Chapter 2
Preliminaries

This chapter introduces some preliminary theoretical concepts that will be useful later in the book. The following notation is used: $\lambda_{\text{min}}\{A\}$ and $\lambda_{\text{max}}\{A\}$, to indicate the smallest and largest eigenvalues, respectively, of a symmetric positive definite matrix $A \in \mathbb{R}^{n \times n}$. The determinant of matrix $A$ is indicated as $|A|$, whereas the notation $|x|$ stands for the absolute value of a real-valued variable $x \in \mathbb{R}$. The norm of a vector $x \in \mathbb{R}^n$ is defined as $\|x\| = \sqrt{x^T x}$, and the induced norm of matrix $A$ is defined as $\|A\| = \sqrt{\lambda_{\text{max}}\{A^T A\}}$. $0_{n \times m}$ is an $n$ by $m$ null rectangular matrix, and $0_n$ is an $n$ by $n$ square null matrix. Finally, $I_n$ is the $n$ by $n$ identity matrix.

2.1 Fundamentals of Nonlinear Systems

A nonlinear dynamic system can be represented by a finite number $n$ of nonlinear differential equations in the so-called state-space form

$$
\dot{x}_1 = f_1(x_1, x_2, \ldots, x_n), \\
\dot{x}_2 = f_2(x_1, x_2, \ldots, x_n), \\
\vdots \\
\dot{x}_n = f_n(x_1, x_2, \ldots, x_n),
$$

where $\dot{x}_i$ represents the derivative of $x_i$ with respect to the time. Variables $x_1, x_2, \ldots, x_n$ stand for the states of the system, which are the variables that give a complete description of the system at any particular time, and $f_1, f_2, \ldots, f_n$ are nonlinear functions describing the overall system dynamics. Under this notation, the state-space system can also be written in the vector form

$$
\dot{x} = f(x),
$$

(2.1)
where $f$ is a $n \times 1$ nonlinear vector function and $x$ is the $n \times 1$ state vector. A particular value of the state vector is called a point, because it corresponds to a point in the state-space or a particular configuration of the dynamic system. The number of states $n$ is called the order of the system. A solution $x(t)$ of the Eq. (2.1), which usually corresponds to a curve in the state-space as time $t$ varies from zero to infinity, is generally known as a state trajectory or a system trajectory [44]. The most important thing in control systems engineering is the behavior of the system trajectories $x(t)$ around some operating point or desired nominal motion, which introduces the notion of stability, a concept that will be seen in detail in the following section.

If a system trajectory corresponds to a single point along its complete time evolution, then such a point is an equilibrium. In other words, a particular value $x^*$ of the state vector is an equilibrium point of the system (2.1) if, for the initial condition $x(0) = x^*$, the system trajectory $x(t)$ remains equal to $x^*$ for all $t$ greater than zero. This also means that the constant vector $x^*$, which is an equilibrium point, satisfies $f(x^*) = 0$. The concept of equilibrium is very important for the study of stability, since many stability problems are naturally formulated with respect to equilibrium points [44].

The nonlinear system (2.1) is said to be autonomous, since $f$ does not depend explicitly on time $t$. On the other hand, the system is called non-autonomous if the nonlinear function vector $f$ explicitly depends on time $t$, such that

$$\dot{x} = f(t, x).$$  \hfill (2.2)

An analogous property for linear systems depends on whether the matrix of coefficients $A$ varies with time or not, being classified as either time-varying or time-invariant, respectively.

Furthermore, it is important to note that the state Eqs. (2.1) and (2.2) are described without explicit presence of an input variable $u$, that is, the so-called unforced state equation. There are two reasons to work with an unforced state equation: (i) The input to the system is assumed to be zero. (ii) The state equation represents the closed-loop dynamics of a feedback system composed of the controller and the plant, with the control input $u$ being a function of the state $x$, a function of time $t$, or both, and therefore disappearing in the closed-loop dynamics [44, 134].

