1

The Control Loop: Characterization and Behavior in Open Loop and Closed Loop

1.1. Introduction

After demonstrating that the plant output disturbance is none other than the effect of the set of so-called "original" disturbances on its output, this chapter develops *tracking-regulation duality* in the operational and frequency domains, applying the most dual approach possible in order to best structure the presentation of the control loop.

After a very detailed and broad definition of transmittance in open loop, which responds to the linear systems coming under engineering, the open loop is used to physically present the stable, unstable or barely oscillating nature of the control loop, as well as the left-hand criterion, which governs this character for a stable open loop with minimum phase shifting.

Transmittances in closed loop, which are the tracking and regulation transmittances, as well as the resulting tracking and regulation functions, which introduce *reference tracking* and *disturbance rejection*, constitute the substrate of tracking-regulation duality.

The structural nature, dictated by such a duality, is then used to study, in turn, the input sensitivity (of the plant), which results in replacing the academic Proportional, Integral and Derivative controller (PID) with the non-academic PID, the behavior and performances of the control loop, as well as its dynamics, before ending with the charts in tracking and regulation, whose duality speaks for itself.

1.2. Definition and terminology

Referring to the dictionary definition of the verb “to track”, a *servo-control* can be defined by its function, which consists of subjecting (or servo-controlling) the output of a plant to a reference, independently of the presence of disturbances.

More precisely, this definition introduces both the idea of a *reference tracking* (up to a factor, which can be a single unit) and the idea of a *disturbance rejection* (consistent with independence from disturbances): these two ideas, which relate to the reference and disturbances, respectively, indeed contribute in a complementary manner to *global reference tracking*, which can be defined as the joint response to the reference and disturbances. Moreover, the nature of a tracking such as this justifies the designation, “servo-control” that has long been, and continues to be, attributed in this field.

A servo-control can also be referred to as a “servo-system”, “corrected-error control” or “control loop” (or, more simply, “control”). Within the framework of our developments, we have a preference for the designation *control loop* because in the genericity of the term “control”, it provides a true reflection of the specific structural nature of a servo-control.

1.3. The plant

1.3.1. *Definition*

The plant is the system to be servo-controlled (see Figure 1.1). It is controlled by the *input* (or *control*) $u(t)$ and disturbed by *disturbances* (said to be *original*) $d'_1(t)$, ... , $d'_q(t)$. The *output*, $y(t)$, depends on the input and disturbances: in fact, the input and disturbances are the *stresses*, and the output, $y(t)$, is *the response to the stresses*.

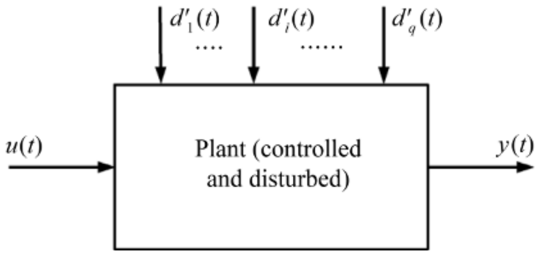

Figure 1.1. *The plant, with output,* $y(t)$, *is controlled by* $u(t)$ *and disturbed by* $d'_1(t)$, ... , $d'_q(t)$

1.3.2. *From a modeling difficulty to a pragmatic approach to disturbances*

The relationship between the stresses and output is a *cause-and-effect relationship*, with the stresses as the cause of the effect on the output.

In accordance with a declination of the stresses, the relationship between the input and output (or *input–output relationship*) is a cause-and-effect relationship that *can be modeled*, definable using a differential equation (which may or not be linear, according to the nature of the plant). While the relationship between a disturbance and the output (or *disturbance–output relationship*) is also a cause-and-effect relationship, it is, however, *difficult (or even impossible) to model*.

It is precisely this *modeling difficulty* that is at the origin of the *pragmatic approach to disturbances*, as presented here.

The idea is to relate the set of original disturbances, $d'_1(t)$, …, $d'_q(t)$, processed by the plant, to the plant output, in the form of an *output disturbance*, $d(t)$.

To this end, let us consider the block diagram of the plant (see Figure 1.2), in which $G(s)$ represents the transmittance relative to the input, $u(t)$, (or *plant transmittance*) and $G'_i(s)$ the transmittance relating to the original disturbance, $d'_i(t)$, that is,

$$G(s) = \frac{Y(s)}{U(s)} \text{ and } G'_i(s) = \frac{Y(s)}{D'_i(s)}, \qquad [1.1]$$

with each of the transmittances defined for a stress, and the others considered to be zero.

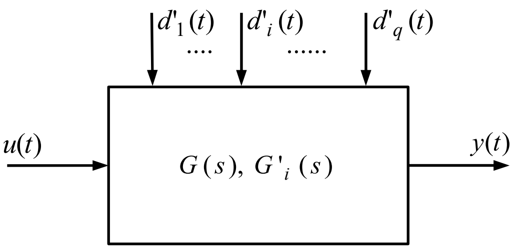

Figure 1.2. *Functional representation of the plant*

The output Laplace transform, $y(t)$, admits an expression of the form:

$$Y(s) = G(s)U(s) + G'_1(s)D'_1(s) + ... + G'_i(s)D'_i(s) + ... + G'_q(s)D'_q(s), \quad [1.2]$$

or:

$$Y(s) = G(s)U(s) + \sum_i G'_i(s)D'_i(s). \quad [1.3]$$

That is, by writing

$$D(s) = \sum_i G'_i(s)D'_i(s): \quad [1.4]$$

$$Y(s) = G(s)U(s) + D(s), \quad [1.5]$$

with this equation reflecting the block diagram in Figure 1.3, in which $d(t) = \mathcal{L}^{-1}[D(s)]$.

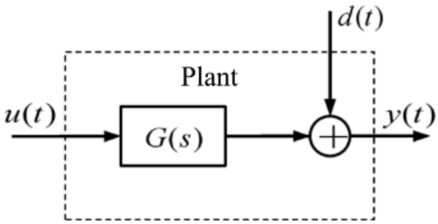

Figure 1.3. *Plant controlled by* $u(t)$ *and disturbed at the output by* $d(t)$

The disturbance, $d(t)$, known as the *plant output disturbance* or, more simply, *disturbance* where there is no ambiguity, therefore represents *the effect on the plant output of the set of original disturbances*.

In conclusion, when the plant is integrated into a control loop, the *control system* (which generates the control, $u(t)$, of the plant) processes, not the original disturbances, but *the effect of these disturbances on the plant output*.

As a warning, it should be specified that if the control output is none other than the plant output, it is nonetheless important (as we will see below) not to confuse *plant output disturbance*, $D(s)$, and *control output disturbance*, $D(s)/(1+\beta(s))$, with $\beta(s)$ designating the open-loop transmittance of the control. Moreover, $1+\beta(s)$ or $1+\beta(j\omega)$ defines the (complex) *efficiency factor*, which *divides* the disturbance, $D(s)$ or $D(j\omega)$, existing at the plant output in the absence of a control system.

1.4. Functional representation of the control loop

1.4.1. *Block diagram and corresponding transfers*

The block diagram in Figure 1.4 provides a functional representation of a corrected error control of a disturbed plant.

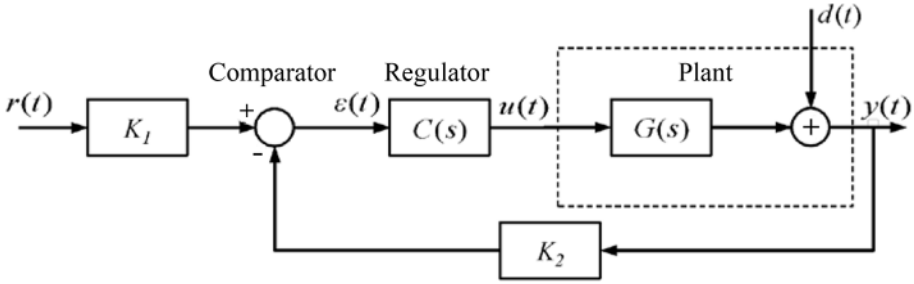

Figure 1.4. *Block diagram of a control loop: the dotted line framing the plant (controlled and disturbed) in compliance with the previous figure, is not indispensable and does not generally feature in the functional representation*

There are two different types of variable involved:

– the *external variables*, which, where stresses are concerned, constitute the reference input (or reference), $r(t)$, and the output disturbance of the plant (or disturbance), $d(t)$, and where the response to these stresses is concerned, the output, $y(t)$, of the plant and the control loop;

– the *internal variables*, which are the error (or gap) signal $\varepsilon(t)$ and the input (or control) $u(t)$ of the plant.

With regard to the symbolic operators of the *direct (or action) chain*, $G(s)$ represents the transmittance of the *system to be servo-controlled (or plant)* and $C(s)$ the transmittance of the *controller (or regulator)*. K_1 and K_2 designate the transfer factors of the *display system* and the *return (or reaction) chain*, generally constituted by the sensor of the output variable: these transfer factors result from the reduction of the transfer functions, $K_1(s)$ and $K_2(s)$, dictated by the significance of their bandwidth with respect to the control loop. The sub-assembly of the control loop, which delivers (or generates) the control law, $u(t)$, from the reference, $r(t)$, and the output, $y(t)$, is called the *system*. In the specific case where K_1 and K_2 are respectively reduced to become one, the control loop is said to have a *unit display and reaction* or, more simply, an elementary form.

1.4.2. *Regarding the gap detector*

The gap detector delivers the error (or gap) signal, $\varepsilon(t)$, from the reference, $r(t)$, and the output, $y(t)$. It comprises three sub-systems: the output comparator, $\varepsilon(t)$, the input display system, $r(t)$, and the input sensor, $y(t)$.

1.4.2.1. *The comparator*

The *comparator*, which delivers the error signal through a *subtraction operation*, is generally electronic in nature, and aims to compare physical values of the same type. In analog, it can be achieved by a differential amplifier with operational amplifier.

1.4.2.2. *The display system*

The *display system* can be a sensor, which transforms the reference value, $r(t)$, into an electronic value, notably a voltage, which is indeed compatible with the nature of the comparator, as defined technologically. If the reference value, $r(t)$, is an electrical voltage to begin with, it can be applied to the non-inverting input of the comparator, either directly or via a potentiometer or a low-power amplifier.

1.4.2.3. *The sensor (of the output variable)*

The *sensor* consists of the sensor itself and its environment. An illustrative example is that of a *position sensor*, which can be achieved by a potentiometer (the sensor itself) supplied by two opposite direct voltages via two identical resistors located on either side of the resistive track of the potentiometer. In the case of a rectilinear position, the track is straight; for an angular position, the track is circular for a rotation of less than one turn, or helical for a rotation of more than one turn.

In light of the elements presented in the above section and generalized here, a *sensor transforms (or converts) one physical value into another physical value*. For example, it transforms a mechanical value into an electrical value in a position or speed control.

In the angular position control of a motor shaft, the sensor constituting the reaction chain converts the shaft's angular position into an electrical voltage. If the sensor converts 1 radian into 5 V, the transfer factor, K_2, is equal to 5 V/rd, that is, $K_2 = 5$ (unitless). If the same sensor is modified with respect to its environment (via the supply voltages or the series resistors) to ensure conversion from 1 radian to 1 V, K_2 is then reduced to 1 V/rd, that is, $K_2 = 1$ (unitless): thus, the sensor output gives (up to 1 *measurement noise* aside) the measurement of the output, $y(t)$, that is, the *output measured* $y_m(t)$, such that (unitless) $y_m(t) = y(t)$.

