 10 & 0.5 \\ 0 & 0 & 1 & 4 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{A.28}$$

The null space is thus:

$$\begin{bmatrix} 6 & 1 \\ -10 & -0.5 \\ -4 & -2 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{A.29}$$

A.1.4 Eigenvalues, eigenvectors

Given a square matrix $\underline{\underline{\mathbf{A}}}$, almost all vectors change direction when multiplied with $\underline{\underline{\mathbf{A}}}$ except then for the eigenvectors. They point in the same direction. The basis equation for the eigenvalues is:

$$\underline{\underline{\mathbf{A}}}\,\underline{\mathbf{x}} = \lambda\,\underline{\mathbf{x}} \tag{A.30}$$

where λ is an eigenvalue of $\underline{\underline{\mathbf{A}}}$. Solving the equation

$$\det\left(\underline{\underline{\mathbf{A}}} - \boldsymbol{\lambda}\,\underline{\underline{\mathbf{I}}}\right) = 0 \tag{A.31}$$

The eigenvectors are found by solving for each eigenvalue the basis equation:

$$\left(\underline{\underline{\mathbf{A}}} - \boldsymbol{\lambda}_i\,\underline{\underline{\mathbf{I}}}\right)\,\underline{\mathbf{v}}_i = 0 \tag{A.32}$$

The matrix with the columns being the eigenvectors is called the eigenvector matrix:

$$\underline{\underline{\mathbf{V}}} := [\underline{\mathbf{v}}_1, \underline{\mathbf{v}}_2, \ldots, \underline{\mathbf{v}}_n]_{\forall i} \tag{A.33}$$

A.1.5 Matrix diagonalisation

If the matrix $\underline{\underline{\mathbf{A}}}$ is square of dimension $n \times n$ and has n distinct eigenvalues, then the Eigenvector matrix is invertible.

$$\underline{\underline{\mathbf{V}}}^{-1}\,\underline{\underline{\mathbf{A}}}\,\underline{\underline{\mathbf{V}}} = \boldsymbol{\Lambda} := \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \tag{A.34}$$

This can be proven easily considering that $\underline{\underline{\mathbf{A}}}\,\underline{\underline{\mathbf{V}}} = \underline{\underline{\mathbf{V}}}\,\mathbf{\Lambda}$ (Strang (2009)):

$$\underline{\underline{\mathbf{A}}}\,\underline{\underline{\mathbf{V}}} =: \underline{\underline{\mathbf{A}}}\begin{bmatrix}\underline{\mathbf{x}}_1 & \dots\underline{\mathbf{x}}_n\end{bmatrix} := \begin{bmatrix}\lambda_1\,\underline{\mathbf{x}}_1 & \dots & \lambda_n\,\underline{\mathbf{x}}_n\end{bmatrix}$$

which in turn is:

$$\begin{bmatrix}\lambda_1\,\underline{\mathbf{x}}_1 & \dots & \lambda_n\,\underline{\mathbf{x}}_n\end{bmatrix} =: \begin{bmatrix}\underline{\mathbf{x}}_1 & \dots\underline{\mathbf{x}}_n\end{bmatrix}\begin{bmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_n\end{bmatrix} := \underline{\underline{\mathbf{V}}}\,\mathbf{\Lambda}$$

A.1.5.1 Multiple eigenvalues

If the matrix has multiple eigenvalues, the diagonalisation is not possible. However other functions can be computed, such as the power of the matrix or, which is essential for the solution of linear ordinary differential equations, the exponential. There exists a canonical representation called the Jordan form, in which multiple eigenvalues along the diagonal form Jordan blocks, which for an arbitrary eigenvalue λ_i have the form:

$$\underline{\underline{\mathbf{J}}}_1 := \begin{bmatrix}\lambda_1 & 1 & 0 & \dots & 0 \\ 0 & \lambda_1 & 1 & \ddots & \vdots \\ \vdots & 0 & \ddots & \ddots & 0 \\ \vdots & & \ddots & \ddots & 1 \\ 0 & 0 & \dots & 0 & \lambda_1\end{bmatrix} \tag{A.35}$$

A.1.5.2 Similar matrices

Simularity transformations are essential in system theory, as they are used to transform systems into alternative state spaces, which for the purpose are more convinient or uncover essential properties. The similarity transformations are defined by:

$$\underline{\underline{\mathbf{M}}} := \underline{\underline{\mathbf{S}}}^{-1}\,\underline{\underline{\mathbf{AS}}}$$

The tranformation matrix $\underline{\underline{\mathbf{M}}}$ may be any square *invertible* matrix. Quadaratic matrices are similar if (Strang (2009)):

they have the ***same****:*	***modified*** $\underline{\underline{\mathbf{S}}}$ *modifies:*
eigenvalues	*eigenvectors*
trace	*nullspace*
determinant	*column space*
rank	*row space*
number of independent eigenvectors	*singular values*
Jordan form	

A.1.6 Matrix differentiation

Let $\underline{\mathbf{y}}$ be a vector function of $\underline{\mathbf{x}}$: $\underline{\mathbf{y}}(\underline{\mathbf{x}})$ and $\underline{\underline{\mathbf{Q}}}$ be a symmetrical matrix, then the scalar quadratic form is:

$$\begin{aligned}
\frac{d\left(\underline{\mathbf{y}}^T(\underline{\mathbf{x}})\,\underline{\underline{\mathbf{Q}}}\,\underline{\mathbf{y}}(\underline{\mathbf{x}})\right)}{d\,\underline{\mathbf{x}}^T} &= \frac{d\,\underline{\mathbf{y}}^T(\underline{\mathbf{x}})}{d\,\underline{\mathbf{x}}}\,\underline{\underline{\mathbf{Q}}}\,\underline{\mathbf{y}}(\underline{\mathbf{x}}) + \underline{\mathbf{y}}^T(\underline{\mathbf{x}})\,\underline{\underline{\mathbf{Q}}}\,\frac{d\,\underline{\mathbf{y}}(\underline{\mathbf{x}})}{d\,\underline{\mathbf{x}}^T} && \text{(A.36)}\\
&= \frac{d\,\underline{\mathbf{y}}^T(\underline{\mathbf{x}})}{d\,\underline{\mathbf{x}}}\,\underline{\underline{\mathbf{Q}}}\,\underline{\mathbf{y}}(\underline{\mathbf{x}}) + \left(\underline{\mathbf{y}}^T(\underline{\mathbf{x}})\,\underline{\underline{\mathbf{Q}}}\,\frac{d\,\underline{\mathbf{y}}(\underline{\mathbf{x}})}{d\,\underline{\mathbf{x}}^T}\right)^T && \text{(A.37)}\\
&= \frac{d\,\underline{\mathbf{y}}^T(\underline{\mathbf{x}})}{d\,\underline{\mathbf{x}}}\,\underline{\underline{\mathbf{Q}}}\,\underline{\mathbf{y}}(\underline{\mathbf{x}}) + \frac{d\,\underline{\mathbf{y}}^T(\underline{\mathbf{x}})}{d\,\underline{\mathbf{x}}}\,\underline{\underline{\mathbf{Q}}}^T\,\underline{\mathbf{y}}(\underline{\mathbf{x}}) && \text{(A.38)}\\
&= 2\,\frac{d\,\underline{\mathbf{y}}^T(\underline{\mathbf{x}})}{d\,\underline{\mathbf{x}}}\,\underline{\underline{\mathbf{Q}}}\,\underline{\mathbf{y}}(\underline{\mathbf{x}}) && \text{(A.39)}
\end{aligned}$$

B

Calculus

B.1 Leibnitz' Rule

$$\begin{aligned}\frac{\partial}{\partial x}\int_{f(x)}^{g(x)} dx'\, F(x,x') := \quad & F(x,g(x))\frac{\partial g(x)}{\partial x} - F(x,f(x))\frac{\partial f(x)}{\partial x} \\ + \quad & \int_{f(x)}^{g(x)} dx'\,\frac{\partial}{\partial x}F(x,x')\,. \qquad \text{(B.1)}\end{aligned}$$

B.2 Euler's Theorem on Homogeneous Functions

Euler's Theorem of Homogeneous Functions 1. *Let* $f(x_1,\ldots,x_k)$ *be a function such that*

$$f(\lambda\, x_1,\ldots,\lambda\, x_k) := \lambda^n\, f(x_1,\ldots,x_k)\,. \qquad \text{(B.2)}$$

then f *is said to be a* homogeneous function of degree n *for which*

$$n\, f(x_1,\ldots,x_k) := \sum_{i:=1}^{k}\frac{\partial\, f(x_1,\ldots,x_k)}{\partial\, x_i}\, x_i\,. \qquad \text{(B.3)}$$

Proof. Differentiation of the homogeneous condition with respect to λ gives

$$\begin{aligned}\frac{d}{d\,\lambda}\, f(\lambda\, x_1,\ldots,\lambda\, x_k) &:= \frac{d}{d\,\lambda}\,\lambda^n\, f(x_1,\ldots,x_k)\,, \\ \sum_{i:=1}^{k}\frac{\partial\, f(\lambda\, x_1,\ldots,\lambda\, x_k)}{\partial\,\lambda\, x_i}\, x_i &:= n\,\lambda^{n-1}\,\lambda^n\, f(x_1,\ldots,x_k)\,.\end{aligned}$$

setting $\lambda = 1$, one obtains :

$$\sum_{i:=1}^{k}\frac{\partial\, f(x_1,\ldots,x_k)}{\partial\,\lambda\, x_i}\, x_i := n\,\lambda^n\, f(x_1,\ldots,x_k)\,.$$

□

B.3 Legendre Transformation Generating New Extensive Properties

Let Φ_1 be a vetorial arbitrary extensive quantity, which is a function of two vectors of extensive quantities Φ_a and Φ_b:

$$\Phi_1 := \Phi_1(\Phi_a, \Phi_b)\,. \tag{B.4}$$

A new extensive quantity Φ_2 is defined:

$$\begin{aligned} \Phi_2 &:= \Phi_2(\varphi, \Phi_b)\,, && \text{(B.5)} \\ &:= \Phi_1 - \varphi\, \Phi_a\,. && \text{(B.6)} \end{aligned}$$

The last equation can be interpreted as a tangent plane to the original function with slopes in the different directions, collected in the Jacobian:

$$\varphi := \frac{\partial\, \Phi_1}{\partial\, \Phi_a^T}\,,$$

which take the role of the new variables.
Differentiating the two equations (B.4) and (B.6) one finds:

$$\begin{aligned} d\Phi_1 &:= \frac{\partial\, \Phi_1}{\partial\, \Phi_a}\, d\Phi_a + \frac{\partial\, \Phi_1}{\partial\, \Phi_b}\, d\Phi_b\,, \\ d\Phi_2 &:= d\Phi_1 - \Phi_a^T\, d\varphi^T - \varphi\, d\Phi_a\,. \end{aligned}$$

Elimination of $d\Phi_1$ gives:

$$d\Phi_2 := \frac{\partial\, \Phi_1}{\partial\, \Phi_b}\, d\Phi_b - \Phi_a^T\, d\varphi^T\,.$$

Applying the Legendre transformation to extensive quantities introduces intensive properties, collected in the Jacobian. The transformation can also be inverted in which case the respective roles are exchanged.

B.3.1 Examples

The Legendre transformations are basic to thermodynamics. For example, let

$$\begin{aligned} \Phi_1 &:= \Phi_1 := U(S, V, \underline{\mathbf{n}})\,, && \text{(B.7)} \\ \Phi_2 &:= \Phi_2 := H(T, V, \underline{\mathbf{n}})\,. && \text{(B.8)} \end{aligned}$$

and

$$\Phi_a := \Phi_a := S\,, \tag{B.9}$$
$$\Phi_b := [V, \underline{\mathbf{n}}^T]^T\,. \tag{B.10}$$

Thus

$$\frac{\partial\,\Phi_1}{\partial\,\Phi_b} := \frac{\partial\,U(S, V, \underline{\mathbf{n}})}{\partial\,[V, \underline{\mathbf{n}}^T]}\,, \tag{B.11}$$
$$:= \left[p, \frac{\partial\,U(S, V, \underline{\mathbf{n}})}{\partial\,\underline{\mathbf{n}}^T}\right]\,, \tag{B.12}$$
$$\varphi := \frac{\partial\,U(S, V, \underline{\mathbf{n}})}{\partial\,S}\,, \tag{B.13}$$
$$:= T\,. \tag{B.14}$$

C

Ordinary Differential Equations

C.1 Linear Ordinary Differential Equations

Linear ordinary differential (l-ODEs) are given by the equation:

$$\dot{\underline{\mathbf{x}}} = \underline{\underline{\mathbf{A}}}\,\underline{\mathbf{x}} + \underline{\underline{\mathbf{B}}}\,\underline{\mathbf{u}}$$

with

$\underline{\mathbf{x}}$	::	state vector	$\in \mathbb{R}^{n}$
$\underline{\mathbf{u}}$	::	input vector	$\in \mathbb{R}^{m}$
$\underline{\underline{\mathbf{A}}}$	::	system matrix	$\in \mathbb{R}^{n\times n}$
$\underline{\underline{\mathbf{B}}}$	::	input matrix	$\in \mathbb{R}^{n\times m}$

C.1.1 Similarity transformation

For the case of the matrix $\underline{\underline{\mathbf{A}}}$ having n distinct eigenvectors and eigenvalues, then we can quite easily find the solution by introducing a similarity transformation, by which we map into another state. For the transformation we introduce a transformation matrix $\underline{\underline{\mathbf{T}}}$:

$$\underline{\mathbf{z}} := \underline{\underline{\mathbf{T}}}\,\underline{\mathbf{x}}$$

then with $\underline{\underline{\mathbf{T}}}$ being invertible:

$$\underline{\mathbf{x}} := \underline{\underline{\mathbf{T}}}^{-1}\,\underline{\mathbf{z}}$$

and

$$\dot{\underline{\mathbf{x}}} := \underline{\underline{\mathbf{T}}}^{-1}\,\dot{\underline{\mathbf{z}}}$$

Then

$$\begin{aligned}\underline{\underline{\mathbf{T}}}^{-1}\,\dot{\underline{\mathbf{z}}} &= \underline{\underline{\mathbf{A}}}\,\underline{\underline{\mathbf{T}}}^{-1}\,\underline{\mathbf{z}} + \underline{\underline{\mathbf{B}}}\,\underline{\mathbf{u}}\,,\\ \dot{\underline{\mathbf{z}}} &= \underline{\underline{\mathbf{T}}}\,\underline{\underline{\mathbf{A}}}\,\underline{\underline{\mathbf{T}}}^{-1}\,\underline{\mathbf{z}} + \underline{\underline{\mathbf{T}}}\,\underline{\underline{\mathbf{B}}}\,\underline{\mathbf{u}}\,.\end{aligned}$$

Setting the transformation matrix $\underline{\underline{\mathbf{T}}} = \underline{\underline{\mathbf{V}}}^{-1}$ with the latter being the eigenvector matrix, we observe that:

$$\underline{\underline{\mathbf{V}}}^{-1}\,\underline{\underline{\mathbf{A}}}\,\underline{\underline{\mathbf{V}}} = \mathbf{\Lambda}$$

Thus

$$\underline{\dot{\mathbf{z}}} = \mathbf{\Lambda}\,\underline{\mathbf{z}} + \underline{\underline{\mathbf{T}}}\,\underline{\underline{\mathbf{B}}}\,\underline{\mathbf{u}}\,,$$

the equations are decoupled and each differential equation can be solved independently:

$$\dot{z}_i = \lambda_i\, z_i + \underline{\mathbf{s}}_i^T\,\underline{\underline{\mathbf{T}}}\,\underline{\underline{\mathbf{B}}}\,\underline{\mathbf{u}}\,.$$

where the row vector $\underline{\mathbf{s}}_i^T$ selects the i^{th} row. The solution of this equation is:

$$z_i(t) := e^{\lambda_i\, t}\, z_i(0) + \int_0^t e^{\lambda_i\,(t-\tau)}\,\underline{\mathbf{s}}_i^T\,\underline{\underline{\mathbf{T}}}\,\underline{\underline{\mathbf{B}}}\,\underline{\mathbf{u}}(\tau)\,d\tau$$

Stacking it up yields the solution to the matrix equation:

$$\underline{\mathbf{z}}(t) := \underline{\underline{\mathbf{e}}}^{\mathbf{\Lambda}\,t}\,\underline{\mathbf{z}}(0) + \int_0^t \underline{\underline{\mathbf{e}}}^{\mathbf{\Lambda}\,(t-\tau)}\,\underline{\underline{\mathbf{T}}}\,\underline{\underline{\mathbf{B}}}\,\underline{\mathbf{u}}(\tau)\,d\tau$$

where:

$$\underline{\underline{\mathbf{e}}}^{\mathbf{\Lambda}\,t} := \begin{bmatrix} e^{\lambda_1\,t} & & \\ & \ddots & \\ & & e^{\lambda_n\,t} \end{bmatrix}$$

Transforming back:

$$\begin{aligned}
\underline{\underline{\mathbf{T}}}\,\underline{\mathbf{x}}(t) &:= \underline{\underline{\mathbf{e}}}^{\mathbf{\Lambda}\,t}\,\underline{\underline{\mathbf{T}}}\,\underline{\mathbf{x}}(0) + \int_0^t \underline{\underline{\mathbf{e}}}^{\mathbf{\Lambda}\,(t-\tau)}\,\underline{\underline{\mathbf{T}}}\,\underline{\underline{\mathbf{B}}}\,\underline{\mathbf{u}}(\tau)\,d\tau \qquad \text{(C.1)}\\
\underline{\underline{\mathbf{V}}}^{-1}\,\underline{\mathbf{x}}(t) &:= \underline{\underline{\mathbf{e}}}^{\mathbf{\Lambda}\,t}\,\underline{\underline{\mathbf{V}}}^{-1}\,\underline{\mathbf{x}}(0) + \int_0^t \underline{\underline{\mathbf{e}}}^{\mathbf{\Lambda}\,(t-\tau)}\,\underline{\underline{\mathbf{V}}}^{-1}\,\underline{\underline{\mathbf{B}}}\,\underline{\mathbf{u}}(\tau)\,d\tau\\
\underline{\mathbf{x}}(t) &:= \underline{\underline{\mathbf{V}}}\,\underline{\underline{\mathbf{e}}}^{\mathbf{\Lambda}\,t}\,\underline{\underline{\mathbf{V}}}^{-1}\,\underline{\mathbf{x}}(0) + \int_0^t \underline{\underline{\mathbf{V}}}\,\underline{\underline{\mathbf{e}}}^{\mathbf{\Lambda}\,(t-\tau)}\,\underline{\underline{\mathbf{V}}}^{-1}\,\underline{\underline{\mathbf{B}}}\,\underline{\mathbf{u}}(\tau)\,d\tau\\
\underline{\mathbf{x}}(t) &:= \underline{\underline{\mathbf{e}}}^{\underline{\underline{\mathbf{A}}}\,t}\,\underline{\mathbf{x}}(0) + \int_0^t \underline{\underline{\mathbf{e}}}^{\underline{\underline{\mathbf{A}}}\,(t-\tau)}\,\underline{\underline{\mathbf{B}}}\,\underline{\mathbf{u}}(\tau)\,d\tau\\
\underline{\mathbf{x}}(t) &:= \underline{\underline{\mathbf{e}}}^{\underline{\underline{\mathbf{A}}}\,t}\,\underline{\mathbf{x}}(0) + \int_0^t \underline{\underline{\mathbf{e}}}^{\underline{\underline{\mathbf{A}}}\,\tau'}\,\underline{\underline{\mathbf{B}}}\,\underline{\mathbf{u}}(t-\tau')\,d\tau'
\end{aligned}$$

which is the general solution for linear, time-constant ODEs, in short also called LTI systems. These systems are widely used, as their solution is readily available through the solution of the eigenvalue problem.

C.1.2 Alternative derivation of the solution in the time & Laplace Domain

The first part of this section discusses the solution of the most fundamental class of linear systems, namely linear systems that are described by sets of ordinary, time-constant differential equations. Two solution methods are introduced. The first is the general solution in the time domain. The second is derived in the Laplace domain.

C.1.3 A Time Domain Approach

Let $y := f(u)$ be a scalar, linear functional that represents our system in an input/output form, that is the output is a function $f(\cdot)$ of the input u. The solution for this linear, time-invariant dynamic model given the initial conditions $x^0 = 0$ is obtained in two steps.

First, the input function $u(t)$ is approximated by a series of pulses (Figure C.1):

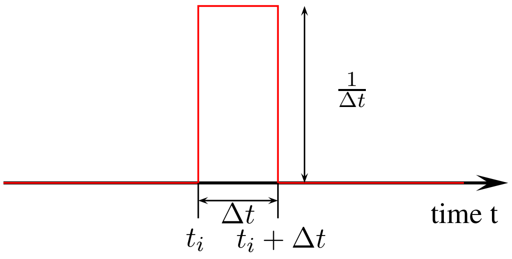


Figure C.1: Pulse function

$$u(-\infty, \infty) = \sum_{i=-\infty}^{\infty} u(t_i)\, \delta_\Delta(t - t_i)\, \Delta t$$

where the pulse function $\delta_\Delta(t - t_i)$ is defined by

$$\delta_\Delta(t - t_i) \;:= \begin{cases} 0 & ; (-\infty < t < t_i) \\ 1/\Delta t & ; (t_i \le t \le t_i + \Delta t) \\ 0 & ; (t_i + \Delta t \le t < \infty) \end{cases}$$

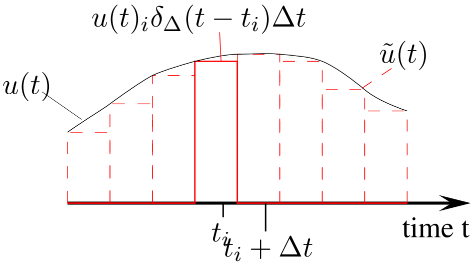


Figure C.2: *Approximation of input function with a series of connected pulses*

$$
\begin{aligned}
y(k) = f(u(-\infty, t_k)) &= f\left(\sum_{i=-\infty}^{k-1} u(t_i)\,\delta_\Delta(t-t_i)\,\Delta t\right)\\
&= \sum_{i=-\infty}^{k-1} u(t_i)\,f\left(\delta_\Delta(t-t_i)\right)\,\Delta t
\end{aligned}
$$

where we took the fact into consideration that $u(t_i)$ is a value and not a function.

Secondly, the transition is made to infinitely narrow pulses. Thus let Δt approach dt

$$
y(t) = \lim_{\Delta t \to dt} \sum_{i=-\infty}^{k-1} u(t_i) f(\delta_\Delta(t-t_i))\Delta t\,,
$$

to obtain the convolution integral:

$$
y(t) = \int_{-\infty}^{t} f(\delta(t-\tau))\,u(\tau)\,d\tau\,, \tag{C.2}
$$

where $\delta(t-\tau)$ the Dirac impulse function :

$$
\delta(t-t_1) = \lim_{\Delta t \to 0} \delta_\Delta(t-t_1)
$$

with the properties

$$
\begin{aligned}
\int_{-\infty}^{\infty} \delta(t-t_1)\,dt &= 1\\
\int_{-\infty}^{\infty} f(t)\,\delta(t-t_1)\,dt &= f(t_1)
\end{aligned}
$$

The result C.2 is the convolution integral, which may be abbreviated by writing :

$$f(\delta(t-\tau)) \star u(\tau)$$

Note that the function $f(\cdot)$ is shifted over the input function $u(\cdot)$ in opposite direction. In the above representation, the function $f(\cdot)$ ”starts“ thereby at the lower limit, whilst the function $u(\cdot)$ ”starts“ at the upper limit. The changing integration variable moves the two functions across each other, so-to-speak. This operation is called convolution.

Defining the impulse response

$$\tilde{g}(t,\tau) := f(\delta(t-\tau))\,,$$

which is equal to $y(t)$ if $u(t) \;:= \delta(t-\tau)$ then

$$y(t) := \int_{-\infty}^{t} \tilde{g}(t,\tau)\, u(\tau)\, d\tau\,.$$

The extension to the multi-variable case is straightforward and results

$$\underline{\mathbf{y}} = \int_{-\infty}^{t} \underline{\underline{\tilde{\mathbf{G}}}}(t,\tau)\, \underline{\mathbf{u}}(\tau)\, d\tau \tag{C.3}$$

$$\begin{array}{lll} \underline{\underline{\tilde{\mathbf{G}}}}(t,\tau) :: & \text{the impulse response matrix} & \\ := & [[\tilde{g}_{i,j}]]_{\forall i,\forall j} & \in \mathbb{R}^{dim(\underline{\mathbf{y}})\times dim(\underline{\mathbf{u}})} \end{array}$$

The impulse response matrix contains the information about the dynamics of the process. The next section will also provide a solution to computing the output given the model and the input. It will also shed some light on the relation between the impulse response matrix and the differential equation model.