A more general representation of an open-loop nonlinear system, that is, the plant of the system under consideration, is the $n$-dimensional first-order vector differential equation

$$\dot{x} = f(t, x, u),$$  \hfill (2.3)

with the vector of nonlinear functions $f \in \mathbb{R}^n$ depending on the time variable $t \in \mathbb{R}^+$, the state vector $x \in \mathbb{R}^n$, and the input vector $u \in \mathbb{R}^p$ to be defined as a function of the state $x$ and time $t$. 
2.2 Fundamental Properties

A generalized vector function \( f(t, x) \) is said to be Lipschitz if it satisfies the \textit{Lipschitz condition}, which is given by the inequality

\[
\| f(t, x) - f(t, y) \| \leq L \| x - y \|, \tag{2.4}
\]

for all \((t, x)\) and \((t, y)\) in some neighborhood of \((t_0, x_0)\), with \(x(t_0) = x_0\). For \( f(t, x) \), being an autonomous and scalar function, the Lipschitz condition can be written as

\[
| f(x) - f(y) | \leq L |x - y|. \tag{2.5}
\]

Then, it is said that \( f \) is a \textit{Lipschitz function}, which implies that on a plot of \( f(x) \) versus \( x \), a straight line joining any two points of \( f(x) \) cannot have a slope whose absolute value is greater than the so-called \textit{Lipschitz constant} \( L \) [134].

The existence and uniqueness of the solution \( x(t) \) for a differential equation system of the form (2.1) and (2.2), for a given initial state \( x(t_0) = x_0 \), with \( t > t_0 \), can be ensured if the right-hand side of (2.1) and (2.2) satisfy the Lipschitz condition in (2.4) [134].

Let \( f \) be a continuously differentiable function at each point on the open interval \((x, y)\). The \textit{Mean Value Theorem} states that there exists a point \( z \) in \((x, y)\) such that [134].

\[
f(y) - f(x) = \frac{\partial f}{\partial x} \bigg|_{x=z} (y - x). \tag{2.6}
\]

Finally, a square matrix \( A \) is called a \textit{Hurwitz matrix} when all its eigenvalues \( \lambda_i \) have a strictly negative real part.

2.3 Concepts of Stability

As mentioned before, the intuitive notion of stability of dynamical systems is a kind of well-behavedness of the system trajectories around a desired operating point (typically an equilibrium) and whether the system trajectories will remain close to the equilibrium if slightly perturbed. But in practical problems, the stability of motion is a much more common concern, that is, the behavior of the system trajectories around a desired motion trajectory. However, the motion stability problem can be transformed into an equivalent stability problem around an equilibrium point. Also, for convenience, it is assumed that the equilibrium point under analysis is at the origin, that is, \( x = 0 \), since any equilibrium point can be shifted to the origin via a change of variables. Therefore, all the definitions of stability will be given as a stability problem around the origin of the state-space [44, 134].
Definition 2.1 The equilibrium point $x = 0$ of the system (2.1) is stable if, for each $\varepsilon > 0$, there is $\delta = \delta(\varepsilon) > 0$ such that
\[
\|x(0)\| < \delta \Rightarrow \|x(t)\| < \varepsilon, \quad \forall t \geq 0.
\] (2.7)
Otherwise, it is unstable.

This definition of stability (also called stability in the sense of Lyapunov) requires the existence of $\delta > 0$ for each $\varepsilon > 0$ and not for some $\varepsilon > 0$. That is, for any value of $\varepsilon$, there must be a value $\delta$, possibly dependent on $\varepsilon$, such that a trajectory starting in a $\delta$ neighborhood of the origin will never leave the $\varepsilon$ neighborhood [134].

Definition 2.2 The equilibrium point $x = 0$ of the system (2.1) is asymptotically stable if it is stable and $\delta$ can be chosen such that
\[
\|x(0)\| < \delta \Rightarrow \lim_{t \to \infty} x(t) = 0.
\] (2.8)
Asymptotic stability means that the equilibrium is stable, and that, in addition, the system trajectories started close to the origin actually converge to it as time $t$ approaches infinity. The set of all points such that trajectories initiated at these points eventually converge to the origin is called the domain of attraction. An equilibrium point which is Lyapunov stable but not asymptotically stable is called marginally stable [44].