If the angular position of the motor shaft is servo-controlled to that of another motor shaft (in a motor-shaft synchronization scenario), the display system of the control can be composed of a sensor similar to that constituting the reaction chain, with the display system having to convert the angular position of the other shaft into an electrical voltage. If the conversion is brought to 1 V for 1 radian, the transfer factor, K_1, is reduced to 1 V/rd, that is, $K_1 = 1$ (unitless), the output of the sensor then giving (up to 1 *measurement noise* aside) the measurement of the reference, $r(t)$, that is, the *measured reference*, $r_m(t)$, such that (unitless) $r_m(t) = r(t)$.

1.4.2.4. *Benefit of transfer functions* $K_1(s)$ *and* $K_2(s)$, *reduced to transfer factors*

The block diagram in Figure 1.4 enables the error signal, $\varepsilon(t)$, to be expressed in the form:

$$\varepsilon(t) = K_1 r(t) - K_2 y(t) \cdot \qquad [1.6]$$

In the frequency band in which the control loop is *effecient (good reference tracking and good disturbance rejection)*, the open-loop gain is high, as is, consequently, the action-chain gain, given that the reaction chain is not supposed to provide any gain (as a sensor is not, in essence, an amplifier).

In other words, for a finite output, $y(t)$, the error signal, $\varepsilon(t)$, is then close to zero, thus enabling us to write:

$$K_1 r(t) - K_2 y(t) = 0\,, \qquad [1.7]$$

that is:

$$y(t) = \frac{K_1}{K_2} r(t) \cdot \qquad [1.8]$$

This relationship demonstrates that the value and variations of $y(t)$ are (as is appropriate for reference tracking and disturbance rejection) like those of $r(t)$, *because the ratio* K_1/K_2 *is a proportionality coefficient, given that* K_1 *and* K_2 *are transfer factors*. Hence, the benefit of such factors, which moreover merely confirm that a comparator is, essentially, supposed to compare values of the same type, which reflect the reference and the output (therefore proportional to $r(t)$ and $y(t)$).

The transfer factors K_1 and K_2, whose ratio, K_1/K_2, determines the proportionality coefficient between $r(t)$ and $y(t)$, thus need to be precise and reliable in order to meet the precision and reliability requirements of the output, $y(t)$. To this end, their achievement will rely more greatly on passive components, which are more precise and reliable than active components.

1.4.2.5. *An absolute reference*

For $K_1 = K_2$ or $K_1 = K_2 = 1$ (in the elementary control loop scenario), the relationship

$$y(t) = (K_1/K_2)\, r(t) \qquad [1.9]$$

is reduced to the new relationship:

$$y(t) = r(t), \qquad [1.10]$$

two relationships that call for a clear distinction to be made.

The first relationship expresses that the *desired output*, $(K_1/K_2)\, r(t)$, reflects (in the proportionality sense) the reference, $r(t)$. The second expresses that the *desired output*, $r(t)$, is the reference itself: the reference, $r(t)$, then appears as an *absolute reference* for the output, $y(t)$.

1.5. Open-loop transmittance

1.5.1. *Definition*

The transmittance in open loop, $\beta(s)$, of a control loop is the opposite of the ratio of the Laplace transforms of the output and input of the open loop, with its output then charged by its input impedance, and the reference, $R(s)$, and the disturbance, $D(s)$, being considered nil (see Figure 1.5), that is:

$$\beta(s) = \left[-\frac{V_s(s)}{V_e(s)} \right]_{R(s)=D(s)=0} . \qquad [1.11]$$

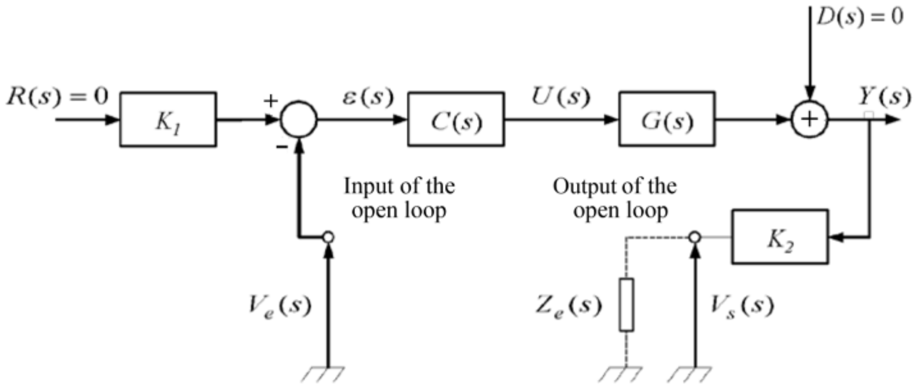

Figure 1.5. *Block diagram, from which the transmittance in open loop, $\beta(s)$, is determined: $Z_e(s)$ designates the impedance presented by the inverting input of the comparator*

1.5.2. *General expression*

Observing Figure 1.5 enables us to express the open-loop output symbolically:

$$V_s(s) = K_2 G(s) C(s)(-V_e(s)) , \qquad [1.12]$$

from which we obtain:

$$\frac{V_s(s)}{V_e(s)} = -K_2 G(s) C(s) , \qquad [1.13]$$

or, given the definition of $\beta(s)$ given by the relationship [1.11]:

$$\beta(s) = C(s)G(s)K_2, \qquad [1.14]$$

a result that reflects the fact that *the open-loop transmittance is none other than the product of the transmittances of the action and reaction chains*.

1.5.3. *Regarding the input impedance of the open loop*

Taking the inverting input impedance of the comparator in the definition of open-loop transmittance into account enables any looped system that is electrical, electronic, mechanical, thermal, etc., in nature to be presented in the form of a control loop block diagram.

One characteristic example is that of the COLPITTS oscillator (see Figure 1.6), which uses:

– an amplifier stage with a transistor connected in common base;

– loaded by a resonant circuit (with resonance frequency $\omega_0 = 1/\sqrt{LC}$);

– looped by a capacitive voltage divider.

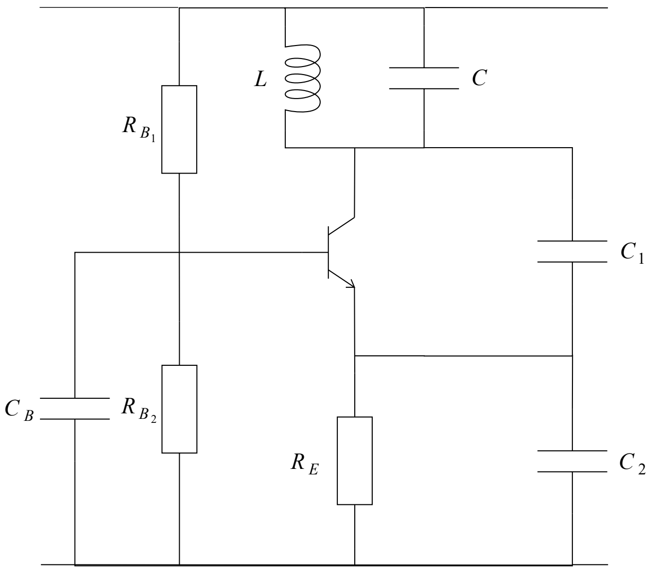

Figure 1.6. *COLPITTS oscillator*

Without taking the input resistance, R_e , of the amplifier stage (see Figure 1.7), into account, the transfer function, $K_2(s)$, of the reaction chain is reduced to the transfer factor, K_2, such that:

$$K_2 = \frac{\frac{1}{C_2 s}}{\frac{1}{C_1 s} + \frac{1}{C_2 s}} = \frac{\frac{1}{C_2}}{\frac{1}{C_1} + \frac{1}{C_2}} = \frac{C_1}{C_1 + C_2}. \quad [1.15]$$

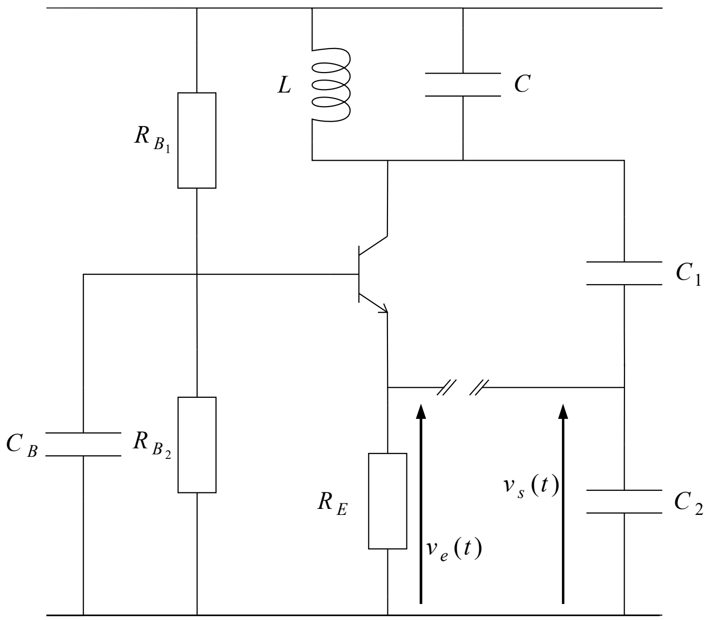

Figure 1.7. *Loop opening with the input resistance of the amplifier stage not taken into account*

Taking the input resistance, R_e, of the amplifier stage (see Figure 1.8) into account enables the transfer, $K_2(s)$, to be expressed in the form:

$$K_2(s) = \frac{R_e \,//\, \frac{1}{C_2 s}}{\frac{1}{C_1 s} + R_e \,//\, \frac{1}{C_2 s}} = \frac{R_e C_1 s}{1 + R_e (C_1 + C_2) s}, \quad [1.16]$$

with R_e of the order of several dozen Ohms, an expression that complies with that relating to the closed loop, the loop opening thus not modifying the operating conditions.

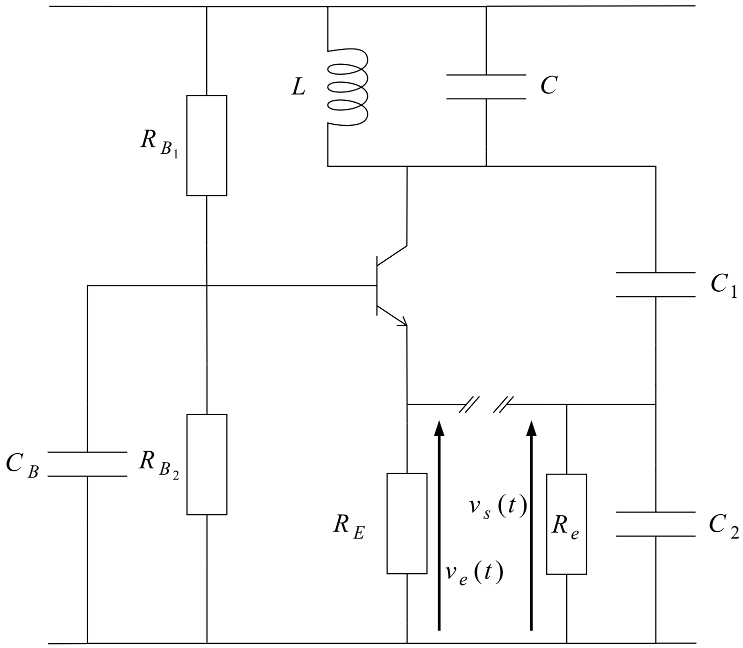

Figure 1.8. *Loop opening with the input resistance of the amplifier stage taken into account:* $R_e = R_E \,//\, (r_{BE}/\beta)$ *with the natural parameters;* $R_e = R_E \,//\, (h_{11}/h_{21})$ *with the hybrid parameters; here* β *designates the transistor current gain with small variations (* $\beta = h_{21}$ *and* $r_{BE} = h_{11}$*)*

1.5.4. *Open-loop Nichols locus: elementary form and characteristic quantities*

Within the framework of a general graphical representation aiming to illustrate the explanations relating to control in a straightforward manner, an *elementary-form open-loop Nichols locus* may be used. It is represented in Figure 1.9 in a graphically truncated form at low and high frequencies. Thus, we attribute to the open-loop Nichols locus, a monotonic and decreasing *"elementary" form (or shape)*, which underlines the *low-pass character* of the corrected plant, $C(j\omega)G(j\omega)$.