C.1.4 General Solution in the Laplace and Time Domain

Through the linearisation of non-linear process models, a linear state-space description is obtained. It describes the propagation of the state in a set of differential equations, one for each state, and a set of algebraic equations linearly combining the inputs and the states to result the outputs:

$$\begin{aligned}\dot{\underline{\mathbf{x}}} &= \underline{\underline{\mathbf{A}}}\,\underline{\mathbf{x}}(t) + \underline{\underline{\mathbf{B}}}\,\underline{\mathbf{u}}(t)\\ \underline{\mathbf{y}} &:= \underline{\underline{\mathbf{C}}}\,\underline{\mathbf{x}}(t) + \underline{\underline{\mathbf{D}}}\,\underline{\mathbf{u}}(t)\end{aligned}$$

In this section, a solution is constructed by first solving the equations in the Laplace domain. Since the Laplace transformation converts the differential equations into algebraic equations, the solution is found readily:

$$\begin{aligned}s\underline{\mathbf{x}}(s) - \underline{\mathbf{x}}(t=0) &= \underline{\underline{\mathbf{A}}}\,\underline{\mathbf{x}}(s) + \underline{\underline{\mathbf{B}}}\,\underline{\mathbf{u}}(s)\\ \underline{\mathbf{y}}(s) &= \underline{\underline{\mathbf{C}}}\,\underline{\mathbf{x}}(s) + \underline{\underline{\mathbf{D}}}\,\underline{\mathbf{u}}(s)\\ (\underline{\underline{\mathbf{I}}}\,s - \underline{\underline{\mathbf{A}}})\,\underline{\mathbf{x}}(s) &= \underline{\mathbf{x}}(t=0) + \underline{\underline{\mathbf{B}}}\,\underline{\mathbf{u}}(s)\\ \underline{\mathbf{x}}(s) &= (\underline{\underline{\mathbf{I}}}\,s - \underline{\underline{\mathbf{A}}})^{-1}\,\big(\underline{\mathbf{x}}(t=0) + \underline{\underline{\mathbf{B}}}\,\underline{\mathbf{u}}(s)\big)\end{aligned}$$

The final result takes the form:

$$\underline{\mathbf{y}}(s) = \underline{\underline{\mathbf{C}}}\,(\underline{\underline{\mathbf{I}}}\,s - \underline{\underline{\mathbf{A}}})^{-1}\,\big(\underline{\mathbf{x}}(t=0) + \underline{\underline{\mathbf{B}}}\,\underline{\mathbf{u}}(s)\big) + \underline{\underline{\mathbf{D}}}\,\underline{\mathbf{u}}(s)\,. \tag{C.4}$$

This approach requires the inversion of a matrix, namely $(\underline{\underline{\mathbf{I}}}s - \underline{\underline{\mathbf{A}}})$. The solution in the time domain is now constructed by transforming the result into the time domain using the inverse Laplace transformation. The result for the first term is:

$$(\underline{\underline{\mathbf{I}}}s - \underline{\underline{\mathbf{A}}})^{-1}\,\underline{\mathbf{x}}(t=0) \xrightarrow{\mathcal{L}^{-1}} \underline{\underline{\mathbf{e}}}^{\underline{\underline{\mathbf{A}}}t}\,\underline{\mathbf{x}}(t=0)$$

And the second term is the convolution integral :

$$(\underline{\underline{\mathbf{I}}}\,s - \underline{\underline{\mathbf{A}}})^{-1}\,\underline{\underline{\mathbf{B}}}\,\underline{\mathbf{u}}(s) \xrightarrow{\mathcal{L}^{-1}} \int_{t=0}^{t} \underline{\underline{\mathbf{e}}}^{\underline{\underline{\mathbf{A}}}(t-\tau)}\,\underline{\underline{\mathbf{B}}}\,\underline{\mathbf{u}}(\tau)\,d\tau$$

The complete solution for a time-constant, linear set of ODEs in the time domain is therefore

$$\underline{\mathbf{x}}(t) = \underline{\underline{\mathbf{\Phi}}}(t)\,\underline{\mathbf{x}}(t=0) + \int_{t=0}^{t} \underline{\underline{\mathbf{\Phi}}}(t-\tau)\,\underline{\underline{\mathbf{B}}}\,\underline{\mathbf{u}}(\tau)\,d\tau \tag{C.5}$$

and

$$\underline{\mathbf{y}}(t) = \underline{\underline{\mathbf{C}}}\,\underline{\underline{\mathbf{\Phi}}}(t)\,\underline{\mathbf{x}}(t=0) + \underline{\underline{\mathbf{C}}}\int_{t=0}^{t} \underline{\underline{\mathbf{\Phi}}}(t-\tau)\,\underline{\underline{\mathbf{B}}}\,\underline{\mathbf{u}}(\tau)\,d\tau + \underline{\underline{\mathbf{D}}}\,\underline{\mathbf{u}}(t) \tag{C.6}$$

with

$$\begin{aligned} \underline{\underline{\mathbf{\Phi}}}(t) \quad &:: \quad \text{fundamental or transition matrix} \\ &:= \quad \underline{\underline{\mathbf{e}}}^{\underline{\underline{\mathbf{A}}}t} \end{aligned}$$

For zero initial conditions $\underline{\mathbf{x}}(t=0) := \underline{\mathbf{0}}$ and $\underline{\underline{\mathbf{D}}} := \underline{\underline{\mathbf{0}}}$ this simplifies to

$$\begin{aligned} \underline{\mathbf{x}}(t) &= \int_{t=0}^{t} \underline{\underline{\mathbf{\Phi}}}(t-\tau)\, \underline{\underline{\mathbf{B}}}\, \underline{\mathbf{u}}(\tau)\, d\tau\,, \\ \underline{\mathbf{y}}(t) &:= \underline{\underline{\mathbf{C}}}\, \underline{\mathbf{x}}(t)\,, \\ &:= \qquad \underline{\underline{\mathbf{C}}} \int_{t=0}^{t} \underline{\underline{\mathbf{\Phi}}}(t-\tau)\, \underline{\underline{\mathbf{B}}}\, \underline{\mathbf{u}}(\tau)\, d\tau\,, \end{aligned}$$

which compares with the result derived in the last section. The relation between the impulse response matrix and the quantities that appear in the differential model is:

$$\underline{\underline{\tilde{\mathbf{G}}}}(t,\tau) \ := \underline{\underline{\mathbf{C}}}\, \underline{\underline{\mathbf{\Phi}}}(t-\tau)\, \underline{\underline{\mathbf{B}}}.$$

As one would expect, the impulse transfer matrix contains information about the dynamics, packed into $\underline{\underline{\mathbf{\Phi}}}$, the input amplification $\underline{\underline{\mathbf{B}}}$, and the link between the state and the output, which is $\underline{\underline{\mathbf{C}}}$.

C.1.5 Stability

The simplest way to derive the stability criterion is to look at the solution in the time domain **??**. The first term is the solution of the autonomous system. If the system is stable, then it should return to its natural steady state without any input after it has been disturbed.

The term

$$\underline{\underline{\mathbf{e}}}^{\underline{\underline{\mathbf{A}}}\,t}\, \underline{\mathbf{x}}(0) := \underline{\underline{\mathbf{V}}}\, \underline{\underline{\mathbf{e}}}^{\mathbf{\Lambda}\,t}\, \underline{\underline{\mathbf{V}}}^{-1}\, \underline{\mathbf{x}}(0)$$

The eigenvector matrix determines the direction, whilst the eigenvalue matrix gives the magnitude of the movement. With imaginary part of the eigenvalue will induce oscillations, the real part of the eigenvalue is determining the magnitude of the change with time. If the real part is positive, the respective exponential function will grow without limit and thus the system is unstable.

The stability criterion is thus:

> **Definition – Stable LTI:** An LTI system is stable iff $\mathbb{R}\left(\lambda_i(\underline{\underline{\mathbf{A}}})\right) < 0\, \forall i$

C.1.6 Steady state

The steady state of a stable system is characterised by no change, thus the derivative of the state with respect to time is zero if the input is constant:

$$\dot{\underline{\mathbf{x}}}\big|_{t:=\infty} = \underline{\underline{\mathbf{A}}}\ \underline{\mathbf{x}}\big|_{t:=\infty} + \underline{\underline{\mathbf{B}}}\,\underline{\mathbf{u}}^0 := 0$$

and thus if the system matrix is invertible:

$$\underline{\mathbf{x}}\big|_{t:=\infty} := -\underline{\underline{\mathbf{A}}}^{-1}\,\underline{\underline{\mathbf{B}}}\,\underline{\mathbf{u}}^0$$

C.2 Nonlinear differential equations

$$\begin{aligned}
\dot{\underline{\mathbf{x}}} &:= \underline{\mathbf{f}}\left(\underline{\mathbf{x}}, \underline{\mathbf{u}}\right) \\
\underline{\mathbf{y}} &:= \underline{\mathbf{g}}\left(\underline{\mathbf{x}}, \underline{\mathbf{u}}\right)
\end{aligned}$$

C.2.1 Properties

- Stiffness: large differences in the eigenvalues of the Jacobian $\left(\frac{\partial\,\underline{\mathbf{f}}(\underline{\mathbf{x}},\underline{\mathbf{u}})}{\partial\,\underline{\mathbf{x}}^T}\right)$. Consider
  - scaling
  - singular perturbation
  - specialised solvers (integrators) mostly implicit solvers that adjust well to the local topology

C.2.2 Similarity transformation

Again, let:

$$\underline{\mathbf{z}} := \underline{\underline{\mathbf{T}}}\,\underline{\mathbf{x}}$$

then with $\underline{\underline{\mathbf{T}}}$ being invertible:

$$\underline{\mathbf{x}} := \underline{\underline{\mathbf{T}}}^{-1}\,\underline{\mathbf{z}}$$

and

$$\dot{\underline{\mathbf{x}}} := \underline{\underline{\mathbf{T}}}^{-1}\,\dot{\underline{\mathbf{z}}}$$

as before. Substitution into the nonlinear model results:

$$\begin{aligned}\dot{\underline{\mathbf{x}}} := \underline{\underline{\mathbf{T}}}^{-1}\,\dot{\underline{\mathbf{z}}} := \underline{\mathbf{f}}\left(\underline{\underline{\mathbf{T}}}^{-1}\underline{\mathbf{x}}, \underline{\mathbf{u}}\right)& \\ \dot{\underline{\mathbf{z}}} &:= \underline{\underline{\mathbf{T}}}\,\underline{\mathbf{f}}\left(\underline{\underline{\mathbf{T}}}^{-1}\underline{\mathbf{x}}, \underline{\mathbf{u}}\right) := \overline{\underline{\mathbf{f}}}\left(\underline{\mathbf{z}}, \underline{\mathbf{u}}\right) \\ \underline{\mathbf{y}} &:= \underline{\mathbf{g}}\left(\underline{\underline{\mathbf{T}}}^{-1}\underline{\mathbf{x}}, \underline{\mathbf{u}}\right) := \overline{\underline{\mathbf{g}}}\left(\underline{\mathbf{z}}, \underline{\mathbf{u}}\right)\end{aligned}$$

Again the input - output behaviour is not affected by the transformation. The requirements obviously are that the transformation must be invertible. It to be a linear transformation is not a requirement, though it is a common case.

C.2.3 Linearisation

Use truncated Taylor expansion to approximate the two functions $\underline{\mathbf{f}}, \underline{\mathbf{g}}$ around a point $\underline{\mathbf{x}}^*, \underline{\mathbf{u}}^*$ in the state-input space and the definition of deviation variables:

$$\begin{aligned}\underline{\mathbf{f}}\left(\underline{\mathbf{x}}, \underline{\mathbf{u}}\right) &\approx \underline{\mathbf{f}}\left(\underline{\mathbf{x}}^*, \underline{\mathbf{u}}^*\right) + \left(\frac{\partial\,\underline{\mathbf{f}}\left(\underline{\mathbf{x}}, \underline{\mathbf{u}}\right)}{\partial\,\underline{\mathbf{x}}^T}\right)_{\underline{\mathbf{x}}^*, \underline{\mathbf{u}}^*} \Delta\underline{\mathbf{x}} + \left(\frac{\partial\,\underline{\mathbf{f}}\left(\underline{\mathbf{x}}, \underline{\mathbf{u}}\right)}{\partial\,\underline{\mathbf{u}}^T}\right)_{\underline{\mathbf{x}}^*, \underline{\mathbf{u}}^*} \Delta\underline{\mathbf{u}} \\ \underline{\mathbf{g}}\left(\underline{\mathbf{x}}, \underline{\mathbf{u}}\right) &\approx \underline{\mathbf{g}}\left(\underline{\mathbf{x}}^*, \underline{\mathbf{u}}^*\right) + \left(\frac{\partial\,\underline{\mathbf{g}}\left(\underline{\mathbf{x}}, \underline{\mathbf{u}}\right)}{\partial\,\underline{\mathbf{x}}^T}\right)_{\underline{\mathbf{x}}^*, \underline{\mathbf{u}}^*} \Delta\underline{\mathbf{x}} + \left(\frac{\partial\,\underline{\mathbf{g}}\left(\underline{\mathbf{x}}, \underline{\mathbf{u}}\right)}{\partial\,\underline{\mathbf{u}}^T}\right)_{\underline{\mathbf{x}}^*, \underline{\mathbf{u}}^*} \Delta\underline{\mathbf{u}}\end{aligned}$$

Deviation variables:

$$\begin{aligned}\underline{\mathbf{z}} &:= \Delta\underline{\mathbf{x}} := \underline{\mathbf{x}} - \underline{\mathbf{x}}^* \\ \underline{\mathbf{v}} &:= \Delta\underline{\mathbf{u}} := \underline{\mathbf{u}} - \underline{\mathbf{u}}^* \\ \underline{\mathbf{w}} &:= \Delta\underline{\mathbf{y}} := \underline{\mathbf{y}} - \underline{\mathbf{y}}^*\end{aligned}$$

Thus:

$$\dot{\underline{\mathbf{x}}} := \dot{\underline{\mathbf{x}}}^* - \Delta\dot{\underline{\mathbf{x}}}$$

Resulting in:

$$\begin{aligned}\dot{\underline{\mathbf{z}}} &= \left(\frac{\partial\,\underline{\mathbf{f}}\left(\underline{\mathbf{x}}, \underline{\mathbf{u}}\right)}{\partial\,\underline{\mathbf{x}}^T}\right)_{\underline{\mathbf{x}}^*, \underline{\mathbf{u}}^*} \underline{\mathbf{z}} + \left(\frac{\partial\,\underline{\mathbf{f}}\left(\underline{\mathbf{x}}, \underline{\mathbf{u}}\right)}{\partial\,\underline{\mathbf{u}}^T}\right)_{\underline{\mathbf{x}}^*, \underline{\mathbf{u}}^*} \underline{\mathbf{v}} \\ \underline{\mathbf{w}} &= \left(\frac{\partial\,\underline{\mathbf{g}}\left(\underline{\mathbf{x}}, \underline{\mathbf{u}}\right)}{\partial\,\underline{\mathbf{x}}^T}\right)_{\underline{\mathbf{x}}^*, \underline{\mathbf{u}}^*} \underline{\mathbf{z}} + \left(\frac{\partial\,\underline{\mathbf{g}}\left(\underline{\mathbf{x}}, \underline{\mathbf{u}}\right)}{\partial\,\underline{\mathbf{u}}^T}\right)_{\underline{\mathbf{x}}^*, \underline{\mathbf{u}}^*} \underline{\mathbf{v}}\end{aligned}$$

C.3 Index of Differential Algebraic Equations

Defining a general DAE:

$$0 := f(x, \dot{x}, t)\,. \tag{C.7}$$

the index (differential index) k of the (non)linear, sufficiently smooth DAE is the smallest k such that

$$0 := f(x, \dot{x}, t)\,, \tag{C.8}$$

$$0 := \frac{d}{d\,t}\, f(x, \dot{x}, t)\,, \tag{C.9}$$

$$\vdots\vdots\vdots \tag{C.10}$$

$$0 := \tfrac{d^k}{d\,t^k} f(x, \dot{x}, t)\,. \tag{C.11}$$

uniquely determines $\dot{x}$ as a continuous function of x and t.

D

Singular Perturbation – An Introduction

D.1 Purpose

In science and engineering one finds often problems where two systems of largely different nature a coupled. One of the sets of the equations describes the ”main” system where the second describes the ”small” system. Often the effects of the small systems may be ignored, but very often too, the small system makes all the difference. Flow systems are typically of this nature in that the boundary layer is very important when describing effects such as lift caused by flow over a profile as it is used in the construction of a wing. The concept, though, may also be applied to time scales such as fast and slow systems. Relevant readings are :

- Generic Lin and Segel (1988)
- Control Kokotovic et al. (1976) and Saksena et al. (1984)

D.2 Problem Definition

For this very simple exposition to singular perturbation, let us define a simple time-constant linear system which describes a system consisting of a slow, main subsystem and a fast second subsystem both being intimately coupled together :

$$\begin{aligned} \dot{\underline{\mathbf{x}}} &= \underline{\underline{\mathbf{A}}}_{11}\,\underline{\mathbf{x}} + \underline{\underline{\mathbf{A}}}_{12}\,\underline{\mathbf{z}}; \qquad && \underline{\mathbf{x}}(0) := \underline{\mathbf{x}}^0 \\ \varepsilon\,\dot{\underline{\mathbf{z}}} &= \underline{\underline{\mathbf{A}}}_{21}\,\underline{\mathbf{x}} + \underline{\underline{\mathbf{A}}}_{22}\,\underline{\mathbf{z}}; \qquad && \underline{\mathbf{z}}(0) := \underline{\mathbf{z}}^0 \\ y &:= \underline{\underline{\mathbf{C}}}_1\,\underline{\mathbf{x}} + \underline{\underline{\mathbf{C}}}_2\,\underline{\mathbf{z}} && \end{aligned}$$

D.3 The Outer Solution

First we assume that the first equation dominates and set and let the small number $\varepsilon \to 0$, thus a pseudo-steady-state assumption is made for the second equation. This yields what is called the outer solution :

$$\begin{aligned} \underline{\dot{\mathbf{x}}_{\mathbf{o}}} &= \underline{\underline{\mathbf{A}}}_{11}\,\underline{\mathbf{x}}_o + \underline{\underline{\mathbf{A}}}_{12}\,\underline{\mathbf{z}}_o; & \underline{\mathbf{x}}(0) &:= \underline{\mathbf{x}}_o^0 \\ \lim_{\varepsilon\to 0} \varepsilon\,\underline{\dot{\mathbf{z}}_{\mathbf{o}}} &= \underline{\underline{\mathbf{A}}}_{21}\,\underline{\mathbf{x}}_o + \underline{\underline{\mathbf{A}}}_{22}\,\underline{\mathbf{z}}_o; & \underline{\mathbf{z}}_o(0) &:= \underline{\mathbf{z}}_o^0 \\ \underline{\mathbf{0}} &= \underline{\underline{\mathbf{A}}}_{21}\,\underline{\mathbf{x}}_o + \underline{\underline{\mathbf{A}}}_{22}\,\underline{\mathbf{z}}_o \end{aligned}$$

thus

$$\underline{\mathbf{z}}_o := -\underline{\underline{\mathbf{A}}}_{22}^{-1}\,\underline{\underline{\mathbf{A}}}_{21}\,\underline{\mathbf{x}}_o$$

Using the result in the first matrix equations yields step-wise the outer solution

$$\begin{aligned} \underline{\dot{\mathbf{x}}_{\mathbf{o}}} &= \underline{\underline{\mathbf{A}}}_{11}\underline{\mathbf{x}}_o + \underline{\underline{\mathbf{A}}}_{12}\,(-\underline{\underline{\mathbf{A}}}_{22}^{-1}\,\underline{\underline{\mathbf{A}}}_{21})\,\underline{\mathbf{x}}_o \\ &= (\underline{\underline{\mathbf{A}}}_{11} - \underline{\underline{\mathbf{A}}}_{12}\,\underline{\underline{\mathbf{A}}}_{22}^{-1}\,\underline{\underline{\mathbf{A}}}_{21})\,\underline{\mathbf{x}}_o \\ &= \underline{\underline{\mathbf{S}}}\,\underline{\mathbf{x}}_o \end{aligned}$$

Integration results in the simple solution

$$\underline{\mathbf{x}}_o(t) = \underline{\underline{\mathbf{e}}}^{\underline{\underline{\mathbf{S}}}\,t}\,\underline{\mathbf{x}}_o^0$$

The output for the outer solution (indicated by a subscribt o) is then

$$\begin{aligned} y_o(t) &:= \underline{\underline{\mathbf{C}}}_1\,\underline{\mathbf{x}}_o(t) + \underline{\underline{\mathbf{C}}}_2\,(-\underline{\underline{\mathbf{A}}}_{22}^{-1}\,\underline{\underline{\mathbf{A}}}_{21})\,\underline{\mathbf{x}}_o(t) \\ &:= (\underline{\underline{\mathbf{C}}}_1 - \underline{\underline{\mathbf{C}}}_2\,\underline{\underline{\mathbf{A}}}_{22}^{-1}\,\underline{\underline{\mathbf{A}}}_{21})\,\underline{\mathbf{x}}_o(t) \\ &:= (\underline{\underline{\mathbf{C}}}_1 - \underline{\underline{\mathbf{C}}}_2\,\underline{\underline{\mathbf{A}}}_{22}^{-1}\,\underline{\underline{\mathbf{A}}}_{21})\,\underline{\underline{\mathbf{e}}}^{\underline{\underline{\mathbf{S}}}\,t}\,\underline{\mathbf{x}}^0 \end{aligned}$$

The outer solution is thus a simple exponential as it was probably expected. This outer solution describes the system approximately in the large time scale, but what about the small time scale, particularly at the beginning of a change ?

D.4 The Inner Solution

The inner solution is constructed by time scaling. The scaling is done such that the scaled time is in the order of the boundary layer. Let the new time be τ:

$$\tau := t/\varepsilon$$

Then

$$\begin{aligned}
\varepsilon^{-1}\,\frac{d\,\underline{\mathbf{x}}_i}{d\,\tau} &= \underline{\underline{\mathbf{A}}}_{11}\,\underline{\mathbf{x}}_i + \underline{\underline{\mathbf{A}}}_{12}\,\underline{\mathbf{z}}_i \\
\frac{d\,\underline{\mathbf{x}}_i}{d\,\tau} &= \varepsilon\,(\underline{\underline{\mathbf{A}}}_{11}\,\underline{\mathbf{x}}_i + \underline{\underline{\mathbf{A}}}_{12}\,\underline{\mathbf{z}}_i) \\
\frac{d\,\underline{\mathbf{x}}_i}{d\,\tau} &\approx 0 \qquad \rightarrow \qquad \underline{\mathbf{x}}_i(\tau) :\approx \underline{\mathbf{x}}^0 \\
\varepsilon\,\varepsilon^{-1}\,\frac{d\,\underline{\mathbf{z}}_i}{d\,\tau} &= \underline{\underline{\mathbf{A}}}_{21}\,\underline{\mathbf{x}}_i + \underline{\underline{\mathbf{A}}}_{22}\,\underline{\mathbf{z}}_i \\
\underline{\mathbf{z}}_i(\tau) &= \underline{\underline{\mathbf{e}}}^{\underline{\underline{\mathbf{A}}}_{22}\,\tau}\,\underline{\mathbf{z}}^0 + \int_0^\tau \underline{\underline{\mathbf{e}}}^{\underline{\underline{\mathbf{A}}}_{22}\,\theta}\,\underline{\underline{\mathbf{A}}}_{21}\,\underline{\mathbf{x}}^0\,d\theta \\
&= \underline{\underline{\mathbf{e}}}^{\underline{\underline{\mathbf{A}}}_{22}\,\tau}\,\underline{\mathbf{z}}^0 + \underline{\underline{\mathbf{A}}}_{22}^{-1}\,\underline{\underline{\mathbf{e}}}^{\underline{\underline{\mathbf{A}}}_{22}\,\tau}\Big|_0^\tau\,\underline{\underline{\mathbf{A}}}_{21}\,\underline{\mathbf{x}}^0 \\
&= \underline{\underline{\mathbf{e}}}^{\underline{\underline{\mathbf{A}}}_{22}\,\tau}\,\underline{\mathbf{z}}^0 + \underline{\underline{\mathbf{A}}}_{22}^{-1}\,(\underline{\underline{\mathbf{e}}}^{\underline{\underline{\mathbf{A}}}_{22}\,\tau} - \underline{\underline{\mathbf{I}}})\,\underline{\underline{\mathbf{A}}}_{21}\,\underline{\mathbf{x}}^0 \\
y_i(\tau) &:= \underline{\underline{\mathbf{C}}}_1\,\underline{\mathbf{x}}^0 + \underline{\underline{\mathbf{C}}}_2\,\left(\underline{\underline{\mathbf{e}}}^{\underline{\underline{\mathbf{A}}}_{22}\,\tau}\,\underline{\mathbf{z}}^0 + \underline{\underline{\mathbf{A}}}_{22}^{-1}\,(\underline{\underline{\mathbf{e}}}^{\underline{\underline{\mathbf{A}}}_{22}\,\tau} - \underline{\underline{\mathbf{I}}})\,\underline{\underline{\mathbf{A}}}_{21}\,\underline{\mathbf{x}}^0\right)
\end{aligned}$$

D.5 Combining the Outer and the Inner Solution

Having the outer and the inner solution available, a combined solution may be constructed by adding the two solutions together and subtracting the common parts of the two :

$$y_c(t) := y_o(t) + y_i(t) - c(t)$$

where the last term represents the common part of the two solutions. In this case, this common part is extremely simple as it is just a constant which can be found easily by analysing the end value :

$$\begin{aligned}
& y_c(t \rightarrow \text{large}) = y_o(t \rightarrow \text{large}) \\
\Rightarrow \quad & y_i(t \rightarrow \text{large}) = c(t)
\end{aligned}$$

Thus

$$\begin{aligned}
\lim_{t\rightarrow\infty} y_i(t) &= \underline{\underline{\mathbf{C}}}_1\,\underline{\mathbf{x}}^0 + \underline{\underline{\mathbf{C}}}_2\,\underline{\underline{\mathbf{A}}}_{22}^{-1}\,(-\underline{\underline{\mathbf{I}}})\,\underline{\underline{\mathbf{A}}}_{21}\,\underline{\mathbf{x}}^0 \\
&= (\underline{\underline{\mathbf{C}}}_1 - \underline{\underline{\mathbf{C}}}_2\,\underline{\underline{\mathbf{A}}}_{22}^{-1}\,\underline{\underline{\mathbf{A}}}_{21})\,\underline{\mathbf{x}}^0
\end{aligned}$$

