UNIVERSIDADE TÉCNICA DE LISBOA
INSTITUTO SUPERIOR TÉCNICO

Fractional Robust System Control

Duarte Pedro Mata de Oliveira Valério
(Mestre)

Dissertação para a obtenção do
Grau de Doutor em Engenharia Mecânica

Orientador: Doutor José Manuel Gutierrez Sá da Costa

Júri

Presidente: Reitor da Universidade Técnica de Lisboa
Vogais: Doutor Blas Manuel Vinagre Jara
Doutor José Manuel Gutierrez Sá da Costa
Doutor José Fernando Alves da Silva
Doutor Miguel Afonso Dias de Ayala Botto
Doutor José António Tenreiro Machado
Doutor Manuel Duarte Ortigueira

Outubro de 2005

Abstract

Fractional order controllers are controllers with a dynamical behaviour described by differential equations including derivatives whose order is not an integer number. This thesis deals with such controllers, especially for the single-input, single-output case, though the multiple-input, multiple-output case is sometimes addressed. Its objectives are to validate and enlarge results in the area of identification and control, and then to show their usefulness by applying them to plants.

New material in this thesis includes a fractional version of Levy's identification method, a digital version of the Crone identification method, tuning rules for fractional PID controllers, and applications of fractional PID controllers and fractional H_2 and H_∞ controllers to the hybrid position / force control of a rigid robot, to the position control of a flexible robot, to the control of several benchmark plants, and to the control of a thermal system.

Keywords

Fractional control; Fractional calculus; Robust control; Fractional PIDs; H_2 and H_∞ control; Flexible robotics

Resumo

Os controladores de ordem fraccionária são controladores cujo comportamento dinâmico é descrito por equações diferenciais que envolvem derivadas cuja ordem não é um número inteiro. Esta tese trata de tais controladores, particularmente para o caso de uma entrada e uma saída, embora o caso de entradas e saídas múltiplas seja por vezes considerado. Os seus objectivos são validar e ampliar resultados na area da identificação e do controlo, e aplicá-los posteriormente a sistemas com o intuito de provar a sua utilidade.

Os novos resultados desta tese incluem uma versão fraccionária do método de identificação de Levy, uma versão digital do método de identificação Crone, regras de sintonia para controladores PID fraccionários, e aplicações de controladores PID fraccionários e de controladores H_2 e H_∞ fraccionários ao controlo híbrido de posição / força de um robô rígido, ao controlo de posição de um robô flexível, ao controlo de vários sistemas paradigmáticos, e ao controlo de um sistema térmico.

Palavras-chave

Controlo fraccionário; Cálculo fraccionário; Controlo robusto; PIDs fraccionários; Controlo H_2 e H_∞; Robótica flexível

Resumo alargado

Os controladores de ordem fraccionária são controladores cujo comportamento dinâmico é descrito por equações diferenciais que envolvem derivadas cuja ordem não é um número inteiro. Esta tese trata de tais controladores, particularmente para o caso de uma entrada e uma saída, embora o caso de entradas e saídas múltiplas seja por vezes considerado. Os seus objectivos são validar e ampliar resultados na area da identificação e do controlo, e aplicá-los posteriormente a sistemas com o intuito de provar a sua utilidade.

A tese acha-se estruturada por capítulos do modo seguinte:

❖ O capítulo 2 resume a teoria do cálculo fraccionário. O cálculo fraccionário é naturalmente muito mais vasto do que o conteúdo deste capítulo. Este compreende apenas os resultados fundamentais necessários para o que se segue. Estes constam da tase para providenciar material de apoio indispensável para que um leitor versado em controlo de sistemas mas não em controlo fraccionário possa compreender o assunto abordado.

❖ O capítulo 3 diz respeito à aplicação do cálculo fraccionário ao controlo. Abordam-se assuntos como a representação de sistemas fraccionários, a sua controlabilidade, observabilidade e estabilidade. Estuda-se um sistema importante, cuja função de transferência é s^{ν}, e estudam-se e comparam-se aproximações suas de ordem inteira, de utilidade para a implementação de controladores fraccionários.

❖ O capítulo 4 descreve métodos para identificar modelos fraccionários a partir de dados experimentais (respostas na frequência e no tempo). Também se considera um método de identificação que resulta em modelos inteiros e é útil para o controlo fraccionário.

❖ O capítulo 5 descreve alguns métodos analíticos para desenvolver controladores de ordem fraccionária. Consideram-se, em particular, o desenvolvimento de controladores Crone (de todas as três gerações), de controladores PID fraccionários e de controladores H_2 e H_∞ fraccionários.

❖ O capítulo 6 expõe algumas aplicações dos métodos dos capítulos anteriores. Os sistemas considerados são um braço robótico vertical, para o qual se pretende um controlo híbrido de posição / força; um robô flexível horizontal; três funções de transferência paradigmáticas, úteis para modelar vários sistemas; e um sistema térmico.

❖ O capítulo 7 extrai conclusões e perspectiva o trabalho futuro nesta área.

Há dois apêndices após o texto principal, onde se recolhem assuntos que estorvariam a exposição se fossem incluídos quando são necessários pela primeira vez:

❖ O apêndice A contém assuntos matemáticos necessários para a teoria do cálculo fraccionário.

❖ O apêndice B descreve brevemente uma caixa de ferramentas para Matlab, intitulada Ninteger, que implementa a maioria dos métodos e algoritmos descritos no texto principal da tese.

Podem-se extrair duas conclusões principais do trabalho contido nesta tese.

A primeira não é original, embora se espere que se ache agora mais cabalmente apoiada: o cálculo fraccionário é uma ferramenta útil para o controlo. É possível modelar muitos sistemas com exactidão por meio de modelos fraccionários, e, embora em tais casos se possam igualmente empregar modelos inteiros, a sua complexidade teria então de ser significativa, ou o seu desempenho insatisfatório. Os controladores fraccionários alcançam bons desempenhos, tanto para sistemas inteiros como fraccionários. Conseguem amiúde um grau significativo de robustez.

Isto não quer evidentemente dizer que os controladores fraccionários sejam sempre melhores. Antes do mais, o que se considera ser melhor depende do caso. Há circunstâncias em que um controlador fraccionário poderá ter o melhor desempenho, mas será também excessivamente complexo; preferir-se-á então certamente um controlador inteiro mais simples com um desempenho inferior mas aceitável.

Mesmo sem levar isso em conta, a superioridade do desempenho não é sempre certa. Os controladores fraccionários são por vezes adequados e têm um desempenho aceitável; outras vezes isso não sucede. A presente tese procura documentar os casos em que se podem empregar os diferentes métodos. Em tais casos, há que ver se os controladores fraccionários têm efectivamente um desempenho aceitável. Se sim, devem-se experimentar igualmente métodos apropriados de desenvolvimento de controladores inteiros, a fim de comparar os resultados, para se escolher entre todas as possibilidades uma que providencie bons resultados e seja simples.

A segunda conclusão é que há ainda muito trabalho a desenvolver nesta área. Esta não é, uma vez mais, uma conclusão nova, embora se espere que a tese amplie o que já era conhecido.

Acknowledgements

This work has been partially supported by Fundação para a Ciência e a Tecnologia, grant number SFRH/BD/2875/2000, using funds of programme POCTI that is part of QCA III.

I wish to acknowledge appreciated support from:

❖ Filipe Morais, for his help with the active noise control in a duct, which he had worked on using conventional controllers (unfortunately this application did not make it into the final version of this thesis);

❖ Professor Heitor Pina, for his help with a numerical method (this acknowledgment is originally found in my Master's thesis, for the numerical method was needed for the toolbox for Matlab I developed back then; it is repeated here because a much improved version of that toolbox still shows up);

❖ João Reis, for his help with the flexible robot's hardware;

❖ Jorge Martins, for the same reason, and for his sharing information related to this system which was modelled and studied by him;

❖ Professor Pedro Jorge Freitas, for his valuable help with a problem of Theory of Numbers;

❖ Sérgio Fernandes, for his help with malfunctioning laboratory equipment;

❖ Professor Tenreiro Machado, for liberally sharing papers and other academic literature: without him, the bibliography would be much shorter, and the thesis would fail to consider several recent achievements.

Last but not least, I wish to thank my supervisor, Professor José Sá da Costa, for drawing my attention to this field in the first place, and then for his continued support. It is only a pity that the habit of thanking the supervisor be so widespread that that is done almost systematically, so that you cannot tell those cases when it is done solely out of a mere sense of obligation (or, even worse, a desire to conjure the otherwise absent supervisor's goodwill) from those when that is genuinely done—as is (I am luckily able to acknowledge) the present one.

Table of contents

Table of figures

Table of tables

Notation and abbreviations

Only the most frequently used notation is to be found here. Other symbols and variables were needed: their definition is found in the immediately preceding text. In particular, subsections 6.1.1 and 6.2.1 begin with lists of variables which are used in the models therein presented.

