Chapter 1

Introduction: An Overview of Classical Sliding Mode Control

A.J. FOSSARD* and T. FLOQUET**
* DERA/CERT/ONERA, Toulouse, France
** ENSEA, Cergy-Pontoise, France

1.1 Introduction and historical account

Sliding mode control has long proved its interests. Among them, relative simplicity of design, control of independent motion (as long as sliding conditions are maintained), invariance to process dynamics characteristics and external perturbations, wide variety of operational modes such as regulation, trajectory control [14], model following [30] and observation [24]. Although the subject has already been treated in many papers [5, 6, 13, 20], surveys [3], or books [7, 17, 28], it remains the object of many studies (theoretical or related to various applications). The main purpose of this chapter is to introduce the most basic and elementary concepts such as attractivity, equivalent control and dynamics in sliding mode, which will be illustrated by examples and applications.

Sliding mode control is fundamentally a consequence of discontinuous control. In the early sixties, discontinuous control (at least in its simplest form of bang-bang control) was a subject of study for mechanics and control engineers. Just recall, as an example, Hamel's work [15] in France, or Cypkin's [27] and Emelyanov's [9] in the USSR, solving in a rigorous way the
problem of oscillations appearing in bang-bang control systems. These first studies, more concerned with analysis and where the phenomena appeared rather as nuisances to be avoided, turned rapidly to synthesis problems in various ways. One of them was related to (time) optimal control, another to linearization and robustness. In the first case, discontinuities in the control, occurring at given times, resulted from the solution of a variational problem. In the second, which is of interest here, the use of a discontinuous control was an a priori choice. The more or less high frequency of the commutations used depended on the goal pursued (linearization), as produced by the beating spoilers used in the early sixties to control the lift of a wing, conception of corrective nonlinear networks enabling them to bypass the Bode's law limitations and, of course, generation of sliding modes. Although both approaches and objectives were at the beginning quite different, it is interesting to note that they turned out to have much in common.

In fact, it was when looking for ways to design what we now call robust control laws that sliding mode was discovered at the beginning of the sixties. For the needs of military aeronautics, and even before the term of robustness was used, control engineers were looking for control laws insensitive to the variations of the system to be controlled. The linear networks used at these time did not bring enough compensation to use high gains required to get parametric insensitivity: they match the Bode's law according to which phase and amplitude effects are coupled and antagonist.

At the beginning of 1962, on B. Hamel's idea, studies of nonlinear compensators were initiated, whose aim was to overcome previous limitations. Typically, these networks, acting on the error signal $x$ of the feedback system, were defined by the relation

$$u = |F_1(x, \dot{x}, ..., |) \sgn(F_2(x, \dot{x}, ..., |))$$

where $| |$ denotes the absolute value and $F_1$ and $F_2$ are appropriate linear filters. Hence the output was discontinuous but modulated by a function of $x$ and its derivatives. Under the simplest form, one had, for instance,

$$u = - |x| \sgn(x + k \dot{x}) \quad (1.1)$$

instead of the classical PD corrector.

It is easy to see that, under the approximation of the first harmonic:

- the equivalent gain of such a network (for a sinusoidal input $x = x_0 \sin \omega t$) is independent of the amplitude $x_0$ and only depends on the pulsation $\omega$ (as a linear network), hence the denomination of pseudo linear network;
it produces a lead phase without any increase (and even decrease) of the dynamics amplitude.

For instance, in the previous case, if \( \varphi(\omega) \) is the phase of \( 1 + kp \) (\( p \) denoting the Laplace operator) at \( \omega \), the real \( Re \) and imaginary parts \( Im \) of the equivalent gain are given by

\[
Re = 1 - \frac{2}{\pi} (\varphi - \sin \varphi \cos \varphi)
\]

\[
Im = \frac{2}{\pi} \sin^2 \varphi
\]

leading to the generalized transfer locus of Figure 1.1 where, for comparison, the loci for a simple PD (dotted line) and a classical lead phase network (thin line) are given. This shows that a lead phase can be obtained (theoretically till \( \frac{\pi}{2} \)), not only without increase of the dynamics rate but also with a small reduction (from 1 to \( \frac{3}{2} \)).

Figure 1.1: Nyquist plots

In fact, it appeared simultaneously in France and in the former USSR, that these laws presented two different aspects:

- pseudo linear compensation: astute combinations of linear and non-linear signals, including commutations, can lead to appreciable advantages while being freed from the disadvantages specific to purely linear systems.
they generate a sliding motion by controlling the evolution of the system through commutations. This mode is certainly nonoptimal but exhibits a rather interesting sensitivity.

1.2 An introductory example

By way of illustration, let us take the simple example of a variable inertia $\frac{\alpha^2}{p^2}$ [1], as shown in Figure 1.2.