D.6 Example

The attached figures show the simulation results for a system :

$$\begin{array}{rcrrcr}
\underline{\underline{\mathbf{A}}}_{11} & := & -5 & \underline{\underline{\mathbf{A}}}_{12} & := & 1 \\
\underline{\underline{\mathbf{A}}}_{21} & := & 1 & \underline{\underline{\mathbf{A}}}_{22} & := & -1 \\
\underline{\underline{\mathbf{C}}}_{1} & := & 1 & \underline{\underline{\mathbf{C}}}_{2} & := & 1 \\
\underline{\mathbf{x}}^{0} & := & 10 & \underline{\mathbf{z}}^{0} & := & 5 \\
\varepsilon & := & 0.01 & & &
\end{array}$$

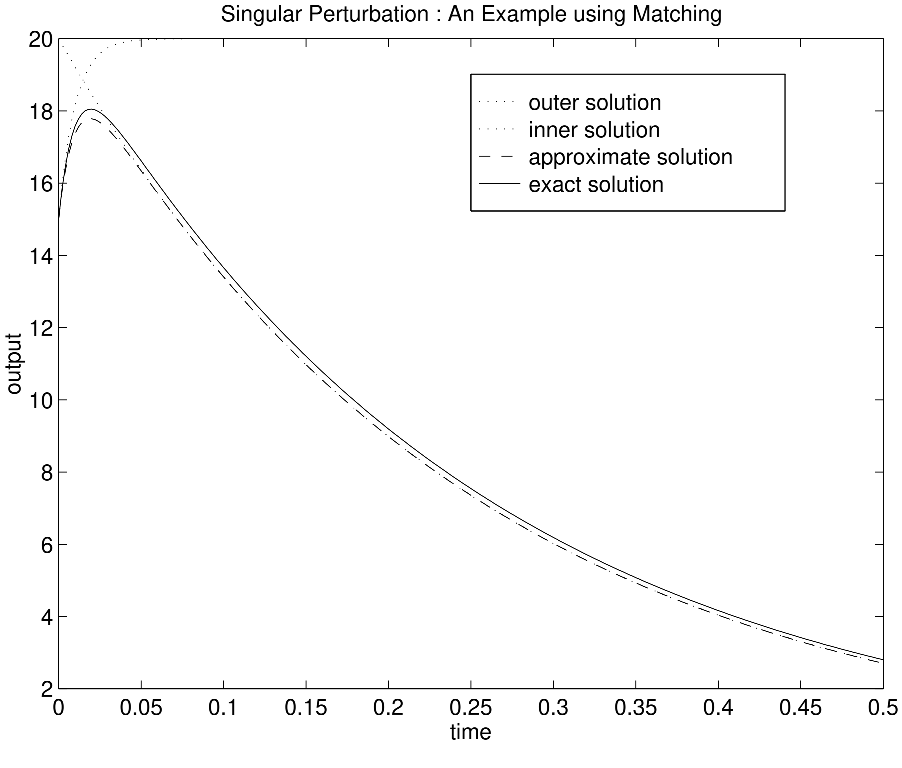


Figure D.1: Inner solution, outer solution and combined solution compared with the exact solution

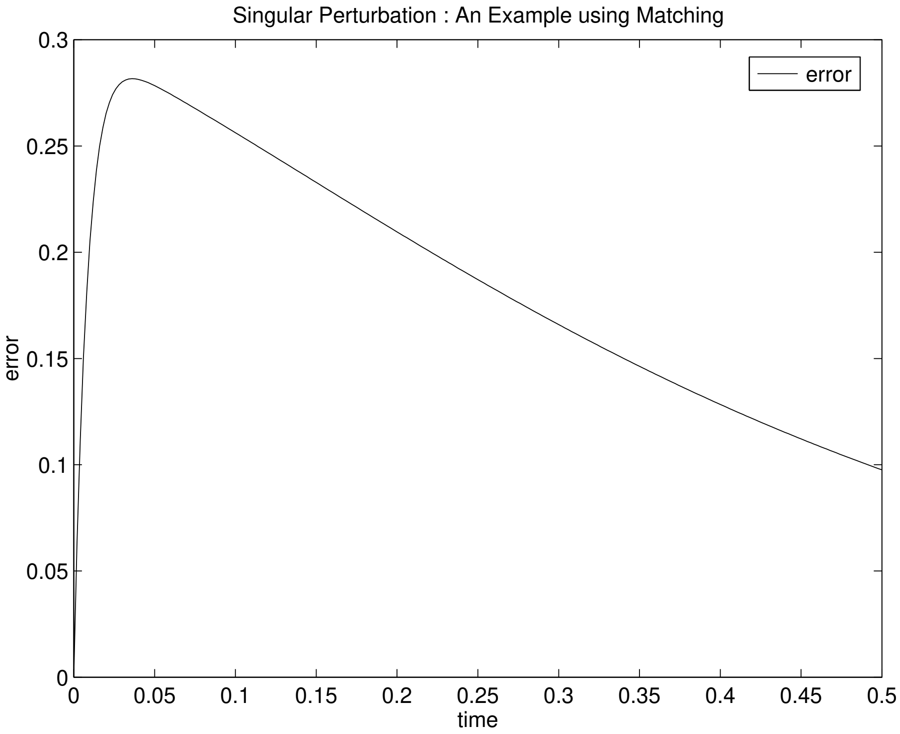


Figure D.2: Error of exact solution and combined solution

D.7 Tikhonov's Theorem

$$\frac{d\underline{\mathbf{x}}}{dt};=\underline{\mathbf{f}}(\underline{\mathbf{x}},\underline{\mathbf{z}},t),$$
$$\mu\frac{d\underline{\mathbf{z}}}{dt};=\underline{\mathbf{g}}(\underline{\mathbf{x}},\underline{\mathbf{z}},t).$$

Taking the limit as $\mu \to 0$ this becomes the "degenerate system"

$$\frac{d\underline{\mathbf{x}}}{dt};=\underline{\mathbf{f}}(\underline{\mathbf{x}},\underline{\mathbf{z}},t),$$
$$\underline{\mathbf{z}}(\underline{\mathbf{x}},t);=\text{root}\left(\underline{\mathbf{g}}(\underline{\mathbf{x}},\underline{\mathbf{z}},t)\right)$$

Note that there may exist more than one root.

Tikhonov's theorem states that as $\mu \to 0$, the solution of the system of two differential equations above approaches the solution of the degenerate system if $\underline{\mathbf{z}}(\underline{\mathbf{x}},t)$ is a stable root of the "adjoined system" $\frac{d\underline{\mathbf{z}}}{dt}=\underline{\mathbf{g}}(\underline{\mathbf{x}},\underline{\mathbf{z}},t)$.

E

Friction factor relations for flow-through systems

E.1 Approximations of the Colebrook equation

Source: Wikipedia

Equation	Author	Year
$\lambda = .0055(1 + (2 \times 10^4 \cdot \frac{\varepsilon}{D} + \frac{10^6}{Re})^{\frac{1}{3}})$	Moody	1947
$\lambda = .094(\frac{\varepsilon}{D})^{0.225} + 0.53(\frac{\varepsilon}{D}) + 88(\frac{\varepsilon}{D})^{0.44} \cdot Re^{-\Psi}$ where $\Psi = 1.62(\frac{\varepsilon}{D})^{0.134}$	Wood	1966
$\frac{1}{\sqrt{\lambda}} = -2\log(\frac{\varepsilon}{3.715D} + \frac{15}{Re})$	Eck	1973
$\frac{1}{\sqrt{\lambda}} = -2\log(\frac{\varepsilon}{3.7D} + \frac{5.74}{Re^{0.9}})$	Jain and Swamee	1976
$\frac{1}{\sqrt{\lambda}} = -2\log((\frac{\varepsilon}{3.71D}) + (\frac{7}{Re})^{0.9})$	Churchill	1973
$\frac{1}{\sqrt{\lambda}} = -2\log((\frac{\varepsilon}{3.715D}) + (\frac{6.943}{Re})^{0.9}))$	Jain	1976
$\lambda = 8[(\frac{8}{Re})^{12} + \frac{1}{(\Theta_1+\Theta_2)^{1.5}})]^{\frac{1}{12}}$ where $\Theta_1 = [2.457\ln[(\frac{7}{Re})^{0.9} + 0.27\frac{\varepsilon}{D}]]^{16}$ $\Theta_2 = [(\frac{37530}{Re})^{16}$	Churchill	1977
$\frac{1}{\sqrt{\lambda}} = -2\log[\frac{\varepsilon}{3.7065D} - \frac{5.0452}{Re}\log(\frac{1}{2.8257}(\frac{\varepsilon}{D})^{1.1098} + \frac{5.8506}{Re^{0.8981}})]$	Chen	1979
$\frac{1}{\sqrt{\lambda}} = 1.8\log[\frac{Re}{0.135Re(\frac{\varepsilon}{D})+6.5}]$	Round	1980

Equation	Author	Year
$\frac{1}{\sqrt{\lambda}} = -2\log(\frac{\varepsilon}{3.7D} + \frac{5.158 log(\frac{Re}{7})}{Re(1+\frac{Re^{0.52}}{29}(\frac{\varepsilon}{D})^{0.7}}$	Barr	1981
$\frac{1}{\sqrt{\lambda}} = -2\log[\frac{\varepsilon}{3.7D} - \frac{5.02}{Re}\log(\frac{\varepsilon}{3.7D} - \frac{5.02}{Re}\log(\frac{\varepsilon}{3.7D} + \frac{13}{Re}))]$ or $\frac{1}{\sqrt{\lambda}} = -2\log[\frac{\varepsilon}{3.7D} - \frac{5.02}{Re}\log(\frac{\varepsilon}{3.7D} + \frac{13}{Re})]$	Zigrang and Sylvester	1982
$\frac{1}{\sqrt{\lambda}} = -1.8\log\left[\left(\frac{\varepsilon}{3.7D}\right)^{1.11} + \frac{6.9}{Re}\right]$	Haaland	1983
$\lambda = [\Psi_1 - \frac{(\Psi_2-\Psi_1)^2}{\Psi_3-2\Psi_2+\Psi_1}]^{-2}$ or $\lambda = [4.781 - \frac{(\Psi_1-4.781)^2}{\Psi_2-2\Psi_1+4.781}]^{-2}$ where $\Psi_1 = -2\log(\frac{\varepsilon}{3.7D} + \frac{12}{Re})$ $\Psi_2 = -2\log(\frac{\varepsilon}{3.7D} + \frac{2.51\Psi_1}{Re})$ $\Psi_3 = -2\log(\frac{\varepsilon}{3.7D} + \frac{2.51\Psi_2}{Re})$	Serghides	1984
$\frac{1}{\sqrt{\lambda}} = -2\log(\frac{\varepsilon}{3.7D} + \frac{95}{Re^{0.983}} - \frac{96.82}{Re})$	Manadilli	1997
$\frac{1}{\sqrt{\lambda}} = -2\log\{\frac{\varepsilon}{3.7065D} - \frac{5.0272}{Re}\log[\frac{\varepsilon}{3.827D} - \frac{4.657}{Re}\log((\frac{\varepsilon}{7.7918D})^{0.9924} + (\frac{5.3326}{208.815+Re})^{0.9345})]\}$	Monzon, Romeo, Royo	2002
$\frac{1}{\sqrt{\lambda}} = 0.8686\ln[\frac{0.4587Re}{(S-0.31)^{\frac{S}{(S+1)}}}]$ where $S = 0.124Re\frac{\varepsilon}{D} + \ln(0.4587Re)$	Goudar, Sonnad	2006
$\frac{1}{\sqrt{\lambda}} = 0.8686\ln[\frac{0.4587Re}{(S-0.31)^{\frac{S}{(S+0.9633)}}}]$ where $S = 0.124Re\frac{\varepsilon}{D} + \ln(0.4587Re)$	Vatankhah, Kouchakzadeh	2008
$\frac{1}{\sqrt{\lambda}} = \alpha - [\frac{\alpha+2\log(\frac{\beta}{Re})}{1+\frac{2.18}{\beta}}]$ where $\alpha = \frac{(0.744\ln(Re))-1.41}{(1+1.32\sqrt{\frac{\varepsilon}{D}})}$ $\beta = \frac{\varepsilon}{3.7D}Re + 2.51\alpha$	Buzzelli	2008

Equation	Author	Year
$\lambda = \frac{6.4}{(\ln(Re)-\ln(1+.001Re\frac{\varepsilon}{D}(1+10\sqrt{\varepsilon}D)))^{2.4}}$	Avci, Kargoz	2009
$\lambda = \frac{0.2479-0.0000947(7-\log Re)^4}{(\log(\frac{\varepsilon}{3.615D}+\frac{7.366}{Re^{0.9142}}))^2}$	Evangleids, Papaevangelou, Tzimopoulos	2010

F

Graph Theory

Graph is used for different objects. Probably most commonly one refers to a graphical representation of data in the form of a plot, for example a x-y plot or a pie-chart, a histogram etc. The term *graph* is also used for a (graphical) representation of a network consisting of a set of nodes and connections. Many problems can convinently be represented by a diagram consisting of nodes (points, circles or other graphical entities) connected by lines or arrows thus forming a network. For nodes one also used the term *vertices* and for the connections one also uses frequently the term *arc*. Graphically this may reflect into a set of points (circles, ellipses or any other object that visualises a body, volume or system) and a set of lines, bars, arrows representing the connections.

Graphs are used in different contents where networks are useful such as Internet representation, English thesaurus (Words, 2016), abstract syntax tree, etc. A number of examples can be found on (Viz, 2016), a webpage that describes a graph visualisation tool. In the context of modelling we have three major uses of the graph theory, namely as a graphical representation of the space taken by the plant, thus a representation of control volumes and their interaction **??**, **??**. In order to handle complexity, that is a very large graph, this is extended to a hierarchical graph representation. Graph theory is also a very handy tool to analyse equations and variables in the form of a bi-partite graph. Mathematical expressions can be mapped into trees consisting of variables and operators uses in computer science to store coded expressions. These are known as abstract syntax trees.

The theory of graphs is old and goes back to Euler publishing a paper on the *Seven Bridges of K´onigsberg* and published in 1736, which is regarded as the first paper in the history of graph theory. The subject being part of discrete mathematics is documented in many text books commonly available from the library. There are also many resources on the web describing the subject (Wikipedia, 2016; MathWorld, 2016). The exposition here is thus only a summary of the material being used in the current context.

F.1 Basics of Graph Theory

Graph, vertice, node & incidence function : A *graph* is a triple

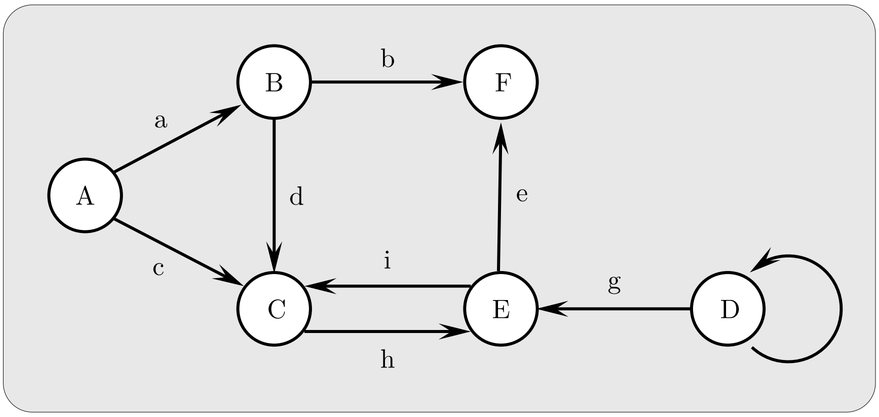


Figure F.1: An example graph G

$G := \{V(G), E(G), f_G\}$ with $V(G)$ being a set of *vertices (nodes)* and $E(G)$ and a set of *edges (arcs)* and an *incidence function* f_G that associates with each edge of G an unordered, not necessarily distinct, pair of vertices of G.

Joint : If e is an edge and u and v are two vertices such that $f(e) = (u, v)$ then e is said to *join* the vertice u with the vertice v.

Ends : The two vertices u, v are called the ends of the edge e.

Connected : Two vertices that are *connected* by an edge are called *adjacent*.

Example :

The example graph G F.1 consists of set of vertices and edges:

$$V(G) := \{A, B, C, D, E, F\}\,, \tag{F.1}$$

$$E(G) := \{a, b, c, d, e, f, g, h, i\}\,. \tag{F.2}$$

$$\nu(G) := |V(G)| = 6\,. \tag{F.3}$$

$$\epsilon(G) := |E(G)| = 9\,. \tag{F.4}$$

The incidence function is then:

$$\begin{aligned} f_E := \{ \quad & (A, B), (B, F), (A, C), (B, C), (E, F), \\ & (D, D), (D, E), (C, E), (E, C) \quad \}\,. \end{aligned} \tag{F.5}$$

Ends : An edge with two distinct *ends* is called a *link*.

Incident : The ends of an edge are said to be *incident* with the edge, and vica versa.

Multiple edge : If there are more than one edge have the same end, the edge is called *multiple* edge.

Loop : An edge that has that connects to the same vertice on both ends is called a *loop*

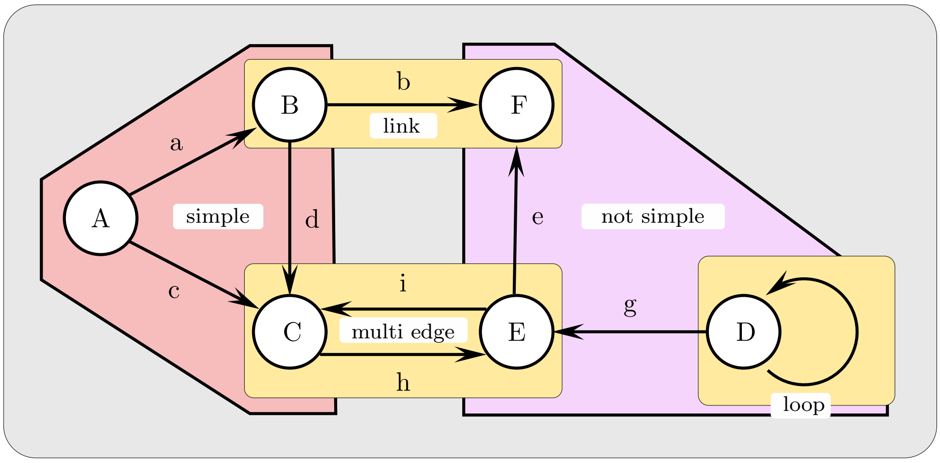


Figure F.2: Basic graph structures

Isomorphic graphs : Two graphs G and H are called *isomorphic* $G \approxeq H$ if there exists bijections $\phi : V(G) \rightarrow V(H)$ and $\Phi : E(G) \rightarrow E(H)$ such that $f_G(e) = (u, v)$ if and only if $f_H(\Phi(e)) = (\phi(u), \phi(u))$. Such a pair (ϕ, Phi) of mappings is called an *isomorphism* between G and H. In layman terms: the structure of the two graphs is the same, whilst the edges and the vertices are labelled differently.

Simple graph : A graph is *simple* if it has no loops and no two of its links join the same pair of vertices.

Complete graph : If each pair of distinct vertices is joined by an edge is called a *complete graph*. Not considering isomorphism, there is only one complete graph with n vertices, which is denoted by K_n.

Empty graph : A graph is *empty* if it contains no edges.

Finite graph : A graph is *finite* if both its vertex set and the edge set are finite.

Trivial graph : A graph with only one vertex is called *trivial* , all others non-trivial.

Identical graphs : Two graphs G and H are called *identical* if $V(G) = V(H)$, $E(G) = E(H)$ and $f(G) = f(H)$

Bipartite graph : A *bipartite graph* is one whose vertex set can be partitioned into two subsets X and Y, so that each edge has one

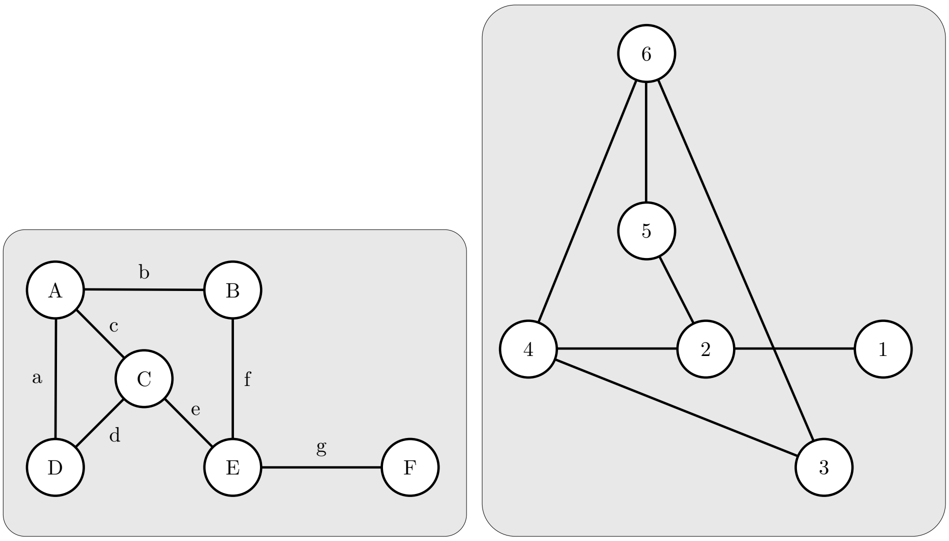


Figure F.3: Two isomorphic graphs with the mappings: $A = 6, B = 5, C = 4, D = 3, E = 2, F = 1$ and $a = h, b = i, c = j, d = k, e = k, f = l, g = m$

h

end in X and the other end in Y. The partition of a graph's vertice $V(G) = (X(G), Y(G)$ is called a *bipartition* of the graph G.

Complete bipartite graph : Is a graph, which is a simple bipartite graph, with the bipartition X and Y in which each vertex of Y is joined to each vertex of Y. If m, n denote the cardinality of the two sets X and Y, respectively, then the graph is denoted by $K_{m,n}$. This concept can be extended to k-partitioned graphs.

Incidence matrix : Any graph can be represented in a $\nu \times \epsilon$ matrix, where $\nu := |V(G)|$ and $\epsilon := |E(G)|$. The incidence matrix of G is the matrix $\underline{\underline{\mathbf{M}}}(G) := [m_{i,j}]$ with $m_{i,j}$ being the number of times that a vertice v_i and an edge e_j are incident.

Adjacency matrix : The adjecancy matrix of the graph G is the $\nu \times \nu$ matrix $\underline{\underline{\mathbf{A}}}(G) := [a_{i,j}]$ in which $a_{i,j}$ is the number of edges joining v_i and v_j

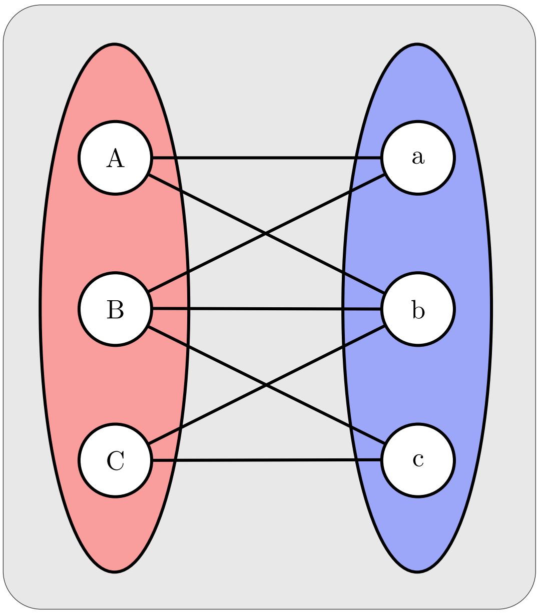


Figure F.4: A bipartite graph showing the two sets on the left and the right. The shown graph is also complete.

Example : For our example above the incidence matrix is:

$$\underline{\underline{\mathbf{M}}} := \begin{array}{c|cccccccc} & a & b & c & d & e & f & g & h \\ \hline A & -1 & & -1 & & & & & \\ B & +1 & -1 & & -1 & & & & \\ C & & & +1 & +1 & & & & -1 \\ D & & & & & & \pm1 & -1 & \\ E & & & & & -1 & & +1 & +1 \\ F & & +1 & & & +1 & & & \end{array} \tag{F.6}$$

and the djacency matrix of the directed graph:

$$\underline{\underline{\mathbf{A}}} := \begin{array}{c|cccccc} & A & B & C & D & E & F \\ \hline A & & 1 & 1 & & & \\ B & & & 1 & & & 1 \\ C & & & & & 1 & \\ D & & & & 1 & 1 & \\ E & & & & & & 1 \\ F & & & & & & \end{array} \tag{F.7}$$

with the rows being the source nodes and the column the sink nodes.

Let $\underline{\underline{\mathbf{U}}}(\underline{\underline{\mathbf{A}}})$, $\underline{\underline{\mathbf{D}}}(\underline{\underline{\mathbf{A}}})$ be the upper triangular and the diagonal matrix extracted from $\underline{\underline{\mathbf{A}}}$ then the adjacency matrix for the un-directed graph is given by $\underline{\underline{\mathbf{U}}} + \underline{\underline{\mathbf{D}}} + \underline{\underline{\mathbf{U}}}^T$, which is symmetrical.

