
DYNAMICAL MODELLING AND CONTROL OF CHEMICAL REACTIVE SYSTEMS

Otacílio Bezerra Leite Neto
Federal University of Ceará
Department of Teleinformatics Engineering
minhotmog@gmail.com

Contents

1	Introduction	1
1.1	Control Systems Engineering	1
1.2	Chemical Reactive Systems	2
1.3	Motivation	3
1.4	Objectives	4
1.5	Chapters Organization	4
2	Dynamical System Analysis	5
2.1	Model from First Principles	5
2.2	Mathematical Models of Systems	9
2.3	Time Response Analysis	15
2.4	Similarity Transformations	20
2.5	Stability, Controlability and Observability	23
2.6	Frequency Response Analysis	26
3	Controller Design	27
3.1	Control Architectures	27
3.2	Full-State Feedback	27
3.3	Regulation and Reference Tracking	27
3.4	Deterministic State Observers	27
4	Optimal Control	29
4.1	General Formulation	29
4.2	Linear Quadratic Regulator (LQR)	29
4.3	Optimal State Estimators	29
4.4	Linear Quadratic Gaussian (LQG)	29
4.5	Robustness and Stability Analysis	29
5	Methodology	31
6	Results and Discussion	33
7	Conclusion	35
	Bibliography	36
	Appendix A - Proof of Theorems	38

Chapter 1

Introduction

This chapter presents the main problem in discussion and the basic concepts concerning its formulation and solutions, which are detailed further in the next chapters. This is a work on Control Theory and its application to Chemical Reactive Systems, therefore the discussion will follow the notation common to the literature of this field, and a "modern" approach to this theory is explored.

The sections are organized as follows: Section 1.1 provides general definitions for control systems engineering, Section 1.2 discuss chemical reactive systems and its importance in both industry and academia, Sections 1.3 and 1.4 describes the motivation and justification of this work, respectively, and Section 1.5 details the subsequent chapters in this document.

1.1 Control Systems Engineering

The discipline of Control Systems Engineering deals with the design of devices, named *controllers*, that are integrated to a physical system (a *dynamical system*, in most cases) in order to impose a desired behavior to this system. To achieve this goal, the discipline covers topics ranging from applied mathematics, such as dynamical systems theory and signal processing, to a more engineering discussion, regarding instrumentation and implementation of these controllers in a real-life plant or individual system.

A system, in a broad physical sense, is defined as a ensemble of interacting components that responds to external stimuli producing a determined dynamical response, and whose individual parts are not able to produce the same functionality by their own. Thus, the first essential element in Control Theory is a mathematical model of the system of interest. One such model is the *Input-Output Representation*, as shown in **fig!!!**, in which an input stimuli, a signal $u(t)$, acts on the system producing an output response, a signal $y(t)$, described by the following differential equation:

$$\alpha_n \frac{d^n y(t)}{dt^n} + \alpha_{n-1} \frac{d^{n-1} y(t)}{dt^{n-1}} + \cdots + \alpha_1 \frac{dy(t)}{dt} + \alpha_0 y(t) = \beta_m \frac{d^m u(t)}{dt^m} + \beta_{m-1} \frac{d^{m-1} u(t)}{dt^{m-1}} + \cdots + \beta_1 \frac{du(t)}{dt} + \beta_0 u(t) \quad (1.1)$$

[fig]

In this representation, the input $u(t)$ is called the *manipulated variable*, since it represents a arbitrary stimuli that can be given directly by human action or a by an automatic controller, while the output $y(t)$ is called the *controlled variable*, since it can only be modified indirectly through $u(t)$. This also leads to a *cause-and-effect* interpretation of the system.

A model can provide a quantitative understanding of the system that is useful both to access some response specifications and to design controllers to modify them based on some requirements. In the case of the model in Equation (1.1), it is possible to calculate the response

$y(t)$, and its derivatives, resulting from any specific action $u(t)$. Besides, a model can be used to perform computer simulations, in order to visualize the dynamical behavior of the system without actually manipulating it, since real experiments could be expensive or even damage the system. Consider, for instance, a schematic and a simulation for a model representing a mass-spring-damper system, shown at !!!!.

[fig]

In this simulation, the *rise time*, *peak time*, *overshoot ratio* and *steady-state value* are examples of response specifications that can be defined to describe the system behavior to a external stimuli (in this case, a constant force of unit magnitude). These specified parameters are characteristic to responses of a class of systems known as *underdamped second-order systems*, that will be discussed further in the document.

A controller is used to calculate, for a time $t \in [t_0, t_N)$, the necessary input $u(t)$ to produce an output $y(t)$ as close as possible to a desired reference signal $r(t)$. There are two common configurations, shown in **Fig!!!**, of how to connect the controller to the system.

[fig]

The configuration in **Fig!!!.a**, known as *Open-Loop Controller*, calculates the action as a function $u(t) = \pi(r, t)$, given an initial condition $y(t_0) = y_0$. In this case, the controller does not observe the output $y(t)$, and relies on the model to guarantee that the system is driven to the reference. Of course, if there are any external disturbances acting on this configuration, or if the model is not reliable enough, it is not possible to guarantee that the requirements are met. Thus, these type of controllers are not suitable for critical applications, and its use is restricted to systems where deviance from the desired reference can be tolerated. [exemplos de aplicações?]

In contrast, the configuration in **Fig!!!.b**, known as *Closed-Loop Controller* or *Feedback Controller*, calculates the action as a function $u(t) = \pi(e, t)$, where $e(t) = r(t) - y(t)$ is the error between the reference and the actual response. Now, the controller will observe the system output, trough some sensor device, and compares it to the desired reference in order to calculate a *corrective action*. This feedback property can make the system reject disturbances while still driving it to the desired reference. Thus, the Feedback Controller became the most popular choice of controller configuration in industry for a wide range of applications, even for critical ones. [exemplos de aplicações? + references]

1.2 Chemical Reactive Systems

A chemical reaction, the transformation of a chemical substance into another, is a process central to chemistry and to nature itself. A reaction equation is a intuitive representation of such transformations. For instance, consider the following equation representing a *synthesis reaction*:



In this equation, the compounds A and $2B$ forms the set of *reactants*, \mathbb{R} , while $3C$ and D forms the set of *products*, \mathcal{P} . The coefficients in such equations are the *stoichiometric numbers*, providing an information about proportionality between the quantity of each substances in the reaction.

Usually the products can be directly used as reactants in another reaction, in which case they can also be referred as a intermediate product (or byproduct), and the equations can be appended in a "series" representation. In this case, each k -th intermediate product forms a set \mathcal{I}_k . In addition to a chain of series reactions, there is also the possibility of different reactions to occur in parallel, in the same system. The combination of these sets of reactants, byproducts, products and reactions are often referred as a *chemical reaction network*, and the associated

equation can be represented in general form as:

$$\begin{cases} \mathbb{R}^{(1)} & \longrightarrow & \mathcal{I}_1^{(1)} & \longrightarrow & \cdots & \longrightarrow & \mathcal{I}_{M_1}^{(1)} & \longrightarrow & \mathcal{P}^{(1)} \\ \mathbb{R}^{(2)} & \longrightarrow & \mathcal{I}_1^{(2)} & \longrightarrow & \cdots & \longrightarrow & \mathcal{I}_{M_2}^{(2)} & \longrightarrow & \mathcal{P}^{(2)} \\ \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\ \mathbb{R}^{(N)} & \longrightarrow & \mathcal{I}_1^{(N)} & \longrightarrow & \cdots & \longrightarrow & \mathcal{I}_{M_N}^{(N)} & \longrightarrow & \mathcal{P}^{(N)} \end{cases} \quad (1.3)$$

Moreover, chemical reactions displays a dynamical behavior concerning the speed at which a reaction occurs. This rate of reaction, its *kinetics*, are dependent on the conditions in the environment, such as temperature and pressure, and on some properties of the reaction itself. In the case of a *isothermal process*, i.e., when the temperature in the environment remains constant, this rate can be calculated as a constant K , leading to a representation on the form:

$$\mathbb{R} \xrightarrow{K} \mathcal{P} \quad (1.4)$$

When the temperature in the environment is not constant, the process is said to be *endothermic* or *exothermic* if, respectively, it consumes or produces energy. The kinetics of the reactions in such processes are usually functions of the temperature which, given an activation energy E , are assumed to follow the Arrhenius equation:

$$K(T) = K_0 e^{-E/T} \quad (1.5)$$

In practice, these chemical reactions are produced by mixing the reactants in some environment with adequate conditions. In order to control the quantities of these substances, actual processes consists in a manipulation of the concentrations of reactants in some container, usually by providing a mass flow of these substances through some fluid. A major interest is to manipulate the reactants in some way to produce a desired concentration of one or more products in the chemical reaction network, allowing this problem to be addressed by a control engineering perspective. A *chemical reactor system*, depicted in **Fig!!!**, is a system where a controller can manipulate the concentration of some reactants to produce a desired concentration of some products.

[img]

When the process is not isothermal, the occurrence of a reaction contributes to the entropy of the environment, and consequently affects the kinetics of the subsequent ones. To compensate for this, practical applications also try to control the conditions in the environment using instruments external to the reactions themselves. Because of the use of the Arrhenius equation to model these reaction rates, this control is usually implemented through a cooling or heating system coupled to the original reactor system, resulting in the schematic on **Fig!!!**.

[img]

1.3 Motivation

The use of automatic controllers to impose a desired behaviour to physical systems is a practice ubiquitous in many engineering fields. In the last years, the price of digital computers have been dropping while their performance have been growing. Consequently, digital controllers have became the central key in developments in important and innovative fields such as aeronautics [reference] and autonomous driving [reference]. In parallel, this theory is also useful to understand and bring inspiration from nature itself since, for instance, the mechanisms for temperature regulation observed in vertebrate animals behave as a feedback controller [Heller et al., 1978].

Most recent developments in Control Theory focus on using Feedback Controllers to achieve *Robust and Optimal Control*. This theory accounts for the design of controllers that deals with

uncertainty, either from the model or from the observation of the system, and are able to achieve the control objectives in a *optimal manner*. Despite being a few decades old, these fields have gained a lot of interest in the last years thanks to recent results in *Machine Learning*, particularly in *Reinforcement Learning*, that are having success in using optimization techniques for artificial agents to control themselves in environments loaded with uncertainty [reference].

Furthermore, the specific application of controlling chemical reactor systems brings benefits from the fact that chemical reactions are present in most biological and industrial processes. In this sense, controllers can be used to guarantee safety constraints, maximize productivity and minimize the use of resources, in such way that is unfeasible without automatic and high performing machinery.

1.4 Objectives

This work aims to provide a self-contained discussion of modelling and control of chemical reactive systems in the perspective of modern control theory. Therefore, the results are focused on *state feedback controllers* modelled in continuous and discrete time, but analysis in the frequency domain is also considered in order to explain some concepts. Several properties of these models, both in the open-loop and closed-loop regime, are summarized in the document and the intention is to have a generalized framework to understand, evaluate and design those systems. Finally, the theory of more advanced methods such as optimal estimation and optimal control is also developed in the same sense.

1.5 Chapters Organization

The chapters of this document are mainly organized in two parts. The first part, comprised by the chapters 2, 3 and 4, builds the necessary theoretical background and provides the mathematical framework for the applications. The second part, comprised by the chapters 5 and 6, describes the experiments and results of applying these methods in real-world applications.

Individually, the chapters are organized as follows: chapter 2 introduces the dynamical models and its several properties with respect to the real system behavior, chapter 3 discusses classical methods in developing automatic controllers and state observers, chapter 4 presents more advanced methods in optimal estimation and optimal control, chapter 5 describes the practical experiments used to validate the previous discussions, chapter 6 summarizes and discusses the results of the experiments and evaluate the several controllers performances and, finally, chapter 7 provides the conclusion of the document and possible future works.

Chapter 2

Dynamical System Analysis

This chapter discusses the mathematical models for dynamical systems and its use in response analysis. The section starts by introducing a procedure to build models from physical principles and presenting equivalent common representations. Next, the response of systems, both in time domain and frequency domain, are analyzed in the light of such models, relating the mathematical structure with the dynamical behavior. Finally, some important properties are defined and proved using this formulation.

2.1 Model from First Principles

A dynamical system is a physical system whose states evolve with time. For this reason, one can represent a dynamical system using the *first principles* from physics itself, and formulate the evolution in time by calculating the rate of change of the states in respect to time. Thus, dynamical models can be equated using differential equations with time derivatives.