$\mathbf{A}^T$	transpose of matrix **A**
$\mathscr{C}(x)$	characteristic or floor of *x*, that is to say, the greatest integer smaller than or equal to *x* (often represented as $[x]$ or $\lfloor x \rfloor$)
Crone	*commande robuste d'ordre non-entier*
${}_cD_x^\nu y(x)$	order ν derivative of function $y(x)$, *c* being the inferior limit of integration and *x* the superior limit (see chapter 1)
$F(a,b;c;x)$	hypergeometric function ${}_2F_1(a,b;c;x)$ (see subsection A.2.3)
${}_pF_q(a_i;b_j;x)$	hypergeometric functions (see subsection A.2.3)
FIR	finite impulse response
IMC	internal model control
j	square root of -1 (but notice that a variable named *j* is sometimes used as dumb index in summations or iterated products)
$\mathscr{L}[f(x)] \equiv F(s)$	Laplace transform of function $f(x)$ (see subsection 2.4)
$\log x$	Napierian logarithm of *x*
$\log_{10} x$	common (decimal) logarithm of *x*
MIMO	multiple input, multiple output
$P(x,z)$	one of the five incomplete gamma functions (see subsection A.2.4)
$Q(x,z)$	one of the five incomplete gamma functions (see subsection A.2.4)
SISO	single-input, single-output
$\operatorname{tr}[\mathbf{A}]$	trace of square matrix **A**, that is to say, sum of all components of **A** in its main diagonal
$\bar{z}$	complex conjugate of *z*, that is to say, $\bar{z} = \operatorname{Re}[z] - j\operatorname{Im}[z]$
$\Gamma(x)$	gamma function (see subsection A.2.1)
$\Gamma(x,z)$	one of the five incomplete gamma functions (see subsection A.2.4)
$\gamma(x,z)$	one of the five incomplete gamma functions (see subsection A.2.4)
$\gamma^*(x,z)$	one of the five incomplete gamma functions (see subsection A.2.4)
ω	frequency

1. Introduction

1.1. Scope

Fractional order control systems are the subject of this thesis. For the benefit of the reader who does not know what that means let us recall some basic definitions.

As elsewhere in Engineering, *system* is the part of the Universe we want to study, bounded by some suitable border (that may exist physically or be imaginary).

In control systems, our purpose is to devise an algorithm for causing certain properties of a system to behave as we want, and to implement it physically in a device termed *controller*. The system to control is often named *plant*.

The *outputs* of the plant are the measurable quantities thereof we want to master. (Such quantities may be directly available, but it may happen that their values only indirectly be known.) In order to control a plant, we need to act upon some of its other measurable quantities, so that it behaves as intended; these quantities we act on are the plant's *inputs*. There are certainly still other properties of the plant that may influence the outputs but which we cannot master: changes thereof will change the outputs if we do not correct the inputs and so render control more difficult—they are hence called *disturbances*.

Two short examples. If we want to control the flow of a tap we will act on the number of turns of its knob. The flow is the output and the number of turns of the knob the input. This is a case of a *single-input, single-output* (SISO) plant. When we drive a car we want to control the speed and the direction of its movement. These are its outputs. The inputs we act on are usually the number of turns of the steering wheel, the position of the three pedals and the gear we put. We cannot act on the direction or speed of the wind but this also affects the car performance; so we will call the wind a disturbance. The car is an example of a *multiple-input, multiple-output* (MIMO) plant. Single-input, multiple-output (SIMO) and multiple-input, single-output (MISO) plants are also possible but are usually dealt with as particular instances of MIMO plants.

Control algorithms are nowadays usually implemented by electronic means such as a computer or a PLC, although purely mechanical controllers have been of current use and in some cases still are (the most famous example in History is probably the Watt regulator[1]). The controller interacts with the plant by means of *actuators* for acting on the inputs and *sensors* for knowing the values of outputs (and sometimes of disturbances also, so as to cope better with them).

When we need not care about the dynamic behaviour of a plant we have a problem of *logical control*. This is for instance the case of a set of traffic lights that are turned on and off in a negligible time as soon as the current is established or cut. (It should be remarked that this does not always mean that the problem is a simpler one.) The name *control systems* is usually kept for when the dynamic behaviour of the plant is important. That is for instance the case of a robot arm which is flexible: if we abruptly stop the motor that drives it, the arm will vibrate for a while before coming to rest. This dynamic behaviour must be taken into account when planning a controller for a trajectory *without* such (probably highly undesired) vibrations.

The dynamic behaviour of a plant is usually described by *differential* or *difference*

[1] Bennett (1996, p. 17).

equations. So is the dynamic behaviour of control algorithms. Sometimes the equation describing the plant includes derivatives whose order is *not* an integer number. Such fractional differential equations have been used to model plants in several different areas, as shall be documented below in section 1.3. Those plants are known as *fractional order plants*.

Control algorithms for fractional order plants may need to have a dynamic behaviour described by fractional derivatives too. Even when the plant may be described using integer order derivatives only, controllers dynamically described by fractional order derivatives may be of use for providing additional efficiency. This is because allowing the order of derivatives to assume values that are not integer may increase the ability to develop a controller closer to what is necessary to fulfil the control specifications. The usual theory of control that makes use of integer derivatives only is but a particular case of *fractional control* that is a generalisation thereof.

This thesis deals with such *fractional order controllers*, especially for the SISO case, though the MIMO case is sometimes addressed. Its objectives are to validate and enlarge results in the area of *identification* (the area that studies how to find models able to describe experimental data) and *control*, and then to show their usefulness by applying them to plants.

1.2. Overview of contents

The material in this thesis is arranged by chapters as follows:

❖ Chapter 2 is an overview of fractional calculus theory. Of course, there is much more to fractional calculus than what is dealt with in this chapter. Its contents are merely the fundamentals needed for what follows. They are included so as to provide the necessary background for a reader acquainted with control systems but not with this particular field of control to be able to understand the thesis.

❖ Chapter 3 addresses the application of fractional calculus to control. Issues such as the representation of fractional plants, their controllability, observability and stability are considered. An important plant, corresponding to transfer function s^ν, is studied, and integer order approximations thereof, useful for implementing fractional controllers, are reviewed and compared.

❖ Chapter 4 describes methods for identifying fractional models from experimental data (frequency and time responses). An identification method providing integer models useful for fractional control is also addressed.

❖ Chapter 5 describes some analytical methods for developing fractional order controllers. In particular, the development of Crone controllers (of all three generations), of fractional PIDs and of fractional H_2 and H_∞ controllers is addressed.

❖ Chapter 6 presents some applications of the methods from previous chapters. The plants addressed are a vertical robotic arm, for which hybrid position / force control is sought; a horizontal flexible robot; three benchmark transfer functions useful for modelling several plants; and a thermal system.

❖ Chapter 7 draws conclusions and presents a perspective of future work in this area.

After the main text, there are two appendixes, collecting material that would check the exposition if included when needed by the first time:

❖ Appendix A contains some material on Mathematics needed for the theory of fractional calculus. The border between what is covered in this appendix and what is not was difficult to establish. Topics not covered during undergraduate courses in Mathematics for future Engineers offered in the Technical University of Lisbon were included. Topics covered in such courses but given marginal attention were also included. And a few crucial well-known results were also included for easy reference. This appendix does not depend on material found elsewhere in the thesis. This means that it can be read right away. Or its contents may be read whenever they are needed for the first time (there are remissions in such cases).

❖ Appendix B shortly describes a toolbox for Matlab, called Ninteger, that implements most of the methods and algorithms described in the thesis's main text. Since it depends on material presented elsewhere, it may be good to read it at the end.

A list of what this thesis does *not* include (and the reader is thus assumed to know) follows:

❖ Integer order calculus in $\mathbb{R}$. (Actually a result from calculus in $\mathbb{R}^n$ is needed to prove Theorem 2.4.)

❖ Fundamentals of (integer) control systems. The previous section explains several basic ideas thereof so as to make clear for any person what this thesis is about, but of course for understanding it thoroughly some further insights into feedback and feedforward control, both in the time and the frequency domains, are necessary. It would be a waste of time, paper and effort to try to include these here when there are such good manuals as Goodwin *et al.* (2001) or Ogata (1995), just to give two examples.

❖ Mathematical tools for handling control systems in the frequency and time domains such as the Laplace and Z transforms.

1.3. What is known?

After all, acres of rubbish are printed daily and no one bothers.

George Orwell, *The freedom of the press*

No thesis appears out of thin air—and this thesis is no exception. Whatever the field of knowledge explored, there are previous contributions laid by many other authors upon which the investigation for the thesis was built. So this section contains a brief overview of the most important ones known to the author. (Readers not previously acquainted with the subject of this thesis may fail to understand some terms in the lists of this section and the next. It is expected that after reading the whole thesis this shall not happen any more.)

❖ The theory of fractional calculus is the object of several texts. The books by Miller *et al.* (1993) and Samko *et al.* (1993) are thorough expositions thereof (the later being more complex than the first). Both begin with a summary of the historical development of this field. The book by Podlubny (1999, p. 1-242) also contains (beyond several applications that shall be mentioned below) all the theory needed for control engineering purposes.

❖ There are also shorter texts on fractional calculus that nevertheless cover all the fundamentals thereof. Such are Wheeler (1997), Lavoie *et al.* (1976) and Loverro (2004).

❖ Then there are short texts that contain but a short introduction to the field. Within this category, the paper by Kleinz *et al.* (2000) is most remarkable because of the pedagogical framework its material is set in.

❖ Continuous approximations of fractional plants are the subject of the following texts. The Crone approximation is described by Oustaloup (1991) and by practically all texts on fractional control (it is by far the most popular). Carlson's approximation was set forth by Carlson *et al.* (1964). The Matsuda approximation is based upon the methods of Matsuda *et al.* (1993). These approximating methods, as well as the continued fraction approximation, are reviewed in Vinagre *et al.* (2000) and Vinagre (2001).