Let G_1 and G_1 be two non-empthy graphs and define the union as a new graph: $G := (V(G_1) \cup V(G_2), E(G_1) \cup E(G_2))$

Subgraph : G_1 and G_2 are subgraphs of G.

Subergraph : G is a supergraph to G_1 and G_2

Disjoint : If $G_1 \cap G_2 := \emptyset$ this being short for $(V(G_1) \cap V(G_2), E(G_1) \cap E(G_2)) = (\emptyset, \emptyset)$ then the two graphs are disjoint.

Induced subgraph : The subgraph $G_1 := G[V_1]$ is called an *induced subgraph* of G if E_1 is the subset of E that have both ends in V_1.

Edge-induced subgraph : The subgraph $G_1 := G[E_1]$ is called an *edge-induced subgraph* of G if V_1 is the subset of V with the set of ends E_1.

Underlying simple graph : The spanning subgraph of G is obtained by deleting all loop and reduce all multiple edges single edges.

Spanning subgraph : Is a subgraph with identical vertice sets. H is a spanning subgraph of H if $V(H) = V(G)$.

Walk : A *walk* W in the graph G is a finite non-null sequence of alternating vertices.

origin, terminus : lines starting with the vertice v_0, called the *ori-*

gin, and ending with vertice v_k, called the *terminus*, thus $W := v_0\, e_1\, v_1\, e_2 \ldots e_k\, v_k$.

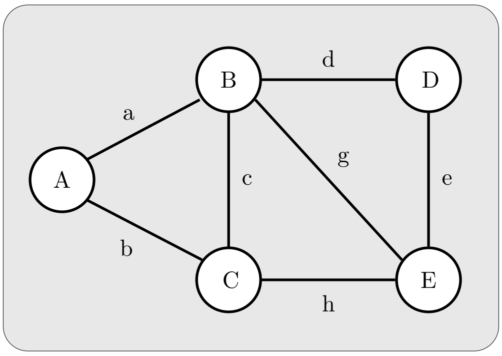
Ends : W is called a *trail* if the edges are distinct.

Path : If in addition the vertices are distinct, the walk is called a *path.*

Closed walk : A walk is *closed* if it is of positive length and the origin and terminus are identical.

Cycle : A *cycle* is a closed walk which has passes through distinct vertices.

Connected graph : Two subgraphs $G[V_a]$ and $G[V_b]$ are *connected* if there exists at least one vertice having ends in each of the sets V_a and V_b. Otherwise the graph is called *disconnected* and the two subgraphs $G[V_a]$ and $G[V_b]$ are called *components* of G.



walk : A-b-C-c-B-d-D-d-B-a-A-b-C

trail: A-b-C-c-B-d-D-e-E-g-B

path: A-a-B-d-D-e-E-f-C

closed walk: A-a-B-d-D-e-E-g-B-c-C-b-A

Figure F.5: A general walk and special walks: trail, path, closed walk and cycle

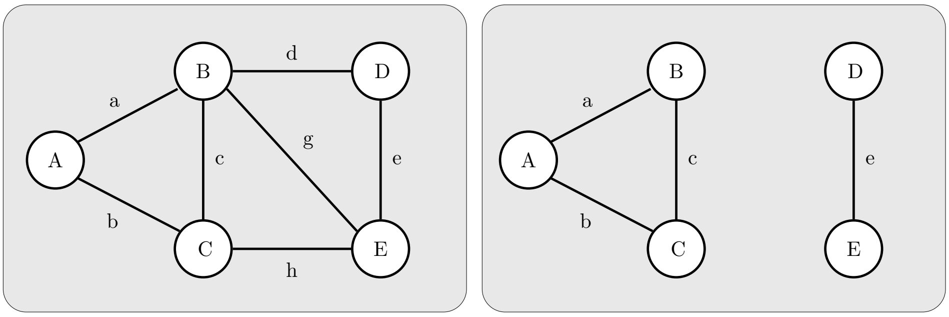


Figure F.6: Connected and disconnected graphs, components

G

Elements of Statistics

G.1 Probabilty

G.1.1 Axiomatic Definition

Definition - Event Space $\mathcal{E}$: An event space is called a probability space or a probability field if every set of observed events A has a probability $P(A)$

Definition - Probability Axioms :

1. Probability is a real number $\in [0,1]$
2. Impossible implies $P(\emptyset) := 0$ and certain implies $P(\mathcal{E}) := 1$
3. For disjunct, mutually exclusive events A and B: $P(A+B) := P(A) + P(B)$
4. For non-mutually exclusive events A and B: $P(A+B) := P(A) + P(B) - P(AB)$

Definition - Conditional Probability : The conditional probability of A given B, that is event B has already occurred and $P(B) \neq 0$, is

$$P(A|B) = \frac{P(AB)}{P(B)}$$

Definition - Random Variable : a function of an event space.

Probability of a random variable to assume a value between a and b is given by

$$P(x \in [a,b]) := \int_0^b p(x)\,dx - \int_0^a p(x)\,dx\,,$$

where $p(x)$ is the probability density function characterising the continuous random variable x. If the random variable is discrete, the integral is replaced by corresponding summations.

G.1.2 Bayes' Theorem

Theorem G.1.1 (Bayes' Theorem)**.** *Let $\{A_i\}$ be a disjunct set of observed events and B an observed event, then for each j:*

$$\begin{aligned} P(A_j|B) &= \frac{P(A_jB)}{P(B)} \\ &= \frac{P(A_j)\,P(B|A_j)}{\sum_1^n P(A_i)\,P(B|A_i))} \end{aligned} \tag{G.1}$$

The theorem is often used with A_j denoting a statement about an unknown phenomenon, whilst B presents the known information about the process. $P(A_j)$ is denoted as prior probability, $P(A_j|B)$ as posterior probability and $P(B|A_j)$ as likelihood.

G.1.3 Distribution Measures

Definition - Modus : the maximum of the probability distribution function.

Definition - Median : location where the cumulative distribution function is 1/2. The most important measures are the mean

$$\begin{aligned} \bar{x} &:= \mathbf{E}\left[x_i\right] \\ &:= \sum_i x_i\,p(x_i) \quad x :: \text{discrete} \\ \bar{x} &:= \mathbf{E}\left[x\right] \\ &:= \int_{-\infty}^{+\infty} x\,p(x)dx \quad x :: \text{continuous} \end{aligned}$$

and the variance

$$\mathrm{var}\,(x) := \mathbf{E}\left[(x_i - \mathbf{E}\left[x_i\right])^2\right]$$

The central moments are defined as

$$\begin{aligned} \mu_k &:= \mathbf{E}\left[(x_i - \mathbf{E}\left[x_i\right])^k\right] \\ &:= \sum_i\,(x_i - \bar{x})^k\;p(x_i) \quad x :: \text{discrete} \\ \mu_k &:= \mathbf{E}\left[(x - \mathbf{E}\left[x\right])^k\right] \end{aligned}$$

$$:= \int_{-\infty}^{+\infty} (x-\bar{x})^k \, p(x)\, dx \quad x :: \text{continuous}$$

The moments are defined as

$$\begin{aligned}
\mu'_k &:= \mathbf{E}\left[(x_i - \mathbf{E}[x_i])^k\right] \\
&:= \sum_i x_i^k \, p(x_i) \quad x :: \text{discrete} \\
\mu'_k &:= \mathbf{E}\left[(x - \mathbf{E}[x])^k\right] \\
&:= \int_{-\infty}^{+\infty} x^k \, p(x)\, dx \quad x :: \text{continuous}
\end{aligned}$$

The second central moment is the *variance*. The third central moment is called *skewness* and the fourth is called *kurtiosis*.

G.1.3.1 Behaviour of Moments

Let $x, y, z \in \mathcal{E}$ three random variables on the same probability space and a, b, c, d arbitrary constants.

$$\begin{aligned}
\mathbf{E}[a\,x + b] &:= a\,\mathbf{E}[x] + b \\
\mathbf{E}[x + y] &:= \mathbf{E}[x] + \mathbf{E}[y] \\
\mathbf{E}[x\,y] &:= \mathbf{E}[x]\ \mathbf{E}[y] \quad \text{if } x \text{ and } y \text{ are uncorrelated}
\end{aligned}$$

$$\begin{aligned}
\text{var}(x) &:= \mathbf{E}\left[(x - \mathbf{E}[x])^2\right] \\
&:= \mathbf{E}\left[x^2\right] - (\mathbf{E}[x])^2 \\
&\geq 0 \\
&= 0 \quad \text{for } x := \ \text{const} \\
\text{var}(a\,x + b) &:= a^2\,\text{var}(x) \\
\text{var}(x + y) &:= \text{var}(x) + \text{var}(y) \\
&\qquad\qquad\qquad\qquad \text{if } x \text{ and } y \text{ are independent} \\
(cov)(x, y) &:= \mathbf{E}[(x - \mathbf{E}[x])\ (y - \mathbf{E}[y])] \\
&:= \mathbf{E}[(x\,y - x\,\mathbf{E}[y] - \mathbf{E}[x]\ y + \mathbf{E}[x]\ \mathbf{E}[y])] \\
&:= \mathbf{E}[x\,y] - \mathbf{E}[x]\ \mathbf{E}[y] \\
\rho(x, y) &:= \frac{(cov)(x, y)}{\sqrt{\text{var}(x)\ \text{var}(y)}}
\end{aligned}$$

$$\begin{aligned}
(cov)\,(x,y)^2 &\leq \mathsf{var}\,(x)\;\,\mathsf{var}\,(y) \\
&= 1 \quad \text{if } x \text{ and } y \text{ on straight line} \\
&= 0 \quad \text{if } x \text{ and } y \text{ are independent} \\
(cov)\,(a\,x+b, c\,y+d) &:= a\,c\,(cov)\,(x,y) \\
(cov)\,(x+y,z) &:= (cov)\,(x,z) + (cov)\,(y,z)
\end{aligned}$$

G.1.3.2 Some Follow-Ups

Given x_i have all the same expectation value:

$$\mathbf{E}\,[\bar{x}] := \mathbf{E}\left[\frac{1}{n}\sum_{i:=1}^{n} x_i\right]$$

Given x and y are independent

$$\mathsf{var}\,(x-y) := \mathsf{var}\,(x) + \mathsf{var}\,(y)$$

Given x_i are uncorrelated and have the same mean and the same variance σ^2

$$\begin{aligned}
\mathsf{var}\,(\bar{x}) &:= \mathsf{var}\left(\frac{1}{n}\sum_{i:=1}^{n} x_i\right)\;, \\
&:= \left(\frac{1}{n}\right)^2\; n\,\mathsf{var}\,(x_i)\;, \\
&:= \frac{1}{n}\,\sigma^2\,.
\end{aligned}$$

Also

$$\begin{aligned}
\mathbf{E}\,[x] &:= \mathbf{E}\,[\mathbf{E}\,[x|y]] \\
\mathsf{var}\,(x) &:= \mathbf{E}\,[\mathsf{var}\,(x|y)] + \mathsf{var}\,(\mathbf{E}\,[x|y])
\end{aligned}$$

G.2 Most Common Distribution Functions

G.2.1 Binomial Distribution

Number of successes in n independent events with probability p:

$$\begin{aligned}
P\,(x=k) &:= \binom{n}{k}\,p^k\,(1-p)^{n-k}\,, \quad k := 0,1,\ldots,n \\
\mathbf{E}\,[x] &:= n\,p \\
\mathsf{var}\,(x) &:= n\,p\,(1-p) = n\,p\,q
\end{aligned}$$

G.2.2 Poisson Distribution

Number of rare events with the expectation of λ:

$$\begin{aligned} P(x=k) &:= e^{-\lambda}\,\frac{\lambda^k}{k!} \qquad k := 0,1,2,\ldots,n \\ \mathbf{E}\left[x\right] &:= \lambda \\ \mathrm{var}\left(x\right) &:= \lambda \end{aligned}$$

Poisson distribution is the limit of the binomial distribution with
$p \to 0, n \to \infty, n\,p \to \lambda$.

G.2.3 Normal Distribution

Idealised distribution of measurement errors and approximation for many other distributions. The probability distribution of the standard normal distribution for $x \in (-\infty, +\infty) := N(0,1)$ is given by:

$$\begin{aligned} p(x) &:= \frac{1}{\sqrt{2\,\pi\sigma}}\,e^{\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} \\ \mathbf{E}\left[x\right] &:= \mu \\ \mathrm{var}\left(x\right) &:= \sigma^2 \end{aligned}$$

G.2.4 Exponential Distribution

Describes processes without memory.

$$\begin{aligned} p(x) &:= \begin{cases} \frac{1}{\mu}\,e^{-\frac{x}{\mu}} & x \geq 0 \\ 0 & \text{else} \end{cases} \\ \mathbf{E}\left[x\right] &:= \mu \\ \mathrm{var}\left(x\right) &:= \sigma^2 \end{aligned}$$

G.2.5 Uniform Distribution

Mostly used as initial condition in recursive processes and random number generation which thereafter are transformed.

$$p(x) := \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & \text{else} \end{cases}$$

$$\mathbf{E}\left[x\right] := \frac{a+b}{2}$$
$$\mathrm{var}\left(x\right) := \frac{b-a}{12}$$

G.3 Essential Statistics

G.3.1 Chi-Square Distribution

Distribution of the sum of squares of ν independent, standard-normal-distributed random variables $x_i \sim N(0,1)$:

$$\sum_{i:=1} \nu\, x_i^2 \sim \chi_\nu^2$$
$$\mathbf{E}\left[x\right] := \nu$$
$$\mathrm{var}\left(x\right) := 2\,\nu$$

- The distribution is continuous in $(0,\infty)$g.
- For $\nu := 2$ it is the exponential distribution with $\mu := 2$
- Approaches the normal distribution for large ν.
- ν :: integer.

G.3.2 Student t Distribution

Distribution of the standardized arithmetic average $(\bar{x}-\mu)/\sigma_{\bar{x}}$ over $n := \nu + 1$ indebendent normally distributed x_i, where the $\sigma_{\bar{x}}$ is the empirical standard deviation of the average:

$$\sigma_{\bar{x}}^2 := \sum_{i:=1}^{n} \frac{(x_i-\bar{x})^2}{n\,(n-1)}$$

More general: Let x and y be independent, $x \sim N(0,1)$ and $y \sim \chi_\nu^2$ then:

$$\frac{x}{\sqrt{\frac{y}{\nu}}} \sim t_\nu$$

- The distribution is continuous in $(0,\infty)$.
- Symmetrical around 0.

- Bell-shaped
- Longer tails than the normal distribution
- Approaches the normal distribution for large ν.
- ν :: integer.

Mostly used to test an average or to compare two averages.

G.3.3 F-Distribution

Distribution of the quotients of two independent estimates of the same variance starting with a normally distributed variable. More generally: Let x and y be independent and $x \sim \chi^2_{\nu_x}$ and $y \sim \chi^2_{\nu_y}$ then:

$$\frac{x/\nu_x}{y/\nu_y} \sim F_{\nu_x,\nu_y}$$

- The distribution is continuous in $(0, \infty)$.
- Approaches the normal distribution for large ν_x and ν_y with a mean of 1 and a variance of $2/\nu_x + 2/\nu_y$.
- ν_x, ν_y :: integer.

Most common use: Variance analysis.

H

Optimisation – An Introduction

H.1 General Problem

$$\begin{aligned} &\min_{\underline{\mathbf{x}}\in\mathbb{R}^n} F(\underline{\mathbf{x}}) \\ &\text{subject to: } c_i(\underline{\mathbf{x}}) = 0, \qquad i := 1, 2, \ldots, e \\ &\qquad\qquad\quad c_i(\underline{\mathbf{x}}) \geq 0, \qquad i := e+1, \ldots, m \end{aligned}$$

Feasible point $\underline{\mathbf{z}}$:	satisfies all constraints.
Feasible region:	$R := \{\underline{\mathbf{z}}_i \vert \forall i\}$
Infeasible problem:	$R := 0$
Optimal point:	$\underline{\mathbf{x}}^*$
δ- Neighbourhood of $\underline{\mathbf{x}}$:	$N(x, \delta)$

Definition - local minimum : The point $\underline{\mathbf{x}}^*$ is a local minimum of the general constraint optimisation problem if $\exists \delta > 0$ such that:

1. $F(\underline{\mathbf{x}})$ is defined on $N(\underline{\mathbf{x}}^*, \delta)$ and
2. $F(\underline{\mathbf{x}}^*) < F(\underline{\mathbf{y}}) \quad \forall\, \underline{\mathbf{y}} \in N(\underline{\mathbf{x}}^*, \delta), \underline{\mathbf{y}} \neq \underline{\mathbf{x}}^*$

The function $F(\underline{\mathbf{x}})$ is smooth and at least twice-continuously differentiable.

H.2 Unconstraint Optimisation

H.2.1 One-Dimensional

The problem reduces to:

$$\boxed{\min_{\underline{\mathbf{x}}\in\mathbb{R}^1} f(x)}$$

Definition - Necessary conditions :

1. $\left.\frac{\partial f(x)}{\partial x}\right|_{x^*} := f_x(x^*) := 0$
2. $\left.\frac{\partial^2 f(x)}{\partial x^2}\right|_{x^*} := f_{xx}(x^*) \geq 0$

To proof the above conditions expand the function $f(x)$ in a Taylor series about the optimal point:

$$f(x^* + \epsilon) := f(x^*) + \frac{1}{2}\,\epsilon^2\, f_{xx}(x^*)\,. \tag{H.1}$$

H.3 Surface search methods

The optimal point is characterised by the first derivative to be zero. Thus if one has the derivative for the objective function as an algebraic expression, the search for the optimal solution reduces to finding the zero for the function, being the derivative of the objective function with respect to the parameters. So one requires a root solver for this problem.

In the case one does not want or can calculate the derivative, one can use a search method: Starting from an initial position, one moves the estimate in the direction of the optimum, which in case of a maximum is to move uphill, whilst in the minimisation problem one would move downhill. This implies a gradient search, in that one moves permanently towards an optimum. The literature reports a number of procedures to find optimal points on a hyper-surface. We shall restrict ourself to a few simple ones that illustrate the main ideas.

H.3.1 One-dimensional algorithm of Davies, Swann & Campey

The idea of the approach is to sequentially reduce the interval within the maximum is located. Two search algorithms are used sequentially. And next a quadratic model is used to iteratively reduce the interval until the convergence criterion is reached. Figure H.1 depicts the 1-dimensional search for an optimum in a first stage: reducing the search interval.

So first we do a search to bracket the maximum into a closer interval. The search begins on one side moving in the positive direction using the initial

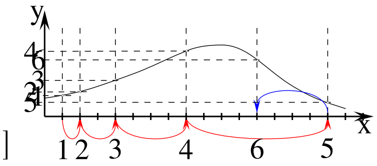


Figure H.1: A simple search method

step size. If in the first step one finds a decrease in the y-value, one starts again, but in the opposite direction. The next step is done in the same direction but double the size. The procedure is continued until the y-value decreases. At that point one halves the current step size and moves that one back. The last 4 points are equally spaced. We choose the one with the largest y-value and the immediate two neighbours, namely the one to the left and the one to the right. In the case, where the the first step gives a decrease and the step in the opposite direction also gives a decrease, one has already found the three values.

Given the three values, a parabolic model is fitted providing an estimate for the maximum.we calculate the y value at the maximum. From these four values we select again the highest and the one to the left and the one to the right, thereby reducing the interval.

Again a parabolic model is fitted and the location of the maximum is calculated by differentiation.

The procedure is repeated until two consecutive evaluation of the location of the maximum meets the conversion criterion.

The method can be refined by using cubic models.

H.3.2 Multi-dimensional search: gradient method

If one wants to get to the top of the mountain, one walks uphill. This is essentially the gradient search:

1. we start from an arbitrary point and find the gradient, for example, by using a local design of experiments, which corresponds to finding an approximate gradient using a local perturbation in all directions.

2. we move along the gradient until the slope levels out, thus becomes zero in this direction.

3. we find the gradient at this location, thus return to point 1

4. break the loop once no progress is made indicating that we are on the top of the mountain.

Note that this procedure always moves orthogonal to the contour one is on. Thus if the mountain has a circular footprint and is regular all the way to the top, thus has a circular contours, then one moves directly to the top. If however, the shape of the contours are distorted, say oval, then one moves to the top on a zigzag path. Scaling of the axis changes the shape of the hill!

H.3.3 Multi-dimensional search: Newton's method

The alternative to the search would be to solve the algebraic optimisation problem using differentiation and a root search method,or if the function is complex to use a Taylor approximation to the second-order term yielding Newton's method:

$$\underline{\mathbf{y}}'(\underline{\mathbf{x}}_1 + \underline{\mathbf{h}}) := \underline{\mathbf{y}}'(\underline{\mathbf{x}}_1) + \underline{\underline{\mathbf{y}}}''\,\underline{\mathbf{h}} \tag{H.2}$$

where the prime indicates the derivative and the double prime the second-order derivative. At the maximum, the derivative is zero, thus

$$\underline{\mathbf{h}} \approx -\left(\underline{\underline{\mathbf{y}}}''\right)^{-1} \underline{\mathbf{y}}' \tag{H.3}$$

and

$$\underline{\mathbf{x}}_{n+1} := \underline{\mathbf{x}}_n - \left(\underline{\underline{\mathbf{y}}}''\right)^{-1} \underline{\mathbf{y}}'(\underline{\mathbf{x}}_n) \tag{H.4}$$

This method is not dependent on the scaling, but has also the zigzag behaviour in complex surfaces mainly when the curvature changes drastically as a function of the direction, like this is the case for the Rosenbrock function or anything that looks like a banana.

H.3.4 Multi-dimensional search: Davidon's method

This method combines the steepest ascent and the Newton method, one of the best gradient methods. The change in the direction is determined by the the product $\underline{\underline{\mathbf{H}}}\,\underline{\mathbf{y}}'$. For the Newton method, the matrix $\underline{\underline{\mathbf{H}}}$ is the inverse of the Hessian, whilst the Davidon's method starts with a unity matrix and adjust it during the search.

The recursive search is defined by:

$$\underline{\mathbf{x}}_{n+1} := \underline{\mathbf{x}}_n - \lambda\,\underline{\underline{\mathbf{H}}}_n\,\underline{\mathbf{y}}'_n \tag{H.5}$$

The variable λ is being determined by a search of the minimum in the given direction $\underline{\underline{\mathbf{H}}}_n\,\underline{\mathbf{y}}'_n$, as described in H.3.1 thus a one-variable search is executed. At the end of each step, the new search direction is being computed:

$$\underline{\underline{\mathbf{H}}}_{n+1} := \underline{\underline{\mathbf{H}}}_n + \underline{\underline{\mathbf{B}}}_n + \underline{\underline{\mathbf{C}}}_n \tag{H.6}$$

$$\underline{\underline{\mathbf{B}}}_n := \underline{\mathbf{\Delta x}}_n \, \underline{\mathbf{\Delta x}}_n^T \left(\underline{\mathbf{\Delta x}}_n^T \, \underline{\mathbf{\Delta y}'}_n \right)^{-1} \tag{H.7}$$

$$\underline{\underline{\mathbf{C}}}_n := -\underline{\underline{\mathbf{H}}}_n \, \underline{\mathbf{\Delta y}'}_n \, \underline{\mathbf{\Delta y}}_n^T \, \underline{\underline{\mathbf{H}}}_n \left(\underline{\mathbf{\Delta y}'}_n^T \, \underline{\underline{\mathbf{H}}}_n \, \underline{\mathbf{\Delta y}'}_n \right)^{-1} \tag{H.8}$$

$$\underline{\mathbf{\Delta x}}_n := \underline{\mathbf{x}}_{n+1} - \underline{\mathbf{x}}_n \tag{H.9}$$

$$\underline{\mathbf{\Delta y}'}_n := \underline{\mathbf{y}'}_{n+1} - \underline{\mathbf{y}'}_n \tag{H.10}$$

$$\tag{H.11}$$

The matrix $\underline{\underline{\mathbf{H}}}$ converges over time to the inverse of the Hessian, thus $\left(\underline{\underline{\mathbf{y}''}}\right)^{-1}$

H.3.5 Multi-dimensional search: Other methods

The literature reports a large number of different methods. Of particular interest are those methods that do not require information about the derivatives, that is first order: the slope and second order: the curvature. These methods generate new search directions based on a linear combination of the present search direction and their results. The catalogue of such methods includes the Rosenbrock method and the often applied Powell method. Both can be found implemented in optimisation packages.

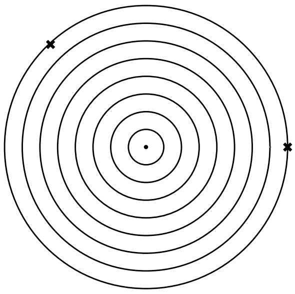


Figure H.2: DIY round-hill climbing

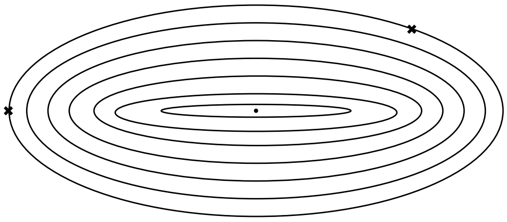


Figure H.3: DIY oval-hill climbing

I

Selective Examples

I.1 Temperature Sensor

I.1.1 Problem setup

Observing the temperature in a process is a common thing to do. For this purpose, one introduces a temperature sensor, called a thermometer. If the desired observation is the temperature of a fluid, one introduces the thermometer into the fluid. If it is a solid, one "sticks" it onto the surface or drills a whole, if this is an option and insert it there. With solids it is also quite common to use a heat-conductive paste. Certainly in these modern times, one may not use a thermometer that requires physical contact, but one may use a infrared sensor. Here we shall have a look at the thermometer-in-fluid case.