Taking as state variables $x_1 = x$, $x_2 = \dot{x}$, the system can be put in the following state space representation

$$\begin{align*}
\dot{x}_1 &= x_2 \\
\dot{x}_2 &= a^2 u
\end{align*}$$

(1.2)

where the control law $u$ is designed as in (1.1) and is given by

$$u = -|x_1| \text{sgn}(x_1 + kx_2)$$

(1.3)

In the following, $\sigma = x_1 + kx_2 = 0$ will be called the switching surface. The term switching illustrates the fact that the control law $u$ commutes while crossing the line $\sigma = 0$.

Then, one can easily see that:

- the phase plane is divided into four regions;
- in regions I and III (where $x_1 \text{sgn}(x_1 + kx_2) > 0$), trajectories are ellipses given by $a^2x_1^2 + x_2^2 = \text{cst}$;
- in regions II and IV (where $x_1 \text{sgn}(x_1 + kx_2) < 0$), trajectories are hyperbolas with asymptotes $x_2 = \pm ax_1$;
- the control only commutes on the boundary surface $x_1 + kx_2 = 0$;
• by a suitable choice of \( k \), all trajectories are directed toward this surface (regardless of which side of the surface they are). Consequently, once it is reached, a new phenomenon appears: the trajectories are “sliding” along this surface.

Figure 1.3: Trajectories in the portrait phase

The classical theory of ordinary differential equations however is unable to explain what occurs here (the solution of the system (1.2) is known to exist and be unique if \( u \) is a Lipschitz function, and so continuous). Consequently, the design of appropriate mathematical tools appears necessary and alternative approaches and construction of solutions can be found in Filippov’s work [11] and in other’s using the theory of differential inclusions [2]. Those results are not developed here since they are the subject of the chapter **Differential Inclusions and Sliding Mode Control**.

To understand more “physically” what is happening, a very simple interpretation can be given just by introducing some kind of imperfections in the switching devices, for instance a time delay \( \tau \). Under such an assumption, the motion proceeds along a succession of small arcs (sequentially ellipsoidal and hyperbolic) between the lines \( x_1 + k x_2 = 0 \) and \( x_1 + k' x_2 = 0 \).
crossing the origin, with

\[ k' = \frac{k - \tau}{1 + a^2k\tau} \]
\[ k'' = \frac{k - \tau}{1 - a^2k\tau} \]

When \( \tau \) tends to zero, the amplitude of these oscillations tends to zero, whereas the frequency increases indefinitely and the representative point “slides” along the line \( x + k\dot{x} = 0 \) (Figure 1.4).

Figure 1.4: Trajectories with time delay

Further important remarks must be made:

In the sliding motion, \( \sigma \equiv 0 \), which implies that the dynamics is now defined by

\[ \dot{x} = -\frac{1}{k}x \]

Therefore, the second-order system behaves then like a first-order system, with time constant \( k \) and independent of the inertia \( a \), and the trajectory will slide along \( \sigma = 0 \) to the origin (thus \( \sigma = 0 \) is also called the sliding surface). Note also that, with the discontinuous control, the system is equivalent to a proportional-derivative feedback associated with an infinite gain.

As \( \sigma = 0 \), \( x_2 + ka^2u = 0 \). On the sliding surface, the motion is consequently the same as if, instead of the discontinuous control, an “equivalent”
continuous control defined by

\[ u_e = -\frac{x_2}{ka^2} \]  

had been used. This equivalent control can be considered as the mean value of the discontinuous control \( u \) on the sliding surface, modulated in width and amplitude. Yet, in sliding motion, the control switches with a high frequency between the values \(-|x_1|\) and \(|x_1|\). This phenomenon is known as chattering and is a drawback of sliding modes (see section 1.3.3).

The latter dynamical behavior is called the ideal sliding mode, that is to say that there exists a finite time \( t_e \) such that for all \( t > t_e \),

\[ s(x(t)) = 0 \]

Of course, the ideal sliding mode along \( x + k\dot{x} = 0 \) only exists for a time-continuous system and without delay, which is not the case in real system. Attention is drawn to the fact that, under sampling, the situation is much more complicated. The problem is beyond the scope of this introductory chapter and the interested reader will find developments in subsequent chapters, for instance Discretization Issues or Sliding Mode Control for Systems with Time Delay.

This simple example allowed us to enhance some characteristics of the sliding phenomenon and it has been shown that the sliding mode was initiated at the first switching. Of course, this is not always the case unless some precautions are taken. For instance, if the discontinuous control

\[ u = -\text{sgn}(x_1 + kx_2) \]

is used instead of (1.3), the sliding mode only occurs in the layer

\[ |x_2| < ka^2 \]

as can be seen in Figure 1.5.

This comes from the fact that the switching surface is known to be attractive if the condition \( ss < 0 \) is fulfilled. This will be detailed in the following sections, as well as the dynamics in sliding motion, the notion of equivalent control, the chattering phenomenon and the robustness properties of the sliding mode.

