Linear systems theory

What are systems? And how do we respresent them/deal with them? That is what this summary is all about. In the first chapter, we look at what systems are, and how we deal with basic linear systems. In the second chapter, we discuss system properties like stability and controllability. Thirdly, we give some thoughts on how we can use feedback in controlling systems. And finally, in the fourth chapter, we'll examine transfer functions in detail. But first, we start by asking the fundamental question: what is systems theory?

1 Basic systems theory principles

1.1 What is systems theory?

A **system** is a part of reality that can be seen as a separate unit. The reality outside the system is known as the **surroundings**. Of course, the system and the surroundings influence each other. The environment influences the system by **input**, denoted by the **input vector** $\mathbf{u}(t)$. Similarly, the system influences the environment by means of the **output** $\mathbf{y}(t)$.

Mathematical **systems theory** (sometimes also called **system theory**) concerns the study and control of systems. In particular, input/output phenomena are examined.

1.2 Modelling principles

Before we can actually concern ourselves with systems, we need to know how we can model them. This is done, using **modelling principles**. The three most important ones are the following principles.

- **Conservation laws** state that certain quantities (like mass or energy) are conserved.
- **Physical laws** describe important relations between variables. (Think of Newton's laws, or the laws of thermodynamics.)
- **Phenomenological principles** are principles that are known from experience. (Examples are Ohm's law of electrical resistance and Fourier's law of heat conduction.)

By using these principles, often a model of the system can be achieved. This model can then be used to examine and control the system.

2 Making differential systems linear

The most easy type of differential systems is the linear one. What are linear differential systems? And how do we get them?

2.1 What are linear differential systems?

A **linear differential system** is a system that can be written as

$$\dot{\mathbf{x}}(t) = A(t)\mathbf{x}(t) + B(t)\mathbf{u}(t), \tag{2.1}$$

$$\mathbf{y}(t) = C(t)\mathbf{x}(t) + D(t)\mathbf{u}(t). \tag{2.2}$$

These equations are known as the **state equation** and the **output equation**, respectively. $\mathbf{x}(t)$ denotes the $(n \times 1)$ **state vector** of the system, $\mathbf{u}(t)$ is the $(m \times 1)$ **input vector** and $\mathbf{y}(t)$ is the $(p \times 1)$ **output vector**. Also, $A(t)$ is the $(n \times n)$ **state matrix**, $B(t)$ is the $(n \times m)$ **input matrix**, $C(t)$ is the $(r \times n)$ **output matrix** and $D(t)$ is the $(r \times m)$ **direct (transmission) matrix**. If the matrices A, B, C and D do not depend on time, then the system is said to be **time-invariant**.

To solve the output $\mathbf{y}(t)$ of the system, two things need to be known. First of all, the initial state of the system $\mathbf{x}(0)$ (often also denoted as $\mathbf{x_0}$) needs to be known. Second, the input $\mathbf{u}(t)$ must be given. If these two things are given (and have the right properties), then the output $\mathbf{y}(t)$ is well defined.

2.2 Linearization

Linear systems are quite important in systems theory. This is mainly because they are easy to work with. Non-linear systems are more difficult to work with. Luckily, non-linear systems can be approximated by a linear system. This is called **linearization**. But how does it work?

Let's suppose we have a non-linear system, described by

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}, \mathbf{u}) \qquad \text{and} \qquad \mathbf{y}(t) = \mathbf{g}(\mathbf{x}, \mathbf{u}). \tag{2.3}$$

We also suppose that we have a solution $\tilde{\mathbf{x}}(t)$ and $\tilde{\mathbf{y}}(t)$ for given initial conditions $\tilde{\mathbf{x}}_\mathbf{0}$ and input $\tilde{\mathbf{u}}(t)$. Now, let's suppose that we have a problem with a slightly different input function $\tilde{\mathbf{u}}(t) + \mathbf{v}(t)$. ($\mathbf{v}(t)$ is thus the deviation from the original input function.) This will then give a solution $\tilde{\mathbf{x}}(t) + \mathbf{z}(t)$ for the state and $\tilde{\mathbf{y}}(t) + \mathbf{w}(t)$ for the output. (Again, $\mathbf{z}(t)$ and $\mathbf{w}(t)$ are deviations!) It can now be shown (using a Taylor-expansion about the original solution) that we can write the system of equations as

$$\dot{\mathbf{z}}(t) = A(t)\mathbf{z}(t) + B(t)\mathbf{v}(t), \tag{2.4}$$
$$\mathbf{w}(t) = C(t)\mathbf{z}(t) + D(t)\mathbf{v}(t). \tag{2.5}$$