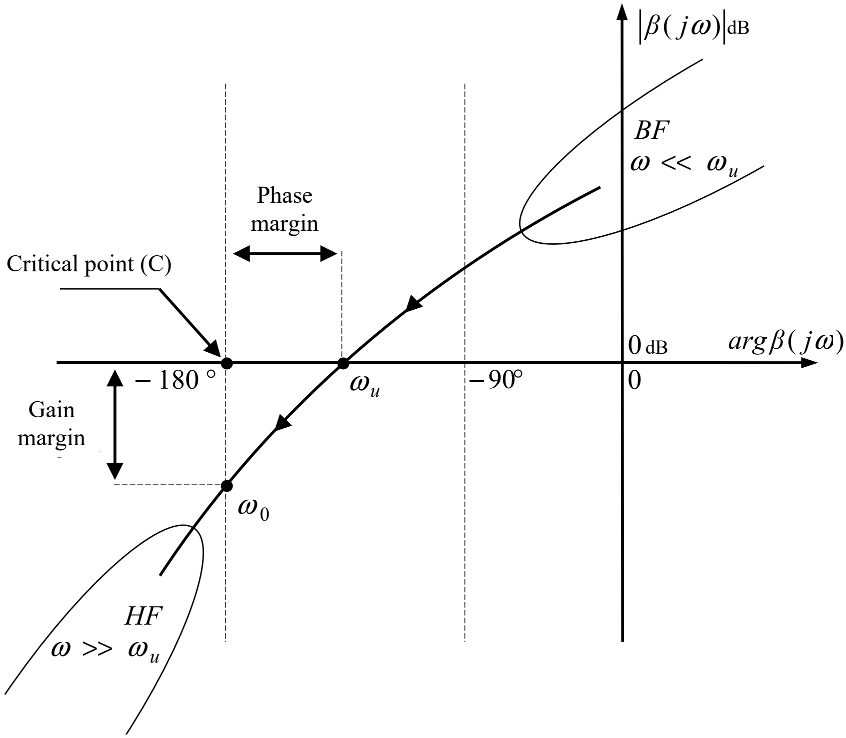

Figure 1.9. *Open-loop Nichols locus said to be of elementary form and associated characteristic quantities*

The form thus defined is widely used in qualitative analysis and reasoning in this field, enabling simple illustration, in open loop, of a certain number of *quantities (including characteristic frequencies)*:

– the *frequency at a phase of –180°*, ω_0, such that $\arg \beta(j\omega_0) = -180°$;

– the *unit gain frequency* (or *transition frequency*), ω_u, such that $|\beta(j\omega_u)| = 1$, that is, $|\beta(j\omega_u)|$ dB = 0;

– the *low frequency area* ($\omega << \omega_u$), for which $|\beta(j\omega)| >> 1$;

– the *high frequency area* ($\omega >> \omega_u$), for which $|\beta(j\omega)| << 1$;

– the *critical point* (0dB, –180°), which corresponds to the *stability limit* of the control (the control barely oscillates if the locus of $\beta(j\omega)$ passes through this point);

– the gain and phase distances to the critical point, which define the *stability margins (gain and phase margins)*, with their sense conditioned by the monotony of the Nichols locus around the critical point (this monotony characterizing the elementary form Nichols locus); indeed, the stability margins, Φ_m and $G_m\mathrm{dB}$, are no longer significant to the distance to the critical point of the Nichols locus of $\beta(j\omega)$, when the latter presents one or more local resonances around frequencies ω_u and ω_0 (see Figure 1.10).

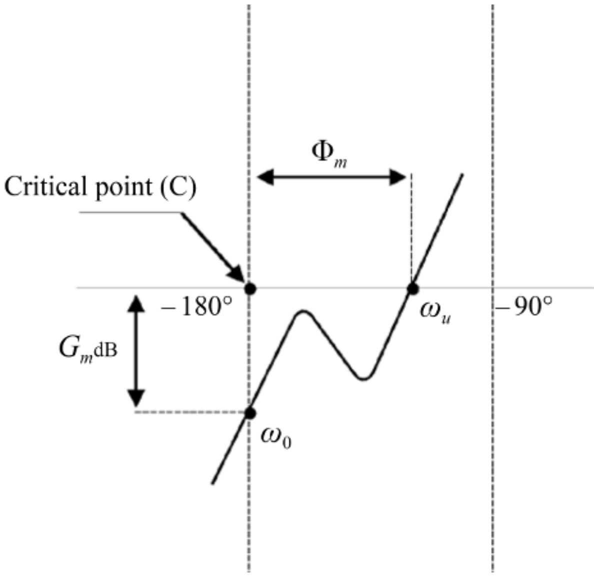

Figure 1.10. *An example of a frequency configuration in open loop, for which the stability margins no longer have a sense*

1.5.5. *Left-hand criterion in the Nichols plane*

1.5.5.1. *Presentation*

The left-hand *criterion* is a *stability criterion* that, in the Nichols plane, is presented as follows: if, when sweeping the open-loop Nichols locus in the increasing ω direction, we keep the critical point, $C\,(0\mathrm{dB}, -180°)$, on the right, the control loop is stable; it is unstable if point C is kept on the left; and finally, it barely oscillates if the Nichols locus of $\beta(j\omega)$ passes through point C.

1.5.5.2. *Demonstration*

The demonstration of the left-hand criterion, as performed here, results from a *physical study* of the stability.

According to the definition of the frequency response in open loop, that is (for $R(j\omega) = D(j\omega) = 0$)

$$\beta(j\omega) = -\frac{V_s(j\omega)}{V_e(j\omega)}, \qquad [1.17]$$

at frequency ω_0, we can write:

$$\beta(j\omega_0) = -\frac{V_s(j\omega_0)}{V_e(j\omega_0)} = -1\times\frac{V_s(j\omega_0)}{V_e(j\omega_0)}, \qquad [1.18]$$

hence:

$$\arg\beta(j\omega_0) = \arg(-1) + \arg\frac{V_s(j\omega_0)}{V_e(j\omega_0)}, \qquad [1.19]$$

that is:

$$\arg\beta(j\omega_0) = -180° + \arg\frac{V_s(j\omega_0)}{V_e(j\omega_0)}. \qquad [1.20]$$

Yet, $\arg\beta(j\omega_0) = -180°$ given the definition of the frequency, ω_0, therefore:

$$\arg\frac{V_s(j\omega_0)}{V_e(j\omega_0)} = 0, \qquad [1.21]$$

which expresses the fact that the input and output sinusoidal signals of the open loop, $v_e(t)$ and $v_s(t)$, vibrate in phase at frequency ω_0, that is $v_e(t) = v_{e_m}\cos(\omega_0 t + \phi)$ and $v_s(t) = v_{s_m}\cos(\omega_0 t + \phi)$; in other words, the output, $v_s(t)$, of the open loop then vibrates in phase with its input, $v_e(t)$; in the declination of the following three scenarios, the input, $v_e(t)$, is assumed to be externally generated (before closing the loop).

With regard to the ratio of amplitudes v_{s_m} and v_{e_m} of $v_s(t)$ and $v_e(t)$, given by the gain

$$\left|\frac{V_s(j\omega_0)}{V_e(j\omega_0)}\right| = |\beta(j\omega_0)|, \qquad [1.22]$$

let us consider the following three scenarios.

Note that $|\beta(j\omega_0)| < 1$, that is, $|\beta(j\omega_0)|\,\text{dB} < 0$, therefore $v_{s_m} < v_{e_m}$; hence, after closing the loop, there is a *convergent oscillation phenomenon* (converging toward zero) after passing through the loop several times, thus demonstrating the control *stability*. Also, $|\beta(j\omega_0)| = 1$, that is, $|\beta(j\omega_0)|\,\text{dB} = 0$, therefore $v_{s_m} = v_{e_m}$; hence, after closing the loop, there is a *self-sustained oscillation phenomenon*, after the first passage through the loop, thus demonstrating the *barely oscillating nature* of the control (the output amplitude of which remains within the linear domain). Finally, $|\beta(j\omega_0)| > 1$, that is, $|\beta(j\omega_0)|\,\text{dB} > 0$, therefore $v_{s_m} > v_{e_m}$; hence, after closing the loop, there is a *divergent oscillation phenomenon* (diverging linearly to infinity) after passing through the loop several times, thus demonstrating the *instability* of the control.

First remark: concerning the three scenarios considered, passing from open loop to closed loop by closing the loop is an operation stemming from thought that through this operation, the output oscillation of the open loop in response to the input oscillation can replace the input oscillation initially applied and then be processed by the loop.

Second remark: concerning the second scenario, it should be emphasized that the *oscillation condition*

$$|\beta(j\omega_0)| = 1 \text{ and } \arg\beta(j\omega_0) = -180°, \qquad [1.23]$$

that is,

$$\beta(j\omega_0) = -1, \qquad [1.24]$$

is in accordance with the oscillation condition defined by the *Barkhausen criterion* (which is best known to electronics engineers), that is,

$$\frac{V_s(j\omega_0)}{V_e(j\omega_0)} = 1, \qquad [1.25]$$

this compliance originating from the relationship:

$$\beta(j\omega_0) = -\frac{V_s(j\omega_0)}{V_e(j\omega_0)}. \quad [1.26]$$

1.5.5.3. *Regarding the spontaneous generation of a self-sustained oscillation*

1.5.5.3.1. Discussion of recorded phenomena

The explanation of the three oscillating phenomena as defined is based:

– on a sinusoidal oscillation, $v_e(t) = v_{e_m} \cos(\omega_0 t + \phi)$, applied at the input of the open loop;

– then on the comparison, at amplitude v_{e_m}, of amplitude v_{s_m} of the sinusoidal oscillation, $v_s(t) = v_{s_m} \cos(\omega_0 t + \phi)$, collected at the output of the open loop;

– with the result of this comparison ($v_{s_m} < v_{e_m}$, $v_{s_m} = v_{e_m}$ or $v_{s_m} > v_{e_m}$) at the origin of the behavior of the control after closing the loop.

While this explanation, which falls within the *context of the loop being opened then closed*, certainly presents the benefit of setting out how the oscillatory phenomena in question manifest themselves in simple terms, this should not lead to the conclusion that these phenomena only arise in such a context (or, more simply, that these phenomena are conditioned by this context).

It is true to say that in reality, we are aware that these phenomena manifest themselves spontaneously, without the loop being opened, therefore within an *exclusively closed-loop context* where no sinusoidal oscillation of frequency ω_0 is applied.

1.5.5.3.2. A legitimate question

So what is the reality, and in particular, what can we say about the spontaneous generation of a self-sustained oscillation (or “self-oscillation”)?

More precisely, how can an oscillator that itself generates a sinusoidal oscillation, of frequency ω_0, become excited? This is certainly a legitimate question, which raises the well-known problem of *oscillator excitation*.

1.5.5.3.3. The answer to this question

The answer to the problem of oscillator excitation indisputably lies in the *noise*, which comprises all of the spectral components, and in particular, that of frequency

ω_0, which, being alone in responding to the oscillation condition, will be alone in giving rise to a self-sustained oscillation.