❖ Digital approximations of fractional plants are the subject of the following texts. Machado (1997) mentions first and second order backwards finite differences approximations based upon MacLaurin series. Machado (1999, 2001) mentions first order backwards finite difference, Tustin and Simpson approximations based upon MacLaurin series. Vinagre *et al.* (2000) mention first order backwards finite differences and Tustin approximations based upon both MacLaurin series and continued fractions. So does Vinagre (2001), together with the application of inversion (as presented in subsection 3.3.9) to the Tustin continued fraction based approximation. Chen and Vinagre (2003) apply a continued fraction expansion to a linear combination of Tustin and Simpson formulas. Tseng (2001) builds an approximation based upon an ideal time-response (though the approach is different from that found below in subsection 3.3.7).

❖ Identification of fractional transfer functions from frequency response data is addressed by Vinagre (2001, p. 133-148) and by Hartley *et al.* (2003).

❖ All three generations of Crone controllers are described by Oustaloup (1991). Oustaloup *et al.* (2000) concentrates on the third generation.

❖ Fractional PID controllers are described by Podlubny (2001, p. 249-250). Several methods have been suggested for tuning its parameters: Vinagre (2001, p. 100-103) and Caponetto *et al.* (2002, 2004) propose methods that resort to analytical considerations, while Monje *et al.* (2004) suggest a numerical minimisation of several criteria. It is worth mentioning that the tuning of *integer* PIDs is addressed by Barbosa *et al.* (2003, 2004a, 2004b) by minimising a penalty function that reflects how far the behaviour of the PID is from that of some desired fractional transfer function, and by Chen *et al.* (2004) with a somewhat similar strategy.

❖ In what concerns H_2 and H_∞ controllers, Malti *et al.* (2003) discuss the reckoning of the H_2 norm of a fractional SISO system (without applying the result to the development of controllers), and Petras *et al.* (2002a) suggest the tuning of H_∞ controllers for fractional SISO systems by numerical minimisation.

❖ Applications of fractional calculus are exceedingly numerous. In what concerns the applications mentioned in chapter 6 of this thesis, the fractional control of rigid robots is the object of Machado *et al.* (1998) and Ferreira *et al.* (2003); and the fractional control of a thermal system is the object of Vinagre *et al.* (2001), Petras *et al.* (2002b; 2002c) and Sabatier *et al.* (2003).

This list fails to mention literal hundreds (if not thousands) of other published texts related to fractional calculus. Surely many are unknown to the author. Among those known, several are not quoted above for the reason put by Eric Blair in the words quoted at the beginning of this section, but more often the reason why they were not mentioned is that their subject is most remote to the one this thesis deals with. Nevertheless a short review of the most recurrent themes showing up in the literature may prove useful, for it shows how fractional calculus has been applied to such different areas as Physics, Engineering, Economy, and Biology.

- ❖ Literature on fractional differential equations is numerous. See for instance the bibliography of Podlubny (1999) for references of some published up to that date.
- ❖ Some papers, such as Lorenzo *et al.* (2001) or Ortigueira (2003) have been published on whether it is reasonable to include in the definition of fractional derivatives a function (termed *initialisation function* or *complementary function*) for ensuring some desirable properties. This question is hanging around since the work of Riemann[2] published back in 1892, though until recently most mathematicians found better *not* to use such initialisations. The issue being still subject to controversy, it shall not be addressed further in this thesis.
- ❖ There are no doubts on what the geometrical interpretation of integer derivatives and integrals is. Fractional derivatives, however, are a different matter, and there is still no indisputable interpretation of what they mean. Machado (2003) and Stanislavsky (2004) attempt probabilistic interpretations; Nigmatullin *et al.* (2004) attempt a geometrical interpretation for complex orders.
- ❖ Several papers deal with fractional operators such as the fractional Fourier transform (this is the case of Candan *et al.* (2003)) or the fractional Laplacian (this is the case of Chen (2002)). Fractional wavelets are also the subject of several papers, among which, for instance, that of Unser *et al.* (2000).
- ❖ Diffusive filters are an approach to fractional derivatives that is not considered in this thesis. See for instance Dauphin (2001) for a review.
- ❖ Fractional finite differences, fractional delays and fractional prediction are the object of Diaz *et al.* (1974), Ortigueira (2000) and Ortigueira *et al.* (2001), among others.
- ❖ Applications of fractional calculus are, as said above, most numerous. A few examples follow.
- ❖ Applications involving diffusion and dispersion are by far the most numerous. For instance, Wang *et al.* (1999) deal with heat diffusion in the soil; Chen and Holm (2003) and Fellah *et al.* (2004) with propagation of acoustic waves; applications to transport of substances by water in soils are found in Lu *et al.* (2002) and Benson *et al.* (2004); del-Castillo-Negrete *et al.* (2004) deal with diffusion in plasmas; fractional random walks are used in a great number of papers, among which Gorenflo *et al.* (2003, 2000). This last reference is an application (among many) to financial questions.
- ❖ Image processing using fractional derivatives is the subject of Mathieu *et al.* (2002), Cooper *et al.* (2003) (this reference being an application of image processing to geophysics) or García-Fiñana *et al.* (2000) (this reference being an application of image processing to biomedicine, another field where several applications are found). Fractional wavelets are also used in this area—see Herrmann (2001), which is another application to geophysics.
- ❖ An example of applications to optics is found in Tajahuerce *et al.* (2000).
- ❖ Viscoelasticity (with applications to modelling of gels, polymers, cement, and so on) is another field with many applications. The first paper on this area was that of Bagley *et al.* (1983).

---

[2] Miller *et al.* (1993, p. 7-8); Samko *et al.* (1993, p. xxviii-xxix).

❖ Fractional derivatives are used as part of a path tracking algorithm for a mobile robot in Suárez *et al.* (2003). Path planning using fractional derivatives is the subject of Oustaloup *et al.* (2003).

❖ Control applications to a flexible transmission (Oustaloup *et al.*, 1995; Valério, 2001a), an active suspension (Lanusse *et al.*, 2003), a buck converter (Calderón *et al.*, 2003) and a hydraulic actuator (Pommier *et al.*, 2003) may also be found in the literature.

It is certainly not excessive to recall once more that the above list is quite far from aiming at completeness, its contents serving solely illustration purposes.

1.4. What is new?

> One smallest new fact obtained in the laboratory, one brick built into the temple of science, far outweights any second-hand exposition which passes an idle hour, but can leave no useful result behind it.
>
> Sir Arthur Conan Doyle, *The lost world*, 5

On the other hand, a thesis presented for obtaining the degree of Doctor is supposed to contain material developed by its author on the subject it deals with. A list, arranged by chapters, of the original contributions of the author for the field in consideration that are presented in this thesis follows:

❖ *Chapter 2.* This chapter presents nothing new. All material therein may be found in any of the several texts on the field mentioned above. Those quoted are the ones the chapter is based upon. It is only expected that this chapter be a short but self-contained introduction.

❖ *Chapter 3.* Fundamentals of fractional order systems found in sections 3.1 and 3.2 are not new, and may be found in Podlubny (1999, p. 243-260), Vinagre (2001, p. 49-88) and Oustaloup (1991). Section 3.3, on approximations, does include some new results in what digital ones are concerned. Above all, no explicit expressions for many of the approximations had ever been provided. Formerly known approximations were listed in the previous section.

❖ *Chapter 4.* The modification of Levy's method in section 4.1 is new. Vinagre (2001) introduces the weights mentioned therein but with an energetically inspired approach. The modifications of the improvements proposed by Sanathanan *et al.* (1963) and Lawrence *et al.* (1979) are also new—yet this last modification relies upon an approach that may be directly used with fractional transfer functions. The continuous version of the identification method in section 4.2 is not new—it was developed to be used together with second generation Crone controllers; see Oustaloup (1991, p. 204-208)—, but its digital version is.

❖ *Chapter 5.* Crone controllers presented in sections 5.1 to 5.3 are not new (see the references above)—the only novelty is the digital identification method already mentioned. In what concerns fractional PIDs, that they may sometimes result from internal model control (as seen in subsection 5.4.1) is a novelty; so are the tuning rules of subsections 5.4.3 to 5.4.6. H_2 and H_∞ controllers are of course no new thing; algorithms for finding the appropriate norms of SISO fractional plants are also not new, as mentioned above (curiously no reference tackles the MIMO case, that is a direct consequence of the SISO); the novelty here is the use of numerical methods for minimising a norm of the control loop.

❖ *Chapter 6.* Plants controlled are taken from Machado *et al.* (1998), Martins (2000, p. 67-69), Nabais (2002, p. 34-41), Morari *et al.* (1989, p. 114-117), Vinagre *et al.* (2001) and Vinagre (2001, p. 252-253). Additionally, relevant material on control of flexible robots is found in Benosman *et al.* (2002).

❖ *Appendix A.* This appendix deals with known Mathematics; so it contains no novelties. (It should be noticed, however, that section A.4 on continued fractions covers an issue most texts on fractional calculus do not bother to enter into, and that some results in that section were adapted to suit the purposes of fractional control.)

❖ *Appendix B.* This appendix describes a computer software, a previous version of which had already been described in Valério (2001b, p. 43-54). Differences for the currently available version are however significant.

1.5. Related publications

Most of the new results listed above were made public in conferences and journals as follows:

❖ Valério *et al.* (2002) presents explicit formulas for four digital approximations of fractional derivatives that use MacLaurin series: those based upon first and second order backwards finite differences and upon Tustin and Simpson formulas.