Definition 2.3 The equilibrium point $x = 0$ of the system (2.1) is exponentially stable if there exist $\alpha > 0$ and $\lambda > 0$ such that
\[
\|x(t)\| \leq \alpha \|x(0)\| e^{-\lambda t}, \quad \forall t \geq 0.
\] (2.9)
This means that the system trajectories started in the domain of attraction of an exponentially stable equilibrium converge to it faster than an exponential function. In fact, the positive number $\lambda$ is called the convergence rate. Note that exponential stability implies asymptotic stability, but the converse is false [44].

Definition 2.4 If the equilibrium $x = 0$ of the system (2.1) satisfies the definition of stability in Definition 2.1, asymptotic stability in Definition 2.2, or exponential stability in Definition 2.3 for any initial condition $x(t)$, then the equilibrium point is said to be globally (asymptotically or exponentially) stable.

Otherwise, the behavior of the system is characterized only locally. That is, we only know how the system trajectories evolve after starting near the equilibrium point, but we have no information about how the system will behave when the initial conditions are some distance away from the equilibrium.

Lyapunov’s direct method can be used to study the stability of systems in a local or global form. The local analysis deals with stability in a neighborhood of the equilibrium point, while the global analysis describes the stability properties in the
whole state-space. This analysis usually involves a locally or globally \textit{positive definite} function, \textit{i.e.}, a strictly positive function, which is null only at the origin.

\textbf{Theorem 2.1} \textbf{(Stability)} If, in a ball $B_r$ which contains the origin, there exists a scalar function $V(x)$ with continuous first partial derivatives along the trajectories of (2.1) such that

- $V(x)$ is positive definite locally in $B_r$,
- $\dot{V}(x)$ is negative semidefinite locally in $B_r$,

then the origin of (2.1) is stable. If the ball $B_r$ can be expanded to be the whole state-space, then the origin is globally stable [44].

\textbf{Theorem 2.2} \textbf{(Asymptotic stability)} If, in a ball $B_r$ which contains the origin, there exists a scalar function $V(x)$ with continuous first partial derivatives along the trajectories of (2.1) such that

- $V(x)$ is positive definite locally in $B_r$,
- $\dot{V}(x)$ is negative definite locally in $B_r$,

then the origin of (2.1) is asymptotically stable. If the ball $B_r$ can be expanded to be the whole state-space and

- $V(x) \to \infty$ as $\|x\| \to \infty$,

then the equilibrium at the origin is globally asymptotically stable [44].

For global asymptotic stability, an additional condition on the function $V$ has to be satisfied: $V(x)$ must be \textit{radially unbounded}, which means that $V(x)$ tends to infinity as $x$ approaches infinity in any direction. In order to prove exponential stability, a minimum dissipation rate must be guaranteed. The following theorem for exponential stability is adapted for the autonomous case from Theorem 4.10 in [134].

\textbf{Theorem 2.3} \textbf{(Exponential stability)} If, in a ball $B_r$ which contains the origin, there exists a scalar function $V(x)$ with continuous first partial derivatives along the trajectories of (2.1) such that

- $k_1 \|x\|^2 \leq V(x) \leq k_2 \|x\|^2$ locally in $B_r$,
- $\frac{\partial V}{\partial x} f(x) \leq -k_3 \|x\|^2$ locally in $B_r$,

with positive constants $k_1$, $k_2$, and $k_3$, then the origin of (2.1) is exponentially stable. If the ball $B_r$ can be expanded to be the whole state-space, then the equilibrium at the origin is globally exponentially stable.

If such a function $V$, which satisfies the conditions in Theorems 2.1, 2.2 or 2.3, exists, then it is called a \textit{Lyapunov function} for the system (2.1), proving that its origin is stable, asymptotically stable or exponentially stable, respectively. Deeper and extended studies on Lyapunov’s direct method for general systems can be found in [40, 44, 134].
2.4 Barbalat’s Lemma

For time-varying systems, it is usually very difficult to find Lyapunov functions with a negative definite derivative. Therefore, asserting asymptotic stability for such systems is not an easy task. Barbalat’s lemma is a very useful tool which helps to ensure asymptotic stability for non-autonomous systems with a negative semidefinite derivative of the Lyapunov function.

Lemma 2.1 (Barbalat) If the differentiable function $f(t)$ has a finite limit as $t \to \infty$, and if $\dot{f}$ is uniformly continuous, then $\dot{f}(t) \to 0$ as $t \to \infty$ [44].