So we insert a sensor into the fluid and measure the temperature of the fluid,but is it really the temperature of the fluid we are measuring? Two quite different issues: First is that we measure some physical quantity that is a function of the temperature, which is the reason the term "obeserving" of the temperature was used. (ii) The reading is providing the state information of the object "sensor" and not the object "fluid". The temperature in the sensor must not be the same as in the fluid if the fluid is changing the temperature relative to the sensor So we have to consider the transport of energy in the form of heat into the sensor and the capacity effect of the sensor. In reality, there is a third main thing to be considered, namely the heat conduction in the sensor's construction towards the outside of the process, such as physical support and wiring. Here we shall focus only on the second aspect, namely the effect of the heat transport fluid-sensor and the capacity effect of the sensor device itself.

I.1.2 A simple model

For the purpose of this example, we assume a very simple abstraction, which is motivated by the fact that we are interested in the sensor dynamics.

Thus let us assume we have a temperature sensor in a liquid and let us define the task of modelling the joint system of fluid and temperature sensor, then we may want to take the view of the temperature sensor to be of uniform temperature, thus we lump the material that makes up the sensor itself into a simple lumped system of uniform with the intensive properties being uniform within the system. Further let us assume that the fluid in which the sensor is immersed can be viewed as consisting of a bulk with uniform temperature and a uniform fluid film around it that acts as the heat-transfer system between the bulk of the fluid and the sensor. Pictorially, this maps into the following graph: The part if interest is the sensor, thus we model the sensor dynamics having already assumed that it can be seen as an internally fast system:

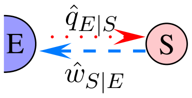


Figure I.1: A simple abstraction of a temperature sensor in a fluid environment

$$\frac{d\,E_S}{d\,t} := \hat{q}_{E|S} - \hat{w}_{S|E}\,.$$

The heat flow model approximates the behaviour of the film by:

$$\begin{aligned} \hat{q}_{E|S} &:= -k_{E|S}\,A_{E|S}\,(T_S - T_E)\,, \qquad \text{(I.1)} \\ k_{E|S}, A_{E|S} &:= \text{given}\,. \end{aligned}$$

and the system volume work term, representing the change of the volume by:

$$\hat{w}_{S|E} := p_S\,\frac{d\,V_S}{d\,t}\,.$$

At this point it is appropriate to make some simplifications and assumptions that affect the energy balance. The first simplification is associated with the fact that the sensor is not moving about, that is, its kinetic energy K_S and potential energy P_S is zero:

$$\begin{aligned} \frac{d\,K_S}{d\,t} &:= 0\,, \\ \frac{d\,P_S}{d\,t} &:= 0\,. \end{aligned}$$

Assuming constant pressure also seems an appropriate thing to do. Introducing the enthalpy:

$$H := U + p\,V\,,$$

and observing that

$$\frac{d\,H}{d\,t} := \frac{d\,U}{d\,t} + \frac{d\,p}{d\,t}\,V + p\,\frac{d\,V}{d\,t}\,,$$

the energy balance reduces to:

$$\frac{d\,H_S}{d\,t} := \hat{q}_{E|S}\,. \tag{I.2}$$

To complete the model we have to provide the link between the temperature of the system and the respective fundamental state, namely the enthalpy in this case:

$$\begin{aligned} H &:= \int_{T_r}^{T} \frac{\partial\,H}{\partial\,T}\,dT\,, \\ &:= \int_{T_r}^{T} C_p(T)\,dT\,. \end{aligned}$$

Assuming we know the heat capacity as a function of time as a product of the known volume, known, constant density and the specific heat capacity in the form of a polynomial with the known parameters $\{a_i\}$:

$$\begin{aligned} C_p(T) &:= V\,\rho\,c_p(T)\,, \\ c_p(T) &:= \sum_i a_i\,T^i\,, \\ a_i, V, \rho &:= \text{given}\,, \end{aligned}$$

the model is completely specified and proper. The dynamics of the sensor are driven by the temperature of the environment, a function of the given conditions and parameters.

To solve the problem the enthalpy-temperature relation must be solved for the temperature, which is probably the most difficult task, as the heat capacity may be given as a polynomial. One one has the temperature, one can compute the heat transfer and consequently the time derivative of the enthalpy from the conservation law.

I.1.3 Putting it into state space notation

The whole idea is to draw the attention to the mathematical problem by considering the nature of the different variables in the context of system theory. The state then is denoted by an x, the inputs u are the effort variables or flows or any quantity that directly relates to this information. The outputs are the observation of the state, directly or indirectly of the systems involved. Thus in our case, the state is first the energy, which then is simplified to internal energy because of fact that the system is stationary in the geometrical space. The next transformation takes advantage of the pressure to be constant, leading to enthalpy being the state.

I.1.3.1 Enthalpy as state

For simplicity reasons we also assume that the specific heat capacity is constant, thus not a function of the temperature. By substituting the linear heat transfer law I.1 into the enthalpy balance I.2 we get:

$$\dot{H}_S = -k_{E|S}\, A_{E|S}\, (T_S - T_E)$$

Next we need to provide an algebraic link between the enthalpy and the temperature measured relative to a reference temperature T_r:

$$H_s := \rho_S\, V_S \int_{T_r}^{T_s} (c_p(\xi))_S \, d\xi$$

Assuming the c_p to be constant then we can readily solve the above integral equation for the upper limit:

$$T_S := \frac{1}{(\rho\, V\, c_p)_S}\, H_S + T_r$$

Substitution yields:

$$\dot{H}_S = -k_{E|S}\, A_{E|S} \left(\frac{1}{(\rho\, V\, c_p)_S}\, H_S + T_r - T_E \right)$$

If we factor out the fraction with the heat capacity,

$$\begin{aligned} \dot{H}_S &= -\frac{k_{E|S}\, A_{E|S}}{(\rho\, V\, c_p)_S} \left(H_S + (\rho\, V\, c_p)_S\, (T_r - T_E)\right) \\ &= -\frac{k_{E|S}\, A_{E|S}}{(\rho\, V\, c_p)_S} \left(H_S - (\rho\, V\, c_p)_S\, (T_E - T_r)\right) \end{aligned}$$

The last term is an enthalpy in terms of the sensor, but with the environment temperature as variable. Remembering that, we can write:

$$\begin{aligned} \dot{H}_S &= -\frac{k_{E|S}\, A_{E|S}}{(\rho\, V\, c_p)_S} \left(H_S - H_W\right) \\ &= -\frac{1}{\tau_{E|S}} \left(H_S - H_W\right) \end{aligned}$$

where in the second equation defined the time constant:

$$\tau_{E|S} := \frac{(\rho \, V \, c_p)_S}{k_{E|S} \, A_{E|S}}$$

With the mass not changing, the state is the enthalpy of the sensor: $x := H$. The input is the enthalpy of the sensor, but computed at the environment temperature: $u := H_W := (\rho \, V \, c_p)_S \; (T_r - T_E)$. In this notation we get:

$$\dot{x}_S = -\Theta_S \; (x_S - u_S)$$

I.1.3.2 Temperature as state

Most people would though not be interested in the enthalpy, but in the temperature, which is not at all surprising: After all that was the purpose of the sensor. In fact the enthalpy representation looks a little strange to most of us. The change requires a state variable transformation to be executed on the differential equation I.2. This is achieved by calculating the time derivative of the enthalpy as a function of the temperature:

$$\begin{aligned}
\frac{d}{d\,t} H_s &:= \frac{d}{d\,t}\left(\rho_S \, V_S \int_{T_r}^{T_s} (c_p(\xi))_S \, d\xi\right) \\
&:= \rho_S \, V_S \frac{d}{d\,t} \int_{T_r}^{T_s} (c_p(\xi))_S \, d\xi \\
&:= \rho_S \, V_S \left(\frac{d\,T_s \, c_p(T_S)_S}{d\,t} - \frac{d\,T_r \, c_p(T_r)_S}{d\,t} + \int_{T_r}^{T_s} \frac{d}{d\,t} (c_p(\xi))_S \, d\xi\right)
\end{aligned}$$

with the third relation being the result of applying the Leibnitz rule B.1. With c_p being constant and the reference temperature being constant, this simplifies to

$$\frac{d}{d\,t} H_s := \rho_S \, V_S \, \dot{T}_s \, (c_p)_S$$

Substitution yields:

$$\rho_S \, V_S \, \dot{T}_s \, (c_p)_S = -k_{E|S} \, A_{E|S} \, (T_S - T_E)$$

and

$$\dot{T}_s = -\frac{k_{E|S} \, A_{E|S}}{\rho_S \, V_S \, (c_p)_S} \, (T_S - T_E)$$

$$= -\frac{1}{\tau_{E|S}} \left(T_S - T_E \right)$$

Using the new definition using systems notation, $x_S := T_S$ and $u_S := T_E$ the same equation as above is obtained in this case.

$$\dot{x}_S = -\Theta_S \left(x_S - u_S \right)$$

The time constant has not changed, but the state and the input: this is now in the state space of the temperature of the sensor and the input is the temperature of the environment.

I.1.3.3 Extensive or intensive state

For the normal consumer the temperature as state and the environment temperature as input looks more familiar and to some extent more logical, though if we consider the numerics, it has also the effect that any error criteria that is applied to the numerical procedure will be on the temperature and not the enthalpy. The error in the temperature may have quite a significant different effect than the error in the enthalpy. In this context it should be considered that it is the enthalpy that is conserved and checks should be done on the conserved quantity to check that the solution is conform with the basic principle. This can be achieved by adding an output equation to the enthalpy representation, which computes the temperature from the enthalpy. It also requires an additional step at the beginning as it is the initial enthalpy to be computed from the initial temperature. So slightly more complicated in the implementation this approach has two advantages:

- No manual substitution is necessary. The sequence of the fundamental equations is merely inverted, thus compute first the temperature from the state enthalpy, then the heat flow and then the right-hand-side of the balance equation becomes available.
- The conservation principle is checked numerically.

I.2 Example: Michaelis Menten kinetics

The reactive system being the subject of analysis is given by:

$$E + S \underset{k_b}{\overset{k_f}{\rightleftharpoons}} C \overset{k_p}{\rightarrow} E + P$$

in which an enzyme (E) and a substrate (S) react to an intermediate (C), an activated substrate producing in a second step a product (P) releasing thereby the enzyme again.

The original work of Michaelis-Menten makes the assumption that the intermediate is in a pseudo-steady state, thus formation and the consumption of the intermediate is fast. From the four kinetic equations:

$$\begin{aligned}
E_1 &:: \dot{c}_S = -k_f \, c_E \, c_S + k_b \, c_C \\
E_2 &:: \dot{c}_E = -k_f \, c_E \, c_S + k_b \, c_C + k_p \, c_C \\
E_3 &:: \dot{c}_C = k_f \, c_E \, c_S - k_b \, c_C - k_p \, c_C \\
E_4 &:: \dot{c}_P = k_p \, c_C
\end{aligned}$$

The second equation is not needed as a simple balance over the enzyme species gives:

$$E_2' :: c_E = c_E^o - c_C$$

as what is not present as enzyme is present in the form of C, which is the activated substrate. Or, which is essentially the same $E_2 + E_3 = 0$

The original work assumes that the first stage is a fast equilibrium, makeing the left-hand-side of E_1 to be zero. Solving the set $\{E_1, E_2', E_4\}$ for $c_C, c_E, \dot{c}_P$ one gets the simplified reaction law:

$$\dot{c}_P = k_p \, c_E^o \, \frac{c_S}{\frac{k_b}{k_f} + c_S}$$

I.3 Evaporating Water from a Glass

I.3.1 Problem Description

We assume a idealised situation, in that a glass with constant diameter is half filled with water. The gas phase of the water up to the glass' top is stagnant and water is diffusing through this stagnant phase into open space above the glass. Latter is assumed to be uniform and not affected by the evaporating water. The room is not saturated with water and the water content is known. You may in general assume to know all material description, including the parameters to be known. The vapour pressure of water in the gas phase of the room is constant and known.

Develop a differential model which can be used to compute on how long it takes to evaporate all the water from the glass. The temperature in the glass may be assumed to be constant and equivalent to the temperature in the room. One may also assume that all physical parameters are known. Note that the diffusion process is much, much faster than the reduction of the water level. Thus look at two time scales one in which the diffusion takes place, with the level of the glass to be constant and then the long time scale in which the level changes. The first step will generate the transfer law for the second part.

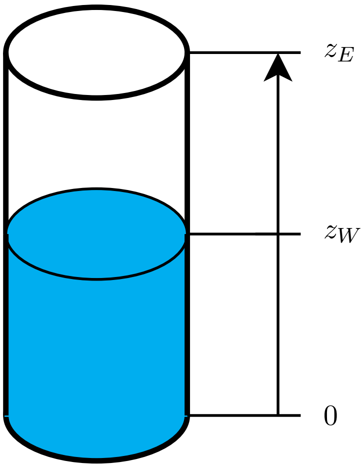


Figure I.2: A half full glass of water in a room with uniform conditions

I.3.2 Solution

The dynamics of the process has at least three different time scales:

- The fastest is the time in which the diffusion profile is being established

- The second is the time scale in which the diffusion is fast compared to the change of the volume of the water in the glass. The justification for this assumption is that the volume of the evaporated water is in 3 order of magnitudes larger in terms of the volume.

- The third is the slowest one in which the level of the water in the glass changes.

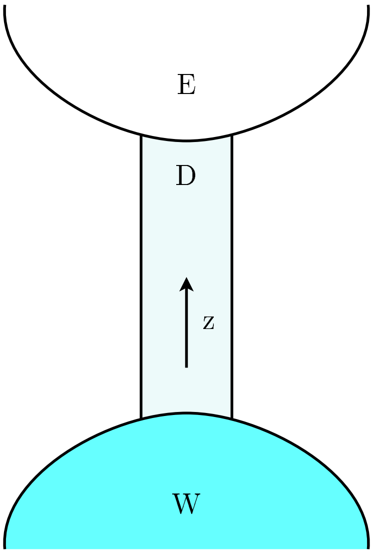


Figure I.3: A possible topology: the water is a single lump (W), which at the small time scale behaves like a reservoir whilst the gas phase (D) is seen as one-dimensionally distributed connecting to the room being modelled as a reservoir (E)

The Diffusion Equation Drawing up either an integral balance or a shell balance, the diffusion law is obtained. Taking the chemical potential as the driving force one obtains Fick's second law in complex form. The "normal" derivation would use the composition as the driving force, which can be seen as the linearised version of the transport with the chemical potential as the driving force. Since we have really only one species to worry about, even though air is diffusing in counter current to replace the water having moved, the scalar version of the diffusion equation is sufficient.

Thus one gets:

$$\frac{\partial\, c(z,t)}{\partial\, t} := c\, \frac{\partial^2\, \mu}{\partial\, z^2}\,. \tag{I.3}$$

Getting the Transfer Law: the Short Time Scale On the second level one assumes that the level of the water is constant, whilst the diffusion is fast. Premultiplying the diffusion equation with the inverse of diffusion matrix and letting the individual diffusion coefficient to go to infinity:

$$\lim_{c \to \infty} c^{-1} \frac{\partial\, c(z,t)}{\partial\, t} := \frac{\partial^2\, \mu}{\partial\, z^2}\,, \tag{I.4}$$

which puts eliminates the state of the diffusion system.

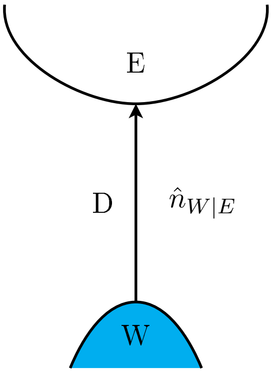


Figure I.4: On the long time scale the diffusion is seen as event-dynamic system, so it is reduced to a pure resistance, whilst the water level is now decreasing

The profile of the driving force along the length is thus a linear function:

$$\mu(z) := a\, z + b\,. \tag{I.5}$$

The two parameters a, b can be readily obtained from the boundary conditions. On the water side, the air is saturated with water damp, whilst on the top end the vapour pressure of water is fixed by the reservoir. Let the saturation pressure be p^* then it can be obtained from the equilibrium condition at the surface:

$$\mu_W := \mu_D(z_W)\,, \tag{I.6}$$

$$\mu^o_W + R T\, \ln 1 := \mu^o_D(z_W) + R T\, \ln x_G(z_W)\,. \tag{I.7}$$

Where the fact that the water is the only species in the water lump has been considered. With p_B being the known barometric pressure and

$$x^*_G(z_W) := \frac{p^*_W}{p_B}\,, \tag{I.8}$$

the saturation vapour pressure can be calculated from the known temperature and the known standard chemical potentials $\mu^o_W, \mu^o_D(z_W)$. The slope a is:

$$a := \frac{x^*_G - x_E}{z_W - z_E}\,. \tag{I.9}$$

The flow through the diffusion system is of interest:

$$\hat{n}_{W|E} := -k\, \frac{\partial\, \mu(z)}{\partial\, z}\,, \tag{I.10}$$

$$:= -k\, a\,, \tag{I.11}$$

$$:= -k\, \frac{x^*_G - x_E}{z_W - z_E}\,. \tag{I.12}$$

Now all is ready for the next bigger time scale:

Finally the Water is Evaporating The rest is simple now. The mass balance for the water lump is drawn up:

$$\dot{n}_W := -\hat{n}_{W|E}\,, \tag{I.13}$$

The flow is as we found:

$$\hat{n}_{W|E} := -k\,\frac{x_G^* - x_E}{z_W - z_E}\,. \tag{I.14}$$

Which must be supplemented with the mapping between the level and the mass:

$$V_W := \frac{n_W}{\rho_W}\,, \tag{I.15}$$

$$:= A_W\, z_W\,. \tag{I.16}$$

This set of equations is to be solved for the secondary state variable in question, namely z_W:

$$z_W := \frac{n_W}{A_W\,\rho_W}\,. \tag{I.17}$$

The rest of the variables are known: $k, z_E, x_E, \mu_L^o, \mu_G^o, R, T$. Thus the resulting set of equations is well defined if one in addition specifies the initial conditions, the problem can be integrated, actually in this case analytically.

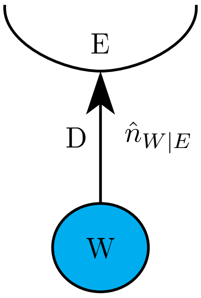


Figure I.5: Finally on the very long time scale, the volume of the water body is changing, thus the level is dropping

I.4 The Mixing Plant

I.4.1 Problem formulation

The mixing plant consists of four vessels: two feed tanks feeding the one mixing tank in the centre, which ejects the product to the storage tank.

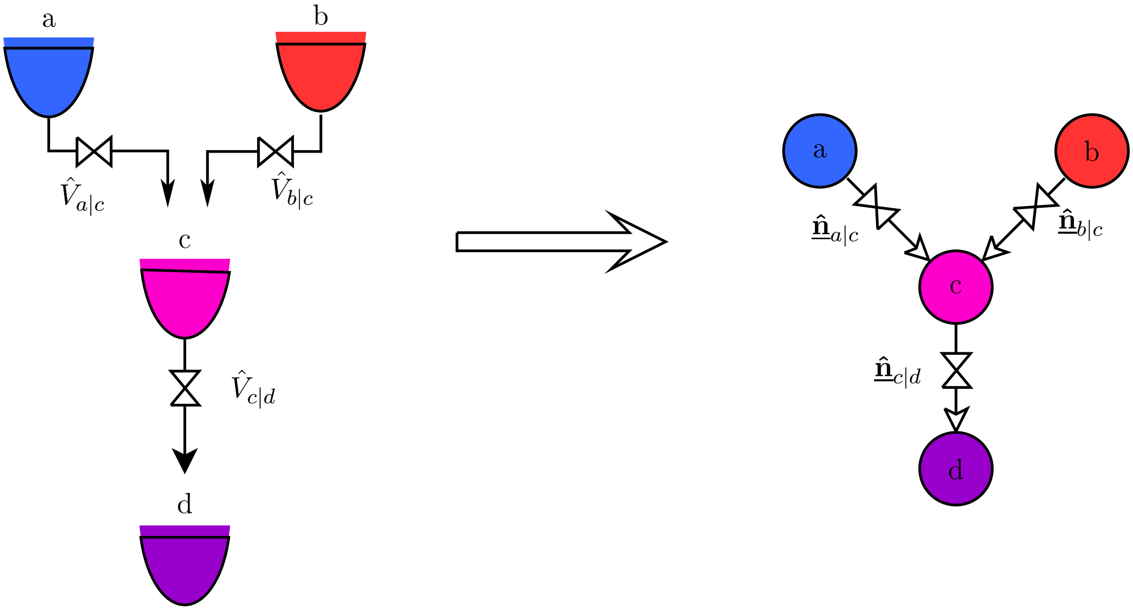

a
b
$\hat{V}_{a|c}$
$\hat{V}_{b|c}$
c
$\hat{V}_{c|d}$
d
a
b
$\hat{\underline{\mathbf{n}}}_{a|c}$
$\hat{\underline{\mathbf{n}}}_{b|c}$
c
$\hat{\underline{\mathbf{n}}}_{c|d}$
d

Figure I.6: A quite common mixing plant: two feed tanks, a mixing tank and a product tank

Generate a ”text book” representation by transforming the component mass balances into the concentration & volume space.

I.4.2 Solution

I.4.2.1 Behaviour: Component Mass Balances

The model of a tank with several inputs and outputs is described as an ideally-stirred tank reactor. The energy balance for the system is not of interest, as no exchange of energy occurs. Thus it is only the component mass balances to be established. The component mass balances are a set of ordinary differential equations in the component mass, which for this task we shall transform into differential equations in the concentration and the volume. Assuming that there is no reaction taking place in any of the tanks, the component mass balances for an arbitrary system S are:

$$\frac{d\,\underline{\mathbf{n}}_S}{d\,t} = \sum_{\forall m} \alpha_{S,m}\,\hat{\underline{\mathbf{n}}}_m + \tilde{\underline{\mathbf{n}}}_S$$

With the $\alpha_{S,m} \in \{-1, 0, +1\}$ giving the reference direction, $\hat{\underline{\mathbf{n}}}_m$ the mass flow m and $\tilde{\underline{\mathbf{n}}}_S$ the reaction dependent transformation rate. The system index $S \in [a, b, c, d]$ and the mass flow index $m \in [a|c, b|c, c|d]$.

I.4.2.2 Transfer

There is no transfer law given, but it is assumed that the volumetric flow is known, thus the transfer is given by:

$$\begin{aligned} \hat{\underline{\mathbf{n}}}_m &:= k_m\,\hat{V}_m\,\underline{\mathbf{c}}_m\,, \\ k_m &:: \text{controlled, thus known}\,, \\ \hat{V}_m &:: \text{known}\,. \end{aligned}$$

The k_m has been introduced merely to demonstrate on where the controller would be connected. In this case, the volumetric flow rate would be the maximum available and this variable would be adjusted by the controller between 0 and 1.

The concentration is the one from the tank the fluid is coming from. Mostly people assume that it may only possibly come from one tank at all time, that is, the flow direction never changes. This may or may not be a valid assumption. Here it seems though reasonable. If this is not the case, then the concentration switches as the volumetric flow changes sign!

I.4.2.3 Reaction

There is no reaction in the tank, thus

$$\tilde{\underline{\mathbf{n}}}_S := \underline{\mathbf{0}}\,.$$

I.4.2.4 State variable transformations

In this section, all variables except the fundamental state, which are the conserved quantities, are to be linked back to the fundamental state and known quantities such as the volumetric flow rate and the density. For the notation, we use s as a generic index for ”system” meaning that the equations really apply to any of them, namely the two feed tanks, the mixing tank and the product tank.

The transfer introduces the concentration. Concentration is defined by :

$$\underline{\mathbf{c}}_S := \frac{\underline{\mathbf{n}}_S}{V_S} \,,$$

Introducing volume, which is a function of the component mass, the basic state:

$$\begin{aligned} V_S &:= \rho_S^{-1}\, m_S \,, \\ \rho_S &:: \text{const} \,. \end{aligned}$$

The density is assumed constant and known. The total mass m_S is obtained as scalar product of the molecular mass vector $\underline{\boldsymbol{\lambda}}$ and the molar masses in the system:

$$m_S := \underline{\boldsymbol{\lambda}}^T\, \underline{\mathbf{n}}_S \,,$$

which completes the set of equations.