❖ Valério *et al.* (2003a, 2003b) address the fractional hybrid position / force control of a vertical robotic arm, found in section 6.1 (the first paper gives all explicit formulas for digital approximations found in section 3.3—though differently organised—and the second is concerned with the optimisation of parameters).

❖ Valério *et al.* (2004a) presents the digital version of the identification method in section 4.2.

❖ Valério *et al.* (2004b) presents version 2.2 of the Matlab toolbox documented in Appendix B.

❖ Valério *et al.* (2004c, 2005d) address the fractional control of a horizontal flexible robot documented in section 6.2.

❖ Valério *et al.* (2005a) presents the comparison of approximations of fractional derivatives found in section 3.3.

❖ Valério *et al.* (2005b) presents the extension of Levy's method to fractional orders found in section 4.1.

❖ Valério *et al.* (2005c, 2005f) present the tuning-rules for fractional PIDs given in section 5.4.

❖ Valério *et al.* (2005e) addresses the tuning of fractional controllers minimising the H_2 norm or the H_∞ norm, as seen in sections 5.5 and 6.4.

In all, the work reported in this thesis was used in:

❖ eight conference papers: Valério *et al.* (2002, 2003a, 2003b, 2004a, 2004b, 2004c, 2005b, 2005c);

❖ one journal paper: Valério *et al.* (2005a);

❖ one chapter in a book: Valério *et al.* (2005d);

❖ two papers submitted to journals (decisions thereon being unknown as of writing): Valério *et al.* (2005e, 2005f).

1.6. Formal details

«I said it in Hebrew—I said it in Dutch—
I said it in German and Greek:
But I wholly forgot (and it vexes me much)
That English is what you speak!»

Lewis Carroll, *The hunting of the Snark* IV, 5

Both the references and the bibliography are shown according to the rules set out in the ISO 690 (1987) standard, as set forth in the Portuguese standard NP 405-1 (1994). Most of the references are relegated to footnotes so as not to hinder the reading.

Graphs were plot and labelled according to the rules laid out by Almeida (2002).

Mathematical issues are, for the sake of clarity, often arranged into a well-established schematic form, involving definitions, lemmas, theorems, corollaries, proofs (the end of which is always clearly marked with Q.E.D.) and remarks. Full mathematical rigour in proofs was sometimes sacrificed to simplicity. This means that, sometimes, restrictive clauses in hypothesis are not made explicit, functions are always assumed to be well-behaved enough for the purpose desired, and so forth. As a consequence, proofs of theorems are often schematic and lack the consideration for details necessary in mathematical texts. Mathematics, being a tool for control, is treated up to the level of what will be needed in the sequence, and not further.

Important formulas are found within a rectangle so as to be easily located.

This thesis was written in English so that the greatest possible number of potential readers may understand it. This is especially important since the scientific community acquainted with the subject the thesis deals with is not very numerous. English being (for the moment) the world’s undisputed *lingua franca*, the problem so plainly put over one century ago by the English mathematician and writer Charles Dogdson in the words quoted above may be avoided.

2. Fractional calculus

‘I just take the train from platform nine and three-quarters at eleven o’clock,’ he read.
His aunt and uncle stared.
‘Platform what?’
‘Nine and three-quarters.’
‘Don’t talk rubbish,’ said Uncle Vernon, ‘there is no platform nine and three-quarters.’

J. K. Rowling, *Harry Potter and the Philosopher’s Stone*, 6

This chapter covers the essentials of the theory of fractional calculus[3]. This theory generalises the notion of derivative for those cases in which the differentiation order is not a natural number; in other words, it gives meaning to the expression $\frac{d^\nu}{dx^\nu} f\left(x\right)$ in those cases where ν is a negative number, a fraction, an irrational, or even an imaginary or complex number. This generalisation may be performed in several ways, leading to several slightly different definitions that do not always lead exactly to the same results[4].

The chapter is organised as follows:

- In section 2.1 derivatives for all integer orders (positive or negative) are defined.
- In section 2.2 a simple particular case is addressed as a preface to what follows.
- Sections 2.3 and 2.4 present the Riemann-Liouville definition for real orders.
- Section 2.5 presents the Caputo definition for real orders.
- Section 2.6 presents the Grünwald-Letnikoff definition for real orders.
- Section 2.7 presents the Riemann-Liouville definition for complex orders.
- In section 2.8 the Grünwald-Letnikoff definition is used for numerically reckoning non-integral derivatives and integrals.
- In section 2.9 the Laplace transforms of fractional derivatives are addressed.

2.1. Integer order calculus

Let us begin with the usual definition of the derivative of a function, which we will denote be means of a functional operator *D* (instead of the notation involving differentials $\frac{d}{dx}$)[5]:

[3] The name ‘non-integer calculus’ is sometimes used instead of ‘fractional calculus’. Of course both are misnames: the theory they label covers integer orders as well as fractional ones; within these, irrational numbers are dealt with exactly as fractions are; and both names completely fail to give a hint about complex numbers being also addressed. For this reason a third name was even coined: ‘generalised calculus’. The most usual name—fractional calculus—was the one retained in this thesis: and hence other names are used such as fractional plants, fractional systems, fractional orders, fractional controllers, fractional PIDs, and so on.

This chapter is an enlarged and improved version of Valério (2001, p. 2-15).

[4] The same happens with integrals: there are several definitions—due to Riemann, Lebesgue, Stieltjes…—which are not exactly equivalent; but for functions well-behaved enough their results do agree.

[5] Should we impose $h > 0$ in (2.1) we would get the definition of left derivative. Should we impose

$$Df(x) \stackrel{\text{def}}{=} \lim_{h\to 0} \frac{f(x)-f(x-h)}{h} \tag{2.1}$$

Let us reckon explicit expressions for higher order derivatives. The first two cases are

$$\begin{aligned} D^2 f(x) &= \lim_{h\to 0} \frac{D^1 f(x) - D^1 f(x-h)}{h} = \\ &= \lim_{h\to 0} \frac{\frac{f(x)-f(x-h)}{h} - \frac{f(x-h)-f(x-2h)}{h}}{h} = \\ &= \lim_{h\to 0} \frac{f(x)-2f(x-h)+f(x-2h)}{h^2} \end{aligned} \tag{2.2}$$

$$\begin{aligned} D^3 f(x) &= \lim_{h\to 0} \frac{D^2 f(x) - D^2 f(x-h)}{h} = \\ &= \lim_{h\to 0} \frac{\frac{f(x)-2f(x-h)+f(x-2h)}{h^2} - \frac{f(x-h)-2f(x-2h)+f(x-3h)}{h^2}}{h} = \\ &= \lim_{h\to 0} \frac{f(x)-3f(x-h)+3f(x+2h)-f(x-3h)}{h^3} \end{aligned} \tag{2.3}$$

These suggest the following general result:

Theorem 2.1: Higher-order derivatives are given by

$$D^n f(x) = \lim_{h\to 0} \frac{\sum_{k=0}^{n} (-1)^k \binom{n}{k} f(x-kh)}{h^n}, \quad n \in \mathbb{N} \tag{2.4}$$

Proof: This is proved by mathematical induction. Expressions (2.1), (2.2) and (2.3) are particular cases of (2.4), so all that is left is to prove the inductive step:

$$\begin{aligned} D\left[D^n f(x)\right] &= \lim_{h\to 0} \frac{\frac{\sum_{k=0}^{n} (-1)^k \binom{n}{k} f(x-kh)}{h^n} - \frac{\sum_{k=0}^{n} (-1)^k \binom{n}{k} f(x-kh-h)}{h^n}}{h} = \\ &= \lim_{h\to 0} \frac{\sum_{k=0}^{n} (-1)^k \binom{n}{k} f(x-kh) - \sum_{k=1}^{n+1} (-1)^{k-1} \binom{n}{k-1} f(x-kh)}{h^{n+1}} \end{aligned} \tag{2.5}$$

$h < 0$ in (2.1) we would get the definition of right derivative.

Since $-(-1)^{k-1} = (-1)^k$ and $\binom{n}{k} + \binom{n}{k-1} = \binom{n+1}{k}$, this becomes[6]

$$DD^n f(x) = \lim_{h\to 0} \frac{(-1)^0 \binom{n}{0} f(x-0h) + \sum_{k=1}^{n} \left[(-1)^k \binom{n+1}{k} f(x-kh) \right] + \ldots \\ \ldots + (-1)^{n+1} \binom{n}{n} f(x-(n+1)h)}{h^{n+1}} \tag{2.6}$$

And since $\binom{n}{0} = 1 = \binom{n+1}{0}$ and $\binom{n}{n} = 1 = \binom{n+1}{n+1}$, this becomes

$$DD^n f(x) = \lim_{h\to 0} \frac{\sum_{k=0}^{n+1} (-1)^k \binom{n+1}{k} f(x-kh)}{h^{n+1}} = D^{n+1} f(x) \tag{2.7}$$

Q.E.D.

We may likewise use ${}_cI_x^n$ to represent indefinite integration, the indexes being the limits of integration. (The integration may of course be performed from *x* to *c*; in that case the operator shall be ${}_cI_x^n f(x)$.) A recursive definition of this functional operator will be

$${}_cI_x^n f(x) \overset{\text{def}}{=} \begin{cases} \int_c^x f(x)\,dx, \text{ if } & n = 1 \\ \int_c^x {}_cI_x^{n-1} f(x)\,dx, \text{ if } & n \in \mathbb{N} \wedge n > 1 \end{cases} \tag{2.8}$$

It is known that indefinite integration and differentiation are inverse operations since[7]

$$D^n \, {}_cI_x^n f(x) = f(x) \tag{2.9}$$

It thus seems reasonable to define[8]

$${}_cD_x^n f(x) \overset{\text{def}}{=} {}_cI_x^{-n} f(x), \quad n \in \mathbb{Z}^- \tag{2.10}$$

For operator D^n to be defined for all natural values of *n* it is still necessary to

[6] It is this property of combinations that makes Pascal's triangle so easy to build.