For the application of Barbalat’s lemma to the analysis of dynamical systems, the following “Lyapunov-like” lemma can be used:

Lemma 2.2 If a scalar function $V(t, x)$ satisfies the following conditions:

- $V(t, x)$ is lower bounded
- $\dot{V}(t, x)$ is negative semidefinite
- $\dot{V}(t, x)$ is uniformly continuous in time

then $\dot{V}(t, x) \to 0$ as $t \to \infty$ [44].

A sufficient condition for a differentiable function to be uniformly continuous is that its derivative be bounded [44].

2.5 Boundedness and Ultimate Boundedness

Let us recall the non-autonomous system introduced in (2.2), for which the origin is not an equilibrium point. Without equilibrium points, it is not possible to talk about stability, since any notion of stability is a property of the equilibrium. However, it is possible to show the boundedness of the system trajectories around the origin, even if it is not an equilibrium.

Definition 2.5 The solutions of (2.2) are

- uniformly bounded if there exists a positive constant $c$, independent of $t_0 > 0$, and for every $a \in (0, c)$, there is $\beta = \beta(a) > 0$, independent of $t_0$, such that

$$\|x(t_0)\| \leq a \Rightarrow \|x(t)\| \leq \beta, \ \forall \ t \geq t_0.$$  \hspace{1cm} (2.10)

- globally uniformly bounded if (2.10) holds for arbitrarily large $a$.

- uniformly ultimately bounded (UUB) with ultimate bound $b$ if there exist positive constants $b$ and $c$, independent of $t_0 \geq 0$, and for every $a \in (0, c)$, there is $T = T(a, b) \geq 0$, independent of $t_0$, such that

$$\|x(t_0)\| \leq a \Rightarrow \|x(t)\| \leq b, \ \forall \ t \geq t_0 + T.$$  \hspace{1cm} (2.11)
2.5 Boundedness and Ultimate Boundedness

- globally uniformly ultimately bounded if (2.11) holds for arbitrarily large $a$.

In the case of autonomous systems, the word “uniformly” may be removed.

Considering a continuously differentiable, positive definite function $V(t, x)$, the following Lyapunov-like theorem can be used for showing uniform boundedness and ultimate boundedness [134].

**Theorem 2.4** Let $D \subset \mathbb{R}^n$ be a domain that contains the origin and $V : [0, \infty) \times D \rightarrow \mathbb{R}$ be a continuously differentiable function such that

$$\alpha_1(\|x\|) \leq V(t, x) \leq \alpha_2(\|x\|),$$

$$\frac{\partial V}{\partial t} + \frac{\partial V}{\partial x} f(t, x) \leq -W(x), \quad \forall \|x\| \geq \mu > 0,$$

$$\forall t \geq 0 \text{ and } \forall x \in D,$$ where $\alpha_1$ and $\alpha_2$ are class $\mathcal{K}$ functions and $W(x)$ is a continuous positive definite function. Take $r > 0$ such that $B_r \subset D$ and suppose that

$$\mu < \alpha_2^{-1}(\alpha_1(r)).$$

Then, there exists a class $\mathcal{KL}$ function $\beta$, and for every initial state $x(t_0)$, satisfying $\|x(t_0)\| \leq \alpha_2^{-1}(\alpha_1(r))$, there is $T \geq 0$ (dependent on $x(t_0)$ and $\mu$) such that the solution of (2.2) satisfies

$$\|x(t)\| \leq \beta(\|x(t_0), t - t_0\|), \quad \forall t_0 \leq t \leq t_0 + T,$$

$$\|x(t)\| \leq \alpha_2^{-1}(\alpha_1(\mu)), \quad \forall t \geq t_0 + T.$$ 

Moreover, if $D = \mathbb{R}^n$ and $\alpha_1$ belongs to class $\mathcal{K}_\infty$, then (2.15) and (2.16) hold for any initial state $x(t_0)$, with no restriction on how large $\mu$ is.

The definitions of class $\mathcal{K}$, $\mathcal{K}_\infty$ and $\mathcal{KL}$ functions can be consulted in [134]. The main differences of Theorems 2.1–2.3 with respect to Theorem 2.4 are that in the latter, the system (2.2) has no equilibrium points and $\dot{V}$ is negative outside a set as given in (2.13).