1.5.5.3.4. Stabilization of the self-sustained oscillation amplitude

To set the amplitude, y_m, of the output oscillation, $y(t) = y_m \cos(\omega_0 t + \phi)$, to a desired value, y_{m_d}, it is necessary for the gain in open loop at frequency ω_0, $|\beta(j\omega_0)|$, to be a function of y_m, such that (see Figure 1.11):

$$|\beta(j\omega_0)| > 1 \text{ for } y_m < y_{m_d}, \qquad [1.27]$$

divergent oscillation phenomenon in transient state;

$$|\beta(j\omega_0)| = 1 \text{ for } y_m = y_{m_d}, \qquad [1.28]$$

self-sustained oscillation phenomenon (or barely oscillating phenomenon) in steady state;

$$|\beta(j\omega_0)| < 1 \text{ for } y_m > y_{m_d}, \qquad [1.29]$$

convergent oscillation phenomenon that gives rise to nothing in steady state.

To this end, it is appropriate to introduce a *nonlinear amplification* into the loop, notably via a nonlinear amplifier incorporated into the controller. Although this amplifier is *linear* for a given amplitude of the input or output, with the output indeed responding sinusoidally to a sinusoidal input, it is *nonlinear* for a variation of the amplitude: indeed, while the input amplitude varies by a given factor, the output amplitude does not vary in the same ratio, even if the output still varies sinusoidally.

An amplifier of this kind serves as an example of an achievement, equipping a *sinusoidal oscillator of order 5/2*, as mentioned in Oustaloup [OUS 83b]. This example is based on an operational amplifier and a *field-effect transistor*, used in variable resistance between drain and source, and controlled by the gate by a negative direct voltage, proportional to the amplitude of the output signal.

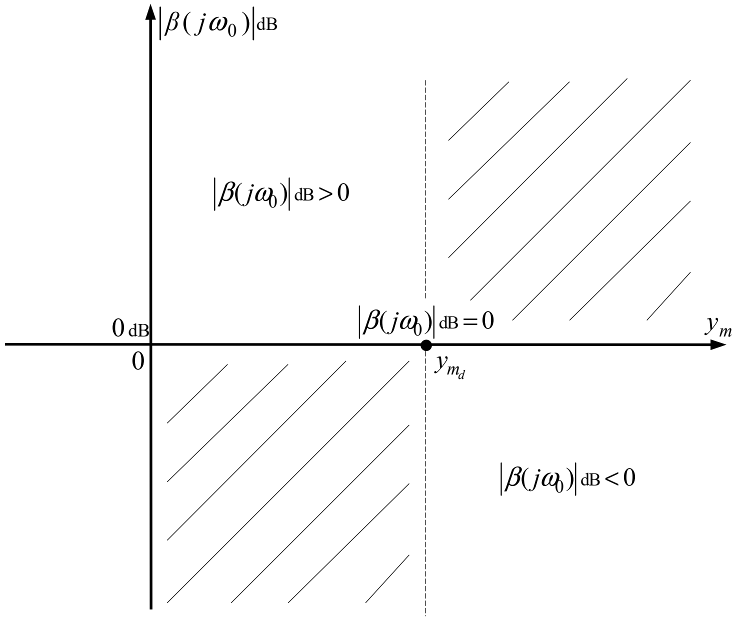

Figure 1.11. *The non-hatched areas define those areas where $|\beta(j\omega_0)|_{dB}$ belongs, enabling y_m to be set at y_{m_d}*

1.6. Closed-loop transmittances

The block diagram in Figure 1.4 enables the Laplace transform of the output, $y(t)$, to be expressed, that is:

$$Y(s) = D(s) + G(s)C(s)\left[K_1 R(s) - K_2 Y(s)\right], \qquad [1.30]$$

or, by gathering the terms in $Y(s)$:

$$Y(s)\left[1 + C(s)G(s)K_2\right] = K_1 C(s)G(s)R(s) + D(s), \qquad [1.31]$$

from which we draw the general expression of the output, $Y(s)$, as a function of stresses $R(s)$ and $D(s)$, that is:

$$Y(s) = K\frac{\beta(s)}{1+\beta(s)}R(s) + \frac{D(s)}{1+\beta(s)}, \qquad [1.32]$$

forming the ratio

$$K = \frac{K_1}{K_2} \qquad [1.33]$$

and using the relationship established above,

$$\beta(s) = C(s)G(s)K_2. \qquad [1.34]$$

The additive nature of the general expression of $Y(s)$ defined by the relationship [1.32] leads to the definition of two transmittances characterizing the influence of the stresses on the output, one relative to $R(s)$ and the other relative to $D(s)$.

1.6.1. *Transmittance in tracking*

1.6.1.1. *Definition and general form*

The *transmittance relating to the reference* (or *transmittance in tracking*) is defined by the ratio:

$$F_r(s) = \left[\frac{Y(s)}{R(s)}\right]_{D(s)=0}, \qquad [1.35]$$

that is, taking [1.32] into account:

$$F_r(s) = K\frac{\beta(s)}{1+\beta(s)}. \qquad [1.36]$$

The general form of expression of $F_r(s)$ thus obtained suggests the definition of a normalized transmittance, that is,

$$\mathbb{F}_r(s) = \frac{F_r(s)}{K} = \frac{\beta(s)}{1+\beta(s)}, \qquad [1.37]$$

known as *reduced transmittance in tracking* or *complementary sensitivity* (noted T or $\overline{S}$ in contemporary literature).

1.6.1.2. *Tracking function*

Transmittance in tracking, $F_r(s)$, characterizes the *tracking function*, which reflects the control loop (output)'s ability to *track* the reference (of course up to a factor K).

The performances of the tracking function can be simulated from the transmittance in open loop, $\beta(s)$, in accordance with the block diagram in Figure 1.12, which represents the cascaded transfer, K, with a *unit-reaction loop*, the action chain of which is formed by $\beta(s)$.

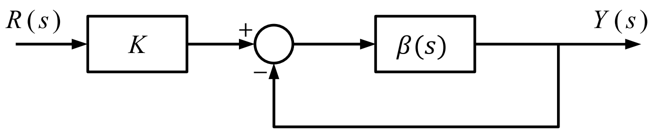

Figure 1.12. *Functional representation of* $F_r(s) = K\beta(s)\left(1+\beta(s)\right)^{-1}$

1.6.2. *Transmittance in regulation*

1.6.2.1. *Definition and general form*

*The transmittance relating to the disturbance (*or *transmittance in regulation)* is defined by the ratio:

$$F_d(s) = \left[\frac{Y(s)}{D(s)}\right]_{R(s)=0}, \qquad [1.38]$$

that is, taking [1.32] into account:

$$F_d(s) = \frac{1}{1+\beta(s)}. \qquad [1.39]$$

The general form of expression of $F_d(s)$ thus obtained is well known under the designation, that is, *sensitivity* (noted S in contemporary literature).

Combining relationships [1.37] and [1.39] gives:

$$F_r(s) + F_d(s) = 1. \qquad [1.40]$$

That is, given the existing notations ($F_r(s) = T = \overline{S}$ and $F_d(s) = S$):

$$T + S = 1, \qquad [1.41]$$

or

$$\overline{S} + S = 1, \qquad [1.42]$$

with these expressions reflecting that $T = \overline{S}$ is the complement to 1 of sensitivity, S, thus justifying the designation, "complementary sensitivity", attributed to T.

1.6.2.2. *Regulation function*

Transmittance in regulation, $F_d(s)$, characterizes the *regulation function*, which reflects the control loop (output)'s ability to *not track* (or *reject*) the disturbance.

The performances of the regulation function can be simulated from the transmittance in open loop, $\beta(s)$, in accordance with the block diagram in Figure 1.13, which directly represents a *unit-action loop*, the reaction chain of which is formed by $\beta(s)$.

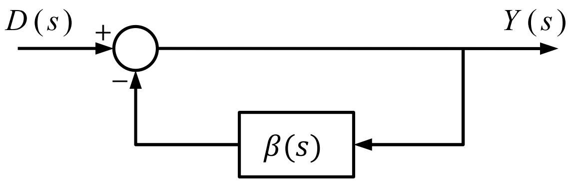

Figure 1.13. *Functional representation of* $F_d(s) = (1 + \beta(s))^{-1}$

1.6.3. *A tracking-regulation dilemma?*

1.6.3.1. *One question*

Given the interest of identifying dilemmas to best manage performance compromises, in the study of closed-loop behavior, is *tracking-regulation duality* not likely to raise questions as to whether or not a tracking-regulation dilemma exists? It is precisely this question of circumstance that forms the basis of the following section.

1.6.3.2. *A well-reasoned response*

At frequencies below ω_u, for which the open-loop gain is greater than one, that is, $|\beta(j\omega)| > 1$, any increase in this gain jointly ensures both:

– a variation toward K of $F_r(j\omega) = K \dfrac{\beta(j\omega)}{1+\beta(j\omega)}$;

– a variation toward zero of $F_d(j\omega) = \dfrac{1}{1+\beta(j\omega)}$.

These two concomittant effects indeed express the joint improvement of:

– the *reference tracking*, via the variation toward K of the frequency response in tracking, $F_r(j\omega)$;

– the *disturbance rejection*, via the variation toward zero of the frequency response in regulation, $F_d(j\omega)$.

Thus, there is an improvement to the reference tracking, accompanied by an improvement to the disturbance rejection, and, conversely, *tracking and regulation are not contradictory requirements*, which excludes the existence of any dilemma formulated in terms of tracking and regulation.

1.7. Input sensitivity

The *(plant) input sensitivity* characterizes the effect of stresses on the plant input.

In tracking, it is defined by the gain:

$$\mathbb{S}_r(\omega) = \left|\frac{U(j\omega)}{R(j\omega)}\right|_{D(j\omega)=0}, \qquad [1.43]$$

and, in regulation, by the gain:

$$\mathbb{S}_d(\omega) = \left|\frac{U(j\omega)}{D(j\omega)}\right|_{R(j\omega)=0}. \qquad [1.44]$$

The gains formulated are indeed significant of the plant input sensitivity to the stresses that are the reference and the disturbance.

1.7.1. *Input sensitivity in tracking*

For $D(j\omega) = 0$, the block diagram in Figure 1.4 enables us to write:

$$U(j\omega) = C(j\omega)\left[K_1 R(j\omega) - K_2 G(j\omega) U(j\omega)\right], \qquad [1.45]$$

that is:

$$U(j\omega)\left[1 + K_2 C(j\omega) G(j\omega)\right] = K_1 C(j\omega) R(j\omega), \qquad [1.46]$$

from which we obtain the input sensitivity in tracking:

$$\mathbb{S}_r(\omega) = \left| \frac{K_1 C(j\omega)}{1 + \beta(j\omega)} \right|, \qquad [1.47]$$

with $1/\mathbb{S}_r(\omega)$ defining the *input immunity in tracking*.

1.7.2. *Input sensitivity in regulation*

For $R(j\omega) = 0$, the block diagram in Figure 1.4 enables us to write:

$$U(j\omega) = C(j\omega)\left(-K_2\left[D(j\omega) + G(j\omega) U(j\omega)\right]\right), \qquad [1.48]$$

that is:

$$U(j\omega)\left[1 + K_2 C(j\omega) G(j\omega)\right] = -K_2 C(j\omega) D(j\omega), \qquad [1.49]$$

from which we obtain the *input sensitivity in regulation*:

$$\mathbb{S}_d(\omega) = \left| \frac{K_2 C(j\omega)}{1 + \beta(j\omega)} \right|, \qquad [1.50]$$

with $1/\mathbb{S}_d(\omega)$ defining the *input immunity in regulation*.