[7] For this reason, indefinite integration is sometimes referred to as anti-differentiation.

[8] Notice that from now on operator *D* sometimes requires indexes, and sometimes not. In general expressions valid for all orders, if such indexes are given but a particular order does not require them, they are to be neglected. When the values of limits are obvious, they are sometimes omitted to alleviate notation.

make

$$D^0 f(x) \overset{\text{def}}{=} f(x) \tag{2.11}$$

This is convenient to ensure a simple form for the law of exponents discussed below.

It must be noticed that the order in which both operators are applied is not irrelevant[9]. Actually, operator D is the *left* inverse of operator I, as (2.9) shows—but it is *not* its *right* inverse, since

$$ {}_cI_x^1 D^1 f(x) = f(x) - f(c) = f(x) - (x-c)^0 f(c) \tag{2.12}$$

$$ {}_cI_x^2 D^2 f(x) = f(x) - f(c) - (x-c) Df(c) = $$
$$ = f(x) - (x-c)^0 f(c) - (x-c) Df(c) \tag{2.13}$$

$$ {}_cI_x^3 D^3 f(x) = f(x) - f(c) - (x-c) Df(c) - \frac{(x-c)^2}{2} D^2 f(c) = $$
$$ = f(x) - (x-c)^0 f(c) - (x-c) Df(c) - \frac{(x-c)^2}{2} D^2 f(c) \tag{2.14}$$

…

$$ {}_cI_x^n D^n f(x) = f(x) - \sum_{i=0}^{n-1} \frac{(x-c)^i}{i!} D^i f(c) \tag{2.15}$$

This is the basis for the law of exponents of operator D^n, which states

$$ {}_cD_x^{-n}\, {}_cD_x^n f(x) = \begin{cases} f(x), & n \in \mathbb{Z}_0^- \\ f(x) - \sum_{k=0}^{n-1} \frac{(x-c)^{n-k-1}}{(n-k-1)!} D^{n-k-1} f(c), & n \in \mathbb{N} \end{cases} \tag{2.16}$$

As a consequence of (2.9), we will have

$$ {}_cD_x^m\, {}_cD_x^n f(x) = {}_cD_x^{m+n} f(x),\ m,n \in \mathbb{N}_0 \vee m,n \in \mathbb{Z}_0^- \vee \left(m \in \mathbb{Z}^+ \wedge n \in \mathbb{Z}^-\right) \tag{2.17}$$

The law of exponents (2.16) shows this would not be true should differentiation precede integration.

2.2. The particular case of the power function

In fractional calculus operator D is generalised so as to also make sense for differentiation orders that are not integer. But let us first consider a simple case where the ideas underlying the generalisation are present and may be easily apprehended.

Let

---

[9] Samko *et al.* (1993), p. 43.

$$f(x) = x^a, \quad x \in \mathbb{R}^+ \tag{2.18}$$

where $a \neq 0$. Then

$$Df(x) = ax^{a-1} \tag{2.19}$$

$$D^2 f(x) = a(a-1)x^{a-2} \tag{2.20}$$

$$D^3 f(x) = a(a-1)(a-2)x^{a-3} \tag{2.21}$$

as long as a *zero* exponent does not appear, of course: from there on, derivatives would be always identically zero. This suggests the following general formula:

Theorem 2.2: The derivatives of *f* are given by

$$D^n f(x) = a(a-1)\ldots\left[a-(n-1)\right]x^{a-n} = x^{a-n}\prod_{i=0}^{n-1}(a-i), \quad n \in \mathbb{N} \tag{2.22}$$

(unless $a-n$ be a negative integer, in which case $D^n f(x) = 0$).

Proof: This is again shown by mathematical induction. (2.19), (2.20) and (2.21) verify (2.22); the inductive step is proved as follows:

$$D\left[D^n f(x)\right] = \left[x^{a-n}\prod_{i=0}^{n-1}(a-i)\right]' = x^{a-n-1}(a-n)\prod_{i=0}^{n-1}(a-i) =$$
$$= x^{a-(n+1)}\prod_{i=0}^{(n+1)-1}(a-i) \tag{2.23}$$

The reason why case $a-n \in \mathbb{Z}^-$ must be considered apart was discussed above. **Q.E.D.**

We may similarly try to find a formula for integration:

$$\int_0^x f(\xi)\,d\xi \equiv \int f(x)\,dx = \frac{1}{a+1}x^{a+1} \tag{2.24}$$

$$\iint f(x)\,dxdx = \frac{1}{(a+1)(a+2)}x^{a+2} \tag{2.25}$$

$$\iiint f(x)\,dxdxdx = \frac{1}{(a+1)(a+2)(a+3)}x^{a+3} \tag{2.26}$$

The integrals above hold as long as the exponent –1 does not appear (recall that $\int x^{-1}dx = \ln x$).

Theorem 2.3: The integrals of *f* are given by

$$ {}_0D_x^n f(x) = \frac{1}{\prod_{i=1}^{-n}(a+i)} x^{a-n}, \quad n \in \mathbb{Z}^- \tag{2.27} $$

as long as $\sim\left(a \in \mathbb{Z}^- \wedge a-n \in \mathbb{N}_0\right)$.

Proof: This is again shown by mathematical induction. (2.24), (2.25) and (2.26) verify (2.22). The inductive step must now be proved, from –1 *backwards*, as follows:

$$ \begin{aligned} D^{-1}\left[D^n f(x)\right] &= \int \left[\frac{1}{\prod_{i=1}^{-n}(a+i)} x^{a-n} \right] dx = \frac{1}{(a-n+1)\prod_{i=1}^{-n}(a+i)} x^{a-n+1} = \\ &= \frac{1}{\prod_{i=1}^{-(n-1)}(a+i)} x^{a-(n-1)} \end{aligned} \tag{2.28} $$

The reason why case $a \in \mathbb{Z}^- \wedge a-n \in \mathbb{N}_0$ must be considered apart was discussed above. **Q.E.D.**

We may now replace the iterated product in (2.22) by the gamma function[10] as follows, making use of (A.40):

$$ \begin{aligned} D^n f(x) &= x^{a-n}\prod_{i=0}^{n-1}(a-i) = (-1)^n x^{a-n}\prod_{i=0}^{n-1}(-a+i) = \\ &= (-1)^n x^{a-n}\frac{\Gamma(a+1)}{\Gamma(a-n+1)}(-1)^n = \frac{\Gamma(a+1)}{\Gamma(a-n+1)} x^{a-n} \end{aligned} \tag{2.29} $$

And we may similarly use (A.40) to transform (2.27) into an expression with function Γ:

$$ \begin{aligned} {}_0D_x^n f(x) &= \frac{1}{\prod_{i=1}^{-n}(a+i)} x^{a-n} = x^{a-n}\frac{1}{\prod_{i=0}^{-n-1}(a+1+i)} = x^{a-n}\frac{1}{\frac{\Gamma(a+1-n)}{\Gamma(a+1)}} = \\ &= x^{a-n}\frac{\Gamma(a+1)}{\Gamma(a+1-n)} \end{aligned} \tag{2.30} $$

The result in (2.29) and (2.30) is the same—and is also valid for $n = 0$. Furthermore, the expression makes sense even for a differentiation order that is not integer, and thus we may reasonably write[11]

[10] Some transcendental functions, among which function Γ, play an important role in generalising calculus to fractional orders. Their definitions and most important properties may be found in Appendix A, section A.2.

[11] Miller *et al.* (1993), p. 2-3.

$$ {}_0D_x^{\nu} f(x) = \frac{\Gamma(a+1)}{\Gamma(a-\nu+1)} x^{a-\nu}, \ \nu \in \mathbb{R} \tag{2.31} $$

as long as $a-\nu \notin \mathbb{Z}^- \wedge a \notin \mathbb{Z}^-$; this is necessary to ensure that function Γ is defined, and also ensures that the exceptions in Theorem 2.2 and Theorem 2.3 will not occur. (It also means that some cases covered by (2.22) and (2.27) are no longer covered by (2.31), but that is the price to pay for having such a general expression.)

As will be seen below, fractional derivatives[12] are always defined making use of function Γ to replace products, just as done above.

2.3. Real order calculus according to the Riemann-Liouville definition

This section generalises operator D for a real order of differentiation[13] according to the most usual definition, Riemann-Liouville's one.

This definition generalises two equalities from integer calculus. The first one is Cauchy's formula.