I.4.2.5 Manipulations

Since we want the differential equations in terms of the concentrations, we start with the variable transformation defining the concentration:

$$\begin{aligned} \dot{\underline{\mathbf{c}}} &:= V_S^{-1}\, \dot{\underline{\mathbf{n}}}_S - V_S^{-2}\, \dot{V}_S\, \underline{\mathbf{n}} \,, \\ &:= V_S^{-1} \left(\dot{\underline{\mathbf{n}}}_S - V_S^{-1}\, \dot{V}_S\, \underline{\mathbf{n}} \right) \,, \\ &:= V_S^{-1} \left(\dot{\underline{\mathbf{n}}}_S - \dot{V}_S\, \underline{\mathbf{c}} \right) \,, \end{aligned}$$

which is a differential equation in the desired new state $\underline{\mathbf{c}}$, but also $\dot{V}_S$ Thus in the next step is to differentiate the equation defining the volume. We also use the assumption that the density is constant in the whole of the plant.

$$\begin{aligned} \dot{V}_S &:= \rho^{-1}\, \dot{m}_S \,, \\ &:= \rho^{-1}\, \underline{\boldsymbol{\lambda}}^T\, \dot{\underline{\mathbf{n}}}_S \,, \\ &:= \rho^{-1}\, \underline{\boldsymbol{\lambda}}^T \left(\sum_{\forall m} \alpha_{S,m}\, \hat{\underline{\mathbf{n}}}_m \right) \,, \\ &:= \sum_{\forall m} \alpha_{S,m}\, \rho^{-1}\, \underline{\boldsymbol{\lambda}}^T\, \hat{\underline{\mathbf{n}}}_m \,, \end{aligned}$$

$$:= \sum_{\forall m} \alpha_{S,m} \, k_m \, \hat{V}_m \, ,$$

For the change in the composition one finds:

$$\begin{aligned}
\underline{\dot{\mathbf{c}}} &:= V_S^{-1} \left(\sum_{\forall m} \alpha_{S,m} \, \underline{\hat{\mathbf{n}}}_m - \left(\sum_{\forall m} \alpha_{S,m} \, \hat{V}_m \right) \underline{\mathbf{c}}_S \right) \, , \\
&:= V_S^{-1} \left(\sum_{\forall m} \alpha_{S,m} \, k_m \, \hat{V}_m \, \underline{\mathbf{c}}_m - \left(\sum_{\forall m} \alpha_{S,m} \, k_m \, \hat{V}_m \right) \underline{\mathbf{c}}_S \right) \, , \\
&:= V_S^{-1} \left(\sum_{\forall m} \alpha_{S,m} \, k_m \, \hat{V}_m \, (\underline{\mathbf{c}}_m - \underline{\mathbf{c}}_S) \right) \, .
\end{aligned}$$

The result is now in generic form and can be applied to any of the tanks. In the case of no inflow, which represents any of the two feed tanks, the concentration change becomes zero, as expected, as the flow concentration $\underline{\mathbf{c}}_m == \underline{\mathbf{c}}_S$.

The ratio $V_S^{-1} \, \hat{V}_m$ are the time constants with respect to the various flows in and out the system S.

Thus the complete model reads:

$$\begin{aligned}
\frac{d\,V_a}{d\,t} &= -k_{a|c} \, \hat{V}_{a|c} \, , \\
\frac{d\,\underline{\mathbf{c}}_a}{d\,t} &= \underline{\mathbf{0}} \, , \\
\frac{d\,V_b}{d\,t} &= -k_{b|c} \, \hat{V}_{b|c} \, , \\
\frac{d\,\underline{\mathbf{c}}_b}{d\,t} &= \underline{\mathbf{0}} \, , \\
\frac{d\,V_c}{d\,t} &= k_{a|c} \, \hat{V}_{a|c} + k_{b|c} \, \hat{V}_{b|c} - k_{c|d} \, \hat{V}_{c|d} \, , \\
\frac{d\,\underline{\mathbf{c}}_c}{d\,t} &= V_c^{-1} \left(k_{a|c} \, \hat{V}_{a|c} \, (\underline{\mathbf{c}}_a - \underline{\mathbf{c}}_c) + k_{b|c} \, \hat{V}_{b|c} \, (\underline{\mathbf{c}}_b - \underline{\mathbf{c}}_c) \right) \, , \\
\frac{d\,V_d}{d\,t} &= k_{c|d} \, \hat{V}_{c|d} \, , \\
\frac{d\,\underline{\mathbf{c}}_d}{d\,t} &= V_d^{-1} \left(k_{c|d} \, \hat{V}_{c|d} \, (\underline{\mathbf{c}}_c - \underline{\mathbf{c}}_d) \right) \, .
\end{aligned}$$

One of the key assumptions in the derivation is that the density is constant. It is left to the reader to derive the equations for the situation where the density is a function of the mole fraction of the mixture.

I.4.2.6 Systems Representation

We define the following vectors:

- state vector $\underline{\mathbf{x}}_S := \begin{bmatrix} V_S \\ \underline{\mathbf{c}}_S \end{bmatrix}$, $S \in \{a, b, c, d\}$

- input vector $\underline{\mathbf{u}} := \begin{bmatrix} k_{a|c} \\ k_{b|c} \\ k_{c|d} \end{bmatrix}$

- output vector $\underline{\mathbf{y}}_S := \underline{\mathbf{x}}_S$, $S \in \{a, b, c, d\}$

- conditions $\underline{\gamma} := \begin{bmatrix} \underline{\mathbf{c}}_a \\ \underline{\mathbf{c}}_b \\ \hat{V}_{a|c} \\ \hat{V}_{b|c} \\ \hat{V}_{c|d} \end{bmatrix}$

- parameters: there are no real parameters. The distinction between parameters and conditions is though not quite sharp. We use the rule that if it is a state that is known than it is a condition.

Remains to use these definitions and rewrite the equations in this new notation:

$$\begin{bmatrix} \dot{x}_a \\ \dot{\underline{\mathbf{x}}}_a \\ \dot{x}_b \\ \dot{\underline{\mathbf{x}}}_b \\ \dot{x}_c \\ \dot{\underline{\mathbf{x}}}_c \\ \dot{x}_d \\ \dot{\underline{\mathbf{x}}}_d \end{bmatrix} = \begin{bmatrix} -u_1\,\gamma_3 \\ \underline{\mathbf{0}} \\ -u_2\,\gamma_4 \\ \underline{\mathbf{0}} \\ u_1\,\gamma_3 + u_2\,\gamma_4 - u_3\,\gamma_5 \\ x_c^{-1}\,\left(u_1\,\gamma_3\,\left(\underline{\mathbf{x}}_a - \underline{\mathbf{x}}_c\right) + u_2\,\gamma_4\,\left(\underline{\mathbf{x}}_b - \underline{\mathbf{x}}_c\right)\right) \\ u_3\,\gamma_5 \\ x_d^{-1}\,\left(u_3\,\gamma_5\,\left(\underline{\mathbf{x}}_c - \underline{\mathbf{x}}_d\right)\right) \end{bmatrix} . \tag{I.18}$$

I.4.2.7 More complex case: variable density

The manipulations can be done slightly differently. Let us recall the equations first:

$$\begin{aligned} \frac{d\,\underline{\mathbf{n}}_S}{d\,t} &= \sum_{\forall m} \alpha_{S,m}\,\hat{\underline{\mathbf{n}}}_m + \tilde{\underline{\mathbf{n}}}_S \\ \hat{\underline{\mathbf{n}}}_m &:= k_m\,\hat{V}_m\,\underline{\mathbf{c}}_m\,, \\ \underline{\mathbf{c}}_S &:= \frac{\underline{\mathbf{n}}_S}{V_S} \\ V_S &:= \frac{n_S}{\rho_S}\,, \\ n_S &:= \underline{\mathbf{e}}^T\,\underline{\mathbf{n}}_S\,, \\ \rho &:= \rho(\underline{\mathbf{n}}) \end{aligned}$$

The objective is to switch the state space from $\underline{\mathbf{n}}$ to $\underline{\mathbf{c}}$ but as the latter is only including intensive variables an extensive must be added, which is V. So in recognition of the latter and the structure of the above algebraic equations we rearrange the algebraic equations slightly and differentiate yielding the equation set:

$$\begin{aligned} \dot{\underline{\mathbf{n}}} &:= \dot{\underline{\mathbf{c}}}\,V + \underline{\mathbf{c}}\,\dot{V} && \to \dot{c} \\ \underline{\mathbf{e}}^T\,\dot{\underline{\mathbf{n}}} &:= \dot{\rho}(\underline{\mathbf{n}})\underline{\mathbf{V}} + \rho\,\dot{V} && \to \dot{V} \end{aligned}$$

$$\dot{\rho(\underline{\mathbf{n}})} := \frac{\partial\,\rho}{\partial\,\underline{\mathbf{n}}^T}\dot{\underline{\mathbf{n}}} \qquad\qquad \rightarrow \dot{\rho}(\underline{\mathbf{n}})$$

The last being trivial we get:

$$\begin{aligned}
\dot{\underline{\mathbf{c}}} &:= V^{-1}\left(\dot{\underline{\mathbf{n}}} - \underline{\mathbf{c}}\,\dot{V}\right) \\
\dot{V} &:= \rho^{-1}\left(\underline{\mathbf{e}}^T\,\dot{\underline{\mathbf{n}}} - \frac{\partial\,\rho}{\partial\,\underline{\mathbf{n}}^T}\dot{\underline{\mathbf{n}}}\,V\right) \\
&:= \rho^{-1}\left(\underline{\mathbf{e}}^T\,\dot{\underline{\mathbf{n}}} - \frac{\partial\,\rho}{\partial\,\underline{\mathbf{n}}^T}\,\dot{\underline{\mathbf{n}}}\right) \\
&:= \rho^{-1}\left(\underline{\mathbf{e}}^T - \frac{\partial\,\rho}{\partial\,\underline{\mathbf{n}}^T}\right)\,\dot{\underline{\mathbf{n}}}
\end{aligned}$$

substituting the right-hand-side of the component mass balances gives the desired result.

It is noteworthy that for a program the remaining substitutions must not be done. Given the state $\underline{\mathbf{c}}, V$ and the conditions and inputs, we first compute the left-hand-side of the component mass balance. This is then used to compute the change in the density, the change in the volume and finally the change in the composition, thereby completing the scheme.

I.5 Marinading a Steak

Diffusion processes are quite common in nature. Actually they are present almost always when solids meet gases or liquids or gases meet liquids. Marinading a steak is thus just a practical example for a large class of processes.

We shall look at a very simple case in which we assume that the marinade consists essentially of water and salt with the salt diffusing into the meat, which in tern is assumed to be essentially stationary water. The geometry of the problem is simplified in that the steak is assumed to have only two active surfaces, namely the two big ones, whilst we assume that the sides are sealed with for example fat. In a first case we place the steak flat on the floor of a pan topping it up with marinade.

I.5.1 Step 0 Abstraction

The process is sketched quickly I.7

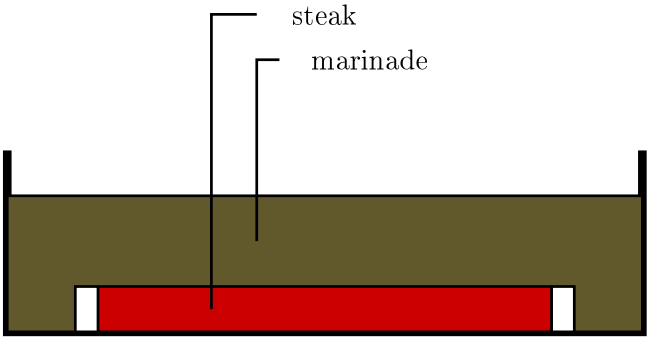


Figure I.7: A steak laying flat in a pan

However, depending on the time scale we choose the abstraction could look quite differently.

In the first case (I.8) we look at a relatively short time scale assuming that the marinade is not changing over time, thus the exchange with the steak is negligible. A diffusion film is assumed to form on the surface of the steak whilst the diffusion really does not penetrate the steak significantly.

In second case (I.9) a larger time scale is considered where the marinade composition is still not changing, but the fluid film is considered unimportant compared to the mixing in the marinade. The steak is modelled as a one-timensional diffusion medium.

In the third case (I.10) the marinade concentration is considered to change.

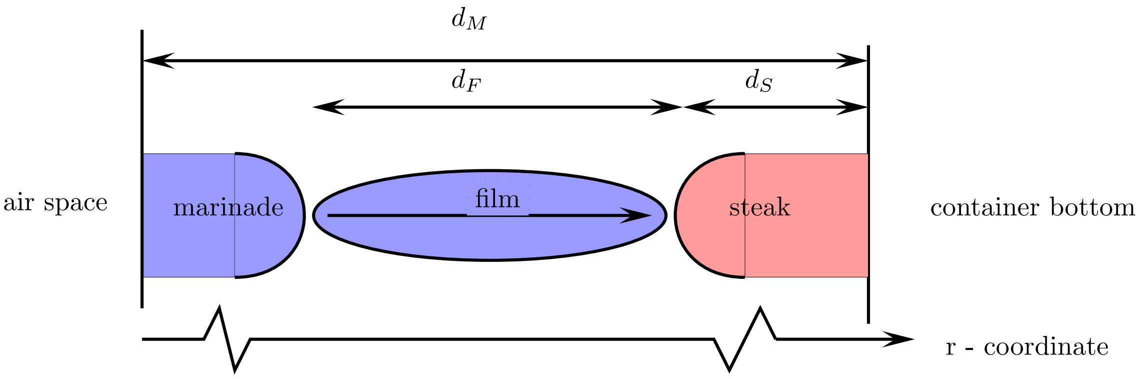


Figure I.8: ***Case 1:*** *Topology assuming the marinade to be well mixed and not changing with time. The transfer system, being the film is 1D distributed*

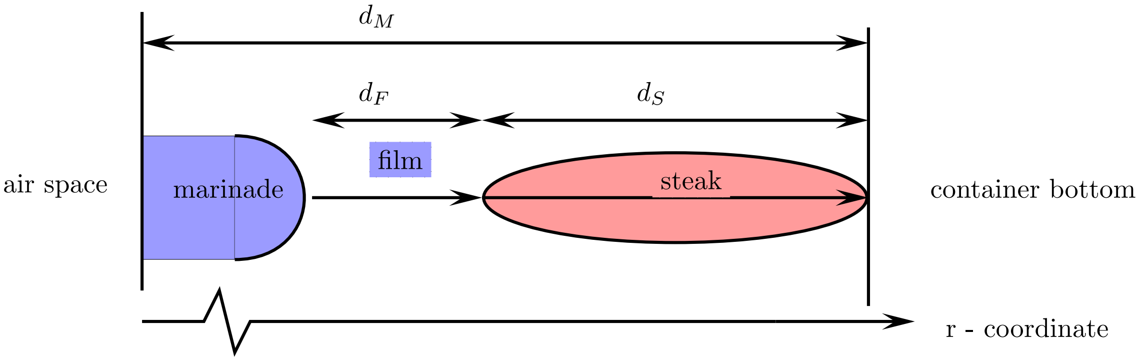


Figure I.9: ***Case 2:*** *Topology assuming the marinade to be well mixed and not changing with time. The film is assumed to be in steady state, whilst the steak is modelled as a 1D distributed system*

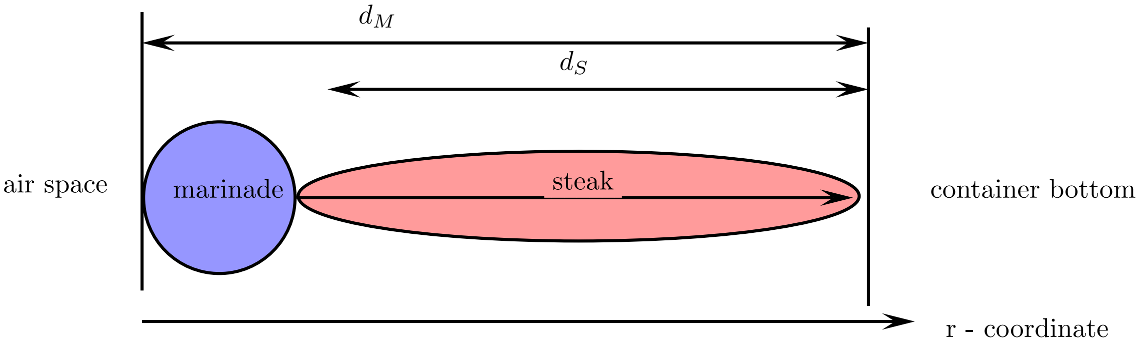


Figure I.10: ***Case 3:*** *Topology assuming the marinade to be well mixed but now changing with time. The steak is 1D distributed*

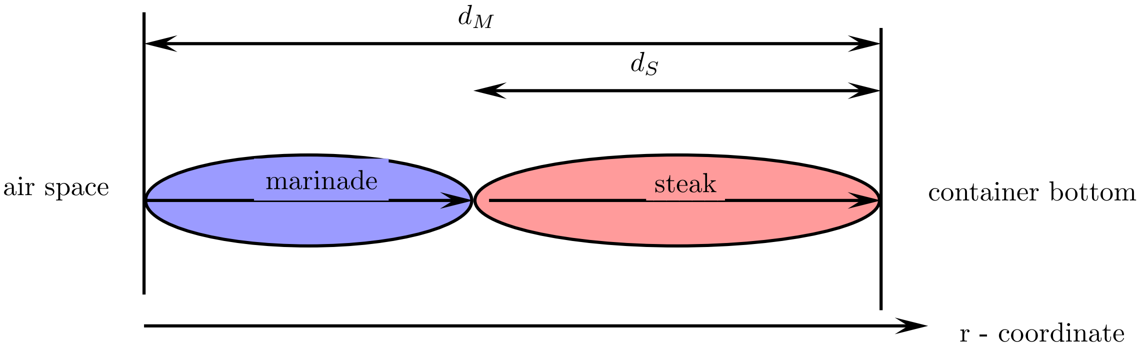


Figure I.11: ***Case 4:*** *Topology assuming both, the marinade and the steak to be 1D distributed*

In the fourth case (I.11) one assumes that the marinade is not moving at all but behaves like a one-timensional diffusion medium.

Below we shall discuss case 3 and 4.

I.5.2 Step 1: Behaviour

We assume the steak to be of uniform thickness d_S and the marinade being of depth d_M. The air space is denoted by A and the container bottom C. Further, we introduce a co-ordinates system, whereby the problem is considered one-dimensional. The co-ordinate is labelled with r see I.8, I.9, I.10, I.11.

I.5.2.1 Case 3

Labelling the marinade with subscript M and the steak with S the behaviour for case 3 is given by:

$$\begin{aligned}
\underline{\dot{\mathbf{n}}}_M &= -\underline{\hat{\mathbf{n}}}_{M|S}\,, \\
\frac{\partial\,\underline{\mathbf{c}}_S}{\partial\,t} &= \underline{\underline{\mathbf{K}}}_S\,\frac{\partial^2\,\underline{\mu}_S}{\partial\,r^2}\,, \\
\text{eq BC} \quad \underline{\mu}_M(d_M - d_S - \epsilon) &= \underline{\mu}_S(d_M - d_S + \epsilon)\,, \\
\text{flow BC} \quad \underline{\hat{\mathbf{n}}}_{M|S}(d_M - d_S - \epsilon) &= \underline{\hat{\mathbf{n}}}_{M|S}(d_M - d_S + \epsilon)\,, \\
\text{flow BC} \quad \underline{\hat{\mathbf{n}}}_{S|C}(d_M) &= \underline{\mathbf{0}}\,.
\end{aligned}$$

The boundary conditions act as coupling equations. At the interface to the marinade, the boundary conditions reflect continuity in the chemical potential and the mass flow, whilst on the other side of the steak the boundary conditions merrely says that the flow is zero.

I.5.2.2 Case 4

For case 4, the well-mixed assumption for the marinade is replaced by a no mixing, purely diffusion assumption:

$$\begin{aligned}
\frac{\partial\,\underline{\mathbf{c}}_M}{\partial\,t} &= \underline{\underline{\mathbf{K}}}_M\,\frac{\partial^2\,\underline{\mu}_M}{\partial\,r^2}\,, \\
\frac{\partial\,\underline{\mathbf{c}}_S}{\partial\,t} &= \underline{\underline{\mathbf{K}}}_S\,\frac{\partial^2\,\underline{\mu}_S}{\partial\,r^2}\,, \\
\text{eq BC} \quad \underline{\mu}_M(d_M - d_S - \epsilon) &= \underline{\mu}_S(d_M - d_S + \epsilon)\,, \\
\text{flow BC} \quad \underline{\hat{\mathbf{n}}}_{M|S}(d_M - d_S - \epsilon) &= \underline{\hat{\mathbf{n}}}_{M|S}(d_M - d_S + \epsilon)\,,
\end{aligned}$$

$$\begin{aligned} \text{flow BC} \quad \underline{\hat{\mathbf{n}}}_{A|M}(0) &= \underline{\mathbf{0}}\,, \\ \text{flow BC} \quad \underline{\hat{\mathbf{n}}}_{S|C}(d_M) &= \underline{\mathbf{0}}\,. \end{aligned}$$

In both cases the initial conditions must be supplemented.

I.5.3 Step 2a: Transport

The transport equation is simply the gradient law in both media:

$$\underline{\hat{\mathbf{n}}}_{M|S} := -\underline{\underline{\mathbf{C}}}\, A \, \frac{\partial \, \underline{\mu}}{\partial \, r}\,.$$

With the transport properties $\underline{\underline{\mathbf{C}}}$ and the boundary surface A being given.

I.5.4 Step 3: Variable Transformation

The transport introduces the chemical potential. Thas what we require is a mapping of the conserved state variables to the chemical potential. Using the model:

$$\underline{\mu} := \underline{\mu}^o + R\,T \, \ln \underline{\mathbf{x}}\,.$$

This introduces the mole fractions, which need to be the result of mapping the component mass:

$$\underline{\mathbf{x}} := n^{-1}\, \underline{\mathbf{n}}\,.$$

And

$$n := \underline{\mathbf{e}}^T \, \underline{\mathbf{n}}\,.$$

With $\underline{\mathbf{e}}^T := [1, 1, \ldots, 1]$.
The distributed models require the concentration, so we add:

$$\begin{aligned} \underline{\mathbf{c}} &:= V^{-1}\, \underline{\mathbf{n}}\,, \\ V &:= \rho^{-1}\, n\,. \end{aligned}$$

Assuming the density ρ to be constant and known completes the transformation defintions.

I.5.5 Step 4: Conditions

There is no reaction taking place and the temperature is assumed to be constant. With the chemical potentials at normal conditions being given, the set of transformations is complete.

I.5.6 Step 6: Manipulations

The partial differential equations are being discretised in the spatial coordinate using a 3-point approximation and using the index k for the points on the regular grid of width Δr (**??**). The discretisations for case 4 is introducing the indexing scheme 0,1,2,. . . , n, n+1,. . . , n+m with point 0 representing the outer surface of the marinade (no flow condition), n representing the boundary to the steak, and n+m the outer surface of the steak (no flow condition). Obviously for case 3 this simplifies by having n = 0.

For the internal points, we thus write:

$$\left.\frac{\partial^2 \underline{\mu}_M}{\partial\, r^2}\right|_k := \frac{\underline{\mu}_{k-1} - 2\,\underline{\mu}_k + \underline{\mu}_{k+1}}{(\Delta\, r)^2}\,.$$

At the extreme points $(0, d_M)$, the equation writes:

$$\left.\frac{\partial^2 \underline{\mu}_M}{\partial\, r^2}\right|_0 := \frac{\underline{\mu}_0 - 2\,\underline{\mu}_1 + \underline{\mu}_2}{(\Delta\, r)^2}\,,$$
$$\left.\frac{\partial^2 \underline{\mu}_M}{\partial\, r^2}\right|_{n+m} := \frac{\underline{\mu}_{n+m-2} - 2\,\underline{\mu}_{n+m-1} + \underline{\mu}_{n+m}}{(\Delta\, r)^2}\,.$$

which is supplemented with the no-flow condition:

$$\hat{\underline{\mathbf{n}}}_{M|S}(0) := \frac{\underline{\mu}_1 - \underline{\mu}_0}{\Delta\, r} := 0\,,$$
$$\hat{\underline{\mathbf{n}}}_{M|S}(d_M) := \frac{\underline{\mu}_{n+m} - \underline{\mu}_{n+m-1}}{\Delta\, r} := 0\,.$$

Alternatively, one can model the extreme boundary points slightly differently by introducing the no flow condition indirectly. One introduces a imaginary point outside the boundary and introduces the flow condition by assuming symmetry at the boundary, thereby implying a zero flow condition at the boundary:

$$\left.\frac{\partial^2 \underline{\mu}_M}{\partial\, r^2}\right|_0 := \frac{\underline{\mu}_{-1} - 2\,\underline{\mu}_0 + \underline{\mu}_1}{(\Delta\, r)^2}\,,$$
$$\left.\frac{\partial^2 \underline{\mu}_M}{\partial\, r^2}\right|_{n+m} := \frac{\underline{\mu}_{n+m-1} - 2\,\underline{\mu}_{n+m} + \underline{\mu}_{n+m+1}}{(\Delta\, r)^2}\,,$$

With the symmetry $\underline{\mu}_{-1} = \underline{\mu}_1$ and $\underline{\mu}_{n+m-1} = \underline{\mu}_{n+m+1}$ the two expressions simplify to:

$$\left.\frac{\partial^2 \underline{\mu}_M}{\partial r^2}\right|_0 := \frac{-2\,\underline{\mu}_0 + 2\,\underline{\mu}_1}{(\Delta\, r)^2}\,,$$

$$\left.\frac{\partial^2 \underline{\mu}_M}{\partial r^2}\right|_{n+m} := \frac{2\,\underline{\mu}_{n+m-1} - 2\,\underline{\mu}_{n+m}}{(\Delta\, r)^2}\,,$$

I.5.6.1 Case 3

Observing that $n := 0$ for case 3 we have the ordinary differential equation describing the behaviour of the tank representing the marinade:

$$\dot{\underline{\mathbf{n}}}_0 := -\underline{\underline{\mathbf{C}}}_S\, A\, \frac{\underline{\mu}_1 - \underline{\mu}_0}{\Delta r}\,.$$

And the matrix equation for the steak:

$$\begin{bmatrix} \dot{\underline{\mathbf{c}}}_1 \\ \dot{\underline{\mathbf{c}}}_2 \\ \vdots \\ \dot{\underline{\mathbf{c}}}_m \end{bmatrix} := \begin{bmatrix} -2\underline{\underline{\mathbf{S}}} & \underline{\underline{\mathbf{S}}} & & & \\ \underline{\underline{\mathbf{S}}} & -2\underline{\underline{\mathbf{S}}} & \underline{\underline{\mathbf{S}}} & & \\ & \ddots & \ddots & \ddots & \\ & & \underline{\underline{\mathbf{S}}} & -2\underline{\underline{\mathbf{S}}} & \underline{\underline{\mathbf{S}}} \\ & & & 2\underline{\underline{\mathbf{S}}} & -2\underline{\underline{\mathbf{S}}} \end{bmatrix} \begin{bmatrix} \underline{\mu}_1 \\ \underline{\mu}_2 \\ \vdots \\ \underline{\mu}_{m-1} \\ \underline{\mu}_m \end{bmatrix} + \begin{bmatrix} \underline{\mu}_0 \\ \underline{\mathbf{0}} \\ \vdots \\ \underline{\mathbf{0}} \\ \underline{\mathbf{0}} \end{bmatrix}\,,$$

,

with $\underline{\underline{\mathbf{S}}} := \Delta r^{-2}\, \underline{\underline{\mathbf{K}}}_S$.