1.7.3. *Input sensitivity in tracking and regulation*

In the case of the elementary control loop ($K_1 = K_2 = 1$), the expressions [1.47] and [1.50] identify, respectively, with the same expression, which defines the *input sensitivity in tracking and regulation*, that is:

$$\mathcal{S}(\omega) = \mathcal{S}_r(\omega) = \mathcal{S}_d(\omega) = \left|\frac{C(j\omega)}{1+\beta(j\omega)}\right|, \qquad [1.51]$$

with $1/\mathcal{S}(\omega)$ defining the *input immunity in tracking and regulation.*

With ω_u designating the unit gain frequency in open loop, $\left(\left|\beta(j\omega_u)\right| = 1\right)$, with for $\omega/\omega_u << 1$, $\left|\beta(j\omega)\right|$ then being large with respect to the unit, $\mathcal{S}(\omega)$ is reduced to:

$$\mathcal{S}(\omega) = \left|\frac{C(j\omega)}{\beta(j\omega)}\right| = \left|\frac{C(j\omega)}{C(j\omega)G(j\omega)}\right| = \frac{1}{\left|G(j\omega)\right|}, \qquad [1.52]$$

that is, in decibels (see Figure 1.14):

$$\mathcal{S}(\omega)\text{dB} = -20\log\left|G(j\omega)\right| = -\left|G(j\omega)\right|\text{dB} \,. \qquad [1.53]$$

For $\omega/\omega_u >> 1$, $\left|\beta(j\omega)\right|$ then being small with respect to the unit, $\mathcal{S}(\omega)$ is reduced to:

$$\mathcal{S}(\omega) = \left|C(j\omega)\right|, \qquad [1.54]$$

that is, in decibels (see Figure 1.14):

$$\mathcal{S}(\omega)\text{dB} = 20\log\left|C(j\omega)\right| = \left|C(j\omega)\right|\text{dB} \,. \qquad [1.55]$$

The gain diagrams relative to [1.53] and [1.55] intersect at frequency ω_u (see Figure 1.14), because from the relationship,

$$\left|C(j\omega_u)\right|\left|G(j\omega_u)\right| = 1, \qquad [1.56]$$

that is,

$$\left|C(j\omega_u)\right|\text{dB} + \left|G(j\omega_u)\right|\text{dB} = 0\,, \qquad [1.57]$$

we obtain:

$$\left|C(j\omega_u)\right|\text{dB} = -\left|G(j\omega_u)\right|\text{dB}\,. \qquad [1.58]$$

The shape of the curve of $-\left|G(j\omega)\right|\text{dB}$ respects the low-pass character of the plant, $G(j\omega)$. With regard to the shape of the curve of $\left|C(j\omega)\right|\text{dB}$, it observes the gain increase with the frequency that accompanies the phase lead of the regulator, $C(j\omega)$, around the frequency, ω_u.

When the frequency, ω_u, increases, the gain diagram of the regulator rises in accordance with the displacement of the intersection point, a rise which, beyond ω_u, is accentuated by the joint growth of the gain slope of the regulator at frequency, ω_u, thus reflecting that the high-frequency input sensitivity increases with ω_u. The growth (with ω_u) of the gain slope of the regulator is explained by the growth (with ω_u) of the regulator phase lead, which must compensate for the growth (with ω_u) of the plant phase-lag to maintain the same phase margin of the control.

Looking to Figure 1.14, it appears that, as a first approximation, the input sensitivity in tracking and regulation is:

– at low frequencies ($\omega << \omega_u$), given exclusively by the *inverse of the plant gain* (thus expressing that the low-frequency input sensitivity does not constitute a constraint for the regulator synthesis);

– at medium frequencies (around ω_u), a function of both the plant and the regulator;

– at high frequencies ($\omega >> \omega_u$), given exclusively by the *regulator gain* (thus expressing that the high-frequency input sensitivity constitutes, in this case, a constraint for the regulator synthesis).

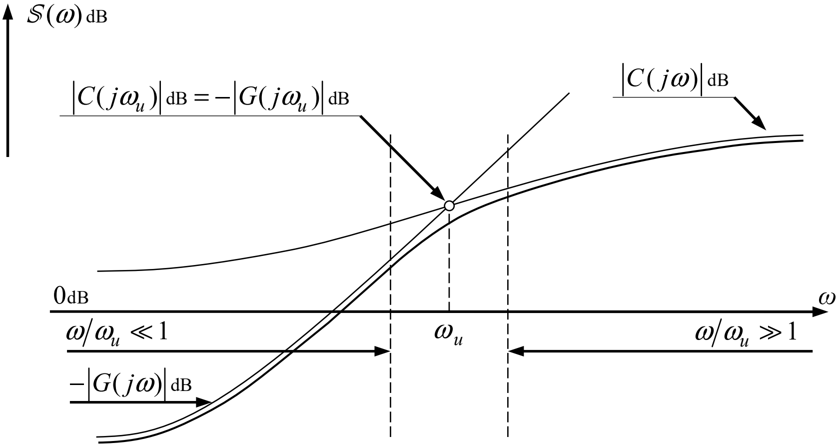

Figure 1.14. *Frequency variation (in decibels) of the input sensitivity in tracking and regulation*

In other words, in the regulator synthesis, *the high-frequency regulator gain needs to be determined with utmost attention*. Notably, it must not tend toward infinity with the frequency, to ensure that the regulator does not produce a high-frequency noise amplifier. Therefore, a "red card" can be attributed to the *academic PID*, whose high-frequency gain tends toward infinity with frequency. The following presentation of this PID offers a brief, simple explanation of this problem and introduces the *non-academic PID* as a solution.

1.7.3.1. *Academic PID*

The "academic" PID controller is defined by the elementary transmittance, in additive form:

$$PID(s) = k + \frac{\omega'_i}{s} + \frac{s}{\omega'_d}, \qquad [1.59]$$

while the three parameters, *k*, ω'_i and ω'_d, are defined as follows:

k : transfer factor of the *proportional action*;

ω'_i : transition (or unit gain) frequency of the *integral action*, ω'_i/s;

ω'_d : transition (or unit gain) frequency of the *differential action*, s/ω'_d.

By factoring as suggested by the relationship [1.59], the controller transmittance can be rewritten in the form:

$$PID(s) = k\left(1 + \frac{\omega'_i}{ks} + \frac{s}{k\omega'_d}\right), \quad [1.60]$$

that is, by writing $\omega_i = \omega'_i / k$ and $\omega_d = k\omega'_d$:

$$PID(s) = k\left(1 + \frac{\omega_i}{s} + \frac{s}{\omega_d}\right), \quad [1.61]$$

from which we obtain the frequency response:

$$PID(j\omega) = k\left(1 + \frac{\omega_i}{j\omega} + j\frac{\omega}{\omega_d}\right), \quad [1.62]$$

a form that calls for the *reduced frequency response*:

$$PID(j\omega)/k = 1 + \frac{\omega_i}{j\omega} + j\frac{\omega}{\omega_d} . \quad [1.63]$$

The new transition frequencies, ω_i and ω_d, such that $\omega_i < \omega_d$, lead to the frequency space being divided into three parts, defined, respectively, by:

$$\omega < \omega_i ,\ \omega_i < \omega < \omega_d \text{ and } \omega > \omega_d , \quad [1.64]$$

with these three sub-spaces enabling the following reductions, which are at the origin of the Bode asymptotic diagrams of $PID(j\omega)/k$.

For $\omega << \omega_i$, and therefore for $\omega << \omega_d$, which expresses $\omega_i / \omega >> 1$ and $\omega/\omega_d << 1$, $PID(j\omega)/k$ is reduced to:

$$PID(j\omega)/k = \frac{\omega_i}{j\omega} , \quad [1.65]$$

an integral action of transition frequency, ω_i .

For $\omega >> \omega_d$, and therefore for $\omega >> \omega_i$, which expresses $\omega/\omega_d >> 1$ and $\omega_i/\omega << 1$, $PID(j\omega)/k$ is reduced to:

$$PID(j\omega)/k = j\frac{\omega}{\omega_d}, \qquad [1.66]$$

a differential action of transition frequency, ω_d.

Finally, for $\omega_i << \omega << \omega_d$, which expresses $\omega_i/\omega << 1$ and $\omega/\omega_d << 1$, $PID(j\omega)/k$ is reduced to:

$$PID(j\omega)/k = 1, \qquad [1.67]$$

a unitary, proportional action.

These three reductions determine the Bode asymptotic diagrams of $PID(j\omega)/k$, as presented by Figure 1.15.

The Bode asymptotic diagrams thus obtained can be interpreted as resulting from the cascading of:

– a frequency-response *proportional-integral controller*

$$PI(j\omega) = 1 + \frac{\omega_i}{j\omega} \qquad [1.68]$$

– a frequency-response *proportional-differential controller*

$$PD(j\omega) = 1 + j\frac{\omega}{\omega_d}, \qquad [1.69]$$

with the characteristic frequencies ω_i and ω_d then being the transitional frequencies of the PI and PD controllers.

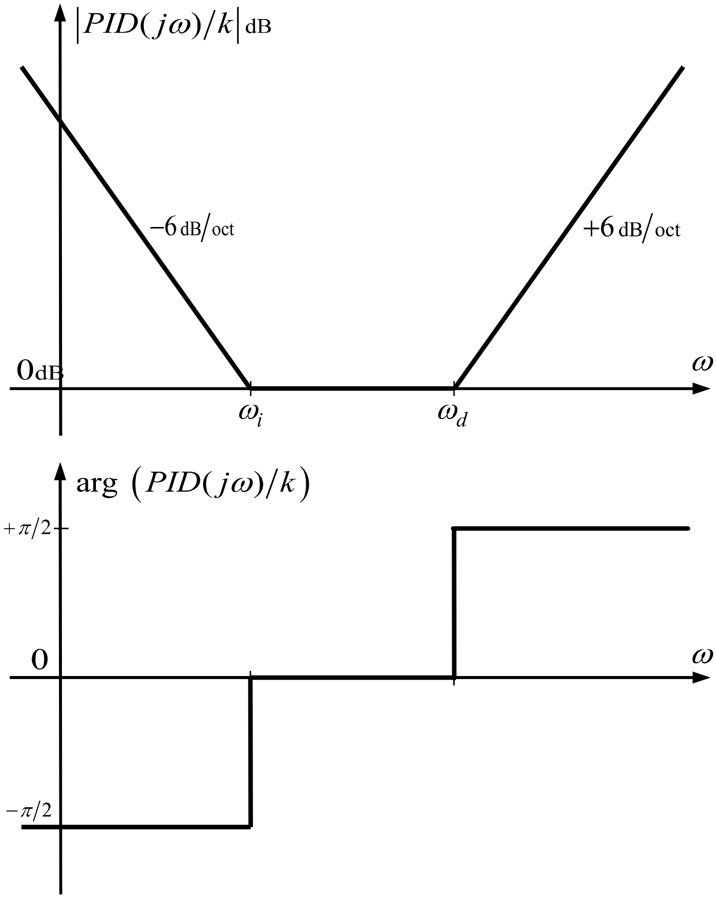

Figure 1.15. *Bode asymptotic diagrams of the academic PID*

Indeed, the product of these frequency responses, that is,

$$PI(j\omega)PD(j\omega) = \left(1 + \frac{\omega_i}{j\omega}\right)\left(1 + j\frac{\omega}{\omega_d}\right), \qquad [1.70]$$

admits, as a development:

$$PI(j\omega)PD(j\omega) = 1 + \frac{\omega_i}{j\omega} + j\frac{\omega}{\omega_d} + \frac{\omega_i}{\omega_d}, \qquad [1.71]$$

or, for $\omega_i / \omega_d << 1$:

$$PI(j\omega)PD(j\omega) = 1 + \frac{\omega_i}{j\omega} + j\frac{\omega}{\omega_d}, \qquad [1.72]$$

or, taking [1.63] into account:

$$PI(j\omega)PD(j\omega) = PID(j\omega)/k, \qquad [1.73]$$

or indeed:

$$k\, PI(j\omega)PD(j\omega) = PID(j\omega). \qquad [1.74]$$

Thus, for a significant ratio between the characteristic frequencies, ω_d and ω_i $\left(\omega_d / \omega_i >> 1\right)$, the *multiplicative form* of the PID controller, that is,

$$k\left(1 + \frac{\omega_i}{j\omega}\right)\left(1 + j\frac{\omega}{\omega_d}\right), \qquad [1.75]$$

is equivalent to its *additive form*, that is:

$$k\left(1 + \frac{\omega_i}{j\omega} + j\frac{\omega}{\omega_d}\right). \qquad [1.76]$$

It is then possible to make use of this *equivalence* to retain and make the best use of the multiplicative form, which is preferable to the additive form, owing to the computational advantages it offers and its closer correspondence to Bode asymptotic diagrams when the characteristic frequencies, ω_d and ω_i approach, or even identify with oneanother, this being confirmed by the computations conducted for $\omega_i = \omega_d = \omega_{id}$ and $\omega = \omega_{id}$.