### 1.3 Dynamics in the sliding mode

#### 1.3.1 Linear systems

Let us consider a linear process, eventually a multi-input system, defined by

\[ \dot{x} = Ax + Bu \]  

(1.5)
where $x \in \mathbb{R}^n$, $u \in \mathbb{R}^m$ and $\text{rank } B = m$.

Let us also define the sliding surface as the intersection of $m$ linear hyperplanes

$$S = \{ x \in \mathbb{R}^n : s(x) = Cx = 0 \}$$

where $C$ is a full rank ($m \times n$) matrix and let us assume that a sliding motion occurs on $S$.

In sliding mode, $s \equiv 0$ and $\dot{s} = CAx + CBu = 0$. Assuming that $CB$ is invertible (which is reasonable since $B$ is assumed to be full rank and $s$ is a chosen function), the sliding motion is affected by the so-called equivalent control

$$u_e = -(CB)^{-1}CAx$$

Consequently, the equivalent dynamics, in the sliding phase, is defined by

$$\dot{x}_e = \left[ I - B(CB)^{-1}C \right] Ax_e = A_e \dot{x}_e \quad (1.6)$$

The physical meaning of the equivalent control can be interpreted as follows. The discontinuous control $u$ consists of a high frequency component ($u_{hf}$) and a low frequency one ($u_s$): $u = u_{hf} + u_s$.

$u_{hf}$ is filtered out by the bandwidth of the system and the sliding motion is only affected by $u_s$, which can be viewed as the output of the low pass filter

$$\tau \dot{u}_s + u_s = u, \quad \tau \ll 1$$
This means that $u_e \simeq u_s$ and represents the mean value of the discontinuous control $u$.

$C$ being full rank, $Cx = 0$ implies that $m$ states of the system can be expressed as a linear combination of the remaining $(n - m)$ states. Thus, in sliding motion, the dynamics of the system evolves on a reduced order state space (whose dimension is $(n - m)$).

It is easy to verify that $A_e$ is independent of the control and has at most $(n - m)$ nonzero eigenvalues, depending on the chosen switching surface, while the associated eigenvectors belong to $\ker(C)$. As $B$ is full rank, there exists a basis where it is equivalent to the matrix

$$
\hat{B} = \begin{bmatrix}
0 \\
B_2
\end{bmatrix}
$$

where $B_2$ is an invertible $(m \times m)$ matrix. Let us decompose the state as $x = [x_1^T, x_2^T]^T$, where $x_1 \in \mathbb{R}^{n-m}$, $x_2 \in \mathbb{R}^m$. Thus, the system (1.5) becomes

$$
\begin{align*}
\dot{x}_1 &= A_{11}x_1 + A_{12}x_2 \\
\dot{x}_2 &= A_{12}x_1 + A_{22}x_2 + B_2u
\end{align*}
$$

and

$$
C = \begin{bmatrix} C_1 & C_2 \end{bmatrix}
$$

the $(m \times m)$ matrix $C_2$ being assumed invertible (which is the necessary and sufficient condition for $CB$ to be invertible since $\det(CB) = \det(C_2B_2)$). Then one can compute $A_e$ as following

$$
A_e = \begin{bmatrix}
A_{11} & A_{12} \\
-C_2^{-1}C_1A_{21} & -C_2^{-1}C_1A_{22}
\end{bmatrix} = \begin{bmatrix}
I & 0 \\
-C_2^{-1}C_1 & I
\end{bmatrix} \begin{bmatrix}
A_{11} - A_{12}C_2^{-1}C_1 & A_{12} \\
0 & 0
\end{bmatrix} \begin{bmatrix}
I & 0 \\
C_2^{-1}C_1 & I
\end{bmatrix}
$$

Under this form, the characteristic polynomial of $A_e$ clearly appears to be

$$
P(A_e) = \lambda^m P(A_{11} - A_{12}C_2^{-1}C_1)
$$

Thus $A_e$ has at least $m$ null eigenvalues and the sliding dynamics is defined by

$$
\begin{align*}
\dot{x}_1 &= (A_{11} - A_{12}C_2^{-1}C_1)x_1 \\
\dot{x}_2 &= -C_2^{-1}C_1x_1
\end{align*}
$$

These last equations are interesting since they show that:
designing \( C \) is analogous to design a state feedback matrix ensuring the desired behavior for the reduced order system \((A_{11}, A_{12})\), provided that the pair \((A_{11}, A_{12})\) is controllable (which is the case if and only if the original pair \((A, B)\) is controllable). Then the problem is a classical one which can be solved by the usual control techniques of direct eigenvalue and eigenvector placement or quadratic minimization [4], [28];