I.5.6.2 Case 4

For case 4 we first have sort out the boundary between the two phases by computing the missing $\underline{\mu}_n$ from the boundary condition:

$$-(\Delta r_M)^{-1}\, \underline{\underline{\mathbf{C}}}_M\, (\underline{\mu}_n - \underline{\mu}_{n-1}) := -(\Delta r_S)^{-1}\, \underline{\underline{\mathbf{C}}}_S\, (\underline{\mu}_{n+1} - \underline{\mu}_n)\,.$$

with $\underline{\underline{\mathbf{R}}} := \frac{\Delta r_S}{\Delta r_M}\, \underline{\underline{\mathbf{C}}}_S^{-1}\, \underline{\underline{\mathbf{C}}}_M$. Which gives:

$$\underline{\mu}_n := \left(\underline{\underline{\mathbf{R}}} + \underline{\underline{\mathbf{I}}}\right)^{-1} \left(\underline{\mu}_{n-1} + \underline{\underline{\mathbf{R}}}\,\underline{\mu}_{n+1}\right)\,.$$

This then substituted into the expressions for the approximations for the two points left and right the boundary:

$$\begin{aligned}
\left.\frac{\partial^2 \underline{\mu}}{\partial r^2}\right|_{n-1} &:= \frac{\underline{\mu}_{n-2} - 2\,\underline{\mu}_{n-1} + \underline{\mu}_{n}}{\Delta r^2}\,,\\
&:= \frac{\underline{\mu}_{n-2} - 2\,\underline{\mu}_{n-1} + \left(\underline{\underline{\mathbf{R}}} + \underline{\underline{\mathbf{I}}}\right)^{-1}\left(\underline{\mu}_{n-1} + \underline{\underline{\mathbf{R}}}\,\underline{\mu}_{n+1}\right)}{\Delta r^2}\,,\\
&:= \frac{\underline{\mu}_{n-2} + \left(\left(\underline{\underline{\mathbf{R}}} + \underline{\underline{\mathbf{I}}}\right)^{-1} - 2\,\underline{\underline{\mathbf{I}}}\right)\underline{\mu}_{n-1} + \left(\underline{\underline{\mathbf{R}}} + \underline{\underline{\mathbf{I}}}\right)^{-1}\underline{\underline{\mathbf{R}}}\,\underline{\mu}_{n+1}}{\Delta r^2}\,,
\end{aligned}$$

and

$$\begin{aligned}
\left.\frac{\partial^2 \underline{\mu}}{\partial r^2}\right|_{n+1} &:= \frac{\underline{\mu}_{n} - 2\,\underline{\mu}_{n+1} + \underline{\mu}_{n+2}}{\Delta r^2}\,,\\
&:= \frac{\left(\underline{\underline{\mathbf{R}}} + \underline{\underline{\mathbf{I}}}\right)^{-1}\left(\underline{\mu}_{n-1} + \underline{\underline{\mathbf{R}}}\,\underline{\mu}_{n+1}\right) - 2\,\underline{\mu}_{n+1} + \underline{\mu}_{n+2}}{\Delta r^2}\,,\\
&:= \frac{\left(\underline{\underline{\mathbf{R}}} + \underline{\underline{\mathbf{I}}}\right)^{-1}\underline{\mu}_{n-1} + \left(\left(\underline{\underline{\mathbf{R}}} + \underline{\underline{\mathbf{I}}}\right)^{-1}\underline{\underline{\mathbf{R}}}\,\underline{\mu}_{n+1} - 2\,\underline{\underline{\mathbf{I}}}\right)\underline{\mu}_{n+1} + \underline{\mu}_{n+2}}{\Delta r^2}\,,
\end{aligned}$$

The equations can be collected into a matrix representation:

$$\underline{\dot{\mathbf{c}}} := \underline{\underline{\mathbf{L}}}\,\underline{\mu}\,,$$

with $$\underline{\dot{\mathbf{c}}} := \begin{bmatrix} \underline{\dot{\mathbf{c}}}_0 \\ \underline{\dot{\mathbf{c}}}_1 \\ \vdots \\ \underline{\dot{\mathbf{c}}}_{n-2} \\ \underline{\dot{\mathbf{c}}}_{n-1} \\ \underline{\dot{\mathbf{c}}}_{n+1} \\ \underline{\dot{\mathbf{c}}}_{n+2} \\ \vdots \\ \underline{\dot{\mathbf{c}}}_{n+m} \end{bmatrix}, \qquad \text{and} \qquad \underline{\mu} := \begin{bmatrix} \underline{\mu}_0 \\ \underline{\mu}_1 \\ \vdots \\ \underline{\mu}_{n-2} \\ \underline{\mu}_{n-1} \\ \underline{\mu}_{n+1} \\ \underline{\mu}_{n+2} \\ \vdots \\ \underline{\mu}_{m+n-1} \\ \underline{\mu}_{m+n} \end{bmatrix},$$

$$\underline{\underline{\mathbf{L}}} := \begin{bmatrix} -2\underline{\underline{\mathbf{M}}} & 2\underline{\underline{\mathbf{M}}} & & & & & & & & \\ \underline{\underline{\mathbf{M}}} & -2\underline{\underline{\mathbf{M}}} & \underline{\underline{\mathbf{M}}} & & & & & & & \\ & \ddots & \ddots & \ddots & & & & & & \\ & & \underline{\underline{\mathbf{M}}} & -2\underline{\underline{\mathbf{M}}} & \underline{\underline{\mathbf{M}}} & & & & & \\ & & & \underline{\underline{\mathbf{M}}} & \underline{\underline{\mathbf{M}}}\underline{\underline{\mathbf{Q}}}_1 & \underline{\underline{\mathbf{M}}}\underline{\underline{\mathbf{Q}}}_2 & & & & \\ & & & & \underline{\underline{\mathbf{S}}}\underline{\underline{\mathbf{Q}}}_2 & \underline{\underline{\mathbf{S}}}\underline{\underline{\mathbf{Q}}}_1 & \underline{\underline{\mathbf{S}}} & & & \\ & & & & & \underline{\underline{\mathbf{S}}} & -2\underline{\underline{\mathbf{S}}} & \underline{\underline{\mathbf{S}}} & & \\ & & & & & & \ddots & \ddots & \ddots & \\ & & & & & & & \underline{\underline{\mathbf{S}}} & -2\underline{\underline{\mathbf{S}}} & \underline{\underline{\mathbf{S}}} \\ & & & & & & & & 2\underline{\underline{\mathbf{S}}} & -2\underline{\underline{\mathbf{S}}} \end{bmatrix}.$$

Where

$$\begin{aligned}
\underline{\underline{\mathbf{M}}} &:= \Delta r_M^{-2}\, \underline{\underline{\mathbf{K}}}_M\,, \\
\underline{\underline{\mathbf{S}}} &:= \Delta r_S^{-2}\, \underline{\underline{\mathbf{K}}}_S\,, \\
\underline{\underline{\mathbf{Q}}}_1 &:= \left(\underline{\underline{\mathbf{R}}} + \underline{\underline{\mathbf{I}}}\right)^{-1} - 2\,\underline{\underline{\mathbf{I}}}\,, \\
\underline{\underline{\mathbf{Q}}}_2 &:= \left(\underline{\underline{\mathbf{R}}} + \underline{\underline{\mathbf{I}}}\right)^{-1} \underline{\underline{\mathbf{R}}}\,.
\end{aligned}$$

I.12 show some results from simulations.

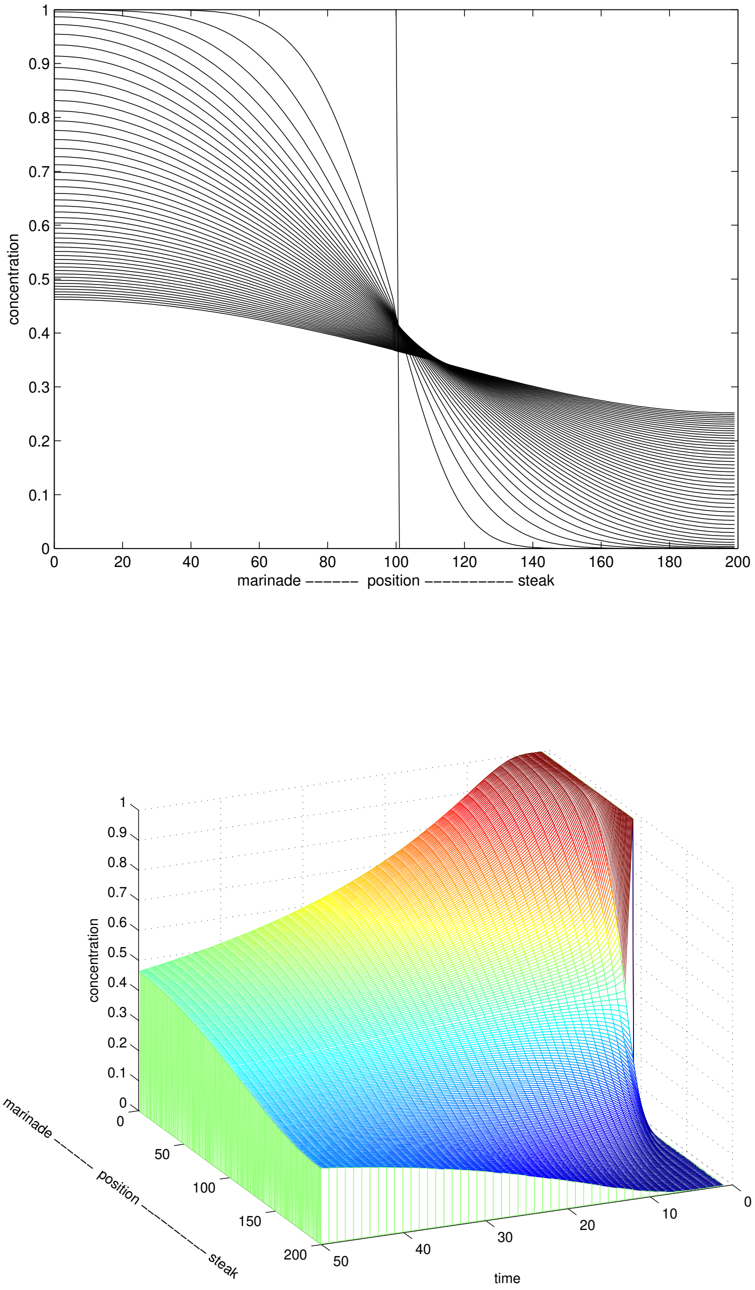


Figure I.12: Mesh plot for marinading steak problem

I.6 First-Order Single-Input-Single-Output System

Single-Input-Single-Output, appreviated with SISO, are systems that are scalar at both ends, so-to-speak, which however does not imply that the state is scalar as well.

The $\{\underline{\underline{\mathbf{A}}}, \underline{\underline{\mathbf{B}}}, \underline{\underline{\mathbf{C}}}, \underline{\underline{\mathbf{D}}}\}$ representation of a generic SISO LTI system is

$$\begin{aligned} \dot{\underline{\mathbf{x}}} &:= \underline{\underline{\mathbf{A}}}\, \underline{\mathbf{x}} + \underline{\mathbf{b}}\, u \\ y &:= \underline{\mathbf{c}}^T\, \underline{\mathbf{x}} + d\, u \end{aligned}$$

The single input is mapped onto the state with the vector $\underline{\mathbf{b}}$ thereby taking the rôle of the matrix $\underline{\underline{\mathbf{B}}}$ and the state is mapped onto the single output by $\underline{\mathbf{c}}^T$ taking the rôle of the matrix $\underline{\underline{\mathbf{C}}}$. The input acting directly on the output is amplified with the scalar d. The transfer function of the SISO is then:

$$g(s) := \underline{\mathbf{c}}^T \left|\underline{\underline{\mathbf{I}}}\, s - \underline{\underline{\mathbf{A}}}\right|^{-1} \operatorname{adj}\left(\underline{\underline{\mathbf{I}}}\, s - \underline{\underline{\mathbf{A}}}\right)\, \underline{\mathbf{b}} + d$$

I.6.1 Scalar-State Case

Making also the state scalar yields the structurally simplest model one can generate without eliminating one or the other system "matrices". The most common additional simplification is the case where d is zero. The transfer function then consists of scalar quantites only and reads:

$$\begin{aligned} g(s) &:= c\,(s-a)^{-1}\, b \\ &:= \frac{b\,c}{-a}\left(-\frac{1}{a}\, s + 1\right)^{-1} \end{aligned}$$

For stable system the $a < 0$, thus the time constant $-\frac{1}{a}$ and the steady-state gain $\frac{b\,c}{-a}$ are positiv.

This first-order sytem is used in various applications as a first approximation for a dynamic behaviour. Applying it to the description of a physical process requires finding two parameters, namely the steady-state gain and the time constant. The identification experiment must excite the system sufficiently dynamic in order to "see" the behaviour of the plant. Probably the most common approach, though not necessarily the best one, is to inject a step and extract the two parameters for the plant's input and the plant's response, called step response.

The fitting is most commonly done manually, meaning on a graph showing the step input and the plant's response. How one can find the two parameters from the response is easy to find from an analysis of the analytical solution.

The solution in the time domain is:

$$y(t) := c\,\exp\{a\,t\}\,x(o) + c\int_0^t \exp\{a\,\theta\}\;b\,u(t-\theta)\,d\theta$$

With $u(t)$ being a step, and assuming that the plant is at zero state initially, the expression simplifies to

$$\begin{aligned} y(t) &:= \frac{c\,b}{a}\,\exp\{a\,\theta\}|_0^t\;u^0 \\ y(t) &:= \frac{c\,b}{a}\,(\exp\{a\,t\} - 1)\;u^0 \end{aligned}$$

For stable plants, that is $a < 0$ the steady state gain is thus:

$$\begin{aligned} k &:= \frac{c\,b}{-a} \\ &:= \frac{c\,b}{|a|} \end{aligned}$$

The tangent at the start of the step response of magnitude u^0 is:

$$\begin{aligned} c\,\dot{x}(t := 0) &:= c\,a\,x(0) + c\,b\,u^0 \\ &:= c\,b\,u^0 \end{aligned}$$

The two asymtodes (tangent at zero and the tangent to the steady state at infinity) to the step response are thus:

$$\begin{aligned} v(t) &:= \frac{c\,b}{|a|}\,u^0 \\ w(t) &:= c\,b\,u^0\,t \end{aligned}$$

and their intersection $t^\times$:

$$\begin{aligned} \frac{c\,b}{|a|}\,u^0 &:= c\,b\,u^0\,t^\times \\ t^\times &:= \frac{1}{|a|} \end{aligned}$$

which is the time constant τ. Interesting is also how far the process has come after n times the time constant:

$$\begin{aligned} y\,(n\,\tau) &:= \frac{c\,b}{a}\left(\exp\left\{a\,n\,\frac{1}{|a|}\right\} - 1\right)\;u^0 \\ &:= \frac{c\,b}{|a|}\,(1 - \exp\{-\,n\})\;u^0 \end{aligned}$$

For

$$\begin{aligned} n &:= 1 && :: (1 - \exp\{-\,1\}) = 0.63 \\ &:= 5 && :: (1 - \exp\{-\,5\}) = 0.99 \end{aligned}$$

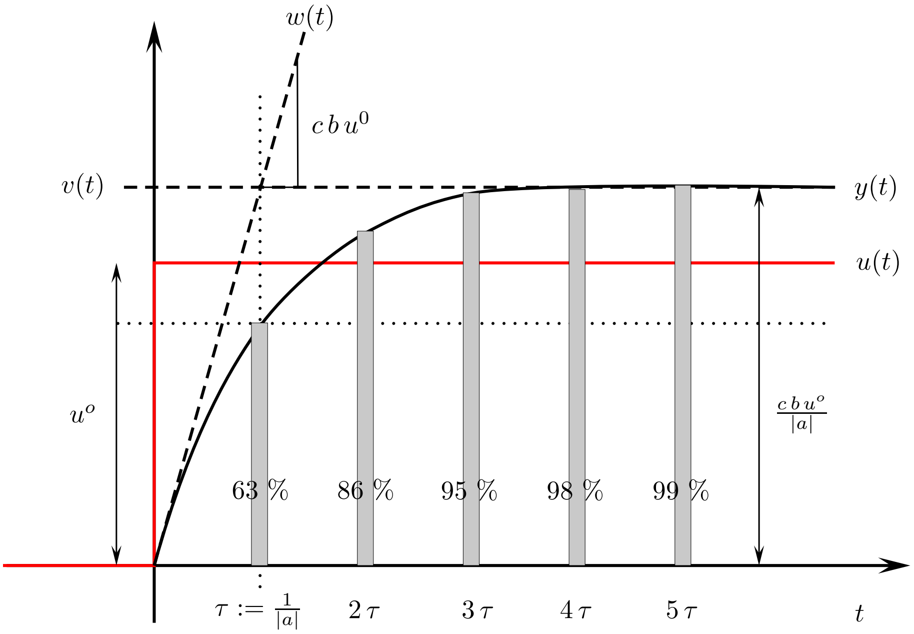


Figure I.13: Properties of a scalar first-order LTI-system

I.6.2 Impulse response

We have seen that the solution in the time domain is:

$$y(t) := c\,\exp\{a\,t\}\,x(o) + c\int_0^t \exp\{a\,(t-\theta)\}\;b\,u(\theta-0)\,d\theta$$

The impulse, a Dirac delta function at time 0 is $\delta\,(t-0)$ is a step to infinity and back within zero time interval and starting with the system at its natural equilibrium position, thus zero.

The solution then looks like:

$$y(t) := c\int_0^t \exp\{a\,(t-\theta)\}\;b\,\delta(\theta-0)\,d\theta$$

The Dirac delta function is only different from 0 at the time the event occurs, thus at time 0. The integral then reduces to:

$$\begin{aligned} y(t) &:= c\int_{-\epsilon}^{+\epsilon} \exp\{a\,(t-\theta)\}\;b\,\delta(\theta-0)\,d\theta \\ &:= c\,b\exp\{a\,t\} \end{aligned}$$

Bibliography

Apostel, L. (1960). Towards the formal study of models in the non-formal sciences from the concept and the role of the model in mathematics and natural and social sciences. In Freudenthal, H., editor, *The concept and the role of the model in mathematics and natural and social sciences*. D.Reidel Publishing Company, Dordrecht, The Netherlands. 1.2

Aris, R. (1963). The fundamental arbitrariness in stoichiometry. *Chemical Engineering Science*, 18(8):554 – 555. 8.2

Aris, R. (1965). Prolegomena to the rational analysis of systems of chemical reactions. *Arch Rational Mech Anal*, 19:81–99. 8.2

Aris, R. (1978). *Mathematical modelling techniques*. Pitman, London. 1.2

Astroem, K. J. and Eykhoff, P. (1971). System identification – a survey. *Automatica*, 7:123–162. 17.1, 17.3, 17.5.1.6

Atkinson, K. E. (1989). *Numerical analysis*. Wiley. 12.1

Bird, R. B., Stewart, W. E., and Lightfood, E. N. (2001). *Transport Phenomena*. Wiley, London. 10.3, 10.5, 11.2.2, 14.2

Bjornbom, P. H. (1977). The relation between the reaction mechanism and the stoichiometric behavior of chemical reactions. *AIChE Journal*, 23(3):285–288. 8.2

Box, G., Hunter, J., and Hunter, W. (2005). *Statistics for Experimenters: Design, Innovation, and Discovery*. Number ISBN 0-471-71813-0. Wiley, 2nd edition. 17.12.1.4

Box, G. E. P., Hunter, W. G., and S, H. J. (1978). *Statistics for experiments – An introduction to design, data analysis, and model building*. John Wiley, New York. 17.12.1

Box, G. E. P. and Tiao, G. C. (1973). *Bayesian inference in statistical analysis*. Addison Wesley. 17.5.2

Breedveld, P. C. (1984). *Physical Systems Theory in Terms of Bond Graphs*. PhD thesis, The Netherlands: University of Twente, Enschede, Netherlands. 5.1

Brenan, K. E., Campell, S. L., and Petzold, L. R. (1989). *Numerical solution of initial-value problems in differential-algebraic equations*. North Holland. 2.4.1

Callen, H. B. (1985). *Thermodynamics and an Introduction to Themostatistics*. Number ISBN 0-471-86256-8. John Wiley & Sons, 2nd edition. 2.4.3, 13.1

Caratheodory, C. (1909). Untersuchung ueber die grundlagen der thermodynamik. *Mathematische Annalen*, 67(1):355–386. One of the key references on axiomatic thermodynamics. 13.1

Chen, G. (2005). *Nanoscale energy transport and conversion: A parallel treatment of electrons, molecules, phonons, and photons*. Oxford University Press. 5.2.1

Courant, R., Friedricks, and Lewy (1928). Existence of solutions for wave equations. *Mathematische Annalen*, 100:32–74. 14.2.1

Deen, W. M. (1998). *Analysis of Transport Phenomena*. Topics in Chemical Engineering. Oxford University Press, New York, Oxford. Nice derivation of the conservation of extensive quantity for a volume in a flow field. 14.2

Denbigh, K. (1971). *The Principles of Chemical Equilibrium*. Cambridge University Press, Cambridge, UK. 6.1.1

Duff, G. F. D. (1956). *Partial differential equations*. University of Toronto Press. 13.1

Eykhoff, P. (1974). *System Identification*. Wiley, New York. 17.1, 17.3, 17.10

Falk, G. and Jung, H. (1959). chapter Axiomatik der Thermodynamik, pages 199–175. 6.1.1

Favache, A. (2009). *Thermodynamics and process control*. PhD thesis, Université catholique de Louvain. 5.1

Favache, A., Dochain, D., and Maschke, B. (2010). An entropy-based formulation of irreversible processes based on contact structures. *Chem Eng Sci*, 65:5204–5216. 5.1

Feynman, R. P., Leighton, R. B., and Sands, M. (1966). *The Feynman Lectures on Physics*. Addison Wesley. 9

Goodwin, G. C. and Payne, R. L. (1977). *Dynamic system identification: Experiment design and data analysis*. Academic Press. 2, 17.5, 17.5.1.4

Grmela, M. and Öttinger, H. (1997). Dynamics and thermodynamics of complex fluids: I.development of a general formalism. *Physical Review E*, 56 (6):6620–6632. 5.1

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. (1986, 2005). *Robust Statistics - The Approach Based on Influence Functions*. Wiley. 17.7.1

Hermann, F. (2004). Entropy from the beginning. *DIREP conference: Teaching and learning physics in new contexts, University of Ostrava*, ISBN 80-7042-378-1:35–40. 10.2

Higbie, R. (1935). The rate of absorption of a pure gas into a still liquid during short periods of exposure. *Trans Am Inst Chem Eng*, 35:36–60. 11.3

Higgins, B. G. and Whitaker, S. (2011). Local, global, and elementary stoichiometry. *AIChE Journal*, pages n/a–n/a. 8.2

Hildebrand, F. B. (1956). *Introduction to numerical analysis*. McGraw-Hill, New York. 12.1

Huber, P. J. (1981, 2004, 2009). *Robust Statistics*. Number ISBN 978-0-470-12990-6. John Wiley & Sons Inc. 17.7.1

Jazwinski, A. H. (1970). *Stochastic Processes and Filter Theory*. Academic Press, New York. 17.9

Johnston, J. and DiNardo, J. (1997). *Econometric methods*. Mc Graw Hill. 17.5.2

Jongschaap, R. and Öttinger, H. (2004). The mathematical representation of driven thermodynamical systems. *Journal of Non-Newtonian Fluid Mechanics*, 120:3–9. 5.1

Kailath, T. (1980). *Linear Systems*. Prentice Hall, Englewook Cliffs, N J, Englewood Cliffs, NJ. 6.1.1

Kalman, R. E. (1963). Mathematical description of linear systems. *Society of Industrial and Applied Mathematics Journal of Control Series A*, 1(2):152–192. 6.1.1, 15.1

Koch, K.-R. (2007). *Introduction to Bayesian Statistics*. Springer. 17.5.2

Kokotovic, P. V., O'Malley, J. R. E., and Sannuti, P. (1976). Singular perturbations and order reduction in control theory – an overview,. *Automatica*, 12(2):123–132. D.1

Lewis, W. K. (1916). The principles of counter-current extraction. *J Ind Eng Chem*, 8(7):825–833. 11.2.2

Lin, C. C. and Segel, L. A. (1988). *Mathematics Applied to Deterministic Problems in the Natural Sciences*. Siam Classics in Applied Mathematics, New York. 14.2.1, D.1