1.7.3.2. *Non-academic PID*

In the case of the academic PID controller, the input sensitivity at high frequencies, given by the controller gain at these frequencies, admits an expression of the form:

$$\underset{\omega \gg \omega_d}{\mathcal{S}(\omega)} = \underset{\omega \gg \omega_d}{\left|PID(j\omega)\right|} = k\frac{\omega}{\omega_d}, \qquad [1.77]$$

a result that is obtained from the multiplicative or additive form of the controller, and that clearly demonstrates that through the controller gain for $\omega >> \omega_d$, the high-frequency input sensitivity tends to infinity with the frequency.

The solution to this problem thus consists of *truncating the differential action of the academic PID* using a first-order low-pass filter, the transitional frequency of which, ω_t, determines the *truncation frequency of the differential action*. The cascading of this low-pass filter with the multiplicative form of the academic PID is then reflected by the following frequency response, which defines the non-academic PID controller:

$$PID(j\omega) = k\left(1 + \frac{\omega_i}{j\omega}\right)\frac{1 + j\dfrac{\omega}{\omega_d}}{1 + j\dfrac{\omega}{\omega_t}}, \qquad [1.78]$$

with $\omega_i < \omega_d < \omega_t$, the Bode asymptotic diagrams of $PID(j\omega)/k$ being presented in Figure 1.16.

The frequency response thus obtained can be interpreted as resulting from the cascading of:

– a *proportional controller* of gain k;

– a frequency-response *proportional-integral controller*

$$1 + \frac{\omega_i}{j\omega}; \qquad [1.79]$$

– a frequency-response *phase-lead controller* (see Chapter 2)

$$\frac{1 + j\dfrac{\omega}{\omega_d}}{1 + j\dfrac{\omega}{\omega_t}}. \qquad [1.80]$$

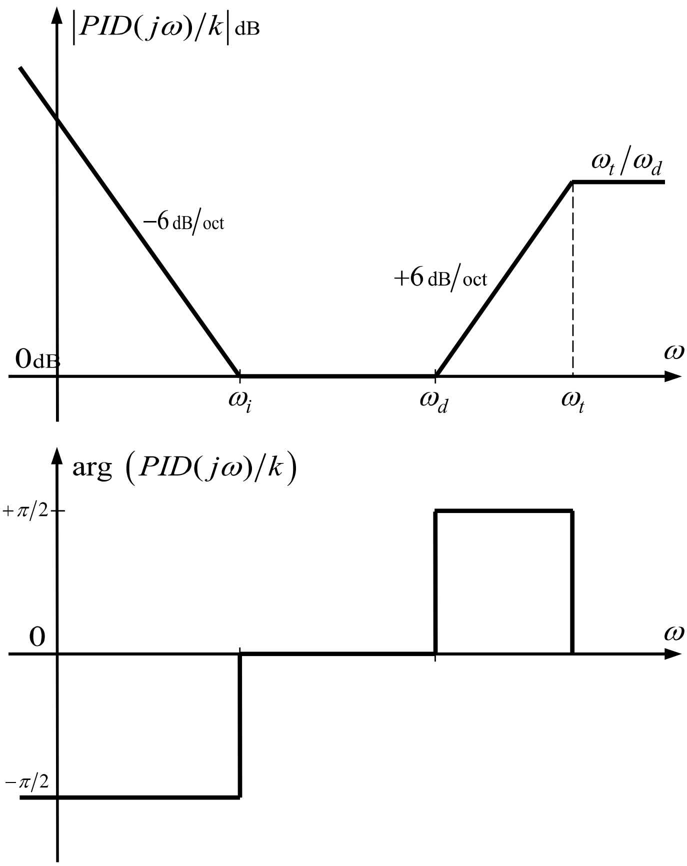

Figure 1.16. *Bode asymptotic diagrams of the non-academic PID*

The input sensitivity at high frequencies, given by the controller gain at these frequencies, then admits a new expression of the form:

$$\underset{\omega \gg \omega_t}{\mathbb{S}(\omega)} = \underset{\omega \gg \omega_t}{\left|PID(j\omega)\right|} = k\,\frac{\omega_t}{\omega_d}\,, \qquad [1.81]$$

this result demonstrating that the high-frequency input sensitivity is now reduced to a constant (in this case $k\,\omega_t/\omega_d$).

1.8. Behavior and frequency performances in tracking and regulation

By way of introduction, in order to fully understand the beginning of the next section, it is important to present the stability degree–precision dilemma, and then specify the principle of bringing it into question, which ensures the robustness of the control when the parametric variations of the plant lead only to open-loop gain variations around the frequency, ω_u.

The stability degree–precision dilemma expresses that any increase in the open-loop gain, aiming to improve precision, leads to a decrease in the stability degree by decreasing the stability margins (significant of the distance to the critical point).

Its calling into question is ensured by an open-loop Nichols locus, which forms, around frequency ω_u, a vertical-line segment, known as the template, which is sufficiently long, with an abscissa between $-\pi/2$ and $-\pi$ (see Figure 1.17): indeed, this straight-line segment, which slides on itself in the event of any gain variation in open loop, ensures the invariance of the phase margin, which alone signifies the distance to the critical point (since it is the smallest stability margin owing to the sufficient length of the template).

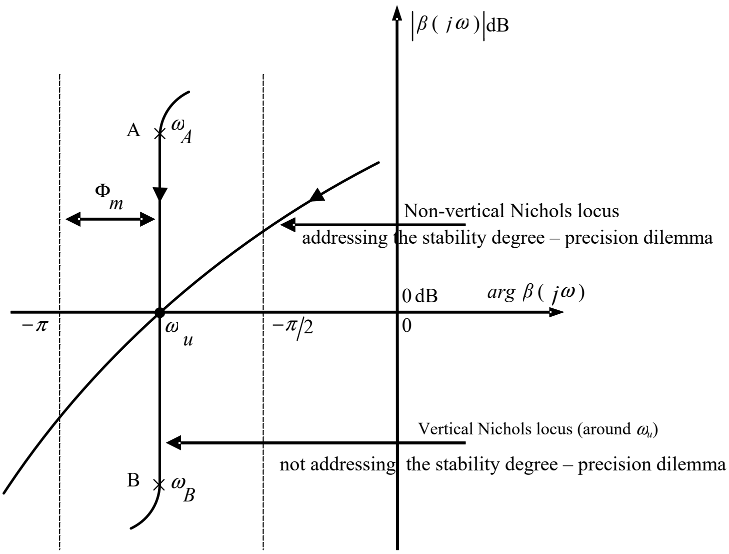

Figure 1.17. *Calling into question of the stability degree–precision dilemma by the (vertical) template of the second-generation CRONE control, and this template being described by:* $\beta(j\omega) = \left(\frac{\omega_u}{j\omega}\right)^n$ *with* $1 < n < 2$ *and* $\omega_A \leq \omega \leq \omega_B$

1.8.1. *Frequency and time behavior*

If in the second-generation CRONE control (see Chapter 5), we consider a template extending across the entire frequency domain, we obtain a frequency response in open loop, $\beta(j\omega)$, which identifies a non-integer integrator, that is:

$$\beta(j\omega) = \left(\frac{\omega_u}{j\omega}\right)^n \text{ with } 1 < n < 2 \qquad [1.82]$$

The general expressions of the closed-loop frequency responses, directly deduced from [1.36] and [1.39], that is:

$$F_r(j\omega) = K\frac{\beta(j\omega)}{1+\beta(j\omega)} \qquad [1.83]$$

and

$$F_d(j\omega) = \frac{1}{1+\beta(j\omega)}, \qquad [1.84]$$

can then be particularized according to the relationships:

$$F_r(j\omega) = \frac{K}{1+\left(j\dfrac{\omega}{\omega_u}\right)^n} \qquad [1.85]$$

and

$$F_d(j\omega) = \frac{\left(j\dfrac{\omega}{\omega_u}\right)^n}{1+\left(j\dfrac{\omega}{\omega_u}\right)^n}. \qquad [1.86]$$

The corresponding gain asymptotic diagrams, represented in Figure 1.18, reflect that:

– first, *the control loop behaves like a low-pass filter in tracking and a high-pass filter in regulation*;

– second, $F_r(j\omega)$ *and* $F_d(j\omega)$ *have the same transitional frequency*, in this case ω_u, expressing that (temporally*) the dynamics rapidity is the same in tracking and regulation*.

The *open-loop transition frequency*, ω_u, therefore determines the *closed-loop transitional frequency*, that is, the *transitional frequency in tracking and regulation*. The existence of the same transitional frequency in both tracking and regulation expresses (frequentially) the identity of the *transmission band in tracking* and the *rejection band in regulation*; these two bands can be referred to using a common designation, namely the *(control) transmission and rejection band*, because, for the control, this frequency band is both a *transmission band* (implied to be in tracking) and a *rejection band* (implied to be in regulation).

The gain asymptotic diagrams in Figure 1.18 suggest a comparative analysis of the corresponding temporal behaviors to be conducted via the unit-step responses in tracking and regulation, $y_r(t)$ and $y_d(t)$.

Concerning the reference or disturbance unit step:

– the *instantaneous transition* of the step generates infinite frequencies at instant zero; as Figure 1.18 indicates, these frequencies are not transmitted in tracking (therefore $y_r(0^+) = 0$), but are entirely transmitted in regulation (therefore $y_d(0^+) = 1$);

– the step *level*, which becomes a continuous signal long after the transition (which can then be forgotten), generates, at infinite instants, a zero frequency; as shown in Figure 1.18, this frequency is transmitted with the factor K in tracking (therefore $y_r(\infty) = K$), but is not transmitted in regulation (therefore $y_d(\infty) = 0$).

Thus, for $t > 0$ (with t varying from zero to infinity), $y_r(t)$ evolves from 0 to K, while $y_d(t)$ evolves from 1 to 0, these evolutions being compliant with the general shape of the step responses of a low-pass filter and a high-pass filter.

In fact, these results are obtained without computation from the principle drawn from our analysis, according to which: “when the time domain is swept to the right (with time varying from zero to infinity), the frequency domain is swept to the left (with frequency indeed varying from infinity to zero)”.