A straightforward procedure to model a system consists in identifying the variables of interest and relating them using conservation laws, such as conservation of mass, conservation of energy or conservation of momentum. The resulting differential equations are in the form:

$$\left(\begin{array}{c} \text{Rate of} \\ \text{Accumulation} \end{array} \right) = \left(\begin{array}{c} \text{Mass/Energy/Momentum} \\ \text{entering the System} \end{array} \right) - \left(\begin{array}{c} \text{Mass/Energy/Momentum} \\ \text{leaving the System} \end{array} \right) \quad (2.1)$$

The choice of which conservation law to use depends on the system itself, since the variables of interest can provide dynamics to the system in many forms. Usually, conservation of mass is used to relate dynamics of concentrations and volumes, or other material variables, while conservation of momentum is often used to relate dynamics of motion. Since energy can be converted on form, the conservation laws of this quantity can be used to model several dynamics, such as the rate of change in heat, electrical charges or velocity of a system.

In the case of a chemical reactor system, the variables of interest are the concentrations of the chemical substances in the system. Hence, the rate of accumulation of a substance can be represented using the mass conservation law, or mass balance:

$$\begin{aligned} \left(\begin{array}{c} \text{Accumulation} \\ \text{of mass} \\ \text{in the system} \end{array} \right) &= \left(\begin{array}{c} \text{Mass} \\ \text{entering} \\ \text{the System} \end{array} \right) - \left(\begin{array}{c} \text{Mass} \\ \text{leaving} \\ \text{the System} \end{array} \right) \\ &= \left[\left(\begin{array}{c} \text{Mass flow} \\ \text{entering} \\ \text{System} \end{array} \right) + \left(\begin{array}{c} \text{Mass} \\ \text{produced} \\ \text{by reactions} \end{array} \right) \right] - \left[\left(\begin{array}{c} \text{Mass flow} \\ \text{leaving} \\ \text{System} \end{array} \right) + \left(\begin{array}{c} \text{Mass} \\ \text{consumed} \\ \text{by reactions} \end{array} \right) \right] \end{aligned} \quad (2.2)$$

Theorem 2.1. (*Mass Balance of Reactors*) Consider a closed isothermal reactor system comprised of a diluted solution of constant volume V , whose reactions are described by a chemical reaction network as in Definition 1.3. The change of mass for any compound A in this reactor is described by the dynamical model:

$$\frac{d(\rho_A)}{dt} = q(\rho_{in}^{(A)} - \rho_{out}^{(A)}) + \left(\sum_{\alpha X \rightarrow \beta A} \frac{1}{\beta} K_{XA}(\rho_X)^\alpha \right) - \left(\sum_{\alpha A \rightarrow \beta X} \frac{1}{\beta} K_{AX}(\rho_A)^\alpha \right) \quad (2.3)$$

where $\rho_{in}^{(A)}$ and $\rho_{out}^{(A)}$ are the densities of A in the flows entering and leaving the system, respectively, and where $\alpha X \rightarrow \beta A$ and $\alpha A \rightarrow \beta X$ represents the reactions in the network between A and any other compound X with density ρ_X , each occurring with kinetic rates K_{XA} and K_{AX} , respectively.

Proof. First of all, as denoted in (2.2), the mass flow of any substance A entering and leaving the system, M_{in} and M_{out} , respectively, given a fluid inflow F_{in} with density $\rho_{in}^{(A)}$ and a fluid outflow F_{out} with $\rho_{out}^{(A)}$, can be calculated as:

$$M_{in} = \rho_{in}^{(A)} F_{in} \quad M_{out} = \rho_{out}^{(A)} F_{out} \quad (2.4)$$

To calculate the mass contribution from the reactions, it is necessarily first to formulate the mass contribution for a single reaction. In this case, consider a reaction between two chemical compounds X and Y , with stoichiometric numbers α and β :



Under the assumption that the reactant is in a dilute solution, the rate of this equation obeys the *law of mass action* [seek reference]. Given a constant kinetic rate K_{XY} , since the system is isothermal, and the volume of the solution as V , the mass of X consumed, $M_{cons}^{(X)}$, and the mass of Y produced, $M_{prod}^{(Y)}$, are given by the power-laws:

$$M_{cons}^{(X)} = \frac{V}{\beta} K_{XY}(\rho_X)^\alpha \quad M_{prod}^{(Y)} = \frac{V}{\alpha} K_{XY}(\rho_X)^\alpha \quad (2.6)$$

where ρ_X and ρ_Y are the respective densities of these compounds. Assuming that the network represents a set of reactions occurring within an chemical solution of volume V , the mass of a substance A that is consumed and produced by the reactions, named respectively M_{cons} and M_{prod} , are given by summing over the contribution of each reaction on the network where A is either a reactant or a product to any other compound X :

$$M_{cons} = V \sum_{\alpha A \rightarrow \beta X} \frac{1}{\beta} K_{AX}(\rho_A)^\alpha \quad M_{prod} = V \sum_{\alpha X \rightarrow \beta A} \frac{1}{\beta} K_{XA}(\rho_X)^\alpha \quad (2.7)$$

Finally, packing all together, the mass balance of any substance A in a isothermal chemical reactive system can be represented by the general dynamical model:

$$\begin{aligned} \left(\begin{array}{c} \text{Accumulation} \\ \text{of mass} \\ \text{in the system} \end{array} \right) &= \left[\left(\begin{array}{c} \text{Mass flow} \\ \text{entering} \\ \text{System} \end{array} \right) + \left(\begin{array}{c} \text{Mass} \\ \text{produced} \\ \text{by reactions} \end{array} \right) \right] - \left[\left(\begin{array}{c} \text{Mass flow} \\ \text{leaving} \\ \text{System} \end{array} \right) + \left(\begin{array}{c} \text{Mass} \\ \text{consumed} \\ \text{by reactions} \end{array} \right) \right] \\ \frac{d(\rho_A V)}{dt} &= \left[\rho_{in}^{(A)} F_{in} + V \sum_{\alpha X \rightarrow \beta A} \frac{1}{\beta} K_{XA}(\rho_X)^\alpha \right] - \left[\rho_{out}^{(A)} F_{out} + V \sum_{\alpha A \rightarrow \beta X} \frac{1}{\beta} K_{AX}(\rho_A)^\alpha \right] \end{aligned} \quad (2.8)$$

Since the system is closed, i.e., there are no leaks or unknown sources of fluids, the assumptions of a constant volume implies that $F_{in} = F_{out} = F$. Normalizing each term by the volume and substituting a new variable $q = F/V$ results in:

$$\frac{d(\rho_A)}{dt} = q(\rho_{in}^{(A)} - \rho_{out}^{(A)}) + \left(\sum_{\alpha X \rightarrow \beta A} \frac{1}{\beta} K_{XA} (\rho_X)^\alpha \right) - \left(\sum_{\alpha A \rightarrow \beta X} \frac{1}{\beta} K_{AX} (\rho_A)^\alpha \right) \quad (2.9)$$

□

Notice some important restrictions to the use of the model just presented. First of all, to calculate the mass contribution of an individual reaction was necessary to use a model which assumes that the reactor system is actually comprised of a dilute solution in some closed container. In industry, this means that the reactor system is actually a tank containing the solution. The inflow and outflow of fluid can be represented by flows through pipes which can be manipulated by some valve or pump. An illustration of such physical system is exhibited at Fig. 2.1a.

Furthermore, this model accounts for a single substance A , but the system is actually a solution of several compounds, each one with a specific concentration. From the model presented, it is visible that it is necessary to compute each concentration ρ_X before actually computing the rate of change in ρ_A . However, from the same model, the computation of the rate of change of any ρ_X may depend on ρ_A itself. Therefore, the change of mass inside the whole system is actually the result of a system of differential equations:

$$\begin{cases} \frac{d(\rho_{X_1})}{dt} = f(\rho_{X_1}, \rho_{X_2}, \dots, \rho_{X_n}, \rho_{in}^{(X_1)}, \rho_{out}^{(X_1)}, t) \\ \frac{d(\rho_{X_2})}{dt} = f(\rho_{X_1}, \rho_{X_2}, \dots, \rho_{X_n}, \rho_{in}^{(X_2)}, \rho_{out}^{(X_2)}, t) \\ \vdots \\ \frac{d(\rho_{X_n})}{dt} = f(\rho_{X_1}, \rho_{X_2}, \dots, \rho_{X_n}, \rho_{in}^{(X_n)}, \rho_{out}^{(X_n)}, t) \end{cases} \quad (2.10)$$

where X_1, X_2, \dots, X_n are the chemical compounds inside the reactor and $f(\cdot)$ is the dynamical model presented in Theorem 2.1.

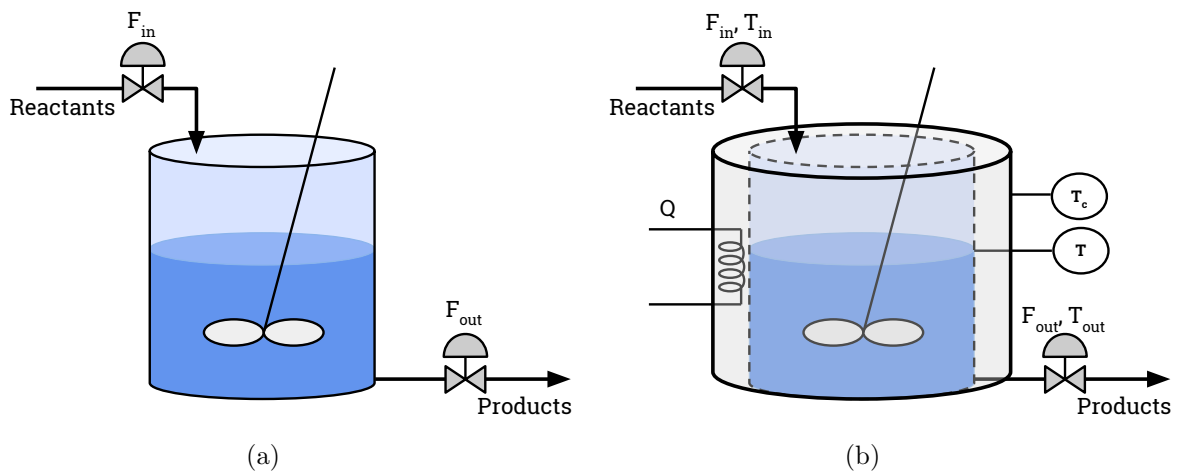


Figure 2.1: schematic representations of industrial reactor tanks for (a) a simple isothermal process and (b) a non-isothermal process with heating/cooling system.

The model discussed so far is very simple and can describe several applications. To account for more complex systems, the same modeling procedure can be applied. For instance, it is

possible to extend the description to account for exothermic and endothermic processes, when the temperature inside the system has a dynamical evolution and the dynamics of the reactions starts to depend on it.

When discussing non-isothermal processes, it is also common to discuss heating or cooling systems that tries to impose certain operational conditions to the reactions, as illustrated in Fig. 2.1b. In exothermic processes, for instance, the heat accumulated in the system tends to grows as the reactions occurs, which can be very dangerous. One approach to regulate the temperature consists in involving the chemical solution, or the container containing it, with a material whose temperature can be manipulated, transferring or absorbing heat by conductance. The temperature of this material can be manipulated by, for instance, running a heated fluid or converting electrical energy to heat energy.

Theorem 2.2. (*Non-isothermal Reactor System*) Consider a closed non-isothermal reactor system comprised of a diluted solution of constant volume V . Assume, also, that the system is involved by a cooling/heating system with capacity Q and that the reactions obeys the Arrhenius equation. The change of mass for any compound A in this reactor is described by the dynamical model:

$$\frac{d(\rho_A)}{dt} = q(\rho_{in}^{(A)} - \rho_{out}^{(A)}) + \left(\sum_{\alpha X \rightarrow \beta A} \frac{1}{\beta} K_{XA} e^{-\frac{E_{XA}}{T}} (\rho_X)^\alpha \right) - \left(\sum_{\alpha A \rightarrow \beta X} \frac{1}{\beta} K_{AX} e^{-\frac{E_{AX}}{T}} (\rho_A)^\alpha \right) \quad (2.11)$$

where E_{XA} and E_{AX} are the activation energy needed for each reaction, and the rest of the parameters are the same as defined in Theorem 2.1. Furthermore, the change of temperature inside the reactor system, T , and in the cooling/heating system, T_C , are described by the dynamical models:

$$\begin{aligned} \frac{d(T)}{dt} &= q(T_{in} - T_{out}) + \eta(T_C - T) + \delta \sum_{\alpha A \rightarrow \beta X} K_{AX} e^{-\frac{E_{AX}}{T}} (\rho_A)^\alpha \Delta H_{AX} \\ \frac{d(T_C)}{dt} &= \gamma Q + \beta(T - T_C) \end{aligned} \quad (2.12)$$

where T_{in} and T_{out} are the temperatures of the fluid inflow and outflow, respectively, ΔH_{AX} is the energy change from each reaction $\alpha A \rightarrow \beta X$ and $\eta, \delta, \gamma, \beta \in \mathbb{R}$ are proportionality parameters specific to the system and environmental conditions.