Ljung, L. (1987). *System Identification Theory for the User*. Prentice Hall Inc. Englewood Cliffs, New Jersey. 17.1, 17.3, 17.5.2, 17.8.1, 17.8.2, 17.8.3, 17.10

Maletinsky, M. (1978). *Identification of Continuous Dynamical Systems with Spline-Type Modulating Function Method*. PhD thesis, ETH, Zuerich, Switzerland, Diss ETH Nr 6206. 17.5.1.7

Maronna, R. A., Martin, R. D., and Yohai, V. J. (2006). *Robust Statistics - Theory and Methods*. John Wiley & Sons, Inc. 17.7.1

MathWorld, W. (2016). Graphs. F

Mrugala, R. (2000). On a special family of thermodynamic processes and their invariants. *Reports on Mathematical Physics*, 46 (3):461–468. 5.1

Nernst, W. (1888). Zur kinetik der in lösung befindlichen k örper; erste abhandlung: Theorie der diffusion. *Z. phys. Chem.*, 2(9):613 – 637. 11.2.1

Nerode, A. (1958). Lineear automaton transformation. *proc Amer math Soc*, pages 541–544. 6.1.1

Öttinger, H. and Grmela, M. (1997). Dynamics and thermodynamics of complex fluids ii. illustrations of a general formalism. *Physical Review E*, 56 (6):6633–6655. 5.1

Preisig, H. A. (1984). *On the Identification of Structurally Simple Dynamic Models for the Energy Distribution in Stirred-Tank Reactor Equipment*. PhD thesis, ETH-Zuerich, Switzerland, Diss ETH Nr 7616. 17.5.1.7

Preisig, H. A. (2010). Constructing and maintaining proper process models. *Comp & Chem Eng*, 34(9):1543–1555. 3.5.2

Preisig, H. A. and Rippin, D. W. T. (1993a). Theory and application of the modulating function method–part i review and theory of the method and theory of the spline-type modulating function method. *Computers & Chemical Engineering*, 17(1):1–16. 17.5.1.7

Preisig, H. A. and Rippin, D. W. T. (1993b). Theory and application of the modulating function method–part ii algebraic representation of maletinsky's spline-type modulating functions. *Computers & Chemical Engineering*, 17(1):17–28. 17.5.1.7

Rajeev, S. G. (2008). A Hamilton-Jacobi formalism for thermodynamics. *Annals of Physics*, 323:2265–2285. 5.1

Rudin, W. (1987). *Real and complex analysis*. Number ISBN 978-0-07-054234-1. McGraw-Hill Book Co. 2.1.1.1

Saksena, V. R., O'Reilly, J., and Kokotovic, P. V. (1984). Singular perturbations and time-scale methods in control theory: Survey 1976-1983. *Automatica*, 20(3):273–293. D.1

Schlichting, H., Gersten, K., Krause, E., Oertel, H. J., and Mayes, C. (2004). *Boundary-Layer Theory*. Number ISBN 3-540-66270-7. Springer, 8th edition edition. 11.2.2

Schwarz, H. R. (1989). *Numerical analysis - a comprehensive introduction*. Wiley. 12.1

Strang, G. (2009). *Introduction to Linear Algebra*. Wellesley-Cambridge Press. 8.2.1, A.1, A.1.2, A.1.5, A.1.5.2

Truesdell, C. (1980). *The tragicomical history of thermodynamics 1822-1854*. Springer-Verlag, New York. 10.2

Truesdell, C. and Toupin, R. (1960). The classical field theories. In Flugge, S., editor, *Handbuch der Physik*, volume III, pages 226–858. Springer Verlag, New York. 8.2

Viz, G. (2016). Graph viz. F

Whitman, W. G. (1923). The two-film theory of absorption. *Chem and Met Eng*, 29:147 ff. 11.2.2

Wikipedia (2016). Graph theory. F

Words, G. (2016). Interactive thesaurus. F

Nomenclature

$\hat{E}$ total energy flow. 106, 107, 123

$\hat{H}$ enthalpy flow. 29, 114

$\hat{K}$ flow of kinetic energy. 108, 124

$\hat{P}$ flow of potential energy. 108, 124

$\hat{U}$ flow of internal energy. 109, 124

$\hat{V}$ volume flow. 69, 109, 124–127, 129, 130, 164

$\hat{\Phi}$ flow of extensive quantity. 35, 36, 54, 57, 69, 70, 181–183

$\hat{\varphi}$ flow of extensive quantity. 67, 68, 103

$\hat{\varphi}$ flux of normed extensive quantity. 58, 59

$\hat{m}$ mass flow in kg/s. 25–28, 43, 84

$\hat{n}$ flow of molar mass. 80, 81

$\hat{n}$ flow of molar mass. 28

$\hat{q}$ heat flow. 29, 30, 65, 66, 69, 106, 107, 109, 114, 119, 140

$\hat{w}$ work flow. 106–109

$\hat{w}^V$ flowof volume work. 123, 124

$\hat{w}^f$ flow of friction work. 123–126

$\underline{\underline{\mathbf{\Gamma}}}$ system matrix input. 225, 226

$\underline{\underline{\mathbf{C}}}$ coefficient matrix. 87–90, 92

$\underline{\underline{\mathbf{F}}}$ incidence matrix of the directed graph of the network – network matrix. 36, 60, 61, 70, 79–81, 103, 106, 162, 165, 186–189

$\underline{\underline{\mathbf{F}}}^m$ mass-network matrix - the incidence matrix of the directed graph for the connected mass network. 43, 47, 50, 106–112, 114

$\underline{\underline{\mathbf{F}}}^n$ molar mass-network matrix - the incidence matrix of the directed graph for the connected molar mass network. 47

$\underline{\underline{\mathbf{F}}}^q$ heat-flow-network matrix - the incidence matrix of the directed graph for the connected heat flow network. 50, 106, 107, 109–112, 114

$\underline{\underline{\mathbf{F}}}^w$ work-flow-network matrix - the incidence matrix of the directed graph for the connected work-flow network. 50, 106–109

$\underline{\underline{\mathbf{K}}}$ diagonal matrix with reaction constants. 99–101

$\underline{\underline{\mathbf{K}}}^q$ diagonal matrix with heat conductivity parameters in the diagonal. 70

$\underline{\underline{\mathbf{N}}}$ stoichiometric matrix. 90, 91, 97–101, 103, 112, 164, 166, 170, 174, 175

$\underline{\underline{\mathbf{P}}}$ permutation matrix. 92, 93, 98, 100

$\underline{\underline{\mathbf{S}}}$ selection matrix. 81, 99, 100, 186, 188, 189

$\underline{\underline{\hat{\phi}}}$ matrix of fluxes of extensive quantities. 162, 163, 165

$\underline{\underline{\boldsymbol{\Omega}}}$ left null space matrix. 187, 188

Ω sample space. 201, 202

Φ extensive quantity. 35, 36, 53, 54, 57–60, 165, 181, 183, 184, 252, 253

$\underline{\underline{\boldsymbol{\Phi}}}$ fundamental matrix. 225, 226

$\underline{\tilde{\mathbf{n}}}$ production rate of molar mass. 83, 84, 87, 97, 101, 164, 165

$\underline{\bar{\boldsymbol{\Psi}}}$ estimator of the nominal model. 201–204, 212

Ψ estimator, a function of the data. 200, 201

Θ parameter. 183, 184, 215, 222–224, 301, 302

$\underline{\mathbf{0}}$ null vector. 87, 98, 188

$\underline{\mathbf{X}}$ vector of extensive quantities namely volume and component mass vector. 156

$\underline{\hat{\mathbf{E}}}$ energy flow vector. 50, 106, 107

$\underline{\hat{\mathbf{H}}}$ vector flow of enthalpy. 110–112

$\underline{\hat{\mathbf{K}}}$ vector flow of kinetic energy. 108

$\underline{\hat{\mathbf{P}}}$ vector flow of potential energy. 108

$\underline{\hat{\mathbf{U}}}$ vector flow of internal energy. 109

$\underline{\hat{\boldsymbol{\Phi}}}$ vector of extensive quantity flow. 36, 55–57, 59–61, 70, 79, 102, 103, 162, 165, 186–188

$\underline{\hat{\varphi}}$ flux of extensive quantity. 56, 57, 68, 102

$\underline{\hat{\phi}}$ flux vector of extensive quantity. 59

$\underline{\hat{\mathbf{m}}}$ vector of mass flows. 43

$\underline{\hat{\mathbf{n}}}$ vector of molar mass flows. 56, 80, 81, 83, 84, 87, 112, 114, 164, 165

$\underline{\hat{\mathbf{q}}}$ heat flow vector. 50, 106, 107, 109–112, 114

$\underline{\hat{\mathbf{w}}}$ work flow vector. 50, 106–109

$\underline{\hat{\mathbf{x}}}$ vector of flow of extensive state. 170, 174, 175

$\underline{\boldsymbol{\Phi}}$ vector of extensive quantities. 36, 55, 60

$\underline{\tilde{\xi}}$ vector of production rate of extent of reaction. 112, 114, 164, 166, 170, 174, 175

$\underline{\tilde{\boldsymbol{\Phi}}}$ production vector of extensive quantity. 102, 103

$\underline{\tilde{\varphi}}$ production rate of extensive quantity normed by volume. 102, 103

$\underline{\tilde{\mathbf{x}}}$ vector of production of extensive state. 170, 174, 175

$\underline{\boldsymbol{\Theta}}$ vector of parameters. 172, 173, 175

$\underline{\boldsymbol{\lambda}}$ vector of molecular masses. 84, 111, 310

$\underline{\boldsymbol{\nu}}$ vector of stoichiometric coefficients. 101, 102, 112

$\underline{\boldsymbol{\pi}}$ vector of effort variables. 57, 70

$\underline{\boldsymbol{\varphi}}$ volume-normed extensive quantity - thus a density. 55–57, 102, 103, 165

$\underline{\dot{\boldsymbol{\Phi}}}$ vector of extensive quantity accumulation. 55, 60, 61, 70, 79, 102, 103, 186, 187

$\underline{\dot{\mathbf{m}}}$ vector of accumulation of mass. 43

$\underline{\dot{\mathbf{n}}}$ vector of accumulated molar mass. 47, 83, 84, 87, 97

$\underline{\dot{\mathbf{x}}}$ vector of time derivative extensive state. 170, 174, 175

$\underline{\mathbf{c}}$ concentration vector. 96, 98–102, 164–166

$\underline{\mathbf{c}}_p$ vector of specific heat capacities. 113, 114

$\underline{\mathbf{f}}$ vector of forces. 122, 123, 129

$\underline{\mathbf{h}}$ vector of partial molar enthalpies. 112, 113

$\underline{\mathbf{n}}$ vector of molar mass in mol/s. 47, 56, 57, 63, 64, 80, 87, 98–100, 111–115, 156–158, 165, 166, 252, 253

$\underline{\mathbf{n}}$ normal vector. 55–57, 59, 162, 165

$\underline{\mathbf{p}}$ vector of properties. 170, 172–175

$\underline{\mathbf{r}}$ vector of spatial co-ordinates. 55, 57, 59, 67, 68, 102, 103, 134, 162, 163, 165

$\underline{\mathbf{s}}$ vector of state variable transforms. 172, 173, 175

$\underline{\mathbf{v}}$ velocity vector. 55–57, 122, 129

$\underline{\mathbf{x}}$ vector of extensive states = primary state. 170, 172, 173, 175–177

α confidence level. 208, 210, 211, 215

α heat diffusivity. 120

α reference co-ordinate indexed with system and flow. 35, 36, 43, 60, 106–109, 114, 162, 170, 174, 175

$\bigtriangledown$ gradient. 122, 123, 128, 129, 133, 142

$\bullet$ pure. 113, 114

χ^2 chi square distribution. 210

$\dot{K}$ accumulation of kinetic energy. 124

$\dot{P}$ accumulation of potential energy. 124

$\dot{\Phi}$ accumulation of extensive quantity. 35, 36, 70, 181–184, 188

μ chemical potential. 64, 69, 94, 134, 141, 142, 157, 158, 305–307

μ dynamic viscosity. 122, 129, 130

$\hat{\mu}$ estimate of the centre of the distribution (median, mean, mode). 216–218

μ moment of a distribution. 284, 285

$\underline{\mu}$ vector of chemical potential. 318–323

ν stoichiometric coefficient. 93, 96, 101, 102

φ intensive property. 252, 253

π effort variable. 35, 67–70, 103, 138, 163, 183, 184

ρ density. 26, 27, 120, 122, 124, 126, 128–130, 299–301, 307, 310, 313, 314

ρ "robustification" function. 219, 220

σ standard deviation. 288

σ^2 variance. 201, 206, 207, 210, 234, 235

$\star$ optimal. 205, 206, 208, 212, 215

τ average collision time. 65, 66

τ time constant. 300, 301

τ time on the streched time axis. 183, 184

root root of the expression in the bracket = 0. 98, 269

$\underline{\theta}^\star$ vector of estimated parameters. 206–209, 211, 213

$\bar{\underline{\theta}}$ vector of parameters of the nominal model. 200–204, 212

$\underline{\theta}$ vector of parameter. 194, 199–201, 204–206, 208, 212–215, 221, 223, 224, 231–234

$\tilde{\xi}$ rate of extent of reaction. 96–101

ε scale – the inverse of the stretching factor in time scaling. 181–184

$\varkappa$ mole fraction. 133, 134, 142

$\dot{\varphi}$ time derivative of the scaled extensive property. 181, 182

φ extensive quantity per volume - a density. 59

φ scaled extensive property. 181

$\underline{\underline{\mathbf{0}}}$ null matrix. 187

$\underline{\mathbf{0}}$ null vector. 202, 225

A area. 27, 119, 129, 130, 164, 307

$\mathcal{A}$ list of atoms. 88

$A(q)$ system polynom in shift operator q. 220–224

a_i coefficient of A-polynomial. 220, 221, 223, 224

$B(q)$ input polynom in shift operator q. 220–224

b_i coefficient of B-polynomial. 220, 221, 223, 224

$\underline{\mathbf{b}}$ parameter's bias term. 212

$\underline{\underline{\underline{\mathbb{C}}}}$ conductivity tensor. 67

$\underline{\underline{\mathbf{C}}}$ conductivity matrix. 57, 68, 163

$\underline{\underline{\mathbf{C}}}$ column selecting matrix. 186, 187

$C(q)$ ouput polynom in shift operator q. 213, 221–224

$\underline{\underline{\mathbf{C}}}$ system matrix ouput. 225, 226

c_i coefficient C-polynomial. 221, 223, 224

c concentration. 84, 85, 87, 96, 99, 100, 133, 134, 141, 142

c conductivity. 67, 103

$\underline{\mathbf{c}}$ conductivity vector. 68

c_V valve constant - a characteristic of the valve. 127

r correlation. 206

c_p specific heat capacity at constant pressure. 30, 66, 111–114, 120, 300, 301

$\mathcal{C}$ conditions. 194

$\underline{\underline{\mathbf{R}}}$ correlation matrix. 206

(cov) covariance function. 285, 286

cov covariance function. 206

D characteristic dimension. 128–130

$\mathcal{D}$ mass diffusivity. 129, 130, 133, 134, 142

$\mathcal{D}$ diffusivity. 152, 153

$\mathcal{D}$ input / output data. 194

$\underline{\underline{\mathbf{D}}}$ norming matrix. 232–234

$D(q)$ D-polynomial in q. 213, 223

d_i coefficient of D polynomial. 224

d diameter. 66

$\underline{\underline{\mathrm{diag}}}$ diagonal matrix of vector. 231–233

$\dot{E}$ accumulation of total energy. 106, 107, 140

E total energy. 50, 106, 123, 124

$e(k)$ error at time k. 220–223

e specific energy. 65, 66

E_A activation energy. 97

$\underline{\underline{\mathbf{\Lambda}}}$ diagonal matrix with eigenvalues in the diagonal. 208

$\underline{\underline{\mathbf{V}}}$ matrix of eigenvectors. 208

e error. 215, 219, 220

$\underline{\mathbf{e}}$ vector of errors. 205, 208–214, 226

$\mathbf{E}$ expectation operator. 201–204, 210–213, 222, 284–288

F F-distribution. 208, 211, 215

$\mathcal{F}$ list of streams. 36

$\underline{\underline{\mathbf{F}}}$ matrix with rows being the vector of nonlinear functions of the input. 204–209, 211–213, 232, 233, 235

$F(q)$ F-polynom in shift operator q. 223, 224

f_i coefficient of F-polynomial. 224

f Darcy friction factor. 129, 130

f force. 123–125, 128

f function - nonlinear. 214, 215

$\underline{\underline{\mathbf{G}}}$ matrix of transfer functions. 198

g gravitation constant. 124, 126

g nonlinear function. 96, 99, 101, 102

$\dot{H}$ accumulation of enthalpy. 109, 110, 114, 119

H enthalpy. 29, 109–111, 120, 252

h height. 27, 28, 124, 126

h partial molar enthalpy. 29, 30, 111–114

h step. 147–151

$\underline{\underline{\mathbf{I}}}$ identity matrix. 47, 202–204, 226, 233

$\underline{\mathbf{I}}$ vector of ones. 202

$\underline{\mathbf{u}}$ vector of inputs. 194, 198–201, 204, 207, 210, 213–215, 225, 226, 231, 232

$\underline{\underline{\mathbf{W}}}$ instrumental variable matrix. 213

$\underline{\underline{\mathbf{J}}}$ Jacobian matrix. 189

$\mathbf{J}$ objective function. 194, 205, 208, 214, 215, 219

K Nernst distribution constant. 142

K kinetic energy. 106, 124

$\underline{\underline{\mathbf{K}}}$ Kalman gain matrix. 226

k^q heat conductivity. 119, 120

k^M diffusivity of momentum. 129, 130

k transfer property : a fomr of conductivity. 69, 70

k^0 reaction constant. 97

k^n mass flow parameter. 69

k^q heat flow parameter. 30, 66, 69

k^r reaction "constant". 96, 97, 99–102

k^V volume flow parameter. 69

k_B Boltzmann constant. 66

L length. 129, 130

$\underline{\underline{\mathbf{M}}}$ Fisher information matrix. 201–204, 235

M model. 194, 200

$\bar{M}$ nominal model. 200, 201

$\mathcal{M}$ set of models. 194

$\dot{m}$ accumulation of mass. 43, 84

m index mass stream. 43

m stream index. 36

m mass in kg. 25–28, 43, 66, 111, 112, 124, 126

$\underline{\mathbf{y}}$ vector of secondary states. 170, 172–177, 189

m distribution median. 217

n_e total degree of freedom. 210, 211

n molar mass. 28, 29, 65, 66, 80, 94, 100, 156–159

n number of data points. 204, 208–211, 215–217, 234, 235

n number of experiments. 233, 234

n number of particles. 65

$\underline{\mathbf{f}}$ nonlinear vector function f. 198, 199, 204, 207, 212

$\underline{\mathbf{g}}$ nonlinear vector function g. 212

k number of parameters. 199, 204, 208–211, 215

O order of magnitude. 148

$\circ$ Hadamard operator :: element by element product. 233

P potential energy. 106, 124

P probability distribution. 201

p pressure. 64, 66, 69, 109–114, 122–130, 157–159, 253

p probability density. 201–204, 214

$\underline{\underline{\mathbf{Q}}}$ matrix of weights - positive definite. 205

q number of repeated experiments. 210, 211

q shift operator. 220–224, 226

R gas constant. 97, 134, 141, 142, 158, 159

$\underline{\underline{\mathbf{R}}}$ row selecting matrix. 186, 187

Re Reynold number. 129, 130

r spatial co-ordinate. 147–153, 164

v random variable (noise). 206, 207

r_x spatial co-ordinate x. 55, 58, 59

r_y spatial co-ordinate y. 55, 58

r_z spatial co-ordinate z. 55, 58

$\mathbb{R}$ real numbers. 199, 204, 239–241, 244, 255, 261, 291, 292

S system boundary. 55–57, 59, 60, 162, 165

S Entropy. 63, 64, 94, 156–158, 252, 253

$\mathcal{S}$ list of species. 88

$\underline{\underline{\mathbf{S}}}$ design-of-experiment matrix. 232–234

s entropy density. 158, 159

s estimated standard deviation. 208, 210, 211, 216, 219

s system index. 43, 50, 60, 61, 103, 106–109, 162, 165, 166

$\underline{\hat{\mathbf{x}}}$ state vector of model. 225, 226

$\underline{\bar{\mathbf{x}}}$ state vector of nominal model. 225, 226

T temperature. 30, 35, 64, 66, 69, 96, 97, 109, 111–114, 119–122, 134, 141, 142, 152, 153, 157–159, 252, 253, 298, 300–302

t time in s. 25–29, 50, 54–59, 102, 103, 111, 112, 114, 120, 122–124, 128, 129, 134, 152, 153, 164–166, 183

$\dot{U}$ accumulation of internal energy. 107, 109, 110, 124

U internal energy. 63, 64, 94, 106, 156–158, 252, 253

U internal energy. 109, 110, 124

u internal energy density. 158, 159

$u(k)$ input at time k. 220–224

u input. 199, 212, 232

u specific internal energy. 66

$\underline{\bar{\mathbf{u}}}$ vector of normed inputs. 231

$\dot{V}$ accumulation of volume. 109, 110

$\underline{\underline{\mathbf{V}}}$ variance co-variance matrix. 201–203, 235, 236

V volume. 26, 27, 43, 54–59, 63, 64, 94, 97–103, 109, 110, 112, 114, 119, 120, 156–159, 164–166, 252, 253

v density volume. 158, 159

v velocity. 65, 66, 124, 126–130, 133, 164

$\underline{\mathbf{v}}$ noise vector v. 200, 201, 225, 226

var variance function. 206, 207, 235

var variance function. 284–288

$\underline{\underline{\mathbf{R}}}$ variance-covariance of output noise. 225, 226

$\underline{\underline{\mathbf{Q}}}$ variance-covariance matrix of input noise. 225

$\underline{\mathbf{w}}$ noise vector w. 200, 225, 226

x spatial x-coordinate. 65, 66, 119, 120, 123–125, 138, 141

x state variable. 146–151, 153

$\underline{\mathbf{x}}$ vector of states. 198

$\underline{\dot{\mathbf{x}}}$ vector time deriviate of state. 188, 189

$\underline{\hat{\mathbf{x}}}$ vector of state flow. 188, 189

$\underline{\tilde{\mathbf{x}}}$ vector of production rates. 188, 189

$\underline{\mathbf{x}}$ state vector. 189

Y list of observed random variables. 216, 217

$\hat{y}$ model output (scalar). 204

$\bar{y}$ output from nominal model. 212

$\hat{y}$ output of model. 207, 212, 214, 215, 221–224, 231, 233

$y(k)$ ouput at time k. 220–224

y output (scalar). 199, 204, 215, 234, 235

y observed stochastic variable. 216–218

$\underline{\hat{\mathbf{y}}}$ vector of model output. 194, 200, 204, 205, 212, 213, 223, 226, 232–234

$\underline{\bar{\mathbf{y}}}$ vector of outputs of the nominal plant/model. 200–204, 212, 225, 226

$\underline{\mathbf{y}}$ vector of outputs. 194, 198, 204–206, 208–211, 214

$\underline{\dot{\mathbf{y}}}$ vector of time derivate of secondary state. 189

$\underline{\bar{\mathbf{z}}}$ nominal omitted output. 212

$\underline{\bar{\mathbf{z}}}$ output of the omitted additive part of the model. 212

Terminology

1D one spatial dimensional. 15, 18, 21

2D two spatial dimensional. 15, 18, 21

3D three spatial dimensional. 15, 18, 21

AE algebraic equations. 9

behaviour usually input/output behaviour - given a suffiently exciting input, the process shows a change in the state, which in turn may be observed as an output. 15, 20, 167

boundary the dual to system. 17

conserved maintains a balance over time. 25

DAE differential algebraic equations. 9

discrete-event dynamic things that just happen in an instance in the context of a given time scale. 5, 18

distributed not-uniform in the sense that the intensive properties are a function of the spatial coordinate - in contrast to lumped systems. 9, 13, 15, 18

environment the part of the universe in which the modelled system is embedded and which is constant, thus consists of reservoirs. 14, 16–18, 20, 25, 26, 29, 30

extensive depends on the size, the extent of the system - mass, energy, momentum, volume etc.. 25

granularity resolution of the model in terms of relative volume size of the simple capacities. 24

lumped uniform in the sense that the intensive properties are not a function of the spatial coordinate - in contrast to distributed systems. 9, 13, 15, 18

ODE ordinary differential equation. 9

PDAE partial differential algebraic equations. 9

PDE partial differential equation. 9

plant often used synonym for system being a processing unit of some kind, not necessarily in the context of chemical or biological operations. 14–21, 23

primitive system primitive in terms of the most fundamental - lumped, distributed, reservoir, boundary. 185, 188

proper model a complete model that has zeor degree of freedom when defining initial and boundary conditions and all parameters. 179

reservoir an infinitely large capacity with some given intensive properties, thus all extensive quantities are infinite. 13

state the variable that forms the foundation of the mathematical representation. 167

steady-state the state does not change, thus is constant in time. 7

system An entity cyclicly defined by its boundary. 13–17, 20, 22, 25, 31, 167

time-scale defines a reange in terms of time in which one observes, models, describes the behaviour of process. 7

token an abstract ”thing” being present in the nodes of the graph and transferred by the arcs - for physical systems tokens are usually the conserved quantities. 20

universe all there is, in the context of modelling this is reduced to all relevant parts of the universe. 17

variable An algebraic symbol that takes a value when applied. 167