Theorem 2.4 (Cauchy's formula): Indefinite integrals are given by

$$ {}_cD_x^{-n} f(x) = \int_c^x \frac{(x-t)^{n-1}}{(n-1)!} f(t)\,dt, \quad n \in \mathbb{N} \tag{2.32} $$

Proof: This is again by mathematical induction. First notice that ${}_cD_x^{-n} f(x)$ may be rewritten as

$$ {}_cD_x^{-n} f(x) = \int_c^x \int_c^{x_1} \int_c^{x_2} \ldots \int_c^{x_{n-1}} f(t)\,dt \ldots dx_3 dx_2 dx_1 \tag{2.33} $$

Dirichlet's equality shows that[14]

$$ \int_c^x \int_c^{x_1} f(x_1,t)\,dt dx_1 = \int_c^x \int_t^x f(x_1,t)\,dx_1 dt \tag{2.34} $$

In particular, if f is a function of t alone,

$$ \int_c^x \int_c^{x_1} f(t)\,dt dx_1 = \int_c^x \int_t^x f(t)\,dx_1 dt = \int_c^x f(t) \int_t^x dx_1 dt = \int_c^x f(t)(x-t)\,dt \tag{2.35} $$

[12] Some authors use the designation *fractional derivatives* for the case $D^{\nu} f(x),\ \nu > 0$ only, and use the designation *fractional integrals* for the case $D^{\nu} f(x),\ \nu < 0$. This is, of course, a distinction that no longer makes sense if $\nu \in \mathbb{C} \setminus \mathbb{R}$. Other authors merge all cases under the name *fractional differintegration*. In this thesis, the phrase *fractional derivatives* is used indiscriminately irrespective of the value of ν.

[13] Or integration. See the previous footnote.

[14] Widder (1947, p. 165).

Applying the same property again[15],

$$\int_c^x \int_c^{x_1} \int_c^{x_2} f(t)\,dt\,dx_2\,dx_1 = \int_c^x \int_c^{x_1} (x_1 - t) f(t)\,dt\,dx_1 = \int_c^x \int_t^x (x_1 - t) f(t)\,dx_1\,dt =$$
$$= \int_c^x f(t) \int_t^x (x_1 - t)\,dx_1\,dt = \int_c^x f(t) \frac{(x-t)^2}{2}\,dt \tag{2.36}$$

(2.35) and (2.36) verify (2.32); the inductive step is as follows:

$$\int_c^x \int_c^{x_1} \frac{(x_1 - t)^{n-1}}{(n-1)!} f(t)\,dt\,dx_1 = \int_c^x \int_t^x \frac{(x_1 - t)^{n-1}}{(n-1)!} f(t)\,dx_1\,dt =$$
$$= \int_c^x f(t) \int_t^x \frac{(x_1 - t)^{n-1}}{(n-1)!}\,dx_1\,dt = \int_c^x f(t) \frac{(x-t)^n}{n!}\,dt \tag{2.37}$$

Q.E.D.

Because of (A.32), Cauchy's formula (2.32) may be rewritten as

$$ {}_cD_x^{-n} f(x) = \frac{1}{\Gamma(n)} \int_c^x (x-t)^{n-1} f(t)\,dt \tag{2.38}$$

This makes sense even if the integration order is not a natural number, and so we may define[16]

$$ {}_cD_x^{\nu} f(x) \overset{\text{def}}{=} \frac{1}{\Gamma(-\nu)} \int_c^x (x-t)^{-\nu-1} f(t)\,dt,\ \nu \in \mathbb{R}^- \tag{2.39}$$

The second equality from integer calculus that will be generalised is (2.17). The case $\nu > 0$ will be defined using the definition for $\nu < 0$ as follows:

$$ {}_cD_x^{\nu} f(x) \overset{\text{def}}{=} D^n \left[{}_cD_x^{\nu-n} f(x) \right],\ n = \min\{k \in \mathbb{N} : k > \nu\},\ \nu \in \mathbb{R}^+ \tag{2.40}$$

In this manner we obtained the complete Riemann-Liouville definition, valid for functions $f(x)$ for which the integral in (2.39) exists. Putting formulas (2.39), (2.40) and (2.11) together[17]:

---

[15] It should be noticed that, should x be the lower limit of integration, the polynomial multiplying $f(t)$ would be $(t - x)$ instead of $(x - t)$. That is why the Riemann-Liouville definition will be slightly different for this case, as seen below in footnote 17.

[16] Miller *et al.* (1993, p. 23-25); Samko *et al.* (1993, p. 33); Podlubny (1999, p. 62 and 65).

[17] A few additional remarks:

The notation above is the most usual one (Miller *et al.*, p. 21 and 36, 1993). The following notations may also be found in the literature (Samko *et al.*, p. 33 and 37, 1993; Lavoie *et al.*, p. 242-244, 1976):

$$
{}_cD_x^\nu f(x) \overset{\text{def}}{=} \begin{cases} \dfrac{1}{\Gamma(-\nu)}\displaystyle\int_c^x (x-\xi)^{-\nu-1} f(\xi)\,d\xi, \text{ if } \nu<0 \\ f(x), \text{ if } \nu=0 \\ D^n\left[{}_cD_x^{\nu-n} f(x)\right], \ n=\min\{k\in\mathbb{N}: k>\nu\}, \text{ if } \nu>0 \end{cases} \tag{2.41}
$$

Remark 1: It is important to stress that operator D is local only if the order is a natural number, that is to say, for usual (integer) derivatives only. In all other cases integration limits are needed, even if the order is positive[18].

Remark 2 (the law of exponents): Just as when the orders are integer, the following equality, similar to (2.9), holds for the Riemann-Liouville definition:

$$
{}_cD_x^\nu\, {}_cD_x^{-\nu} f(x) = f(x), \quad \nu\geq 0 \tag{2.42}
$$

Once more, however, if the operators' order is changed the equality holds no more[19]:

$$
\begin{aligned} &{}_cD_x^{-\nu}\, {}_cD_x^\nu f(x) = \\ &= f(x) - \sum_{k=0}^{n-1}\frac{(x-c)^{\nu-k-1}}{\Gamma(\nu-k)} D^{n-k-1}\, {}_cD_x^{\nu-n} f(c), \ n=\min\{m\in\mathsf{N}: m>\nu>0\} \end{aligned} \tag{2.43}
$$

2.4. More on the Riemann-Liouville definition

It is possible to justify the Riemann-Liouville definition (2.40) for $\nu\in\mathbb{R}^-$ by other means, always to keep valid properties that hold for an integer order. One of those

---

$$
{}_cD_x^\nu f(x) \equiv \left(D_{c+}^\nu f\right)(x) \equiv D_{x-c}^\nu f(x) \equiv \frac{d^\nu}{d(x-c)^\nu} f(x)
$$

$$
{}_xD_c^n f(x) \equiv \left(D_{c-}^\nu f\right)(x) \equiv D_{c-x}^\nu f(x) \equiv \frac{d^\nu}{d(c-x)^\nu} f(x)
$$

Instead of always calling this operator ${}_cD_x^\nu$ Riemann-Liouville operator, some authors use different names depending on the limits of integration. For instance, Miller *et al.* (1993, p. 21-22) call it Riemann-Liouville operator only when $c=0$, Riemann operator when $c>-\infty$, Liouville operator when $c=-\infty$, and Weil transform (represented as ${}_xW_{+\infty}^\nu$) if the integration is performed from x up to $+\infty$.

Miller *et al.* (1993, p. 352-356), and Samko *et al.* (1993, p. 173-174), have tables of derivatives and integrals for several usual functions using this definition.

If x is the lower limit of integration, and as a consequence of the fact mentioned in footnote 15, we shall have (Samko, p. 37, 1993), instead of (2.41),

$$
{}_xD_c^\nu f(x) \overset{\text{def}}{=} \begin{cases} \dfrac{1}{\Gamma(-\nu)}\displaystyle\int_x^c (\xi-x)^{-\nu-1} f(\xi)\,d\xi, \text{ if } \nu<0 \\ f(x), \text{ if } \nu=0 \\ (-1)^n D^n\left[{}_cD_x^{\nu-n} f(x)\right], \ n=\min\{k\in\mathbb{N}: k>\nu\}, \text{ if } \nu>0 \end{cases}
$$

[18] Wheeler (1997, p. 4-7); Lavoie *et al.* (1976, p. 244).
[19] Samko *et al.* (1993, p. 44-48); Miller *et al.* (1993, p. 57-63).

other justifications makes use of the Laplace transform and is of interest for what follows. Recall that the Laplace transform is defined as[20]

$$\mathscr{L}\left[f(t)\right] \equiv F(s) \overset{\text{def}}{=} \int_0^{+\infty} f(t)e^{-st}dt \tag{2.44}$$

The following results will be needed:

Theorem 2.5[21]: The Laplace transform of an integer derivative is given by

$$\mathscr{L}\left[\frac{d^n}{dt^n}f(t)\right] = s^n F(s) - \sum_{k=1}^{n} s^{n-k}\frac{d^{k-1}}{dt^{k-1}}f(0), \quad n \in \mathbb{N} \tag{2.45}$$

Theorem 2.6 (Convolution theorem)[22]: The following equality holds if the Laplace transforms therein exist:

$$\mathscr{L}\left[\int_0^t f(t-\tau)g(\tau)d\tau\right] = \mathscr{L}\left[f(t)\right]\mathscr{L}\left[g(t)\right] \tag{2.46}$$

Lemma 2.1: The Laplace transform of the power function is

$$\mathscr{L}\left[t^\lambda\right] = s^{-1-\lambda}\Gamma(\lambda+1) \tag{2.47}$$

Proof: By definition,

$$\mathscr{L}\left[t^\lambda\right] = \int_0^{+\infty} t^\lambda e^{-st}dt \tag{2.48}$$

The primitive involved in (2.48) is[23]

$$\int t^\lambda e^{-st}dt = -s^{-1-\lambda}\Gamma(1+\lambda, st) \tag{2.49}$$

as may be easily verified:

$$\frac{d}{dt}\left[-s^{-1-\lambda}\Gamma(1+\lambda,st)\right] = -s^{-1-\lambda}s\frac{d}{d(st)}\int_{st}^{+\infty} e^{-y}y^\lambda dy =$$
$$= -s^{-\lambda}\left[-e^{-st}(st)^\lambda\right] = t^\lambda e^{-st} \tag{2.50}$$

---

[20] It is expedient to assume that function *f* is defined for positive values of *t* only; this will be done in all that follows whenever Laplace transforms are considered. The question also arises of knowing, for each particular function, the values of *s* for which the transform exists. This will never be addressed below since it is not necessary for anything that follows. See Hildebrand (1976, p. 53-76) or Widder (1947, p. 365-423) about the Laplace transform and the resulting method for solving differential equations.