A proof of this theorem can be found in Appendix A. The non-isothermal reactor system is a more general model that accounts for the fact that the temperature of the environment is usually not constant. From this assumption, the flow entering and leaving the system are also not assumed to have the same temperature that the fluid inside the reactor. In practical applications, the temperature of the fluid inflow can either be manipulated or measured, where the temperature of the fluid outflow is actually assumed to be equal to the temperature inside the reactor. In addition, the proportionality constants are not functions of any dynamical variable, so they can be calculated before the operation of the system by using the properties of the materials and containers.

In the case of this model, the rate of changes in the chemical concentrations depends on the temperature through the Arrhenius equation. However, the temperature of the reactor itself depends on those concentrations. So, as noted in Equation (2.10), the dynamical model of the entire reactor is a system of differential equations relating all those quantities.

2.2 Mathematical Models of Systems

The last section presented the foundation for modeling a dynamical system using first principles from physics. Although it was a well-defined formulation, the resulting models are not guaranteed to be practical in a mathematical sense. This is due to the fact that the differential equations, as evidenced in Equation (2.11), are usually nonlinear functions of the variables of interest, and the analysis of such functions are quite more challenging.

In the perspective of control theory, there are two main formats for the model of a system: the *Input-Output* (IO) and the *State-Space* (SS) representations, both depicted in Fig. 2.2.

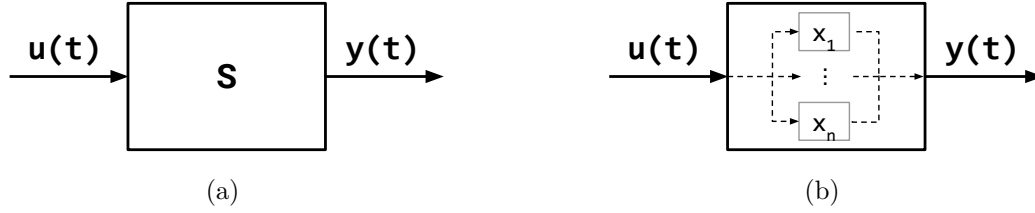


Figure 2.2: graphical interpretation of (a) Input-Output models and (b) State-Spaces models.

Definition 2.1. (Input-Output Representation) An Input-Output (IO) representation of a dynamical system with $p \geq 1$ output variables, represented by $\mathbf{y}(t) : \mathbb{R} \rightarrow \mathbb{R}^p$, and $r \geq 1$ input variables, represented by $\mathbf{u}(t) : \mathbb{R} \rightarrow \mathbb{R}^r$, is the system of differential equations:

$$\begin{cases} h_1 \left(y_1, \dot{y}_1, \dots, y_1^{(n_1)}, u_1, \dot{u}_1, \dots, u_1^{(m_{11})}, u_2, \dot{u}_2, \dots, u_2^{(m_{12})}, \dots, u_r, \dot{u}_r, \dots, u_r^{(m_{1r})}, t \right) = 0 \\ h_2 \left(y_2, \dot{y}_2, \dots, y_2^{(n_2)}, u_1, \dot{u}_1, \dots, u_1^{(m_{21})}, u_2, \dot{u}_2, \dots, u_2^{(m_{22})}, \dots, u_r, \dot{u}_r, \dots, u_r^{(m_{2r})}, t \right) = 0 \\ \vdots \\ h_p \left(y_p, \dot{y}_p, \dots, y_p^{(n_p)}, u_1, \dot{u}_1, \dots, u_1^{(m_{p1})}, u_2, \dot{u}_2, \dots, u_2^{(m_{p2})}, \dots, u_r, \dot{u}_r, \dots, u_r^{(m_{pr})}, t \right) = 0 \end{cases} \quad (2.13)$$

where:

$$\begin{aligned} \dot{y}(t) &= \frac{dy(t)}{dt}, \quad \ddot{y}(t) = \frac{d^2y(t)}{dt^2}, \quad \dots, \quad y^{(n)}(t) = \frac{d^ny(t)}{dt^n} \\ \dot{u}(t) &= \frac{du(t)}{dt}, \quad \ddot{u}(t) = \frac{d^2u(t)}{dt^2}, \quad \dots, \quad u^{(n)}(t) = \frac{d^nu(t)}{dt^n} \end{aligned}$$

An input-output representation is a simple model that describes the entire system using only two types of variables, and its derivatives. Therefore, the dimension of these variables and the order of derivatives at each differential equation provides the information about the structure of the model. For instance, in the case where $p = r = 1$ the system can be classified as a *Single-Input Single-Output* (SISO) configuration, or it can be classified as a *Multiple-Input Multiple-Output* (MIMO) configuration if $p, r > 1$.

This model presents a cause-and-effect interpretation of the system where the direct relationship between the input and output signal, and its derivatives, are equated as if the system was a processing unit. In practice, the input signals $\mathbf{u}(t)$ are the manipulated variables of the system, where the output signals $\mathbf{y}(t)$ are the observations of the controlled variables. This representation, then, brings an easy visualization on how a desired system behavior can be achieved by applying a specific input signal, posing as a practical framework for designing controllers.

Definition 2.2. (State-Space Representation) A State-Space (SS) representation of a system with $n \geq 1$ states variables, represented by $\mathbf{x}(t) : \mathbb{R} \rightarrow \mathbb{R}^n$, for $p \geq 1$ output variables, represented by $\mathbf{y}(t) : \mathbb{R} \rightarrow \mathbb{R}^p$, and $r \geq 1$ input variables, represented by $\mathbf{u}(t) : \mathbb{R} \rightarrow \mathbb{R}^r$, is given by the systems of state and output equations:

$$\begin{array}{ll} \text{State Equations:} & \text{Output Equations:} \\ \left\{ \begin{array}{l} \dot{x}_1(t) = f_1(x_1, \dots, x_n, u_1, \dots, u_r, t) \\ \dot{x}_2(t) = f_2(x_1, \dots, x_n, u_1, \dots, u_r, t) \\ \vdots \\ \dot{x}_n(t) = f_n(x_1, \dots, x_n, u_1, \dots, u_r, t) \end{array} \right. & \left\{ \begin{array}{l} y_1(t) = g_1(x_1, \dots, x_n, u_1, \dots, u_r, t) \\ y_2(t) = g_2(x_1, \dots, x_n, u_1, \dots, u_r, t) \\ \vdots \\ y_p(t) = g_p(x_1, \dots, x_n, u_1, \dots, u_r, t) \end{array} \right. \end{array} \quad (2.14)$$

or, in the matrix form:

$$\begin{cases} \dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), t) \\ \mathbf{y}(t) = \mathbf{g}(\mathbf{x}(t), \mathbf{u}(t), t) \end{cases} \quad (2.15)$$

The State-Space representation is yet another formulation for a dynamical model, but centered in the concept of *state variables*. In a formal definition, the set of state variables is the smallest set of linearly independent variables that can unequivocally determine the value of all the states variables given an initial state $\mathbf{x}(t_0)$ and a forcing function $\mathbf{u}(t)$, for any time $t \geq t_0$. In a physical perspective, however, these variables accounts for quantities that can describe the dynamics of the system, such as position or velocities, or they are latent variables that somehow stores intrinsic information about the system behavior. In comparison to the Input-Output representation, this formulation poses a simpler mathematical model, since it is composed by a system of ordinary differential equations and a system of algebraic equations. However, this model presents a semantical improvement over the latter since the inclusion of the state variables expands the internal description of the system.

In the light of these two formulations, a system can be classified in respect to the model mathematical structure. There are five main properties used for this classification: if the system is causal or non-causal, linear or nonlinear, dynamical or instantaneous, time-invariant or time-varying and with or without delay. The necessary and sufficient conditions for each one of these properties are summarized in Table 2.1.

	Input-Output	State-Space
Causal	$m_{ij} \leq n_k$ $i \in [1, \dots, p], j \in [1, \dots, r]$	Always causal
Linear	$h_i(\cdot) = \sum_{j=0}^{n_i} y^{(j)} + \dots$ $\dots + \sum_{k=1}^r \sum_{l=0}^{m_{ik}} u_k^{(l)}$ $i \in [1, 2, \dots, p]$	$f_i = \mathbf{a}_i(t)\mathbf{x}(t) + \mathbf{b}_i(t)\mathbf{u}(t), i = 1, 2, \dots, n$ $g_j = \mathbf{c}_j(t)\mathbf{x}(t) + \mathbf{d}_j(t)\mathbf{u}(t), j = 1, 2, \dots, p$ $\mathbf{a}_i, \mathbf{c}_j \in \mathbb{R}^{1 \times n}$ and $\mathbf{b}_i, \mathbf{d}_j \in \mathbb{R}^{1 \times r}$
Dynamical	$n_i > 0$ or $m_{jk} > 0$ $i, j \in [1, \dots, p], k \in [1, \dots, r]$	$n > 0$
Time-Invariant	$h_i(y_i(t), \dots, u_1(t), \dots, u_r(t)) = 0$ $i \in [1, 2, \dots, p]$	$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t))$ $\mathbf{y}(t) = \mathbf{g}(\mathbf{x}(t), \mathbf{u}(t))$
Without-Delay	All the signals share the same arguments	All the signals share the same arguments

Table 2.1: necessary and sufficient conditions for different classes of models.

This work focus on dynamical linear systems, since its models are the most well studied in the control theory community. In reality, a physical system is always causal, nonlinear and time-varying [seek references], but the models can assume different properties with fairly accuracy. The benefit of linear systems is that it obeys the superposition principle, and a linear combination of the inputs directly causes the exact same linear combination of the individual outputs. Under the assumption of a linear system, a nice result is that the vectorial functions $\mathbf{f}(\cdot)$ and $\mathbf{g}(\cdot)$ of the State-Space representation in (2.23) reduces to a simple matrix form.

Definition 2.3. (Linear State-Space Representation) A State-Space representation describing a linear system with state vector $\mathbf{x}(t) : \mathbb{R} \rightarrow \mathbb{R}^n$, output vector $\mathbf{y}(t) : \mathbb{R} \rightarrow \mathbb{R}^p$ and input vector $\mathbf{u}(t) : \mathbb{R} \rightarrow \mathbb{R}^r$ is given by the system of equations:

$$\begin{cases} \dot{\mathbf{x}}(t) = \mathbf{A}(t)\mathbf{x}(t) + \mathbf{B}(t)\mathbf{u}(t) \\ \mathbf{y}(t) = \mathbf{C}(t)\mathbf{x}(t) + \mathbf{D}(t)\mathbf{u}(t) \end{cases} \quad (2.16)$$

where $\mathbf{A}(t) : \mathbb{R} \rightarrow \mathbb{R}^{n \times n}$, $\mathbf{B}(t) : \mathbb{R} \rightarrow \mathbb{R}^{n \times r}$, $\mathbf{C}(t) : \mathbb{R} \rightarrow \mathbb{R}^{p \times n}$ and $\mathbf{D}(t) : \mathbb{R} \rightarrow \mathbb{R}^{p \times r}$. In the case of a time-invariant linear system, these matrices becomes constants.

This formulation has the advantages that the time response of the system can be easily calculated and that the analysis of the dynamics follows well-established results from linear algebra applied to the matrices $\mathbf{A}(t)$, $\mathbf{B}(t)$, $\mathbf{C}(t)$ or $\mathbf{D}(t)$, as well as for the vectors $\mathbf{x}(t)$ and $\mathbf{u}(t)$. Furthermore, the physical interpretation of the system through the state variables becomes straightforward in this model, even in the case of latent variables.

In addition to the State-Space representation, the linear assumption also benefits Input-Output representations. One major analytical tool that can be used in these cases is to transform this model to a frequency domain, using a linear transform operator, in order to simplify the solution for the differential equations. The most popular choice of transformation is the *Laplace transform*, $\mathcal{L}\{h(t)\}$, which converts functions in time to functions in complex frequencies. Using the properties of this operator, differential equations are converted to simple algebraic equations.

Theorem 2.3. (Transfer Function) Given a linear model for a SISO system, with initial conditions $\mathbf{y}(0^-) \equiv \mathbf{u}(0^-) \equiv 0$, in the Input-Output formulation:

$$\alpha_n \frac{d^n y(t)}{dt^n} + \dots + \alpha_1 \frac{dy(t)}{dt} + \alpha_0 y(t) = \beta_m \frac{d^m u(t)}{dt^m} + \dots + \beta_1 \frac{du(t)}{dt} + \beta_0 u(t) \quad (2.17)$$

Its transfer function, in the Laplace domain, is calculated as:

$$G(s) = \frac{Y(s)}{U(s)} = \frac{\beta_m s^m + \beta_{m-1} s^{m-1} + \dots + \beta_1 s + \beta_0}{\alpha_n s^n + \alpha_{n-1} s^{n-1} + \dots + \alpha_1 s + \alpha_0} \quad (2.18)$$

An indirect result of this is that the SS representation can be converted to the IO representation using the Laplace transform operator, leading to a notion of equivalence between the two representations. Notice that the extension to the MIMO case is straightforward: computing the transfer function between each pair of input and output results in the matrix $\mathbf{G} \in \mathbb{C}^{n \times m}$.