In other words, we have linearized the system! The matrices A, B, C and D can be found, using

$$A(t) = \frac{\partial f}{\partial x}(\tilde{\mathbf{x}}, \tilde{\mathbf{u}}) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial f_n}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix}, \qquad B(t) = \frac{\partial f}{\partial u}(\tilde{\mathbf{x}}, \tilde{\mathbf{u}}) = \begin{bmatrix} \frac{\partial f_1}{\partial u_1} & \cdots & \frac{\partial f_1}{\partial u_m} \\ \vdots & & \vdots \\ \frac{\partial f_n}{\partial u_1} & \cdots & \frac{\partial f_n}{\partial u_m} \end{bmatrix}, \tag{2.6}$$

$$C(t) = \frac{\partial g}{\partial x}(\tilde{\mathbf{x}}, \tilde{\mathbf{u}}) = \begin{bmatrix} \frac{\partial g_1}{\partial x_1} & \cdots & \frac{\partial g_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial g_n}{\partial x_1} & \cdots & \frac{\partial g_n}{\partial x_n} \end{bmatrix}, \qquad D(t) = \frac{\partial f}{\partial u}(\tilde{\mathbf{x}}, \tilde{\mathbf{u}}) = \begin{bmatrix} \frac{\partial g_1}{\partial u_1} & \cdots & \frac{\partial g_1}{\partial u_m} \\ \vdots & & \vdots \\ \frac{\partial g_n}{\partial u_1} & \cdots & \frac{\partial g_n}{\partial u_m} \end{bmatrix}. \tag{2.7}$$

In other words, you first find the derivative matrices (like $\frac{\partial f}{\partial x}$). You then insert the original solution $\tilde{\mathbf{x}}, \tilde{\mathbf{u}}$ into this matrices, and the linearization is complete!

However, it is very important to remember that you are dealing with deviations from the original solution. In other words, your quantities $\mathbf{v}$, $\mathbf{z}$ and $\mathbf{w}$ are not the real physical quantities. And, to make things even worse, books often use the symbols $\mathbf{u}$, $\mathbf{x}$ and $\mathbf{y}$ to indicate both the normal physical quantities, and the deviations from the original solution. Always make sure that you know what each symbol exactly means!

2.3 Solving linear differential systems

Let's suppose we have a linear differential system $\dot{\mathbf{x}}(t) = A(t)\mathbf{x}(t) + B(t)\mathbf{u}(t)$. We now would like to find a solution. To do this, we need to perform four steps.

1. Find the n independent solutions $\xi_\mathbf{1}(t), \xi_\mathbf{2}(t), \ldots, \xi_\mathbf{n}(t)$ of the so-called **autonomous state equation $\dot{\mathbf{x}} = A(t)\mathbf{x}$**.

2. Assemble the **fundamental matrix** $Y(t)$, according to

$$Y(t) = \begin{bmatrix} \xi_{\mathbf{1}}(t) & \xi_{\mathbf{2}}(t) & \ldots & \xi_{\mathbf{n}}(t) \end{bmatrix}. \tag{2.8}$$

3. Assemble the **transition matrix** $\Phi(t, s)$ according to

$$\Phi(t, s) = Y(t) Y^{-1}(s). \tag{2.9}$$

   This matrix has several interesting properties, like

$$\Phi(t, t) = I, \qquad \Phi(t_2, t_0) = \Phi(t_2, t_1)\Phi(t_1, t_0) \qquad \text{and} \qquad \Phi^{-1}(t, s) = \Phi(s, t). \tag{2.10}$$

4. Find the solution $\mathbf{x}(t)$ of the state of the system, using

$$\mathbf{x}(t) = \Phi(t, t_0)\, \mathbf{x_0} + \int_{t_0}^{t} \Phi(t, s)\, B(s)\, u(s)\, ds \tag{2.11}$$

3 Time-invariant systems

A **time-invariant system** is a system described by $\dot{\mathbf{x}}(t) = A\mathbf{x}(t) + B\mathbf{u}(t)$. In other words, the matrices do not depend on time. How do we deal with this kind of systems?