[21] Hildebrand (1976, p. 58-59); Widder (1947, p. 377).

[22] Hildebrand (1976, p. 63-65); Widder (1947, p. 380).

[23] See the Appendix A, subsection A.2.4, about the incomplete gamma function Γ.

and so (2.48) becomes

$$\mathscr{L}\left[t^{\lambda}\right]=\left[-s^{-1-\lambda}\Gamma\left(1+\lambda,st\right)\right]_{t=0}^{+\infty}=-s^{-1-\lambda}\int_{+\infty}^{+\infty}e^{-y}y^{\lambda}dy+s^{-1-\lambda}\int_{0}^{+\infty}e^{-y}y^{\lambda}dy \tag{2.51}$$

The first term is zero and the second, by definition (A.34), is the right-hand member of (2.47). **Q.E.D.**

Corollary: Thanks to (A.32), (2.47) reduces, for integer powers, to

$$\mathscr{L}\left[t^{n}\right]=\frac{n!}{s^{n+1}} \tag{2.52}$$

We may now give a second justification of (2.40) as follows. Let us consider the following system for a natural *n*:

$$\begin{cases} D^{n}y\left(x\right)=f\left(x\right) \\ D^{k}y(0)=0, \quad k=0,1,\ldots,n-1 \end{cases} \tag{2.53}$$

It may be solved by means of the Laplace transform, becoming, by application of (2.45),

$$s^{n}Y(s)=F(s)\Leftrightarrow Y(s)=s^{-n}F(s) \tag{2.54}$$

Now using (2.46) and (2.52) we shall have

$$y\left(x\right)=\int_{0}^{x}\frac{\left(x-t\right)^{n-1}}{\left(n-1\right)!}f\left(t\right)dt=\frac{1}{\Gamma\left(n\right)}\int_{0}^{x}\left(x-t\right)^{n-1}f\left(t\right)dt \tag{2.55}$$

Since $f(x)$ is the *n*-th derivative of $y(x)$, $y(x)$ is the *n*-th indefinite integral of $f(x)$. Thus[24]

$${}_{0}D_{x}^{-n}f\left(x\right)=\frac{1}{\Gamma\left(n\right)}\int_{0}^{x}\left(x-t\right)^{n-1}f\left(t\right)dt \tag{2.56}$$

But this formula makes sense even when *n* is not a natural number. By allowing the integration to be performed from some point other than 0, we obtain formula (2.39) again.

2.5. Real order calculus according to the Caputo definition

After establishing (2.39), it would also have been possible to generalise operator *D* for positive orders as follows:

[24] Miller *et al.* (1993, p. 25 and 28).

$$ {}_cD_x^\nu f(x) \overset{\text{def}}{=} {}_cD_x^{\nu-n}\left[D^n f(x)\right],\ n=\min\{k\in\mathbb{N}:k>\nu\},\ \nu\in\mathbb{R}^+ \tag{2.57} $$

This definition, due to Caputo[25], no longer corresponds to a generalisation of (2.17), but leads to simpler Laplace transform formulas, as shall be seen below in subsection 2.9, and that is why it is sometimes preferred to the Riemann-Liouville one.

The complete definition is

$$ {}_cD_x^\nu f(x) \overset{\text{def}}{=} \begin{cases} \dfrac{1}{\Gamma(-\nu)}\displaystyle\int_c^x (x-\xi)^{-\nu-1} f(\xi)\,d\xi,\ \text{if}\ \ \nu<0 \\ f(x),\ \text{if}\ \ \nu=0 \\ {}_cD_x^{\nu-n}\left[D^n f(x)\right],\ n=\min\{k\in\mathbb{N}:k>\nu\},\ \text{if}\ \ \nu>0 \end{cases} \tag{2.58} $$

2.6. Real order calculus according to the Grünwald-Letnikoff definition

This section generalises operator *D* for a real order of differentiation or integration according to the definition introduced by Grünwald and Letnikoff[26].

That definition stems from formula (2.4), which gives the same results as

$$ D^n f(x) = \lim_{h\to 0}\frac{\sum_{k=0}^{m}(-1)^k\binom{n}{k}f(x-kh)}{h^n},\quad m,n\in\mathbb{N}\wedge m>n \tag{2.59} $$

This is because, for $k,n\in\mathbb{N}:k>n$, we have

$$ \binom{n}{k} = \frac{\Gamma(n+1)}{\Gamma(k+1)\Gamma(n-k+1)} = \frac{n!}{k!\infty} = 0 \tag{2.60} $$

Definition (2.59) may be generalised to define

$$ D^\nu f(x) = \lim_{\substack{h\to 0\\ m\to+\infty}}\frac{\sum_{k=0}^{m}(-1)^k\binom{\nu}{k}f(x-kh)}{h^\nu} \tag{2.61} $$

Two things should be taken into account. First, the upper limit of the summation *m* has to be taken up to infinity because, since ν need not be integer, terms will not be zero from some value of *k* on. Second, it would be good to know which integration indexes result from (2.61). This question is easy to answer for $\nu=-1$; (A.43) shows we will have

[25] Podlubny (1993, p. 78-81).
[26] Miller *et al.* (1993, p. 38-42); Samko *et al.* (1993, p. 385-386); Podlubny (1999, p. 43-55).

$$D^{-1}f(x) = \lim_{\substack{h\to 0\\ m\to+\infty}} \frac{\sum_{k=0}^{m}(-1)^k \frac{(-1)^k\Gamma(k+1)}{\Gamma(k+1)\Gamma(1)} f(x-kh)}{h^{-1}} = \lim_{\substack{h\to 0\\ m\to+\infty}} \sum_{k=0}^{m} f(x-kh)h \tag{2.62}$$

Now this is the Riemann definition of integral ${}_cD_x^{-1} f(x)$ if

$$h = \frac{x-c}{m} \tag{2.63}$$

Similarly, for orders other than –1, we will have[27]

$$ {}_cD_x^{\nu} f(x) \overset{\text{def}}{=} \lim_{\substack{h\to 0\\ mh=x-c}} \frac{\sum_{k=0}^{m}(-1)^k \binom{\nu}{k} f(x-kh)}{h^{\nu}} \tag{2.64}$$

We now need the following result, which is quoted without proof:

Lemma 2.2[28]**:** The following limit holds:

$$\lim_{k\to+\infty} \frac{\Gamma(k-\nu)}{k^{-\nu-1}\Gamma(k+1)} = 1 \tag{2.65}$$

Theorem 2.7: Definitions (2.64) and (2.39) are the same, provided that conditions are verified for both to exist[29].

Proof: Using (A.43) in (2.64), we get

$$ {}_cD_x^{\nu} f(x) = \lim_{\substack{h\to 0\\ mh=x-c}} \frac{\sum_{k=0}^{m}(-1)^k \frac{(-1)^k\Gamma(k-\nu)}{\Gamma(k+1)\Gamma(-\nu)} f(x-kh)}{h^{\nu}} =$$

$$= \lim_{\substack{h\to 0\\ mh=x-c}} \sum_{k=0}^{m} \frac{h^{-\nu-1}k^{-\nu-1}}{\Gamma(-\nu)} h \frac{\Gamma(k-\nu)}{k^{-\nu-1}\Gamma(k+1)} f(x-kh) \tag{2.66}$$

Using (2.65), this becomes

$$ {}_cD_x^{\nu} f(x) = \frac{1}{\Gamma(-\nu)} \lim_{\substack{h\to 0\\ mh=x-c}} \sum_{k=0}^{m} (kh)^{-\nu-1} f(x-kh)h \tag{2.67}$$

By the Riemann definition of integral this becomes (2.39). **Q.E.D.**

---

27 Wheeler (1997, p. 9-12); Lavoie *et al.* (1976, p. 247-248).

28 Lavoie *et al.* (1976, p. 248-249).

29 Podlubny (1993, p. 62-63); Lavoie *et al.* (1976, p. 248-249). This last reference carefully justifies the steps of the proof sketched below.

2.7. Complex order calculus according to the Riemann-Liouville definition

Thanks to (A.41), the Riemann-Liouville definition of operator D may be generalised to allow for a complex order[30] with only a slight change in definitions (2.39) and (2.40):

$$ {}_cD_x^z f(x) \overset{\text{def}}{=} \begin{cases} \frac{1}{\Gamma(-z)} \int_c^x (x-\xi)^{-z-1} f(\xi)\, d\xi, \text{ if } \operatorname{Re} z < 0 \\ {}_cD_x^n \left[{}_cD_x^{z-n} f(x) \right],\ n = \min\{k \in \mathbb{N} : k > \operatorname{Re} z\},\ \text{if } \operatorname{Re} z > 0 \end{cases} \tag{2.68} $$

The integral of the previous formula is divergent if z is purely imaginary; so this case must be handled as follows:

$$ {}_cD_x^z f(x) \overset{\text{def}}{=} D^1\, {}_cD_x^{z-1} f(x),\ z \in \mathbb{C} : \operatorname{Re} z = 0 \tag{2.69} $$

The expression above includes the identity operator found when $z = 0$.