Theorem 2.4. (*Passage from SS to IO*) Consider a linear and time-invariant system in State-Space form with initial states $\mathbf{x}(0^-) \equiv 0$ and represented as:

$$\begin{cases} \dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{u}(t) \\ \mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) + \mathbf{D}(t) \end{cases} \quad (2.19)$$

The equivalent system in Input-Output representation is given by the transfer function:

$$\mathbf{G}(s) = \mathbf{Y}(s)\mathbf{U}^{-1}(s) = \mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} + \mathbf{D} \quad (2.20)$$

Proof. Applying the Laplace transfer in both sides of the equation:

$$\begin{aligned} \mathcal{L}\{\dot{\mathbf{x}}(t)\} &= \mathbf{A}\mathcal{L}\{\mathbf{x}(t)\} + \mathbf{B}\mathcal{L}\{u(t)\} \\ s\mathbf{X}(s) - \dot{\mathbf{x}}(0^-) &= \mathbf{A}\mathbf{X}(s) + \mathbf{B}\mathbf{U}(s) \\ (s\mathbf{I} - \mathbf{A})\mathbf{X}(s) &= \mathbf{B}\mathbf{U}(s) \\ \mathbf{X}(s) &= (s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}\mathbf{U}(s) \end{aligned} \quad (2.21)$$

By applying the same procedure in the output equations, using the previous result, the Laplace transform of the output is:

$$\begin{aligned} \mathcal{L}\{\mathbf{y}(t)\} &= \mathbf{C}\mathcal{L}\{\mathbf{x}(t)\} + \mathbf{D}\mathcal{L}\{u(t)\} \\ \mathbf{Y}(s) &= \mathbf{C}\mathbf{X}(s) + \mathbf{D}\mathbf{U}(s) \\ \mathbf{Y}(s) &= \mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}\mathbf{U}(s) + \mathbf{D}\mathbf{U}(s) \\ \mathbf{Y}(s) &= (\mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} + \mathbf{D})\mathbf{U}(s) \end{aligned} \quad (2.22)$$

In conclusion, it is possible to obtain the transfer function matrix $\mathbf{G}(s) = \mathbf{Y}(s)\mathbf{U}^{-1}(s) = \mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} + \mathbf{D}$ as an equivalent representation of the same system. \square

Despite of the discussion about the benefits of linear models, it is necessary to account for the fact that physical systems will present, in most situations, nonlinear behavior. For this reason, some effort must be done to develop a linear model that can describe the nonlinear behavior with certain accuracy, even if over some small region of the space. With this motivation, a technique for *linearization* of a nonlinear model is detailed below.

Theorem 2.5. (*Linearization by Taylor Expansion*) Consider a nonlinear time-invariant system:

$$\begin{cases} \dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t)) \\ \mathbf{y}(t) = \mathbf{g}(\mathbf{x}(t), \mathbf{u}(t)) \end{cases} \quad (2.23)$$

Given steady-state operating points \mathbf{x}_o , \mathbf{y}_o and \mathbf{u}_o , the dynamics of the system in the neighborhood of these points can be represented by the linear model:

$$\begin{cases} \Delta\dot{\mathbf{x}}(t) = \mathbf{A}\Delta\mathbf{x}(t) + \mathbf{B}\Delta\mathbf{u}(t) \\ \mathbf{y}(t) = \mathbf{C}\Delta\mathbf{x}(t) + \mathbf{D}\Delta\mathbf{u}(t) \end{cases} \quad (2.24)$$

where

$$\mathbf{A} = \left. \frac{\partial \mathbf{f}}{\partial \mathbf{x}} \right|_{\mathbf{x}_o, \mathbf{y}_o, \mathbf{u}_o} \quad \mathbf{B} = \left. \frac{\partial \mathbf{f}}{\partial \mathbf{u}} \right|_{\mathbf{x}_o, \mathbf{y}_o, \mathbf{u}_o} \quad \mathbf{C} = \left. \frac{\partial \mathbf{g}}{\partial \mathbf{x}} \right|_{\mathbf{x}_o, \mathbf{y}_o, \mathbf{u}_o} \quad \mathbf{D} = \left. \frac{\partial \mathbf{g}}{\partial \mathbf{u}} \right|_{\mathbf{x}_o, \mathbf{y}_o, \mathbf{u}_o} \quad (2.25)$$

and

$$\Delta\mathbf{x}(t) = \mathbf{x}(t) - \mathbf{x}_o \quad \Delta\mathbf{y}(t) = \mathbf{y}(t) - \mathbf{y}_o \quad \Delta\mathbf{u}(t) = \mathbf{u}(t) - \mathbf{u}_o \quad (2.26)$$

Proof. Consider a system represented by state equations $\mathbf{f}(\cdot)$ and output equations $\mathbf{g}(\cdot)$, with steady-state operation points \mathbf{x}_o , \mathbf{y}_o and \mathbf{u}_o . Now, consider a very small perturbation $\Delta \mathbf{u}(t)$ in the input signal around these operation points. This perturbation will result in small changes in the state and output variables:

$$\mathbf{x}(t) = \mathbf{x}_o + \Delta \mathbf{x}(t) \quad \mathbf{u}(t) = \mathbf{u}_o + \Delta \mathbf{u}(t) \quad \mathbf{y}(t) = \mathbf{y}_o + \Delta \mathbf{y}(t) \quad (2.27)$$

This results in the following configuration on the State-Space:

$$\begin{cases} \frac{d(\mathbf{x}_o + \Delta \mathbf{x}(t))}{dt} = \mathbf{f}(\mathbf{x}_o + \Delta \mathbf{x}(t), \mathbf{u}_o + \Delta \mathbf{u}(t)) \\ \mathbf{y}_o + \Delta \mathbf{y}(t) = \mathbf{g}(\mathbf{x}_o + \Delta \mathbf{x}(t), \mathbf{u}_o + \Delta \mathbf{u}(t)) \end{cases} \quad (2.28)$$

where $d(\mathbf{x}_o + \Delta \mathbf{x}(t))/dt = d(\Delta \mathbf{x}(t))/dt$, since \mathbf{x}_o is constant. The perturbed variables are very close to the operation points, hence the functions $\mathbf{f}(\cdot)$ and $\mathbf{g}(\cdot)$ can be approximated by a Taylor series expansion, yielding:

$$\begin{cases} \frac{d(\Delta \mathbf{x}(t))}{dt} = \mathbf{f}(\mathbf{x}_o, \mathbf{u}_o) + \left. \frac{\partial \mathbf{f}}{\partial \mathbf{x}} \right|_{\mathbf{x}_o, \mathbf{y}_o, \mathbf{u}_o} \Delta \mathbf{x}(t) + \left. \frac{\partial \mathbf{f}}{\partial \mathbf{u}} \right|_{\mathbf{x}_o, \mathbf{y}_o, \mathbf{u}_o} \Delta \mathbf{u}(t) + \text{high order terms} \\ \mathbf{y}_o + \Delta \mathbf{y}(t) = \mathbf{g}(\mathbf{x}_o, \mathbf{u}_o) + \left. \frac{\partial \mathbf{g}}{\partial \mathbf{x}} \right|_{\mathbf{x}_o, \mathbf{y}_o, \mathbf{u}_o} \Delta \mathbf{x}(t) + \left. \frac{\partial \mathbf{g}}{\partial \mathbf{u}} \right|_{\mathbf{x}_o, \mathbf{y}_o, \mathbf{u}_o} \Delta \mathbf{u}(t) + \text{high order terms} \end{cases} \quad (2.29)$$

Since the steady-state condition implies zero variation, it is possible to assume $\mathbf{f}(\mathbf{x}_o, \mathbf{u}_o) = 0$ and $\mathbf{g}(\mathbf{x}_o, \mathbf{u}_o) = 0$, since they are ordinary differential equations. Truncating in the first order terms and substituting $\Delta \mathbf{x}(t) = \mathbf{x}(t) - \mathbf{x}_o$, $\Delta \mathbf{u}(t) = \mathbf{u}(t) - \mathbf{u}_o$ and $\Delta \mathbf{y}(t) = \mathbf{y}(t) - \mathbf{y}_o$ results in:

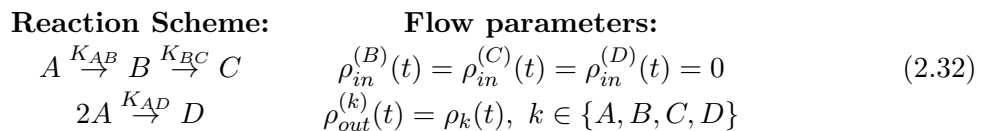
$$\begin{cases} \frac{d(\Delta \mathbf{x}(t))}{dt} = \left. \frac{\partial \mathbf{f}}{\partial \mathbf{x}} \right|_{\mathbf{x}_o, \mathbf{y}_o, \mathbf{u}_o} (\mathbf{x}(t) - \mathbf{x}_o) + \left. \frac{\partial \mathbf{f}}{\partial \mathbf{u}} \right|_{\mathbf{x}_o, \mathbf{y}_o, \mathbf{u}_o} (\mathbf{u}(t) - \mathbf{u}_o) \\ \mathbf{y}(t) - \mathbf{y}_o = \left. \frac{\partial \mathbf{g}}{\partial \mathbf{x}} \right|_{\mathbf{x}_o, \mathbf{y}_o, \mathbf{u}_o} (\mathbf{x}(t) - \mathbf{x}_o) + \left. \frac{\partial \mathbf{g}}{\partial \mathbf{u}} \right|_{\mathbf{x}_o, \mathbf{y}_o, \mathbf{u}_o} (\mathbf{u}(t) - \mathbf{u}_o) \end{cases} \quad (2.30)$$

Finally, since all the Jacobians involved are actually matrices of appropriate dimensions, the final linear approximation of the system is the SS model given by:

$$\begin{cases} \Delta \dot{\mathbf{x}}(t) = \mathbf{A} \Delta \mathbf{x}(t) + \mathbf{B} \Delta \mathbf{u}(t) \\ \mathbf{y}(t) = \mathbf{C} \Delta \mathbf{x}(t) + \mathbf{D} \Delta \mathbf{u}(t) \end{cases} \quad (2.31)$$

□

Note that this method can be used to convert non-linear models obtained from the first-principle procedure to a desirable state-space form, since those models are usually in the non-linear state-space representation. Consider for instance, a reactor system describing an isothermal process that follows the Van de Vusse reaction scheme [seek reference]:



Consider $\mathbf{x} = [\rho_A, \rho_B, \rho_C, \rho_D]^T$ and $u = q$, and consider also that the states are perfectly observed so, $\mathbf{y} = \mathbf{x}$. Using Theorem 2.1, the state-equation for the nonlinear state-space representation of this reactor is:

$$\begin{cases} \dot{x}_1(t) = u(t)(\rho_{in}^{(A)} - x_1(t)) - (K_{AB}x_1(t) + K_{AD}(x_1(t))^2) \\ \dot{x}_2(t) = -u(t)x_2(t) + K_{AB}x_1(t) - K_{BC}x_2(t) \\ \dot{x}_3(t) = -u(t)x_3(t) + K_{BC}x_2(t) \\ \dot{x}_4(t) = -u(t)x_4(t) + \frac{1}{2}K_{AD}(x_1(t))^2 \end{cases} \quad (2.33)$$

To select steady-state operation points \mathbf{x}_o and \mathbf{u}_o , it is necessary to find values for $\mathbf{x}(t)$ and $u(t)$ that satisfy the system $\dot{\mathbf{x}}(t) = \mathbf{0}$. This can be done analytically, in some cases, numerically or by simply simulating the nonlinear system and measuring these quantities when the states displays zero variation. Finally, the linear matrices $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$ are obtained by deriving each one of these equations in respect to each state and input, resulting:

$$\mathbf{A} = \begin{bmatrix} -u_o - K_{AB} - 2K_{AC}x_{1o} & 0 & 0 & 0 \\ K_{AB} & -u_o - K_{BC} & 0 & 0 \\ 0 & K_{BC} & -u_o & 0 \\ K_{AD}x_{1o} & 0 & 0 & -u_o \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} \rho_{in}^{(A)} - x_{1o} \\ -x_{2o} \\ -x_{3o} \\ -x_{4o} \end{bmatrix} \quad (2.34)$$

$$\mathbf{C} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \mathbf{D} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Notice that the response of this model represents the deviations around the given operation points. If the errors in approximating the non-linearities of the actual system are somehow tolerable, this model can still be used as a nominal model to represent all the operations, given that the response is vertically “corrected” as $\mathbf{x}(t) = \Delta\mathbf{x}(t) + \mathbf{x}_o$ (and for initial states $\Delta\mathbf{x}_0 = \mathbf{x}(t_0) - \mathbf{x}_o$ and input signal $\Delta\mathbf{u}(t) = \mathbf{u}(t) - \mathbf{u}_o$). The simulation of both the nonlinear and the linear systems are displayed in Fig. 2.3, with the dashed line the linear response of a system given $\mathbf{x}_o = [6.19, 1.09, 0.60, 1.05]^T$ and $u_0 = 3.03$.