3.1 The solution for time-invariant systems

Time-invariant systems are a lot more easy to solve than time-dependent systems. Normally, it can be very time-consuming to find the transition matrix Φ. However, for time-invariant systems, we simply have

$$\Phi = e^{(t-s)A}. \tag{3.1}$$

By the way, the exponential of a matrix is defined as an infinite series, according to

$$e^{tA} = I + tA + \frac{t^2 A^2}{2!} + \frac{t^3 A^3}{3!} + \ldots = \sum_{i=0}^{\infty} \frac{t^i A^i}{i!}. \tag{3.2}$$

There's just one problem. To find e^{tA}, we have to compute an infinite series. And this usually takes quite a long time. But luckily, there are some tricks. But before we examine those tricks, we need to recap on linear algebra.

3.2 Linear algebra recap

Let's suppose we have an $n \times n$ matrix A. This matrix has k eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$. Each of these eigenvectors λ_i has an **algebraic multiplicity** m_i and a **geometric multiplicity** g_i. The algebraic multiplicity m_i is the number of times which λ_i appears as a root of the **characteristic polynomial** $\det(\lambda I - A)$. The geometric multiplicity g_i is the number of eigenvectors $\mathbf{q_{ij}}$ corresponding to the eigenvalue λ_i. In other words, it is the number of linearly independent solutions $\mathbf{q_{ij}}$ to the equation

$$(\lambda_i I - A)\, \mathbf{q_{ij}} = \mathbf{0}. \tag{3.3}$$

The geometric multiplicity is never bigger than the algebraic multiplicity. So, $g_i \leq m_i$. Also, the sum of all algebraic multiplicities equals the size of the matrix n. So, $\sum_{i=1}^{k} m_i = n$.

We say that the matrix A is **diagonalizable** if there exists an invertible matrix T such that $T^{-1}AT = D$, where D is a diagonal matrix. It can be shown that A is only diagonalizable, if $g_i = m_i$ for all eigenvalues λ_i. If this is indeed the case, then D is the matrix of eigenvalues $D = \operatorname{diag}(\lambda_1, \ldots, \lambda_k)$. In other words, it is the matrix with the k eigenvalues on its diagonal. (By the way, if an eigenvalue λ_i has a multiplicity of m_i, then it also appears m_i times in D. So, D is an $n \times n$ matrix.) Similarly, T is the matrix with as columns the corresponding eigenvectors $\mathbf{q_i}$.

3.3 Solutions for diagonalizable systems

Now it's time to find an alternate expression for e^{tA}. Let's suppose that A is diagonalizable. Then, we have $T^{-1}AT = D$. It can now be shown that

$$e^{tA} = Te^{t\left(T^{-1}AT\right)}T^{-1} = Te^{tD}T^{-1} = T \begin{bmatrix} e^{\lambda_1 t} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & e^{\lambda_k t} \end{bmatrix} T^{-1}. \tag{3.4}$$

Let's denote the rows of T by $\mathbf{w_1}, \ldots, \mathbf{w_n}$. The columns are still denoted by $\mathbf{q_1}, \ldots, \mathbf{q_n}$.We can now also write

$$A = \sum_{i=1}^{n} \lambda_i \mathbf{q_i} \mathbf{w_i}, \qquad \text{and similarly,} \qquad e^{tA} = \sum_{i=1}^{n} e^{t\lambda_i} \mathbf{q_i} \mathbf{w_i}. \tag{3.5}$$

Let's suppose that the system has no input. (Thus, $\mathbf{u}(t) = \mathbf{0}$.) The solution $\mathbf{x}(t)$ of the system is now known as the **free response** of the system. It is given by

$$\mathbf{x}(t) = e^{(t-t_0)A}\mathbf{x_0} = \sum_{i=1}^{n} e^{(t-t_0)\lambda_i} \mathbf{q_i} \mathbf{w_i} \mathbf{x_0} = \sum_{i=1}^{n} \mu_i e^{(t-t_0)\lambda_i} \mathbf{q_i}, \tag{3.6}$$

where $\mu_i = \mathbf{w_i}\mathbf{x_0}$. (Remember that $\mathbf{w_i}$ is a row vector.) The above equation implies an interesting fact. The free response of a system $\mathbf{x}(t)$ can be decomposed along the eigenvectors $\mathbf{q_i}$. The solution corresponding to one eigenvector $\mathbf{q_i}$ is called a **mode** of the system. A system will be in mode i, if the initial vector $\mathbf{x_0}$ is aligned with the eigenvector $\mathbf{q_i}$.