- the dynamics only depends on the matrix \( A_{11}, A_{12} \), and not on \( A_{21}, A_{22} \). For a single-input system, this means in particular, that if the system is written under the canonical controllability form,

\[
\dot{x} = \begin{pmatrix}
0 & 1 & \cdots & 0 & 0 \\
\vdots & \ddots & \ddots & \vdots & \vdots \\
\vdots & \ddots & 1 & \vdots & \vdots \\
0 & \cdots & \cdots & 0 & 1 \\
-a_0 & \cdots & \cdots & -a_{n-1}
\end{pmatrix} x + \begin{pmatrix}
0 \\
\vdots \\
\vdots \\
0 \\
1
\end{pmatrix} u
\]

then the sliding dynamics is independent from the parameters \( a_i \) of the system.

Note that this remark can be generalized to multi-input systems. However, observe that, for this kind of system, the design of the control law is more complex than in the single-input case as the required sliding motion must take place at the intersection of the \( m \) switching surfaces. Broadly speaking, at least three strategies can be considered:

- the first one uses a hierarchical procedure where the system is gradually brought to the intersection of all the surfaces. Denoting \( S_1, \ldots, S_m \) the \( m \) linear hyperplanes such that \( S = \bigcap_{i=1}^{m} S_i \), and starting from an arbitrary initial condition, the control \( u_1 \) is designed to induce a sliding mode on the surface \( S_1 \), for any control \( u_2, \ldots, u_m \). This done, the second control \( u_2 \) (while the system is still sliding on \( S_1 = 0 \)) leads to \( S_1 \cap S_2 \) and generates a sliding mode on this surface, and so on till a sliding motion takes place at the intersection of the \( m \) switching surfaces (Figure 1.6);

- another solution lies in reducing the system in \( m \) single-input subsystems such that every surface \( S_i \) only depends on the \( i^{th} \) component of the discontinuous part of the control.

These first two policies lead to a rather simple procedure. However this implies a high prompting and wear of the actuators of the system since
the control commutes at many more points of the state space than those constituting the sliding surface $\mathcal{S}$. Situations where one control drives the state away from the required intersection by imposing a sliding motion on a subset of surfaces can also occur. A way to face these problems is to make the sliding motion appear only at the intersection of all the manifolds. The control is continuous at the crossing of any separate surface and discontinuous only at the intersection of all of them. For this, the following control laws were proposed (see [7]) [called the unit vector approach],

\[
  u = u_e - \frac{\rho C x}{\|Cx\|}
\]

or

\[
  u = u_e - \frac{\rho M x}{\|Nx\|}
\]

where the matrix $M$ and $N$ are such that

\[
  \ker M = \ker N = \ker C
\]

### 1.3.2 Nonlinear systems

Let us now consider the following nonlinear system affine in the control:

\[
  \dot{x} = f(x) + g(x)u(t)
\]

and a set of $m$ switching surfaces

\[
  \mathcal{S} = \{ x \in \mathbb{R}^n : s(x) = [s_1(x), \ldots, s_m(x)]^T = 0 \}
\]
An extension of the previous results leads to:

- the associated equivalent control

\[ u_e = - \left[ \frac{\partial s}{\partial g(x)} \right]^{-1} \frac{\partial s}{\partial f(x)} \]

obtained by writing that \( \dot{s}(x) = \frac{\partial s}{\partial x} [f(x) + g(x)u(t)] = 0; \)

- the resulting dynamics, in sliding mode

\[ \dot{x}_e = \left[ I - g(x_e) \left[ \frac{\partial s}{\partial x} g(x_e) \right]^{-1} \frac{\partial s}{\partial x} \right] f(x_e) \]

Note that \( \sigma \) must be designed such that \( \frac{\partial s}{\partial x} g(x) \) is regular.

However, it is clear that, outside specific cases, the determination of the switching surfaces, in order to get a prescribed dynamics, is not as easy as in the linear case. One of these specific cases is when the system (1.7) can be transformed into the so-called regular form [18], [19]:

\[ \begin{align*}
\dot{x}_1 &= f_1(x_1, x_2) \\
\dot{x}_2 &= f_2(x_1, x_2) + g_2(x_1, x_2)u
\end{align*} \]

with \( x_1 \in \mathbb{R}^{n-m}, x_2 \in \mathbb{R}^m \) and \( g_2 \) regular. Suppose that the control problem is to stabilize the system at a prescribed point with the following dynamics

\[ \dot{x}_1 = f(x_1, h(x_1)) \]

Defining \( s(x) = x_2 - h(x_1) \) and a control \( u \) such that a sliding mode occurs on \( s = 0 \) solves the problem, and the resulting sliding motion then evolves on a reduced order manifold of dimension \((n - m)\) \( x_2 \) can be viewed as the input of the subsystem whose state is \( x_1 \). This can be illustrated by the example of the two-arm manipulators which can be found in [25]. Yet, the transformation of the system into the regular form can induce complex diffeomorphisms. An alternative is to proceed by pseudo linearization as in [21].