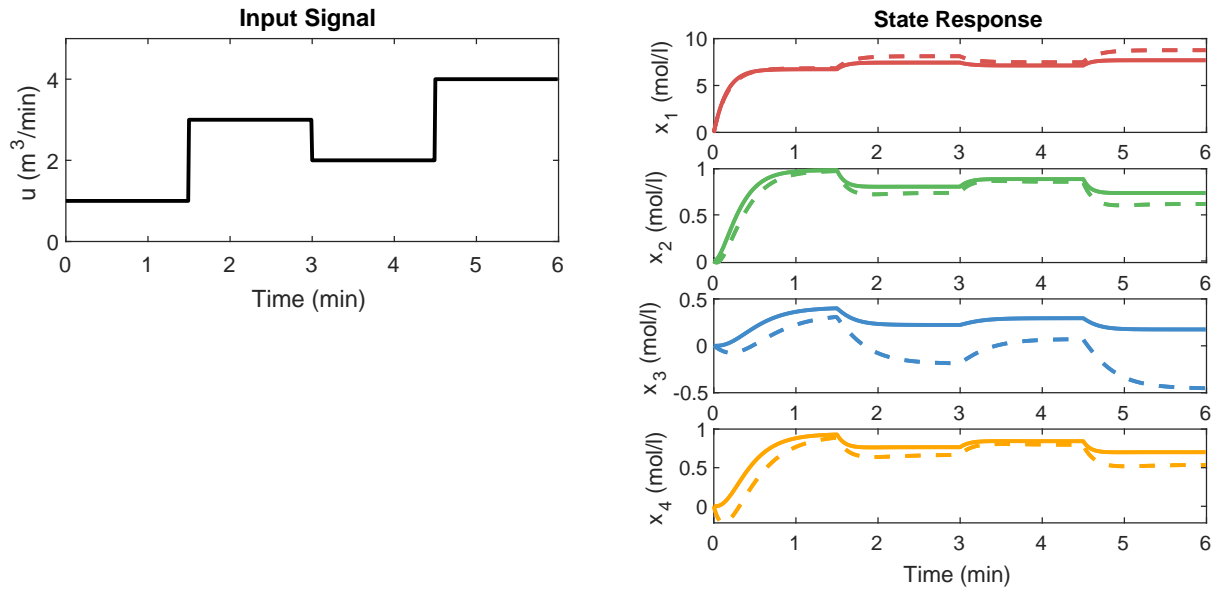


Figure 2.3: comparison between the nonlinear and linear response of the reactor system.

2.3 Time Response Analysis

Once that a model is well-established, it is possible both to simulate the system and to analyze its response, both for a natural or forced regime. This section, then, focus on developing a quantitative understanding of a system behavior through a linear model. The results are focused on continuous-time response of linear and time-invariant (LTI) systems in the form:

$$\begin{cases} \dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) \\ \mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) + \mathbf{D}\mathbf{u}(t) \end{cases} \quad (2.35)$$

First of all, it is necessary so access some properties of the matrix \mathbf{A} , which describes the natural evolution of the states.

Definition 2.4. (State-Transition Matrix) Consider a system in State-Space representation with matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$. Its *State-Transition Matrix*, $e^{\mathbf{A}t} \in \mathbb{R}^{n \times n}$, is the converging series:

$$e^{\mathbf{A}t} = \mathbf{I} + \mathbf{A}t + \frac{\mathbf{A}^2 t^2}{2!} + \frac{\mathbf{A}^3 t^3}{3!} + \cdots = \sum_{k=0}^{\infty} \frac{\mathbf{A}^k t^k}{k!} \quad (2.36)$$

This matrix is central to the computation of the system time response, as it will be shown shortly. Since the matrix \mathbf{A} is a square matrix, this operation leads to some useful properties.

Theorem 2.6. Consider a matrix exponential as in Definition 2.4, for a matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$. Then, the following properties holds:

$$1) \frac{d(e^{\mathbf{A}t})}{dt} = \mathbf{A}e^{\mathbf{A}t} \quad 2) e^{\mathbf{A}t}e^{\mathbf{A}\tau} = e^{\mathbf{A}(t+\tau)} \quad 3) e^{-\mathbf{A}t}e^{\mathbf{A}t} = e^{\mathbf{A}t}e^{-\mathbf{A}t} = \mathbf{I} \quad (2.37)$$

The proof is direct from Definition 2.4, and a detailed discussion of these properties is omitted. Given these results, the calculation of the time response of a system in SS representation becomes straightforward.

Theorem 2.7. (Lagrange Formula) Consider a LTI system in state-space representation. Its response for any time $t \geq t_0$, initial state $\mathbf{x}(t_0)$ and input signal $\mathbf{u}(t)$ is given by the solutions of the state and output equation:

$$\begin{cases} \mathbf{x}(t) = e^{\mathbf{A}(t-t_0)}\mathbf{x}(t_0) + \int_{t_0}^t e^{\mathbf{A}(t-\tau)}\mathbf{B}\mathbf{u}(\tau)d\tau \\ \mathbf{y}(t) = \mathbf{C}e^{\mathbf{A}(t-t_0)}\mathbf{x}(t_0) + \mathbf{C} \int_{t_0}^t e^{\mathbf{A}(t-\tau)}\mathbf{B}\mathbf{u}(\tau)d\tau + \mathbf{D}\mathbf{u}(t) \end{cases} \quad (2.38)$$

Proof. First of all, consider a system in SS representation with state equation as defined in Equation (2.23). Multiplying both sides by $e^{-\mathbf{A}t}$:

$$\begin{aligned} e^{-\mathbf{A}t}\dot{\mathbf{x}}(t) &= e^{-\mathbf{A}t}(\mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t)) \\ e^{-\mathbf{A}t}\dot{\mathbf{x}}(t) - \mathbf{A}e^{-\mathbf{A}t}\mathbf{x}(t) &= e^{-\mathbf{A}t}\mathbf{B}\mathbf{u}(t) \end{aligned} \quad (2.39)$$

Using the first assertive in Theorem 2.6, it is easy to see that $d[e^{-\mathbf{A}t}\mathbf{x}(t)]/dt = e^{-\mathbf{A}t}\dot{\mathbf{x}}(t) - \mathbf{A}e^{-\mathbf{A}t}\mathbf{x}(t)$. Substituting this result in (2.39) and integrating both sides from t_0 to t :

$$\begin{aligned} \frac{d(e^{-\mathbf{A}t}\mathbf{x}(t))}{dt} &= e^{-\mathbf{A}t}\mathbf{B}\mathbf{u}(t) \\ e^{-\mathbf{A}t}\mathbf{x}(t)|_{t_0}^t &= \int_{t_0}^t e^{-\mathbf{A}t}\mathbf{B}\mathbf{u}(t)dt \\ e^{-\mathbf{A}t}\mathbf{x}(t) - e^{-\mathbf{A}t_0}\mathbf{x}(t_0) &= \int_{t_0}^t e^{-\mathbf{A}\tau}\mathbf{B}\mathbf{u}(\tau)d\tau \end{aligned} \quad (2.40)$$

Multiplying both sides by $e^{\mathbf{A}t}$ and using the second and third assertive from Theorem 2.6, the state response can be calculated as:

$$\begin{aligned} e^{\mathbf{A}t}(e^{-\mathbf{A}t}\mathbf{x}(t) - e^{-\mathbf{A}t_0}\mathbf{x}(t_0)) &= e^{\mathbf{A}t} \int_{t_0}^t e^{-\mathbf{A}\tau}\mathbf{B}\mathbf{u}(\tau)d\tau \\ \mathbf{I}\mathbf{x}(t) - e^{\mathbf{A}(t-t_0)}\mathbf{x}(t_0) &= \int_{t_0}^t e^{\mathbf{A}(t-\tau)}\mathbf{B}\mathbf{u}(\tau)d\tau \\ \mathbf{x}(t) &= e^{\mathbf{A}(t-t_0)}\mathbf{x}(t_0) + \int_{t_0}^t e^{\mathbf{A}(t-\tau)}\mathbf{B}\mathbf{u}(\tau)d\tau \end{aligned} \quad (2.41)$$

Finally, substituting (2.41) into the output equation leads to:

$$\begin{aligned} \mathbf{y}(t) &= \mathbf{C} \left(e^{\mathbf{A}(t-t_0)}\mathbf{x}(t_0) + \int_{t_0}^t e^{\mathbf{A}(t-\tau)}\mathbf{B}\mathbf{u}(\tau)d\tau \right) + \mathbf{D}\mathbf{u}(t) \\ \mathbf{y}(t) &= \mathbf{C}e^{\mathbf{A}(t-t_0)}\mathbf{x}(t_0) + \mathbf{C} \int_{t_0}^t e^{\mathbf{A}(t-\tau)}\mathbf{B}\mathbf{u}(\tau)d\tau + \mathbf{D}\mathbf{u}(t) \end{aligned} \quad (2.42)$$

□

When discussing the response of a system, the focus is actually directed to the state equation describing its dynamics, since the output equation only represents an observation of the system through the states. In this sense, the Lagrange formula exposes the nice characteristic of linear systems that the total response is a composition of two separated actions:

$$\mathbf{x}(t) = \mathbf{x}_n(t) + \mathbf{x}_f(t) \quad (2.43)$$

where the natural response, $\mathbf{x}_n(t)$, corresponds to the state-transition matrix multiplication term and the forced response, $\mathbf{x}_f(t)$, corresponds to the integral term. This concept is illustrated in Fig. 2.4 for a state of the model derived in Equation (2.34) given an unitary step input. In this plot, the solid line represents the total response, whereas the dashed and dotted lines represents the natural and forced response, respectively. It is easy to verify that the decomposition of the total response is equal to the sum of those independent components.

The only problem remaining to fully characterize the dynamical response of a system is to compute the state-transition matrix, which can be done in several ways [Moler and Loan, 2003]. A specific method, known as the *Sylvester expansion*, is an analytical solution that brings an interesting understanding of the system behavior through the state-state matrix \mathbf{A} .

Theorem 2.8. (*Sylvester expansion*) Consider a matrix exponential function $f(\mathbf{A}) = e^{\mathbf{A}t}$ for any square matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$, whose distinct eigenvalues are $\lambda \in \mathbb{R}^m$, $m \leq n$, with associated multiplicity vector $\nu \in \mathbb{R}^m$ such that $\sum_j \nu_j = n$. The result of this function can be expanded as:

$$f(\mathbf{A}) = e^{\mathbf{A}t} = \beta_0(t)\mathbf{I} + \beta_1(t)\mathbf{A} + \beta_2(t)\mathbf{A}^2 + \dots + \beta_{n-1}(t)\mathbf{A}^{n-1} = \sum_{i=0}^{n-1} \beta_i(t)\mathbf{A}^i \quad (2.44)$$

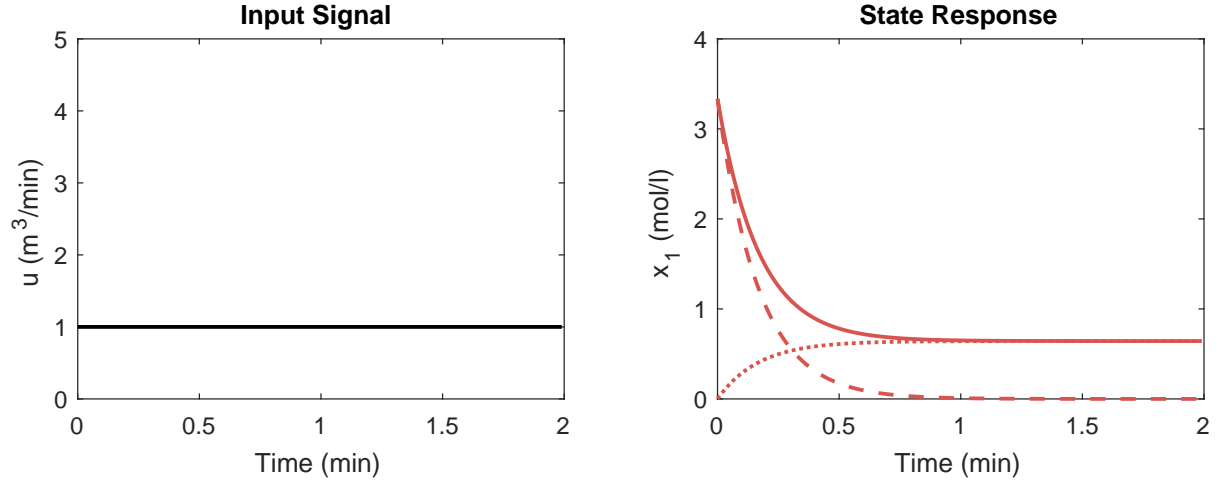


Figure 2.4: illustration of the total response decomposition for linear systems, based on the model from Equation (2.34) given $\mathbf{x}_o = [6.19, 1.09, 0.60, 1.05]^T$ and $u_0 = 3.03$.

where $\beta_i(t) : \mathbb{R} \rightarrow \mathbb{R}$, $i \in [1, 2, \dots, n-1]$, are the functions of time that solve the linear system:

$$\mathbf{V}\boldsymbol{\beta} = \boldsymbol{\eta} \quad (2.45)$$

for the parameter matrix $\boldsymbol{\beta} = [\beta_0(t), \beta_1(t), \dots, \beta_n(t)]^T$, and the vector of modes $\boldsymbol{\eta} = [\eta_1, \eta_2, \dots, \eta_m]^T$ and the confluent Vandermonde matrix $\mathbf{V} = [V_1, V_2, \dots, V_m]^T$ given as

$$\eta_i = [e^{\lambda_i t} \quad t e^{\lambda_i t} \quad t^2 e^{\lambda_i t} \quad \dots \quad t^{\nu_i-1} e^{\lambda_i t}]^T$$

$$V_j = \begin{bmatrix} 1 & \lambda_j & \lambda_j^2 & \dots & \lambda_j^{(\nu_j-1)} & \dots & \lambda_j^{n-1} \\ 0 & 1 & 2\lambda_j & \dots & (\nu_j-1)\lambda_j^{(\nu_j-1)} & \dots & (n-1)\lambda_j^{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & (\nu_j-1)! & \dots & \frac{(n-1)!}{(n-\nu_j)!} \lambda_j^{n-\nu_j} \end{bmatrix}$$

The proof of this expansion is somewhat extensive, but a detailed explanation can be found in [seek reference]. Basically, the expansion in Equation (2.44) is a direct application of the Cayley-Hamilton theorem [seek reference] and the linear system in Equation (2.45) is obtained through the Sylvester's formula [seek reference].