2.6 Feedback Linearization

Let us introduce a single-input–single-output (SISO) nonlinear system of the form

$$\dot{x} = f(x) + g(x)u,$$ (2.17)
and a scalar output function

\[ y = h(x). \tag{2.18} \]

Feedback linearization is based on the idea of the existence of a state feedback control law \( u \) that algebraically transforms the nonlinear system dynamics into an (fully or partly) equivalent linear system. In a simplified way, feedback linearization means canceling the nonlinearities in a system so that the closed-loop dynamics turn out to be linear. All the motion control algorithms presented in this book have their basis in the feedback linearization of single-input–single-output nonlinear systems.

Consider the single-input–single-output system (2.17)–(2.18), where \( f, g \) and \( h \) are sufficiently smooth (this means that all its partial derivatives are defined and continuous) in a domain \( D \subset \mathbb{R}^n \). The first time derivative of the output function \( y \) in (2.18) along the trajectories of the state equation (2.17) is given by

\[
\dot{y} = \frac{\partial h}{\partial x} [f(x) + g(x)u] \overset{\Delta}{=} L_f h(x) + L_g h(x)u,
\]

where

\[
L_f h(x) = \frac{\partial h}{\partial x} f(x)
\]

is called the Lie derivative of \( h \) with respect to \( f \). This means the derivative of \( h \) along the trajectories of the system \( \dot{x} = f(x) \). The notation \( L_f^i \) is used for repeating \( i \) times the calculation of the derivative along the same vector field. This notation is very convenient for defining the general expression of the feedback linearization controller \( u \) such that

\[
u = \frac{1}{L_g L_f^{\rho-1} h(x)} [-L_f^\rho h(x) + v] \tag{2.19}
\]

reduces the input–output map to

\[
y^{(\rho)} = v, \tag{2.20}
\]

with the integer \( \rho \) being the relative degree of the system, that is, the output function \( h(x) \) must be derived \( \rho \) times until it obtains an expression relating the input \( u \) with the output \( y \). In other words, the nonlinear system (2.17)–(2.18) is said to have relative degree \( \rho, 1 \leq \rho \leq n \), in a region \( D_0 \subset D \) if

\[ L_g L_f^{i-1} h(x) = 0, \quad i = 1, 2, \ldots, \rho - 1; \quad L_g L_f^{\rho-1} h(x) \neq 0,
\]

for all \( x \in D_0 \) [134].

In particular, when the state equation is completely linearized through the input transformation \( u \), that is, the relative degree \( \rho \) is equal to the order of the system \( n \), it is known as full state or input-state linearization. On the other hand, when only the input–output map is linearized while the state equation is only partly linearized (that is, the relative degree \( 1 \leq \rho < n \)), it is known as input–output linearization.
Finally, the open-loop output dynamics in (2.20) are a simple linear chain of $\rho$ integrators, which describe the relationship between the output and the new input $v$ to be defined. This new control input may be chosen as

$$v = -k_0 y - k_1 \dot{y} - \ldots - k_{\rho-1} y^{(\rho-1)},$$

leading to the closed-loop output dynamics

$$y^{(\rho)} + k_{\rho-1} y^{(\rho-1)} + \ldots + k_1 \dot{y} + k_0 y = 0,$$

with the positive constants $k_i$ chosen so that the polynomial $p^\rho + k_{\rho-1} p^{\rho-1} + \ldots + k_0$ has all its roots strictly in the left-half complex plane. This ensures that $y(t) \to 0$ with an exponential rate of convergence. However, the selection of $v$ is not unique.

The simpler form of output dynamics in (2.20) can also be used in the development of robust, adaptive or neural networks-based nonlinear controllers, as will be seen in Chaps. 6 and 9.