1.3.3 The chattering phenomenon

An ideal sliding mode does not exist in practice since it would imply that the control commutes at an infinite frequency. In the presence of switching imperfections, such as switching time delays and small time constants in the actuators, the discontinuity in the feedback control produces a particular
dynamic behavior in the vicinity of the surface, which is commonly referred to as chattering [Figure 1.7].

This phenomenon is a drawback as, even if it is filtered at the output of the process, it may excite unmodeled high frequency modes, which degrades the performance of the system and may even lead to unstability [16]. Chattering also leads to high wear of moving mechanical parts and high heat losses in electrical power circuits. That is why many procedures have been designed to reduce or eliminate this chattering. One of them consists in a regulation scheme in some neighborhood of the switching surface which, in the simplest case, merely consists of replacing the signum function by a continuous approximation with a high gain in the boundary layer: for instance, sigmoid functions (see [23]) or saturation functions as shown in Figure 1.8. However, although the chattering can be removed, the robustness of sliding mode is also compromised. Another solution to cope with chattering is based on the recent theory of higher-order sliding modes (see Chapter 3).
The real motion near the surface can be seen as the superposition of a "slow" movement, along the surface, and a "fast" one, perpendicular to this surface (the chattering phenomenon). To put in a prominent position these two movements, let us consider again our introductive example and let us approximate, in an \( \varepsilon \)-neighborhood of the surface, the signum function by a saturation function whose slope is \( \frac{1}{\varepsilon} \). Taking \( \varepsilon \) as a (small) perturbation parameter, the behavior in the boundary layer can be described, under the standard singularly perturbed form, by

\[
\begin{align*}
\dot{x}_1 &= x_2 \\
\dot{x}_2 &= -a^2 x_1 (x_1 + k x_2)
\end{align*}
\]

The slow motion is defined by setting \( \varepsilon = 0 \), hence

\[
\dot{x}_{1s} = x_{2s} = -\frac{1}{k} x_{1s}
\]

and

\[
x_{1s} = x_{10} e^{-\frac{t}{\varepsilon}}
\]

\[
x_{2s} = -\frac{1}{k} x_{10} e^{-\frac{t}{\varepsilon}}
\]

with \( x_{10} \) being the value of \( x_1 \) at point \( M_1 \) (see Figure 1.9). As it has been seen in section 1.2, this corresponds to the dynamics in the sliding motion.

In the time scale \( \frac{t}{\varepsilon} \), the fast motion is defined by

\[
\dot{x}_{2f} = -a^2 x_{10}^2 - a^2 k x_{10} x_{2f}
\]

that is

\[
x_{2f} = -\frac{1}{k} x_{10} (1 - e^{-\frac{a^2 k t}{\varepsilon}}) + x_{10} e^{-\frac{a^2 k t\varepsilon}{\varepsilon}}
\]

and the global motion is approximated by

\[
x_1 = x_{1s} = x_{10} e^{-\frac{t}{\varepsilon}}
\]

\[
x_2 = x_{2s} + x_{2f} - x_{20s} = -\frac{1}{k} x_{10} e^{-\frac{t}{\varepsilon}} + \left( x_{20} + \frac{1}{k} x_{10} \right) e^{-\frac{a^2 k t\varepsilon}{\varepsilon}}
\]

which gives the trajectories in Figure 1.9.

### 1.4 Sliding mode control design

#### 1.4.1 Reachability condition

It has been said that, in the sliding, the motion was independent from the control. Nonetheless, it is obvious that the control must be designed such
Figure 1.9: a) Singular perturbed motion $\varepsilon = 0$; b) Real motion

that it drives the trajectories to the switching surface and maintains it on this surface once it has been reached. The local attractivity of the sliding surface can be expressed by the condition

$$\lim_{s \to 0} \frac{\partial s}{\partial z} (f + gu) < 0 \quad \text{and} \quad \lim_{s \to 0} \frac{\partial s}{\partial z} (f + gu) > 0$$

or, in a more concise way,

$$s \ddot{s} < 0$$

which is called the reachability condition [17].