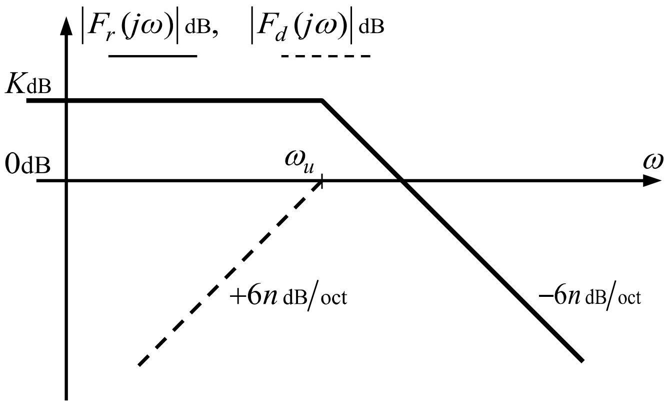

Figure 1.18. *Gain asymptotic diagrams in tracking and regulation*

1.8.2. *Frequency performances*

Given that the control loop presents two very distinct behaviors in tracking and regulation (those of low-pass and high-pass filters), it is useful to distinguish between the frequency performances of $F_r(j\omega)$ and $F_d(j\omega)$, notably the cut-off frequency and the resonance frequency and ratio.

To define these performances in accordance with the dual context of low-pass and high-pass filters, it is appropriate to define beforehand, for each type of filter, the *reference gain*, which is the gain at *a particular frequency, characteristic of the frequencies that the filter is supposed to transmit*. Thus, for a low-pass filter (which is supposed to transmit low frequencies), the reference gain is the *zero-frequency gain*, and for a high-pass filter (which is supposed to transmit high frequencies), the reference gain is the *infinite-frequency gain*: these two gains serve as a reference in the following dual definitions.

1.8.2.1. *Cut-off frequency*

In tracking, the cut-off frequency at –6dB, ω_c, satisfies the equation:

$$|F_r(j\omega_c)| = \frac{1}{2}|F_r(j0)| \qquad [1.87]$$

In regulation, it satisfies the equation:

$$\left|F_d(j\omega_c)\right| = \frac{1}{2}\left|F_d(j\infty)\right| \qquad [1.88]$$

In both cases, the frequency gain, ω_c, is equal to half of the reference gain.

1.8.2.2. *Resonance frequency*

In tracking, the resonance frequency, ω_r, is such that:

$$\left|F_r(j\omega_r)\right| = \left|F_r(j\omega)\right|_{\max}. \qquad [1.89]$$

In regulation, it is such that:

$$\left|F_d(j\omega_r)\right| = \left|F_d(j\omega)\right|_{\max}. \qquad [1.90]$$

In both cases, the frequency gain, ω_r, is equal to the maximum gain.

1.8.2.3. *Resonance ratio*

In tracking, the resonance ratio, Q, is defined by the ratio:

$$Q = \frac{\left|F_r(j\omega_r)\right|}{\left|F_r(j0)\right|} = Q_r. \qquad [1.91]$$

In regulation, it is given by the ratio:

$$Q = \frac{\left|F_d(j\omega_r)\right|}{\left|F_d(j\infty)\right|} = Q_d. \qquad [1.92]$$

In both cases, the resonance ratio is the ratio between the maximum gain and the reference gain.

1.8.3. *Regarding the effect of the increase in frequency ω_u*

When the frequency ω_u increases, with the transmission and rejection band also increasing (see section 1.8.1), the output of the control loop (and therefore of the plant) then varies more rapidly, although the plant bandwidth (low-pass system) has not for its part increased. So what about the plant input?

The answer to this question indisputably lies in the following property (see section 1.8.3.1), a property according to which *the plant input increases with the output rapidity*, this input increase moreover being in accordance with the rise in input sensitivity, as shown in Figure 1.14 when ω_u increases.

1.8.3.1. *A property of low-pass systems*

The faster the output of a low-pass system varies, the more the input is subject to significant stresses. In other words, for a low-pass system (therefore for the plant), *the more rapid the output, the more significant the input.*

1.8.3.2. *An illustration of this property*

One notable illustration of this property is undoubtedly a gamma RC circuit, whose output varies according to a step.

For an elementary plant, let us therefore consider a gamma RC circuit (see Figure 1.19), the input–output model of which is obtained from the electrical equation:

$$u(t) = R\,i(t) + y(t), \qquad [1.93]$$

that is, given that

$$i(t) = C\frac{dy(t)}{dt} \qquad [1.94]$$

$$u(t) = RC\frac{dy(t)}{dt} + y(t)\,, \qquad [1.95]$$

a well-known differential equation describing the dynamic behavior between $u(t)$ and $y(t)$.

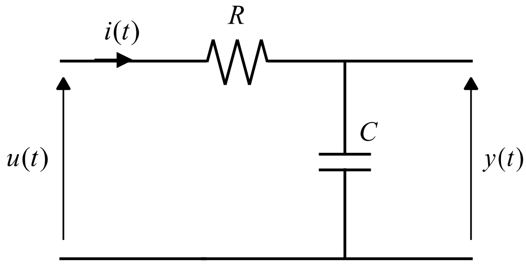

Figure 1.19. *Gamma RC circuit*

The goal is to determine the input, $u(t)$, which ensures a step variation at the output, $y(t)$, with $y(t)$ then presenting a *discontinuity of finite amplitude*, notably unitary for a unit step, that is, $y(t) = 0$ for $t < 0$ and $y(t) = 1$ for $t > 0$.

With $y(t)$ thus being particularized, the differential equation of the RC circuit becomes:

$$u(t) = RC\,\delta(t) + y(t)\,, \qquad [1.96]$$

the presence of the Dirac impulse, $\delta(t)$, expressing that $u(t)$ thus presents a *discontinuity of infinite amplitude*.

1.8.3.3. *A significant conclusion*

For the output of a low-pass system to present a finite amplitude discontinuity, the input must present a discontinuity of infinite amplitude: in other words, owing to the infinitely rapid character of a discontinuity, *an infinitely rapid output of finite amplitude requires an infinitely rapid input of infinite amplitude*.

1.9. Dynamics in tracking and regulation

The general form of transmittances $F_r(s)$ and $F_d(s)$, given by the relationships [1.36] and [1.39], enables us to conduct an immediate comparison of the dynamics in tracking and regulation. Indeed, the general relationship between the step responses in tracking and regulation simply needs to be established, notably by expressing the step response in regulation as a function of the step response in tracking (the latter being better known).

The Laplace transform of the unit step response in regulation, $y_d(t)$, is written as:

$$Y_d(s) = F_d(s)\frac{1}{s} = \frac{1}{1+\beta(s)}\frac{1}{s}\,, \qquad [1.97]$$

that is:

$$Y_d(s) = \left[1 - \frac{\beta(s)}{1+\beta(s)}\right]\frac{1}{s} = \frac{1}{s} - \frac{F_r(s)}{K}\frac{1}{s}\,, \qquad [1.98]$$

or:

$$Y_d(s) = \frac{1}{s} - \frac{1}{K} Y_r(s), \qquad [1.99]$$

with $Y_r(s)$ designating the Laplace transform of the unit step response in tracking, $y_r(t)$.

In the time domain, the symbolic equation [1.99] admits, as an original equation:

$$y_d(t) = u(t) - \frac{1}{K} y_r(t), \qquad [1.100]$$

that is, for $t > 0$ (given that $u(t)$ is then unitary):

$$y_d(t) = 1 - \frac{1}{K} y_r(t). \qquad [1.101]$$

This result expresses that $y_d(t)$ is a *linear function* of $y_r(t)$ and vice versa. These responses therefore have the same nature, aperiodic or oscillatory, and their transience is characterized by the *same dynamic performances*, same reduced first overshoot, same damping ratio and same natural frequency, thus revealing the *same dynamics* in tracking and regulation.

In the case of the elementary control loop ($K = 1$), the relationship [1.101] becomes:

$$y_r(t) + y_d(t) = 1, \qquad [1.102]$$

with the sum of the responses then reduced to a simple constant; in this case: one. Figure 1.20 illustrates this scenario with the graphical construction of $y_d(t)$, starting from $y_r(t)$. The initial values, $y_r(0^+) = 0$ and $y_d(0^+) = 1$, corresponding to high-frequency behaviors, as well as the end values, $y_r(\infty) = 1$ and $y_d(\infty) = 0$, corresponding to low-frequency behaviors, show that the control loop behaves like a low-pass filter in tracking and like a high-pass filter in regulation.

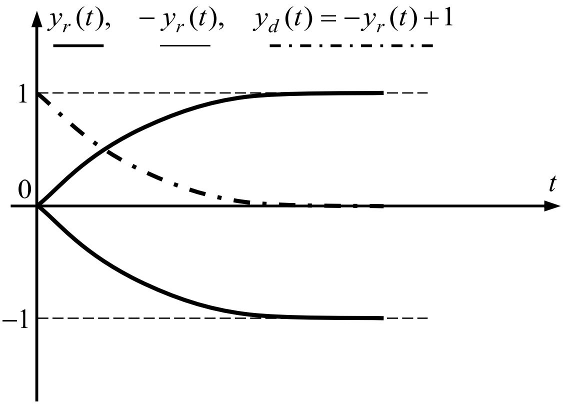

Figure 1.20. *Passage from* $y_r(t)$ *to* $y_d(t)$

1.10. Charts in tracking and regulation

1.10.1. *Chart in tracking: Nichols chart*

1.10.1.1. *Definition*

Starting from the open-loop frequency-response Nichols locus $\beta(j\omega)$, the Nichols chart (plotted in the Nichols plane) enables us to determine, for each frequency ω, the modulus (in decibels) and the argument (in degrees) of the reduced frequency response in tracking, $\mathbb{F}_r(j\omega)$.

To obtain the modulus and argument of the frequency response in tracking, $F_r(j\omega)$, starting from those of $\mathbb{F}_r(j\omega)$, we simply need to use the passage relationship:

$$F_r(j\omega) = K\mathbb{F}_r(j\omega), \quad [1.103]$$

from which the gain and phase are obtained:

$$\left|F_r(j\omega)\right| \mathrm{dB} = \left|\mathbb{F}_r(j\omega)\right| \mathrm{dB} + K \,\mathrm{dB} \quad [1.104]$$

and

$$\arg F_r(j\omega) = \arg \mathbb{F}_r(j\omega) \cdot \quad [1.105]$$

1.10.1.2. *Construction*

Let us write:

$$\beta(j\omega) = A(\omega)\, e^{j\psi(\omega)} \qquad [1.106]$$

and

$$\mathbb{F}_r(j\omega) = B(\omega)\, e^{j\varphi(\omega)}. \qquad [1.107]$$

The general form of $\mathbb{F}_r(j\omega)$ expressed as a function of $\beta(j\omega)$ then becomes:

$$B(\omega)e^{j\varphi(\omega)} = \frac{A(\omega)\, e^{j\psi(\omega)}}{1 + A(\omega)\, e^{j\psi(\omega)}}, \qquad [1.108]$$

from which we deduce the modulus and argument of $\mathbb{F}_r(j\omega)$ as a function of $A(\omega)$ and $\psi(\omega)$:

$$B(\omega) = \frac{A(\omega)}{\left[1 + A^2(\omega) + 2A(\omega)\cos\psi(\omega)\right]^{1/2}} \qquad [1.109]$$

and

$$\varphi(\omega) = \arctan\frac{\sin\psi(\omega)}{A(\omega) + \cos\psi(\omega)} \qquad [1.110]$$

The Nichols chart is obtained by plotting in the Nichols plane:

– the loci of the coordinate points, $A\mathrm{dB}$, $\psi°$, for which

$$B\mathrm{dB} = 20\log\frac{A}{\left(1 + A^2 + 2A\cos\psi\right)^{1/2}} = \mathrm{cte}, \qquad [1.111]$$

with these loci known as *amplitude (or gain) contours*;

– the loci of the coordinate points, AdB , $\psi°$, for which

$$\varphi° = \arctan \frac{\sin\psi}{A + \cos\psi} = \text{cte} , \qquad [1.112]$$

with these loci known as *phase contours*.