The concepts used for SISO systems can be extended for multi-input–multi-output (MIMO) systems. In the MIMO case, the following square systems (systems with the same number of inputs and outputs) are considered:

$$\dot{x} = f(x) + G(x)u,$$  
$$y = h(x),$$

where $x$ in the state vector, $u \in \mathbb{R}^m$ is the vector of control inputs, $f(x) \in \mathbb{R}^n$ is a smooth vector field, $G(x) \in \mathbb{R}^{n \times m}$ is the matrix $[g_1(x), g_2(x), \ldots, g_m(x)]$ with $g_i(x) \in \mathbb{R}^n$ being a smooth vector field, $y \in \mathbb{R}^m$ is the vector of output functions, and $h(x) \in \mathbb{R}^m$ is a vector of scalar functions $h(x)$.

Similar to the SISO case, in order to obtain the input–output linearization of the MIMO system, it is necessary to differentiate the output $y$ until the inputs $u_i$ appear. Assuming that the relative degree $r_i$ is the smallest integer such that at least one of the inputs $u_i$ appears, with $\{i = 1, 2, \ldots, m\}$, then the open-loop output dynamics for the MIMO case can be written as

$$
\begin{bmatrix}
y_1^{(r_1)} \\
\vdots \\
y_m^{(r_m)}
\end{bmatrix} = 
\begin{bmatrix}
L_{r_1}^1 h_1(x) \\
\vdots \\
L_{r_m}^m h_m(x)
\end{bmatrix} + E(x)
\begin{bmatrix}
u_1 \\
\vdots \\
u_m
\end{bmatrix},
\tag{2.23}
$$

where the matrix $E(x) \in \mathbb{R}^{m \times m}$, which is defined as

$$E(x) = 
\begin{bmatrix}
L_{g_1} L_{f}^{r_1-1} h_1 & \ldots & L_{g_m} L_{f}^{r_1-1} h_1 \\
\vdots & \ddots & \vdots \\
L_{g_1} L_{f}^{r_m-1} h_m & \ldots & L_{g_m} L_{f}^{r_m-1} h_m
\end{bmatrix},
\tag{2.24}
$$
is called the decoupling matrix for the MIMO system. If the decoupling matrix is non-singular in a region around the origin, then the input transformation

\[ u = -E^{-1} \begin{bmatrix} L_{r_1} h_1(x) \\ \vdots \\ L_{r_m} h_m(x) \end{bmatrix} + E^{-1} \begin{bmatrix} v_1 \\ \vdots \\ v_m \end{bmatrix} \]  

(2.25)

yields a linear differential relation between the output \( y \) and the new input \( v \),

\[ \begin{bmatrix} y_1^{(r_1)} \\ \vdots \\ y_m^{(r_m)} \end{bmatrix} = \begin{bmatrix} v_1 \\ \vdots \\ v_m \end{bmatrix}. \]  

(2.26)

Since the above input–output relation is decoupled, the decoupling control law given in (2.25) can be used to design tracking or stabilization controllers by using SISO design on each \( y_i - v_i \) relation [44].

Finally, the concept of relative degree for the MIMO system can be formalized as follows:

**Definition 2.6** The system (2.21)–(2.22) is said to have relative degree \((r_1, \ldots, r_m)\) at \( x_0 \) if there exists a neighborhood \( B \) of \( x_0 \) such that \( \forall x \in B \),

- \( L_{g_i} L^k h_j(x) = 0, \quad 0 \leq k \leq r_i - 1, \quad 1 \leq i, j \leq m \),
- \( E(x) \) in non-singular.

The total relative degree of the system is defined by \( r = r_1 + \cdots + r_m \) [44].

### 2.7 Artificial Neural Networks

An artificial neural network is a mathematical representation of a biological neural network. This mathematical model describes the biological process of information processing through the nervous system and its basic processing unit, the neuron. The output \( y(t) \) of each neuron is equal to the resulting value of the nonlinear activation function \( \sigma(\cdot) \), whose argument is equal to the sum of the weighted input signals \( x_1(t), x_2(t), \ldots, x_n(t) \) multiplied by each weighting coefficient \( v_1(t), v_2(t), \ldots, v_n(t) \), respectively, plus the firing threshold or bias \( v_0 \) of the neuron. Therefore, the mathematical model of a neuron can be expressed as

\[ y(t) = \sigma(v^T x), \]  

(2.27)

where \( x = [1 \ x_1 \ x_2 \ \ldots \ x_n]^T \in \mathbb{R}^{n+1} \) is the augmented input vector and \( v(t) = [v_0 \ v_1 \ v_2 \ \ldots \ v_n]^T \) is the augmented weight vector. This basic processing unit is also called a perceptron [60] and its graphical description is given in Fig. 2.1.
A neural network is composed of an array of several neurons distributed in one or more layers, whose inputs and outputs are all interconnected. The first layer, where the input signals $x_i(t)$ are applied, is known as the input layer; the final layer, where the output signals $y_i(t)$ are acquired, is known as the output layer; and all the layers in between are known as hidden layers.