Example 1 In a way of illustration, let us consider a dc-motor modeled by the following transfer function

$$Y(p) = \frac{1}{p(p + 1)} U(p)$$

that is, in a state-space representation:

$$\begin{align*}
\dot{x}_1 &= x_2 \\
\dot{x}_2 &= -x_2 + u \\
y &= x_1
\end{align*}$$

Let us assume that the sliding surface is designed as

$$s = x_2 + \alpha x_1 = 0, \quad \alpha > 0$$
Thus

\[ \dot{s} = (\alpha - 1)x_2 + u \quad (1.12) \]

Using the control law \( u = -k \text{sgn}(s) \), \( k > 0 \), the reachability condition is satisfied in the domain

\[ \Omega = \{ x : |(\alpha - 1)x_2| < k \} \]

since

\[ ss < |s| \left(|(\alpha - 1)x_2| - k \right) < 0 \]

One should note that condition (1.10) is not sufficient to ensure a finite time convergence to the surface. Indeed, in the latter example, the control

\[ u = (1 - \alpha)x_2 - ks \]

provides \( \dot{s} = -ks \), but the convergence to \( s = 0 \) is only asymptotic since

\[ s(t) = s(0)e^{-kt} \]

where \( s(0) \) is the initial value of \( s \). Condition (1.10) is often replaced by the so-called \( \eta \)-reachability condition

\[ s\dot{s} \leq -\eta |s| \quad (1.13) \]

which ensures a finite time convergence to \( s = 0 \), since by integration

\[ |s(t)| - |s(0)| \leq -\eta t \]

showing that the time required to reach the surface, starting from initial condition \( s(0) \), is bounded by

\[ t_e = \frac{|s(0)|}{\eta} \]

In a practical way, the control law is generally displayed as \( u = u_e + u_d \) where \( u_e \) is the equivalent control (allowing us to cancel the known terms on the right hand side of (1.12)) and where \( u_d \) is the discontinuous part, ensuring a finite time convergence to the chosen surface.

The example (1.11) was simulated using the following control law

\[ u = (1 - \alpha)x_2 - k \text{sgn } s \]

where the term \( (1 - \alpha)x_2 \) represents the equivalent control (since \( \dot{s} = 0 \) implies \( u + (\alpha - 1)x_2 = 0 \)). One can also note that the \( \eta \)-reachability condition is satisfied. Figures 1.10 and 1.12 show obviously that the sliding
motion takes place after about 1.3 sec. Indeed, after this time, the dynamics of the system is represented by the reduced order system given by the chosen surface, i.e.:

\[ \dot{x}_1 = -\alpha x_1 = x_2 \]

and the control switches at high frequency. In Figure 1.12 one can see that the equivalent control, in sliding motion, represents the mean value of the control \( u \). The portrait phase, in Figure 1.11, illustrates the two steps of the dynamics behavior: first, a parabolic trajectory before the surface is reached (which is called the reaching phase) and then the sliding along the designed line \( s = 0 \) \( (x_2 = -\alpha x_1) \) to the origin.

![Figure 1.10: Evolution of the states versus time \( x_1 \) (dotted) and \( x_2 \) (solid)](image)

1.4.2 Robustness properties

An important feature of sliding mode control is its robustness properties with respect to uncertainties. In the case of invariant and nonperturbed systems, recall first that the use of a continuous component, equal to \( u_e \), allows the use of a discontinuous component as small as desired. Indeed, for the sake of simplicity, consider the linear system (1.5) and choose the following controller

\[ u = u_e - k (CB)^{-1} \text{sgn}(s) \]
Figure 1.11: Portrait phase of the sliding motion

Figure 1.12: Discontinuous and equivalent control
with \( u_e = -(CB)^{-1}CAx \). This implies
\[
ss = sC\dot{x} = s[CAx + CBu_e - k\text{sgn}(s)] = -k|s| < 0
\]
and consequently \( k \) might be taken high enough when the trajectory is far from the switching surface (so that the reaching time is short) and then as small as desired in order to limit the chattering.

Actually the use of a large enough discontinuous signal is necessary to complete the reachability condition despite parametric uncertainties and exogenous perturbations. Still, to be as simple as possible, consider the system under the canonical controllable form but with parametric uncertainties \( \Delta a_i \)

\[
\dot{x} = \begin{pmatrix}
0 & 1 & \cdots & 0 & 0 \\
\vdots & \ddots & \ddots & \vdots & \vdots \\
\vdots & \ddots & 1 & \vdots & \vdots \\
0 & \cdots & \cdots & 0 & 1 \\
-a_0 - \Delta a_0 & \cdots & \cdots & -a_{n-1} - \Delta a_{n-1}
\end{pmatrix} x + \begin{pmatrix}
0 \\
\vdots \\
\vdots \\
0 \\
1
\end{pmatrix} u
\]

where the \( \Delta a_i \) are all supposed to be bounded such that
\[
a_i^- < |\Delta a_i| < a_i^+
\]

Let the switching surface be
\[
s = [c_0 \ c_1 \ c_{n-2} \ 1]x = 0
\]
(corresponding to the sliding dynamics \( p^{n-1} + c_{n-2}p^{n-2} + \ldots + c_0 = 0 \)).