From the expansion presented in Theorem 2.8, it is possible to understand the relationship between the state-transition matrix $e^{\mathbf{A}t}$ and each eigenvalue λ of the matrix \mathbf{A} . First of all, notice that the formulation of the linear system that defines the parameters $\beta_0(t), \beta_1(t), \dots, \beta_{n-1}(t)$ implies that each one of these functions are linear combinations of the exponentials $e^{\lambda_i t}$ for each eigenvalue λ_i , $i = 1, 2, \dots, m$. These exponentials are known as the *modes* of the matrix \mathbf{A} . Since the Sylvester expansion is linear in those parameters, it is possible to conclude that the state-transition matrix, and consequently the response of a system, is a linear combination of the modes.

Consider, for the sake of illustration, the same reactor model from Equation (2.34), but considering only the first two states (since they are independent of the others). Let $\mathbf{x}_o = [6.19, 1.09]^T$ and $u_0 = 3.03$, just as before. The resulting matrix \mathbf{A} and state-transition matrix

e^{At} are shown below, where is easy to see that $\lambda =$:

$$\mathbf{A} = \begin{bmatrix} -5.93 & 0 \\ 0.83 & -4.70 \end{bmatrix} \quad e^{At} = \begin{bmatrix} e^{-5.93t} & 0 \\ 0.68e^{-4.7t} - 0.68e^{-5.93t} & e^{-4.70t} \end{bmatrix} \quad (2.46)$$

A simulation in time, shown in Figure !!!, shows the evolution of each element of e^{At} for a given time interval. Considering only the natural response $\mathbf{x}_n(t)$, i.e., setting $\mathbf{u}(t) \equiv 0, \forall t \in [t_0, t)$, it is evident that the actual response of each state in the system is a row-wise weighted sum of these elements, where the weights are given by the initial state $\mathbf{x}(t_0)$. Therefore, the (i, j) element of this matrix describes how the j -th state affects the response of the i -th state.

[fig]

Now, attention must be drawn to the case where the eigenvalues are not real, but complex and conjugate.

Theorem 2.9. Consider the same matrix \mathbf{A} defined in Theorem 2.8. Consider, now, that this matrix has two distinct eigenvalues $\lambda_c, \lambda'_c \in \mathbb{C}$ in the form $\lambda_c, \lambda'_c = \alpha \pm j\omega$. In this case, the linear system solved by the parameters $\beta_0(t), \beta_1(t), \dots, \beta_{n-1}(t)$ will have two equations:

$$\begin{cases} \beta_0 + \alpha\beta_1 + \alpha^2\beta_2 + \dots + \alpha^{n-1}\beta_{n-1} &= e^{\alpha t} \cos(\omega t) \\ 0 + \omega\beta_1 + \omega^2\beta_2 + \dots + \omega^{n-1}\beta_{n-1} &= e^{\alpha t} \sin(\omega t) \end{cases} \quad (2.47)$$

Proof. Consider the matrix \mathbf{A} with eigenvalues as specified and the Sylvester expansion as presented. In this case, there will be two equations in the system:

$$\begin{cases} \beta_0 + \lambda_c\beta_1 + \lambda_c^2\beta_2 + \dots + \lambda_c^{n-1}\beta_{n-1} &= e^{\lambda_c t} \\ \beta_0 + \lambda'_c\beta_1 + \lambda'^2_c\beta_2 + \dots + \lambda'^{n-1}_c\beta_{n-1} &= e^{\lambda'_c t} \end{cases} \quad (2.48)$$

Since the eigenvalues are complex and conjugate, it has that $\lambda_c + \lambda'_c = 2\Re[\lambda_c]$ and $\lambda_c - \lambda'_c = 2j\Im[\lambda_c]$. Moreover, the Euler identity $e^{\alpha \pm j\omega} = e^{\alpha t}(\cos(\omega t) \pm j\sin(\omega t))$ shows that $e^{\lambda} + e^{\lambda'} = 2e^{\alpha t} \cos(\omega t)$ and $e^{\lambda} - e^{\lambda'} = 2e^{\alpha t} \sin(\omega t)$. Finally, the first row of 2.49 is obtained by summing the two rows of (2.48) and dividing by 2 and the second row of (2.49) is obtained by subtracting the first row by the second row of (2.48) and dividing by $2j$. \square

From the same reasons stated before, this result implies that the actual response of the system will have sinusoidal components that produces oscillations in the response. The modes associated with complex and conjugate eigenvalues are classified as *pseudo-periodic*, since they are composed by an exponential growth (or decay) enveloping a sinusoidal function. Consider, again for the sake of illustration, the following system in SS representation:

$$\mathbf{A} = \begin{bmatrix} \alpha & \alpha \\ \alpha & \alpha \end{bmatrix} \quad (2.49)$$

The eigenvalues of this system are $\boldsymbol{\lambda} = [1, 1]$. Solving the linear system defined in the Sylvester expansion, considering the equivalent form of Theorem 2.9, yields $\boldsymbol{\beta} = [1, 1]$. The resulting state-transition matrix is:

$$e^{At} = \begin{bmatrix} \alpha & \alpha \\ \alpha & \alpha \end{bmatrix} \quad (2.50)$$

The elements of the resulting state-transition matrix are simulated for an specific time-interval and shown in Figure !!! . It is possible to see, in this case, the pseudo-periodic behavior of the complex conjugate modes, where the dashed lines represents the exponential envelope.

[fig]

The previous discussion on the modes of matrix \mathbf{A} introduced the importance of the eigenvalues of this matrix in analyzing the system response. The analysis, however, was focused on the natural response $\mathbf{x}_n(t)$, whereas the forced response $\mathbf{x}_f(t)$ was deliberately excluded. Now, the analysis focus the opposite case.

Consider, for instance, the state response for the two systems described by the matrices $(\mathbf{A}^{(1)}, b^{(1)})$ and $(\mathbf{A}^{(2)}, b^{(2)})$ given below. Considering initial states $\mathbf{x}^{(1)}(t_0) \equiv \mathbf{x}^{(2)}(t_0) \equiv 0$ and the input signals $\mathbf{u}^{(1)}(t) \equiv \mathbf{u}^{(2)}(t) \equiv 1, t \in [t_0, \infty)$, the time evolution of these states are given in !!!.

$$\mathbf{A}^{(1)} = [\alpha] \quad b^{(1)} = [\alpha] \quad \mathbf{A}^{(2)} = \begin{bmatrix} \alpha & \alpha \\ \alpha & \alpha \end{bmatrix} \quad b^{(2)} = \begin{bmatrix} \alpha \\ \alpha \end{bmatrix} \quad (2.51)$$

[Fig]

Those systems are comprised by, basically, just one aperiodic and one pseudo-periodic mode, respectively. Therefore, it is possible to consider that those are characteristic response of the modes to a unitary step input. The forced response to an unitary step is the default evolution used in literature to characterize the behavior of modes (and ultimately, of systems) in respect to some specifications defined below [seek reference].

Definition 2.5. (Unit-Step Response Specifications) Given an aperiodic mode $e^{\lambda t}$ of the matrix \mathbf{A} associated with the eigenvalue $\lambda \in \mathbb{R}$, its contribution to the response has a time constant (τ) defined as

$$\tau = -\frac{1}{\lambda} \quad (2.52)$$

Furthermore, given pseudo-periodic modes $e^{\lambda t}$ and $e^{\lambda' t}$ of the matrix \mathbf{A} associated with the eigenvalues $\lambda, \lambda' = \alpha \pm j\omega$, their contribution to the response has a time constant (τ), a natural frequency (ω_n) and a damping coefficient (ζ) defined as:

$$\tau = -\frac{1}{\alpha} \quad \omega_n = \sqrt{\alpha^2 + \omega^2} \quad \zeta = -\frac{\alpha}{\omega_n} \quad (2.53)$$

The specifications just defined provides a quantitative way to describe the response of a system in terms of time and frequencies. The time constant, for instance, is a quantity that represents the time needed for the mode to lost 63% of its initial value, since $e^{\lambda\tau} = e^{-1} = 0.37$. A greater value of a time constant indicates that the system is able to “discharge” energy faster. The damping coefficient, by its turn, provides an information about the intensity of the peak in the pseudo-periodic responses, which is known as *overshoot* (or *undershoot* in the case that of a negative peak). From the perspective of control theory, these are some of the specifications used to define desirable transient responses to a controlled system.

Notice that the response specifications are always functions of the real and imaginary parts of the discussed eigenvalues. This brings the possibility of a visualization in the complex plane to interpret how each eigenvalue contributes to the total response. A straightforward notion is that the closer an eigenvalue is to the imaginary axis, the faster is its contribution. Similarly, the furthest an eigenvalue is to the real axis, the more oscillatory is its contribution. Finally, a vector from the origin of the plane to a complex eigenvalue has a norm equal to the natural frequency (ω_n) and the cosine of the angle formed with the imaginary axis is equal to the damping factor (ζ). Finally, a simulation of the contributions from different eigenvalues are shown in !!!!.

[fig]

2.4 Similarity Transformations

A state-space representation can be interpreted through a system of coordinates. A state, in this context, represents a vector as visualized through this reference. Under the assumption of a linear time-invariant system, there is an intuition that is possible to change the representation of the states by changing this system of coordinates through some linear transformation, obtaining a different model to the same system. This is the motivation for the discussion in this section.

First of all, consider a brief reflection about the geometrical interpretation of a State-Space model for a physical system [give attributions]. Let a state-vector in an arbitrary time t be $\mathbf{x}(t) = [1, 3]^T$, defined in the \mathbb{R}^2 space, as shown in Figure !!!a in a Cartesian coordinate system. It is possible to associate to this vector an orthonormal basis $\mathbf{I} \in \mathbb{R}^2$, the 2-dimensional identity matrix, such that the vector $\mathbf{x}(t)$ observed through this basis is itself:

$$\mathbf{I}\mathbf{x}(t) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \end{bmatrix} = \mathbf{x}(t) \quad (2.54)$$

The representation of a vector with respect to a basis can be interpreted geometrically as the points where lines parallel to the components of the basis, drawn from this vector, intercepts the standard basis of the coordinate system. These parallel lines are the dashed lines shown in Figure !!! In the case of the orthonormal basis, for instance, those lines intercept the axis in the same values that were attributed to the original vector.

This concept provides the possibility to associate any other arbitrary basis to represent a state-vector and visualize the states through this perspective. When the basis is not orthogonal, a change in the state-vector with a direction parallel to a component of the basis, i.e., a change in only one element of this vector, produces a change in other directions if observed through the original orthonormal basis. This is an interesting result, since it raises the represent the interactions between two state variables through the basis reference, making it possible to observe mutual changes in those variables from a single direction. Figure !!!b illustrates the vector $\mathbf{x}(t)$ as referenced through the basis $\mathbf{Q} = [\mathbf{q}_1, \mathbf{q}_2]^T$, for $\mathbf{q}_1 = [3, 1]^T$ and $\mathbf{q}_2 = [2, 2]^T$, given as:

$$\mathbf{x}(t) = \begin{bmatrix} 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 3 & 2 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} -1 \\ 2 \end{bmatrix} = \mathbf{G}\mathbf{z}(t) \quad (2.55)$$

[fig]

In the State-Space formulation, the matrix \mathbf{A} represents a linear function that maps state-vectors from \mathbb{R}^n to itself. When applying a new basis to represent the state-vectors, it is intuitive that the mapping performed by this function also changes to still represent the same linear combination of the states. Hence it is necessary to formulate an equivalent model in those conditions.