In particular, we are especially interested in the two-layer perceptron consisting of a single hidden layer with $L$ neurons, all fed by the same input signals $x(t) \in \mathbb{R}^{n+1}$, an output layer with $m$ neurons producing $y(t) \in \mathbb{R}^m$ outputs, the vector of the hidden-layer activation functions $\sigma(\cdot) = [\sigma_1(\cdot) \sigma_2(\cdot) \ldots \sigma_L(\cdot)]^T \in \mathbb{R}^{L+1}$, the vector of the output-layer activation functions $\bar{\sigma}(\cdot) = [\bar{\sigma}_1(\cdot) \bar{\sigma}_2(\cdot) \ldots \bar{\sigma}_m(\cdot)]^T \in \mathbb{R}^m$, an input weights matrix $V \in \mathbb{R}^{[n+1] \times L}$ given by

$$V = \begin{bmatrix} v_{01} & v_{02} & \ldots & v_{0L} \\ v_{11} & v_{12} & \ldots & v_{1L} \\ \vdots & \vdots & \ddots & \vdots \\ v_{n1} & v_{n2} & \ldots & v_{nL} \end{bmatrix}, \quad (2.28)$$

and an output weights matrix $W \in \mathbb{R}^{[L+1] \times m}$ given by

$$W = \begin{bmatrix} w_{01} & w_{02} & \ldots & w_{0m} \\ w_{11} & w_{12} & \ldots & w_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ w_{L1} & w_{L2} & \ldots & w_{Lm} \end{bmatrix}. \quad (2.29)$$

The two-layer neural network is graphically described in Fig. 2.2 and its mathematical model is given as

$$y(t) = \bar{\sigma}(W^T \sigma(V^T x)). \quad (2.30)$$
Fig. 2.2 Graphical description of a two-layer perceptron

Considering the output-layer activation functions \( \hat{\sigma}(\cdot) \) as a vector of linear functions, the general mathematical model of the neural network given in (2.30) can be simplified as

\[
y(t) = W^T \sigma(V^T x).
\]

This expression will be useful later to introduce the function approximation property of neural networks.

### 2.7.1 Universal Function Approximation Property

The universal function approximation property of neural networks has been extensively studied [60], and it has proved to be of fundamental importance in control applications. The basic result of the universal approximation of neural networks having at least two layers, since one-layer neural networks do not generally have a universal approximation capability, says that any smooth function \( f(x) \) can be approximated arbitrarily closely on a compact set using a neural network with appropriate weights.

**Property 2.1** Let \( f(x) : \mathbb{R}^n \to \mathbb{R}^m \) be a smooth function. Then, given a compact set \( S \subseteq \mathbb{R}^n \) and a positive number \( \varepsilon_N \), there exists a two-layer neural network with an
2.7 Artificial Neural Networks

input vector $x \in \mathbb{R}^{n+1}$, an input weights matrix $V \in \mathbb{R}^{[n+1] \times L}$, a vector of activation functions $\sigma \in \mathbb{R}^L$, and an output weights matrix $W \in \mathbb{R}^{[L+1] \times m}$, such that

$$f(x) = W^T \sigma(V^T x) + \varepsilon,$$

with an approximation error $\|\varepsilon\| < \varepsilon_N$ for all $x \in S$, for some sufficiently large number $L$ of hidden-layer networks.

This property ensures the existence of a neural network with “appropriated weights” that approximates $f(x)$, but it does not show how to determine such weights. This issue is not an easy one, but it can be solved using back-propagation tuning and weight adaptation laws derived from the analysis of the closed-loop system trajectories, as will be demonstrated in Chaps. 6 and 9.