3.4 The Jordan form

We now know what to do with systems if A is diagonalizable. But what if A is not diagonalizable? In this case, we don't have $T^{-1}AT = D$. Instead, we will use $T^{-1}AT = J$. In this equation, J is the so-called **Jordan form** of A. It has a block-diagonal structure $J = \operatorname{diag}(J_1, J_2, \ldots, J_k)$. Every submatrix J_i is an $m_i \times m_i$ matrix and corresponds to the eigenvalue λ_i. In fact, it has the form

$$J_i = \begin{bmatrix} \lambda_i & \gamma_{i1} & 0 & \cdots & 0 \\ 0 & \lambda_i & \gamma_{i2} & \ddots & \vdots \\ 0 & 0 & \lambda_i & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \gamma_{i(m_i-1)} \\ 0 & \cdots & 0 & 0 & \lambda_i \end{bmatrix}. \tag{3.7}$$

In this equation, some of the values γ_{ij} are 0, while others are 1. This, in fact, depends on how T is build up. We can remember that $T^{-1}AT = J$ and thus $AT = TJ$. If the individual columns of T are written as $\mathbf{q_1}, \ldots, \mathbf{q_n}$, then we have

$$A\mathbf{q_i} = \lambda \mathbf{q_i} + \gamma \mathbf{q_{i-1}}, \tag{3.8}$$

with λ the corresponding eigenvalue in J. Note that, if $\gamma = 0$, then $\mathbf{q_i}$ is an eigenvector of A, corresponding to the eigenvalue λ. If, instead, $\gamma = 1$, then we have

$$\mathbf{q_{i-1}} = (A - \lambda I)\,\mathbf{q_i}. \tag{3.9}$$

The vector $\mathbf{q_i}$ is known as a **generalized eigenvector**. It can be derived from the (possibly generalized) eigenvector $\mathbf{q_{i-1}}$.

We can now see a relation between the form of J and T. The zeroes in J correspond with the positions of the normal eigenvectors in T. Similarly, the ones in J correspond to the positions of the generalized eigenvectors in T.

4 System response

We now know how the state of a system behaves. But what about the output? That's what we'll look at now.

4.1 The impulse response

The output of the system is given by

$$\mathbf{y}(t) = C(t)\mathbf{x}(t) + D(t)\mathbf{u}(t) = C(t)\Phi(t,t_0)\mathbf{x_0} + \int_{t_0}^{t} C(t)\Phi(t,s)B(s)\mathbf{u}(s)\,ds + D(t)\mathbf{u}(t). \tag{4.1}$$

To simplify matters, we define $K(t,s)$ such that

$$K(t,s) = C(t)\Phi(t,s)B(s). \tag{4.2}$$

Let's assume that there is some time t_0 for which $\mathbf{x_0} = \mathbf{0}$. This eliminates the left term in the output equation. What we remain with is some sort of **mapping function**: we map the output onto the input, without having to know anything about the state. This thus gives us an **external description** of the system.

Now, let's assume that $D(t) = 0$. Also, the input function is given by $\mathbf{u(t)} = \delta(t - t_1)\mathbf{e_i}$. In this relation, δ denotes the **unit impulse function** and $\mathbf{e_i}$ denotes the i'th basis vector. The output $\mathbf{y}(t)$ is now given by

$$y(t) = \int_{t_0}^{t} K(t,s)\delta(s-t_1)\mathbf{e_i}\,ds = K(t,t_1)\mathbf{e_i} = \text{the } i\text{'th column of } K(t,t_1). \tag{4.3}$$

The matrix $K(t,t_1)$ can thus be seen as the response to an impulse function. For this reason, $K(t,t_1)$ is known as the **impulse response matrix**.

4.2 The step response

After examining the impulse function, we will now look at the step function. This time, we assume that the input is given by $\mathbf{u}(t) = H(t-t_1)\mathbf{e_i}$, where $H(t)$ is known as the **unit step function**. This results in an output

$$\mathbf{y}(t) = \int_{t_0}^{t} K(t,s)H(s-t_1)\mathbf{e_i}\,ds = \int_{t_1}^{t} K(t,s)\mathbf{e_i}\,ds = S(t,t_1)\mathbf{e_i}. \tag{4.4}$$

The matrix $S(t,t_1)$ is known as the **step response matrix**. It is related to $K(t,s)$ according to

$$S(t,t_1) = \int_{t_1}^{t} K(t,s)\,ds \qquad \text{and} \qquad \frac{d}{ds}S(t,s) = \frac{d}{ds}\int_{s}^{t} K(t,\tau)\,d\tau = -K(t,s). \tag{4.5}$$

You might wonder why the minus sign on the right side is present. In short, this is because s is the lower limit of the integration.