Theorem 2.10. (Similarity Transformation) Consider a system in SS representation described by the matrices $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$ and a nonsingular transformation matrix $\mathbf{P} \in \mathbb{R}^{n \times n}$. An equivalent representation for the transformation $\mathbf{z}(t) = \mathbf{P}\mathbf{x}(t)$ is:

$$\begin{cases} \dot{\mathbf{z}}(t) = \tilde{\mathbf{A}}\mathbf{z}(t) + \tilde{\mathbf{B}}\mathbf{u}(t) \\ \mathbf{y}(t) = \tilde{\mathbf{C}}\mathbf{z}(t) + \tilde{\mathbf{D}}\mathbf{u}(t) \end{cases} \quad (2.56)$$

where:

$$\tilde{\mathbf{A}} = \mathbf{P}\mathbf{A}\mathbf{P}^{-1} \quad \tilde{\mathbf{B}} = \mathbf{P}\mathbf{B} \quad \tilde{\mathbf{C}} = \mathbf{C}\mathbf{P}^{-1} \quad \tilde{\mathbf{D}} = \mathbf{D} \quad (2.57)$$

Proof. Consider a SS representation and any nonsingular matrix $\mathbf{P} \in \mathbb{R}^{n \times n}$. Making $\mathbf{z}(t) = \mathbf{P}\mathbf{x}(t)$ leads to $\mathbf{x}(t) = \mathbf{P}^{-1}\mathbf{z}(t)$. Substituting this in the state equation results in:

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) \\ \mathbf{P}^{-1}\dot{\mathbf{z}}(t) &= \mathbf{A}\mathbf{P}^{-1}\mathbf{z}(t) + \mathbf{B}\mathbf{u}(t) \\ \mathbf{I}\dot{\mathbf{z}}(t) &= \mathbf{P}\mathbf{A}\mathbf{P}^{-1}\mathbf{z}(t) + \mathbf{P}\mathbf{B}\mathbf{u}(t) \\ \dot{\mathbf{z}}(t) &= \tilde{\mathbf{A}}\mathbf{z}(t) + \tilde{\mathbf{B}}\mathbf{u}(t)\end{aligned}\tag{2.58}$$

Thus, substituting $\mathbf{x}(t) = \mathbf{P}^{-1}\mathbf{z}(t)$ in the output equation results in:

$$\begin{aligned}\mathbf{y}(t) &= \mathbf{C}\mathbf{x}(t) + \mathbf{D}\mathbf{u}(t) \\ \mathbf{y}(t) &= \mathbf{C}\mathbf{P}^{-1}\mathbf{z}(t) + \mathbf{D}\mathbf{u}(t) \\ \mathbf{y}(t) &= \tilde{\mathbf{C}}\mathbf{z}(t) + \tilde{\mathbf{D}}\mathbf{u}(t)\end{aligned}\tag{2.59}$$

□

By this theorem is it clear that, after transforming the state-vector, the entire dynamical model $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$ changes. This would imply that the analysis of the original system, for its properties and time response, is not valid for the similar transformed system. However, as discussed before, these transformations accounts for the same system when observed through a different reference basis. Therefore, it is expected that the model presents the same analysis results, as demonstrated below.

Theorem 2.11. *Consider a system in State-Space form with matrix \mathbf{A} . Consider also a similarity transformation $\mathbf{z}(t) = \mathbf{P}\mathbf{x}(t)$ that results in a similar matrix $\tilde{\mathbf{A}} = \mathbf{P}\mathbf{A}\mathbf{P}^{-1}$. In this case, \mathbf{A} and $\tilde{\mathbf{A}}$ have the same set of eigenvalues.*

Proof. The eigendecomposition problem is defined as:

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}\tag{2.60}$$

From the similarity transformation, $\tilde{\mathbf{A}} = \mathbf{P}\mathbf{A}\mathbf{P}^{-1}$ leads to $\mathbf{A} = \mathbf{P}^{-1}\tilde{\mathbf{A}}\mathbf{P}$. Substituting this in the eigendecomposition:

$$\begin{aligned}\mathbf{P}^{-1}\tilde{\mathbf{A}}\mathbf{P}\mathbf{v} &= \lambda\mathbf{v} \\ \tilde{\mathbf{A}}\mathbf{P}\mathbf{v} &= \lambda\mathbf{P}\mathbf{v} \\ \tilde{\mathbf{A}}\tilde{\mathbf{v}} &= \lambda\tilde{\mathbf{v}}\end{aligned}\tag{2.61}$$

Concluding that the matrices \mathbf{A} and $\tilde{\mathbf{A}}$ shares the same set of eigenvalues λ . □

It is clear from past results that the fact that both the matrices shares the same set of eigenvalues directly implies that they provide the same dynamical responses and, as will be shown later, the same general properties. Therefore, the similarity transformation consists in a method to produce new State-Spaces representations that emphasizes some geometrical perspective of the model, hopefully some perspective that helps analyze a specific property, without actually changing the relationship of the original model with the physical system.

The use of similarity transformations can also benefits the computation of functions of the matrices of the State-Space representation, given that they impose a desirable structure to this matrix. With this motivation, a popular transformation that provides a new representation with computational advantages is the Similarity transformation.

Theorem 2.12. (Diagonalization) Consider a n -dimensional system in State-Space form represented by the matrices $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$, such that the matrix \mathbf{A} has n distinct real eigenvalues, i.e., $\lambda \in \mathbb{R}^n$. Performing the transformation $\mathbf{z}(t) = \mathbf{V}\mathbf{x}(t)$, where $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n]$ is the modal matrix composed by the eigenvectors $\mathbf{v}_i \in \mathbb{R}^n$, $i \in [1, \dots, n]$, of matrix \mathbf{A} , the resulting transformed matrix $\mathbf{\Lambda} = \mathbf{V}\mathbf{A}\mathbf{V}^{-1}$ is diagonal.

Proof. Since the eigenvalues are real and distinct, the eigenvectors must be linearly independent, proving that the inverse \mathbf{V}^{-1} always exist and that it is a feasible transformation matrix. Using the identity for the eigendecomposition of matrix \mathbf{A} :

$$\begin{aligned}
 \lambda \mathbf{v} &= \mathbf{A} \mathbf{v} \\
 [\lambda_1 \mathbf{v}_1 \quad \lambda_2 \mathbf{v}_2 \quad \dots \quad \lambda_n \mathbf{v}_n] &= [\mathbf{A} \mathbf{v}_1 \quad \mathbf{A} \mathbf{v}_2 \quad \dots \quad \mathbf{A} \mathbf{v}_n] \\
 [\mathbf{v}_1 \quad \mathbf{v}_2 \quad \dots \quad \mathbf{v}_n] \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix} &= \mathbf{A} [\mathbf{v}_1 \quad \mathbf{v}_2 \quad \dots \quad \mathbf{v}_n] \\
 [\mathbf{v}_1 \quad \mathbf{v}_2 \quad \dots \quad \mathbf{v}_n] \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix} &= [\mathbf{v}_1 \quad \mathbf{v}_2 \quad \dots \quad \mathbf{v}_n]^{-1} \mathbf{A} [\mathbf{v}_1 \quad \mathbf{v}_2 \quad \dots \quad \mathbf{v}_n] \\
 \mathbf{\Lambda} &= \mathbf{V}^{-1} \mathbf{A} \mathbf{V}
 \end{aligned} \tag{2.62}$$

Concluding that $\mathbf{\Lambda}$ is a diagonal matrix which elements are the eigenvalues of matrix \mathbf{A} . \square

This result can be easily extended to the case where the eigenvalues are conjugate complex pairs, but each pair is distinct. A model with a diagonal matrix \mathbf{A} has the nice property that the evolution of the states are decoupled, in the sense that each state evolution is a linear function of itself. A geometrical interpretation of this transformation is that the eigenvectors of this matrix produces a basis that encode information about the interaction between those states, while the interaction in the original formulation was the linear combination of the modes.

In this diagonalization procedure, the resulting elements of the matrix $\mathbf{\Lambda}$ are the very own eigenvalues of the original matrix \mathbf{A} , which allows for a direct interpretation of the system response by just visualizing this matrix. Furthermore, it is easy to verify that the state-transition matrix for the transformed matrix, $e^{\mathbf{\Lambda}t}$, is also a diagonal matrix whose elements are the modes of the system, and can be easily computed:

$$\begin{aligned}
 e^{\mathbf{\Lambda}t} &= \sum_{k=0}^{\infty} \frac{\mathbf{A}^k t^k}{k!} = \sum_{k=0}^{\infty} \frac{t^k}{k!} \begin{bmatrix} \lambda_1^k & 0 & \dots & 0 \\ 0 & \lambda_2^k & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n^k \end{bmatrix} = \begin{bmatrix} \frac{t^k \lambda_1^k}{k!} & 0 & \dots & 0 \\ 0 & \frac{t^k \lambda_2^k}{k!} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{t^k \lambda_n^k}{k!} \end{bmatrix} \\
 &= \begin{bmatrix} e^{\lambda_1 t} & 0 & \dots & 0 \\ 0 & e^{\lambda_2 t} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & e^{\lambda_n t} \end{bmatrix}
 \end{aligned} \tag{2.63}$$

In the case that the eigenvalues are not all distinct, it may be not possible to design a modal matrix in the same way, since the eigenvectors could not be all linearly independent, and the matrix would not form a basis. In those cases, however, it is still possible to design a generalized modal matrix that transform the original matrix \mathbf{A} to a quasi-diagonal matrix \mathbf{J} , where there will be decoupled block of states. This similarity transformation is known as the Jordan form [seek reference], and it generalizes the notion of diagonalization for any arbitrary matrix.

2.5 Stability, Controlability and Observability

When moving from a discussion of the models in the perspective of dynamical system analysis to a perspective of control theory, it is necessary to define and analyze some important properties of a system. These properties are characteristic to the system, analyzed through a model, but they account directly to questions relating the control objectives and the instrumentation, as it will be shown later.

The first property to be discussed consists in the stability of a system. An instable system, as the name suggests, is a system whose response does not converge to a specific value and rather oscillates or grows unbounded. In physical scenarios, unstable systems are problematic, since their response to external stimuli can result in dangerous situations to itself and, maybe, to the environment around it. Because of this, determining the stability of a system is a crucial procedure into analyzing a system that will be controlled. Under the several quantitative methods to determine if a system is indeed stable, given a mathematical model, a popular and practical one is the Bounded-Input Bounded-Output (BIBO) stability criteria.

Definition 2.6. (BIBO Stability) A dynamical system is defined as BIBO stable if every bounded input stimuli $|\mathbf{u}(t)| \leq \epsilon < \infty$ produces in it a bounded output response $|\mathbf{y}(t)| \leq \delta < \infty$.

The main result behind this criteria is that the natural response of a system should vanish as time evolves, i.e., $\mathbf{x}(t) \equiv 0$ as $t \rightarrow \infty$. This result is very intuitive from Theorem 2.7, since the vanishing of the natural response implies that $e^{\mathbf{A}t} \equiv 0$ as $t \rightarrow \infty$ and the forced response is expected to be bounded, if $\mathbf{u}(t)$ is bounded. Since the state-transition matrix is a linear combination of the modes, and the modes are exponential functions of the eigenvalues of the matrix \mathbf{A} , it is possible to determine a condition for stability in the light of these quantities.

Theorem 2.13. (BIBO Stability in SS) A system in State-Space form, represented by a matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ with eigenvalues $\boldsymbol{\lambda} \in \mathbb{R}^n$, is BIBO stable if and only if $\operatorname{Re}[\lambda_i] < 0$, $\forall i \in [1, 2, \dots, n]$.

Proof. Left as an exercise □

First of all, note that this criteria depends only on the eigenvalues of matrix \mathbf{A} , so the stability property of a system is invariant to any similarity transformation, since \mathbf{A} and any transformed matrix $\tilde{\mathbf{A}} = \mathbf{P}\mathbf{A}\mathbf{P}^{-1}$ shares the same eigenvalues. From the previous results it is also known that the real part of the eigenvalues, independent of their multiplicity or domain, appears as the arguments of the exponential functions that are the system modes. Therefore, an eigenvalue with a negative real part will produce a mode that is a exponential decay, as this theorem indicates. By the same argument, if the matrix \mathbf{A} has at least one eigenvalue $\lambda_j = 0$ such that $\operatorname{Re}[\lambda_i] \leq 0$, $\forall i \in [1, 2, \dots, n]$, then the mode associated with this eigenvalue is a constant and the natural response becomes bounded. This configuration is known as a

marginally stable condition, in the BIBO perspective. The time response of a 2-nd order system is shown Figure !!! for three different poles configurations, given an unitary step.