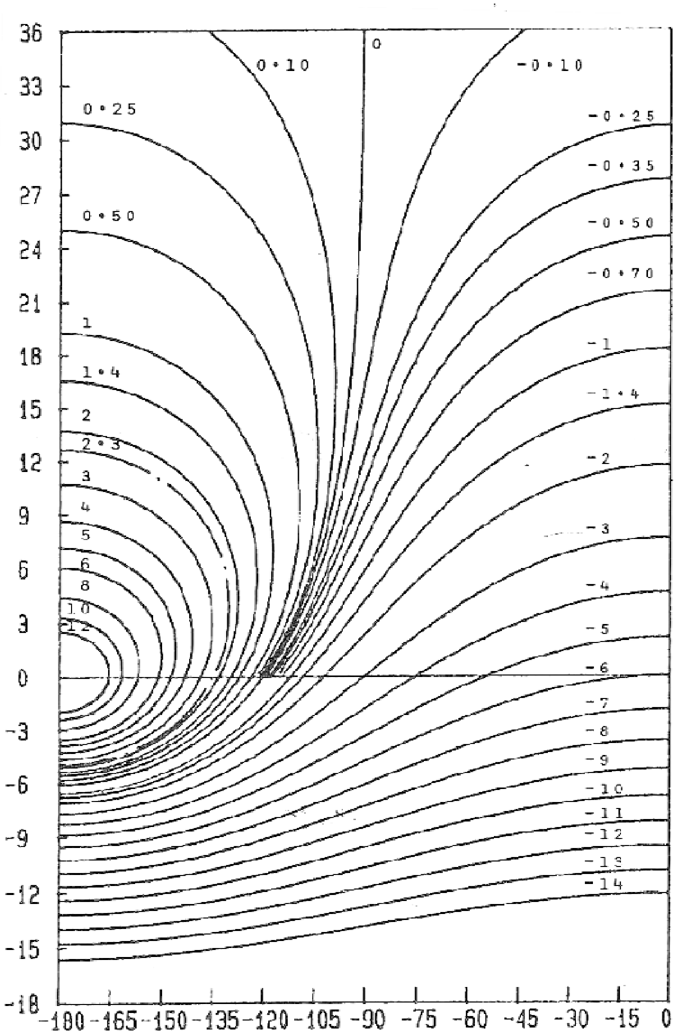

Figure 1.21. *Amplitude contours of the Nichols chart*

These loci are composed of two families of curves presenting a periodicity of 360°. The amplitude contours are represented in Figure 1.21, limiting ψ to the domain most widely used in practice, that is, [0, –180°], with the domain [–180°, –360°] for its part symmetrical about the vertical axis of abscissa –180°.

1.10.1.3. *Use*

Let there be an open-loop frequency-response Nichols locus plotted on the Nichols chart. If a point, M, of this locus corresponding to the frequency, ω_0, can be found, for example, at the intersection of contours $\lambda = -9\text{dB}$ and $\varphi = -60°$, of the Nichols chart, -9dB and $-60°$ constitute, respectively, the modulus and argument of $\mathbb{F}_r(j\omega)$ at frequency ω_0.

1.10.2. *Chart in regulation: dual of the Nichols chart*

1.10.2.1. *Definition*

From the open-loop frequency-response Nichols locus $\beta(j\omega)$, the dual of the Nichols chart (plotted in the Nichols plane) enables us to deduce, for each frequency ω, the modulus (in decibels) and the argument (in degrees) of the frequency response in regulation, $F_d(j\omega)$.

1.10.2.2. *Construction*

By writing

$$\beta(j\omega) = A(\omega)\, e^{j\psi(\omega)} \tag{1.113}$$

and

$$F_d(j\omega) = B(\omega)\, e^{j\varphi(\omega)}, \tag{1.114}$$

the general expression of $F_d(j\omega)$ given as a function of $\beta(j\omega)$ is particularized according to the equation:

$$B(\omega)e^{j\varphi(\omega)} = \frac{1}{1 + A(\omega)\, e^{j\psi(\omega)}}, \tag{1.115}$$

from which we deduce the modulus and argument of $F_d(j\omega)$ as a function of $A(\omega)$ and $\psi(\omega)$:

$$B(\omega) = \frac{1}{\left[1 + A^2(\omega) + 2A(\omega)\cos\psi(\omega)\right]^{1/2}} \qquad [1.116]$$

and

$$\varphi(\omega) = -\arctan\frac{A(\omega)\sin\psi(\omega)}{1 + A(\omega)\cos\psi(\omega)}. \qquad [1.117]$$

The dual of the Nichols chart is obtained by plotting in the Nichols plane:

– the loci of the coordinate points, AdB , $\psi°$, for which:

$$B\text{dB} = 20\log\frac{1}{\left(1 + A^2 + 2A\cos\psi\right)^{1/2}} = \text{cte}\,, \qquad [1.118]$$

with these loci known as *amplitude (or gain) contours*;

– the loci of the coordinate points, AdB, $\psi°$, for which:

$$\varphi° = -\arctan\frac{A\sin\psi}{1 + A\cos\psi} = \text{cte}\,, \qquad [1.119]$$

with these loci known as *phase contours*.

As with the Nichols chart, these loci are composed of two families of curves having a periodicity of 360°. The amplitude contours are represented in Figure 1.22, limiting ψ to the domain [0, –180°], with the domain [–180°, –360°] for its part being symmetrical about the vertical axis of abscissa –180°.

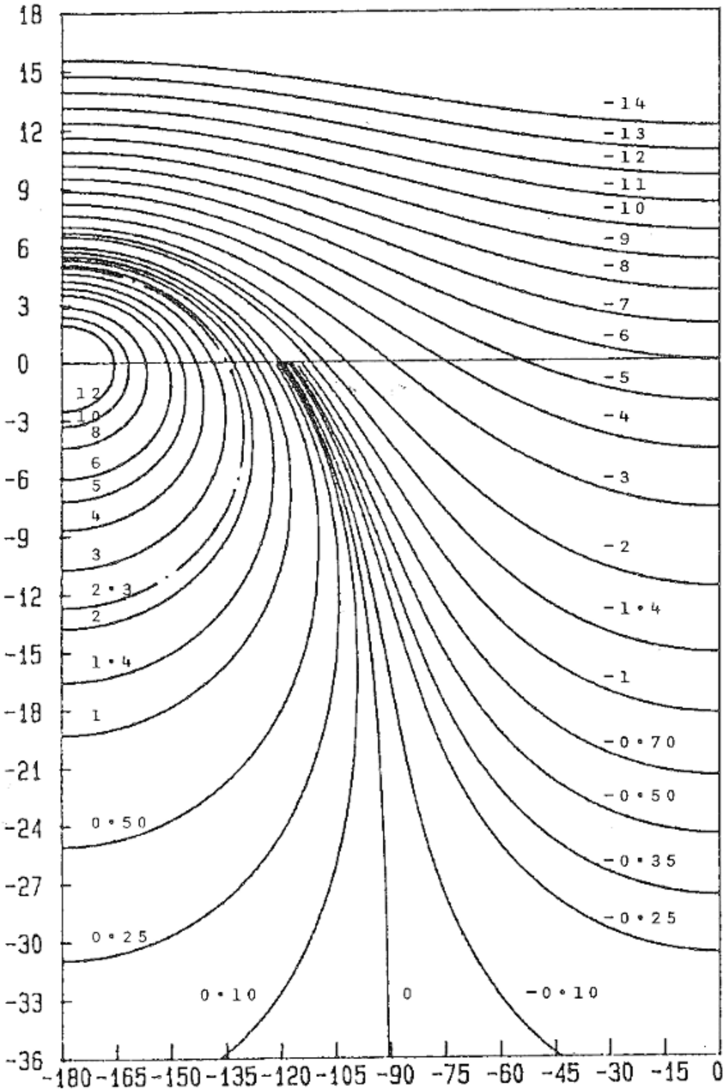

Figure 1.22. *Amplitude contours of the dual of the Nichols chart*

1.10.2.3. *Use*

This type of chart is used in the same way as the Nichols chart. Given an open-loop frequency-response Nichols locus plotted on this chart, if a point, *M*, of this locus corresponding to the frequency, ω_0, can be found, for example, at the intersection of contours $\lambda = -0.5\text{dB}$ and $\varphi = 10°$, -0.5dB and $10°$ constitute, respectively, the modulus and argument of $F_d(j\omega)$ at frequency ω_0.

1.10.3. *Passage from the Nichols chart to its dual*

According to the relationships [1.111] and [1.112], the amplitude and phase contours of the Nichols chart are such that:

$$B_r(A,\psi)\text{dB} = 20\log\frac{A}{\left(1+A^2+2A\cos\psi\right)^{1/2}} = \text{cte} \qquad [1.120]$$

and

$$\varphi°_r(A,\psi) = \arctan\frac{\sin\psi}{A+\cos\psi} = \text{cte}\,. \qquad [1.121]$$

Likewise, according to relationships [1.118] and [1.119], the amplitude and phase contours of the dual chart are such that:

$$B_d(A,\psi)\text{dB} = 20\log\frac{1}{\left(1+A^2+2A\cos\psi\right)^{1/2}} = \text{cte} \qquad [1.122]$$

and

$$\varphi°_d(A,\psi) = -\arctan\frac{A\sin\psi}{1+A\cos\psi} = \text{cte}\,. \qquad [1.123]$$

For a given value of ψ, let us change A to $1/A$ in relationships [1.120] and [1.121], which, in the Nichols plane, comes down to passing from the point $(A\text{dB}, \psi)$ to its symmetry around the axis of arguments $(-A\text{dB}, \psi)$. The set of expressions [1.120] to [1.123] thus enables the passage relationships to be established:

$$B_d(A,\psi)\text{dB} = B_r(1/A,\psi)\text{dB} \qquad [1.124]$$

and

$$\varphi°_d(A,\psi) = -\varphi°_r(1/A,\psi). \qquad [1.125]$$

These relationships reveal that between the Nichols chart and its dual, there is a symmetry with respect to the axis of arguments, on the one hand, between the amplitude contours of the same parameter and, on the other hand, between the phase contours of opposite parameters. In addition, the dual chart can be simply deduced from the Nichols chart: starting from the latter, the sign of the scales on the ordinate and the parameters of the phase contours simply need to be changed.

The dual chart (or the dual of the Nichols chart) can therefore be interpreted as the *inverse of the Nichols chart* (or *inverted Nichols chart*), the origin of which is given in the following section.

1.10.4. *A little background*

The use of the dual chart as a means of studying the regulation function was published in 1978 by Oustaloup [OUS 78a].

A recent bibliography search reveals that while the use of this chart in regulation is not called into question, in itself, the chart's "new" nature dates back beyond 1978, because we can find it in the book [GIB 63].

In this book, this chart, referred to as the "inverted Nichols chart", or "Lohcin chart", enables the determination, from the Nichols locus of $\beta(j\omega)$, for each frequency ω, of the modulus (in decibels) and the argument (in degrees) of the frequency response between the error signal, $\varepsilon(t)$, and the reference, $r(t)$, of an elementary control loop of zero disturbance, $d(t)$.

Again, in his book, Gibson refers to an article by Hill [HIL 62]. In this article, Hill points out that similar ideas were expressed in an article by Balchen [BAL 56].

In this context, it is worth remembering that the Nichols plane (decibel gain on the ordinate, degree phase on the abscissa) and the Nichols chart plotted in this plane date back to 1947, with the Black plane (gain on the abscissa, degree phase on the ordinate) and the Black chart plotted in this plane dating back to 1934. In fact, in the Black plane, from which the Nichols plane originates, the Cartestian coordinates correspond directly to the polar coordinates of a point on the Nyquist locus dating back to 1932. The Bode diagrams that separate the gain and phase frequency plots date back to 1938.