The control law is chosen as follows
\[
u = \sum_{i=1}^{n} k_{i-1}x_i - k_n\text{sgn}(s)
\]

The \( \eta \)-reachability condition, (1.13), can be satisfied by two ways, and thus despite the uncertainties:

- if constant gains are set as \( k_0 = a_0, k_i = a_i - c_{i-1}, i = 1, \ldots, n-1 \), one gets \( ss = -\sum_{i=1}^{n} \Delta a_{i-1}x_is - k_n|s| \)

and thus setting
\[
k_n > \eta + \sum_{i=1}^{n} |\Delta a_{i-1}x_i|
\]
is sufficient to satisfy (1.13). The magnitude of the discontinuity in the control is a function of the state and of the uncertainties on the process. The control law is easy to design but the discontinuity can be important (and consequently the chattering).

- another solution relies on using commuting gains. Taking \( k_0 = \dot{k}_0 + a_0, k_i = \dot{k}_i + a_i - c_{i-1}, i = 1, \ldots, n - 1 \) leads to

\[
ss = \sum_{i=2}^{n}(\dot{k}_{i-1} - \Delta a_{i-1})x_is - k_n |s|
\]

and the condition \( s\dot{s} < -\eta |s| \) can be satisfied by choosing \( k_n = \eta \) as a small positive scalar and

\[
\dot{k}_{i-1} = \begin{cases} 
\alpha_{i-1} & \text{if } x_is > 0 \\
\alpha^i_{i-1} & \text{if } x_is \leq 0 
\end{cases} \quad i = 1, \ldots, n
\]

The structure of the control law is a little more complex but the amplitude of the discontinuity in the control is reduced.

Sliding modes are also known to be insensitive to exogenous perturbations satisfying the so-called matching condition (originally stated and proved by Drazenovic in [6]), that is to say that these perturbations act exactly in the input channels. Considering the perturbed linear system

\[
\dot{x} = Ax + Bu + \Delta(x, t)
\]

where \( \Delta \) is an unknown but bounded function, the matching condition means that the sliding mode is insensitive to the uncertain function \( \Delta \) if it is in the range space of the input matrix \( B \): that is, there exists a known matrix \( D \) and an unknown function \( \delta \) such that \( \Delta = D\delta \) and \( \text{rank}[B\ D] = \text{rank} B \). Indeed, it is easy to show that, in that case,

\[
\left(I - B\ (CB)^{-1} C\right)\Delta = 0
\]

since

\[
\left(I - B\ (CB)^{-1} C\right)B = 0
\]

and thus the dynamics in sliding motion remains independent of the exogenous input \( \Delta \) (\( \dot{x}_e = \left[I - B\ (CB)^{-1} C\right]Ax = A_e x \)).

It is important to note that the system only becomes insensitive to those perturbations during sliding mode but remains affected by the perturbations during the reaching phase (that is to say before the sliding surface has been reached).
1.5 Trajectory and model following

In the previous sections, variable structure control and sliding modes have been designed for regulation purposes but they can also be used for trajectory and model following.

1.5.1 Trajectory following

Without going into the details, and with the aim of outlining the interest of sliding mode controls in trajectory following, let us consider a simple linear single-input system

$$y^{(n)} + a_{n-1}y^{(n-1)} + \ldots + a_1 \dot{y} + a_0 y = u$$

written in the canonical controllable representation

$$\dot{x} = \begin{pmatrix} 0 & 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & 1 & \vdots & \vdots \\ 0 & \cdots & 0 & 1 \\ -a_0 & \cdots & \cdots & \cdots & -a_{n-1} \end{pmatrix} \begin{pmatrix} x \\ u \end{pmatrix}$$

where $x = [y, \dot{y}, \ldots, y^{(n-1)}]^T$.

Assume that the control problem is to constrain the output $y$ to follow a prescribed trajectory $y_d(t)$ and set

$$x_d = [y_d, \dot{y}_d, \ldots, y_d^{(n-1)}]^T$$

Defining the sliding surface to be $s(t) = C(x - x_d)$ and designing a control law leading to a sliding motion on this surface gives $\dot{x} = x_d$. It should be noted, in comparison with the regulation case, that here, the surface is time-varying and that the dynamics of the response is imposed by the desired trajectory (and not by the coefficients of the surface).