[img]

Later chapters will discuss the possibility of stabilize an unstable system through a controller. There are, however, some restrictions to the possibility of controlling or not the states of a system, which includes the necessity of discussing *controllability*. The controllability of a system states whether it is possible to calculate an input signal that drive the system to any arbitrary state, given some time restriction, or not. This property accounts exclusively for this possibility, in the sense that it does not account for the operational feasibility of actually applying this input signal into a physical system, since it may need more energy than an actuator can produce.

Definition 2.7. (Controllability) A system in State-Space form with matrices (\mathbf{A}, \mathbf{B}) is said to be controllable if, for any initial state $\mathbf{x}(t_0) = \mathbf{x}_0$ and terminal state $\mathbf{x}(T) = \mathbf{x}_T$, $T < \infty$, there exists an input signal $\mathbf{u}(t)$, $t \in [t_0, T]$, that can transfer $\mathbf{x}(t_0)$ to $\mathbf{x}(T)$. Otherwise, the system is said to be uncontrollable.

There are several methods to analyze the controllability of a system given a mathematical model and the definition above [seek references]. A popular criteria introduces the concept of a controllability matrix and has a nice geometrical interpretation.

Theorem 2.14. (Controllability in SS) Consider a system in linear State-Space form with matrices (\mathbf{A}, \mathbf{B}) and the controllability matrix $\mathbf{C} \in \mathbb{R}^{n \times nr}$ defined as:

$$\mathbf{C} = [\mathbf{B} \quad \mathbf{AB} \quad \mathbf{A}^2\mathbf{B} \quad \dots \quad \mathbf{A}^{n-1}\mathbf{B}] \quad (2.64)$$

The system is controllable if and only if \mathbf{C} has full row rank.

Proof. Left to the interested reader. □

An intuition behind this theorem is that the full row rank condition implies that \mathbf{C} has n linearly independent columns, therefore these columns can be used as a basis for what is known as the *controllable subspace*. To better understand that, consider the following forced solution $\mathbf{x}_f(t)$ given by the Lagrange formula:

$$\mathbf{x}_f(t) = \int_0^t e^{\mathbf{A}(t-\tau)} \mathbf{B} \mathbf{u}(\tau) d\tau \quad (2.65)$$

From the Cayley-Hamilton theorem, shown in the Sylvester expansion at Theorem 2.8, it is possible to represent $e^{\mathbf{A}(t-\tau)}$ as a linear combination of scalars $\beta_i(t-\tau)$ and powers of the matrix \mathbf{A}^i , $i \in [0, 1, \dots, n-1]$. Thus, using this theorem and substituting $\tau_2 = t - \tau$ for an easier manipulation, the forced response can be represented as:

$$\begin{aligned} \mathbf{x}_f(t) &= \int_0^t \left(\sum_{i=0}^{n-1} \beta_i(\tau_2) \mathbf{A}^i \right) \mathbf{B} \mathbf{u}(t - \tau_2) d\tau_2 = \sum_{i=0}^{n-1} (\mathbf{A}^i \mathbf{B}) \int_0^t \beta_i(\tau_2) \mathbf{u}(t - \tau_2) \\ &= \sum_{i=0}^{n-1} (\mathbf{A}^i \mathbf{B}) \tilde{\beta}_i(\mathbf{u}, t) \end{aligned} \quad (2.66)$$

where $\mathbf{A}^i \mathbf{B}$ are the columns of the matrix \mathbf{C} and $\tilde{\beta}_i(\mathbf{u}, t)$ is a function that depends only on the input signal $\mathbf{u}(t)$ and time t . This result implies that the unforced response $\mathbf{x}_f(t)$ is a linear combination of the columns of \mathbf{C} . If \mathbf{C} has n linearly independent columns, then this linear combination spans the entire n -dimensional space, i.e., the state space, and any desirable state vector \mathbf{x}^* can be reached. If the column rank of \mathbf{C} is less than n , then only a subspace of smaller dimension can be reached through the forced response.

[Example]

While the discussion on controllability concerns the possibility of driving a system to a desirable state through an actuator signal, there is also the necessity to discuss the possibility of determining the internal state of a system given the output signal $\mathbf{y}(t)$. This property, known as *observability*, comes from the fact that any output of a system, related to the states through the matrix \mathbf{C} , may be a combination of states, and that some states may not even be present in the output signal. In conclusion, it is necessary to know if it is possible to reconstruct $\mathbf{x}(t)$ directly through $\mathbf{y}(t)$.

Definition 2.8. Observability A system in State-Space form with matrices (\mathbf{A}, \mathbf{C}) is said to be observable if, given an input signal $\mathbf{u}(t)$ and output signal $\mathbf{y}(t)$, over the interval $t \in [t_0, T]$, it is possible to uniquely determine the value of the initial state $\mathbf{x}(t_0)$. Otherwise, the system is said to be unobservable.

Similarly to the controllability property, there are several ways to analyze the observability of a system, given a mathematical model [seek references]. A popular criteria introduces the concept of a observability matrix.

Theorem 2.15. (Observability in SS) Consider a system in linear State-Space form with matrices (\mathbf{A}, \mathbf{C}) and the observability matrix $\mathbf{O} \in \mathbb{R}^{nq \times n}$ defined as:

$$\mathbf{O} = \begin{bmatrix} \mathbf{C} \\ \mathbf{CA} \\ \mathbf{CA}^2 \\ \vdots \\ \mathbf{CA}^{n-1} \end{bmatrix} \quad (2.67)$$

The system is observable if and only if \mathbf{O} has full column rank.

Proof. Left to the interested reader. □

The interpretation of this theorem follows the same intuition of before: if the matrix \mathbf{O} has full column rank, then it can be used as a basis to span a subspace with the same dimension as the state-space. In fact, the two proofs are similar and there is a direct relationship between controllability and observability, known as the Theorem of Duality [seek reference].

[Example]

The geometrical interpretations of both theorems states a practical solution for the cases where a system is uncontrollable or unobservable. In the first case, the solution would be to add specific actuators to the system, in the condition that they are linearly independent between themselves and the ones actually in operation. The same procedure can be done to solve the unobservable problem, but adding more sensors instead. Those procedures would change the matrices \mathbf{B} and \mathbf{C} , without changing the system dynamics, and could ensure the necessary

conditions in matrices \mathcal{C} and \mathcal{O} . Of course, the implementation of such instrumentation may not be practical, due to technical or economical constraints.

2.6 Frequency Response Analysis

Although the response of dynamical systems are naturally perceived in time, there are advantages of analyzing the models in a frequency domain perspective.

Chapter 3

Controller Design

a

3.1 Control Architectures

a

3.2 Full-State Feedback

a

3.3 Regulation and Reference Tracking

a

3.4 Deterministic State Observers

a

Chapter 4

Optimal Control

a

4.1 General Formulation

a

4.2 Linear Quadratic Regulator (LQR)

a

4.3 Optimal State Estimators

a

4.4 Linear Quadratic Gaussian (LQG)

a

4.5 Robustness and Stability Analysis

Chapter 5

Methodology

a

Chapter 6

Results and Discussion

a

Chapter 7

Conclusion

a

Bibliography

- [Heller et al., 1978] Heller, H. C., Crawshaw, L. I., and Hammel, H. T. (1978). The thermostat of vertebrate animals. *Scientific American*, 239(2):102–115.
- [Moler and Loan, 2003] Moler, C. and Loan, C. V. (2003). Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. *SIAM Review*, 45(1):3–49.

Appendix A - Proof of Theorems

Chapter 2

Proof. (**Theorem 2.2**) Considering that the reactions obeys the Arrhenius equation, the dynamical model for the non-isothermal reactor is a direct application of Theorem 2.1 when assuming that the kinetics are in the form:

$$K_{XY} = K_{XY}(T) = K_{XY}e^{-\frac{E_{XY}}{T}} \quad (7.1)$$

with T being the temperature in Kelvins and E_{XY} being the activation energy needed for a reaction $X \rightarrow Y$ to occur.

In respect to the change in temperatures, the model is obtained by a analysis of the conservation of heat, an energy that obeys this principle:

$$\left(\begin{array}{c} \text{Accumulation} \\ \text{of heat} \end{array} \right) = \left(\begin{array}{c} \text{Heat entering} \\ \text{the System} \end{array} \right) - \left(\begin{array}{c} \text{Heat leaving} \\ \text{the System} \end{array} \right) \quad (7.2)$$

Consider the temperature inside the reactor. Since the system is closed, and involved by the heating/cooling system, this quantity is a result of a change in heat given by the flow of fluid entering and leaving the system, the direct transfer of heat by conduction from the contact with the cooling/heating system and the entropy contribution from the reactions. In this case, the conservation law becomes:

$$\begin{aligned} \left(\begin{array}{c} \text{Accumulation} \\ \text{of heat} \end{array} \right) &= \left(\begin{array}{c} \text{Heat entering} \\ \text{the System} \end{array} \right) - \left(\begin{array}{c} \text{Heat leaving} \\ \text{the System} \end{array} \right) \\ &= \left(\begin{array}{c} \text{Heat transfer} \\ \text{from mass-flow} \end{array} \right) + \left(\begin{array}{c} \text{Heat transfer} \\ \text{from regulator} \end{array} \right) + \left(\begin{array}{c} \text{Entropy} \\ \text{contribution} \\ \text{from reactions} \end{array} \right) \end{aligned} \quad (7.3)$$

From Fourier's law [seek reference], the transfer of heat by convection between the fluids, h_{conv} , and by the conduction between the systems, h_{cond} , obeys the proportionality:

$$h_{conv} \sim T_{in} - T_{out} \quad h_{cond} \sim T_C - T \quad (7.4)$$

The entropy contribution from a reaction $\alpha X \rightarrow \beta Y$, denoted as S_{XY} , is proportional to the concentration of ρ_X consumed multiplied by the energy that it liberates or absorbs:

$$S_{XY} \sim K_{XY}e^{-\frac{E_{XY}}{T}}(\rho_X)^\alpha \Delta H_{XY} \quad (7.5)$$

All the proportionality can be turned into equalities by imposing real constant factors that are calculated without the dynamical variables. In the case of the heat transfer by convection, the change in temperature is directly given by the ratio of volume entering the system. Plugging all together, and summing the entropy contribution from each reaction, the accumulation of heat is modeled as:

$$\frac{d(T)}{dt} = q(T_{in} - T_{out}) + \eta(T_C - T) + \delta \sum_{\alpha A \rightarrow \beta X} K_{AX}e^{-\frac{E_{AX}}{T}}(\rho_A)^\alpha \Delta H_{AX} \quad (7.6)$$

Consider the heating/cooling system involving the reactor system. From the choice of design of this apparatus, it is possible to directly manipulate the temperature of its material using a cooling/heating capacity Q up to a real factor of γ . Similar to the temperature inside the container, there is the conduction of heat between the healing/cooling system and the walls of the container for the reactor. Therefore, the model of heat accumulation for this quantity is given by:

$$\frac{d(T_C)}{dt} = \gamma Q + \beta(T - T_C) \quad (7.7)$$

□

Proof. (**Theorem 2.3**) Applying the Laplace transform in both sides of the equation, and using some properties of this operator, the result is:

$$\begin{aligned} \mathcal{L} \left\{ \alpha_n \frac{d^n y(t)}{dt^n} + \dots + \alpha_1 \frac{dy(t)}{dt} + \alpha_0 y(t) \right\} &= \mathcal{L} \left\{ \beta_m \frac{d^m u(t)}{dt^m} + \dots + \beta_1 \frac{du(t)}{dt} + \beta_0 u(t) \right\} \\ \alpha_0 Y(s) + \sum_{i=1}^n \alpha_i \left(s^i Y(s) - \sum_{j=0}^{i-1} s^j \frac{d^j y(0^-)}{dt^j} \right) &= \beta_0 U(s) + \sum_{k=1}^m \beta_k \left(s^k U(s) - \sum_{l=0}^{k-1} s^l \frac{d^l u(0^-)}{dt^l} \right) \end{aligned} \quad (7.8)$$

By the assumption that all the initial conditions are zero, $y(0^-) = u(0^-) = 0$:

$$\alpha_0 Y(s) + \sum_{i=1}^n \alpha_i s^i Y(s) = \beta_0 U(s) + \sum_{k=1}^m \beta_k s^k U(s) \quad (7.9)$$

Factoring the left side by $\mathbf{Y}(s)$ and the right side by $\mathbf{U}(s)$, and doing some basic algebra, the transfer function $G(s)$, between these variables is:

$$G(s) = \frac{Y(s)}{U(s)} = \frac{\beta_m s^m + \beta_{m-1} s^{m-1} + \dots + \beta_1 s + \beta_0}{\alpha_n s^n + \alpha_{n-1} s^{n-1} + \dots + \alpha_1 s + \alpha_0} \quad (7.10)$$

□