It should also be noted that this idea can be enlarged to nonlinear multi-input systems. Consider for instance the system

$$\begin{align*}
\dot{x}_1 &= 3x_1 + x_2^2 + x_1 \dot{x}_2 \cos x_2 + u_1 \\
\dot{x}_2 &= x_1^3 - x_2 \cos x_1 + u_2
\end{align*}$$

whose outputs are

$$\begin{align*}
y_1 &= x_1 \\
y_2 &= x_2
\end{align*}$$
The control problem is to constrain these outputs to follow trajectories corresponding to second order responses with respect to step inputs. It is sufficient to take the sliding surfaces

\[ s_i = c_i e_i + \dot{e}_i, \quad i = 1, 2 \]

with \( e_i = x_i - x_{id} \), and to generate controls \( u_i \) such that \( s_i \dot{s}_i < 0 \). Taking

\[ u_1 = k_{11} \dot{e}_1 + \alpha_{11} x_1 + \alpha_{12} x_2^2 + \alpha_{13} x_1 \dot{x}_2 + \dot{x}_{id} - k_1 \text{sgn} s_1 \]

gives

\[ \dot{s}_1 = (c_1 + k_{11}) \dot{e}_1 + (3 + \alpha_{11}) x_1 + (1 + \alpha_{12}) x_2^2 + (\cos x_2 + \alpha_{13}) x_1 \dot{x}_2 - k_1 \text{sgn} s_1 \]

so that with \( k_{11} = -c_1, \alpha_{11} = -3, \alpha_{12} = -1 \)

\[ s_1 \dot{s}_1 = (\cos x_2 + \alpha_{13}) x_1 \dot{x}_2 s_1 - k_1 |s_1| \]

Thus, taking

\[ \alpha_{13} = -\text{sgn}(x_1 \dot{x}_2 s_1) \]

implies

\[ s_1 \dot{s}_1 < 0, \forall k_1 > 0 \]

The control \( u_2 \) can be designed similarly such that \( s_2 \dot{s}_2 < 0 \). Then each output follows the predefined trajectories.

### 1.5.2 Model following

Variable structure control and sliding mode can also be used for model following, that is to control the process in such a way that it behaves like a given model (of the same order). The idea is to force a sliding motion on the surfaces

\[ S = K_e (x_m - x) = K_e x_e = 0 \]

where \( x \) and \( x_m \) are respectively, the process and model state vectors. It is easy to see that, in sliding motion, the error dynamics is given by

\[ \dot{x}_e = (1 - \Theta) A K_e \]

with \( \Theta = B (K_e B)^{-1} K_e \).

Except for the case of perfect matching, which supposes that

\[ \text{rank } [B, B_m] = \text{rank } [B, A_m - A] = \text{rank } B \]

there exists a steady-state error which can be computed by the equation

\[
\begin{pmatrix}
(1 - \Theta) A & K_e \\
K_e & 0
\end{pmatrix}
\begin{pmatrix}
x_e \\
K_e
\end{pmatrix} =
\begin{pmatrix}
-(1 - \Theta) A_m^{-1} B_m \bar{u}_m \\
0
\end{pmatrix}
\]
where \([1 - \Theta) A]_T\) denotes the matrix constituted by the \((n - m)\) independent lines of \((1 - \Theta) A\).

In the general case, when the conditions cannot be met, one will only focus on the outputs and integrators to be added on the error \(y_m - y\) so that the steady state error is null (Figure 1.13).

By way of illustration, let us consider the following case of a process given by the transfer function

\[
G(s) = \frac{4p(s + 1)}{s^2 + 4\delta s + 4}
\]

where \(\rho\) and \(\delta\) are parameters which may vary. The control problem is to follow a model corresponding to

\[
G_m(s) = \frac{1}{s^2 + 1.4s + 1}
\]

The following figure shows the results of simulations enhancing the fact that the model following scheme is able to cope with important parametric variations. In Figure 1.14, continuous variations of \(\delta = \frac{1}{2}\) and \(\rho = 2 + \frac{3}{2}t\) (that is to say, for the span time of 9 seconds, \(\delta\) is varying from 0 to 1 and \(\rho\) from 2 to 14).

As far as the problem of model following is concerned, variable structure control laws using sliding modes can also be found in [12], [29] or [30].
1.6 Conclusion

In this introductory chapter, the basic properties and interests of sliding modes have been enhanced. Since this technique involves differential equations with discontinuous right-hand sides, the concept of solution needs to be redefined and alternative approaches to the classical ordinary differential equation theory must be developed. One concerns differential inclusions and is presented in Chapter 2. The main benefits of sliding mode control are the invariance properties and the ability to decouple high-dimensional problems into sub-tasks of lower dimensionality. However, it has been shown that imperfections in switching devices and delays were inducing a high-frequency motion called chattering (the states are repeatedly crossing the surface rather than remaining on it), so that no ideal sliding mode can occur in practice. Yet, solutions have been developed to reduce the chattering and so that the trajectories remain in a small neighborhood of the surface, like the higher-order sliding modes developed in Chapter 3. The continuous case has been considered in this introduction, but the problems induced by sliding modes under sampling and in the presence of delays are treated in Chapters 8, 10, 11.

The control problem given here was a regulation one and the illustrative examples were quite simple. However, sliding modes find their application in many other areas such as observers (Chapter 4), output feedback (Chapter 5) or trajectory following (Chapter 6), and in practical applications such
as robotics [Chapter 13] and control of induction motors [Chapter 14].

References