In short:

$$ {}_cD_x^z f(x) \overset{\text{def}}{=} \begin{cases} \frac{1}{\Gamma(-z)} \int_c^x (x-\xi)^{-z-1} f(\xi)\, d\xi, \text{ if } \operatorname{Re} z < 0 \\ D^1\, {}_cD_x^{-1+z} f(x),\ \text{if } z \in \mathbb{C} : \operatorname{Re} z = 0 \\ f(x),\ \text{if } z = 0 \\ D^n \left[{}_cD_x^{z-n} f(x) \right],\ n = \min\{k \in \mathbb{N} : k > \operatorname{Re} z\},\ \text{if } \operatorname{Re} z > 0 \end{cases} \tag{2.70} $$

2.8. Numerical evaluation of fractional order derivatives

Fractional derivatives are often difficult to find analytically. Numerical approximations are consequently needed. The simplest one makes use of the Grünwald-Letnikoff definition. Formula (2.64) may be used for numerically evaluating approximate values of fractional order integrals and derivatives using some suitably chosen value of h as follows[31]:

$$ {}_cD_x^\nu y(x) \approx \frac{1}{h^\nu} \sum_{k=0}^{\mathscr{E}\left(\frac{x-c}{h}\right)} (-1)^k \binom{\nu}{k} y(x-kh) \tag{2.71} $$

The series is truncated with a number of terms that increases with the decrease of h.

The following result is quoted without proof:

[30] Samko *et al.* (1993), p. 38.
[31] Miller *et al.* (1993, p. 39); Podlubny (1999, p. 200).

Theorem 2.8 (Short memory principle[32]): If function f is bounded in $[c;x]$— that is to say, if a value M exists verifying

$$\left|f(\xi)\right| < M, \quad \forall \xi \in [c;x] \tag{2.72}$$

—, then the approximation

$${}_{c}D_{x}^{\nu} f(x) \approx {}_{x-L}D_{x}^{\nu} f(x) \tag{2.73}$$

results in an error ε bounded by

$$|\varepsilon| \leq \frac{ML^{-\nu}}{\left|\Gamma(1-\nu)\right|} \tag{2.74}$$

Remark 1: Using formula (2.73) together with some suitably chosen value of L is expedient when $x \gg c$ and the number of terms in the summation of formula (2.71) gets very large.

Remark 2: From (2.74) results that, for the error to be smaller than a certain desired value $|\varepsilon|$, we must have

$$L \geq \sqrt[\nu]{\frac{M}{\left|\varepsilon\,\Gamma(1-\nu)\right|}} \tag{2.75}$$

2.9. Laplace transform

Operator D turns out to have Laplace transforms that follow rules quite similar to those valid for integer orders. The different definitions of the operator, however, cause slight differences in the initial conditions that show up in Laplace transforms.

Theorem 2.9[33]: The Laplace transform of the Riemann-Liouville definition of D is

$$\mathscr{L}\left[{}_{0}D_{x}^{\nu} f(x)\right] = s^{\nu} F(s), \; \nu \leq 0 \tag{2.76}$$

$$\mathscr{L}\left[{}_{0}D_{x}^{\nu} f(x)\right] = s^{\nu} F(s) - \sum_{k=0}^{n-1} s^{k} \, {}_{0}D_{x}^{\nu-k-1} f(0), \quad n-1 < \nu \leq n \in \mathbb{N} \tag{2.77}$$

Proof: The result is obvious for $\nu = 0$. For negative orders,

$$\mathscr{L}\left[{}_{0}D_{x}^{\nu} f(x)\right] = \mathscr{L}\left[\frac{1}{\Gamma(-\nu)}\int_{0}^{x} (x-\xi)^{-\nu-1} f(\xi)\, d\xi\right] \tag{2.78}$$

---

[32] Podlubny (1999, p. 203).

[33] Miller *et al.* (1993, p. 69 and 123); Podlubny (1999, p. 104-105).

Using Theorem 2.6, this becomes

$$\mathscr{L}\left[{}_0D_x^\nu f(x)\right]=\frac{1}{\Gamma(-\nu)}\mathscr{L}\left[x^{-\nu-1}\right]\mathscr{L}\left[f(x)\right] \tag{2.79}$$

and using Lemma 2.1 we get

$$\mathscr{L}\left[{}_0D_x^\nu f(x)\right]=\frac{1}{\Gamma(-\nu)}s^{-1+\nu+1}\Gamma(-\nu-1+1)F(s)=s^\nu F(s) \tag{2.80}$$

For positive orders,

$$\mathscr{L}\left[{}_0D_x^\nu f(x)\right]=\mathscr{L}\left[D^n\,{}_0D_x^{\nu-n} f(x)\right] \tag{2.81}$$

and this, by Theorem 2.5, is

$$\mathscr{L}\left[{}_0D_x^\nu f(x)\right]=s^n s^{\nu-n}F(s)-\sum_{k=1}^{n}s^{n-k}D^{k-1}\,{}_0D_x^{\nu-n}f(0)=$$
$$=s^\nu F(s)-\sum_{k=0}^{n-1}s^{n-k-1}\,{}_0D_x^{\nu-n+k}f(0)=s^\nu F(s)-\sum_{k=0}^{n-1}s^{k}\,{}_0D_x^{\nu-k-1}f(0) \tag{2.82}$$

Q.E.D.

Theorem 2.10[34]**:** The Laplace transform of the Caputo definition of *D* is

$$\mathscr{L}\left[{}_0D_x^\nu f(x)\right]=s^\nu F(s),\ \nu\le 0 \tag{2.83}$$
$$\mathscr{L}\left[{}_0D_x^\nu f(x)\right]=s^\nu F(s)-\sum_{k=0}^{n-1}s^{\nu-k-1}D^k f(0),\ n-1<\nu\le n\in\mathbb{N} \tag{2.84}$$

Proof: The sole difference for Theorem 2.9 is found in the positive order case. Instead of (2.81) we shall have

$$\mathscr{L}\left[{}_0D_x^\nu f(x)\right]=\mathscr{L}\left[{}_0D_x^{\nu-n}D^n f(x)\right]=s^{\nu-n}\left[s^nF(s)-\sum_{k=1}^{n}s^{n-k}D^{k-1}f(0)\right]=$$
$$=s^\nu F(s)-\sum_{k=0}^{n-1}s^{\nu-k-1}D^k f(0) \tag{2.85}$$

Q.E.D.

Remark: It is because initial conditions in (2.84) involve integer order derivatives of function *f* only that the Caputo definition of *D* is preferred to the Riemann-Liouville definition in applications where integer order derivatives of involved variables have well-established physical meanings and can be easily obtained by experimental means[35].

[34] Podlubny (1999, p. 80 and 106).
[35] See Ortigueira *et al.* (2005) for a discussion of the problems that using the Riemann-Liouville

Theorem 2.11[36]: The Laplace transform of the Grünwald-Letnikoff definition of *D* is

$$\mathscr{L}\left[{}_0D_x^{\nu} f\left(x\right)\right] = s^{\nu} F\left(s\right) \tag{2.86}$$

Remark: The resolution of fractional order differential equations by means of the Laplace transform is done in the same way as with integer order differential equations[37].

definition brings.

[36] Podlubny (1999, p. 104 and 106-108). The demonstration is skipped since it is wearisome and has to deal with technicalities related to the sense in which integrals converge.

[37] Miller *et al.* (1993, p. 133-152); Podlubny (1999, p. 137-147). Miller *et al.* (1993, p. 321-330), presents several Laplace transforms of functions that appear often when solving differential equations.

3. Fractional calculus in control

This chapter deals with the application of fractional calculus to control. It is organised as follows:

❖ Section 3.1 addresses fractional order plants and describes usual means of representing them, viz., fractional transfer functions and fractional state-space representations. Issues like controllability, observability and stability are handled.

❖ Section 3.2 analyses a simple yet important fractional transfer function, s^ν, in respect of time and frequency responses.

❖ Section 3.3 addresses the issue of approximating fractional order transfer functions by means of integer order functions. The several continuous and discrete approximations are expounded and their goodness assessed with respect to the closeness of their time and frequency responses to the ideal ones.

3.1. Fractional order plants

3.1.1. Fractional transfer functions

There are many dynamic systems whose behaviour can be modelled by means of differential equations involving fractional derivatives. (See section 1.3 for a list of *some* of them together with bibliographical references.) Applying Laplace transforms to such equations, and assuming zero initial conditions, causes transfer functions with non-integer[38] powers of the Laplace transform variable s to appear.

Since numbers for engineering purposes are known up to a finite number of significant figures only, such non-integer powers are in reality expected to be fractional. Complex derivatives are not used for control purposes (save as a trick for obtaining a more compact notation when some controllers are devised, as shall be seen below in section 5.3).

Now although any general expression of the form

$$F(s) = \frac{\sum_{i=1}^{M} b_i s^{q_i}}{\sum_{j=1}^{N} a_j s^{p_j}}, \quad M, N \in \mathbb{N}_0 \wedge a_j, b_i, p_j, q_i \in \mathbb{R} \tag{3.1}$$

may appear, it is not unusual that all exponents of s be multiples of some value:

$$F(s) = \frac{\sum_{i=0}^{M} b_i s^{iq}}{\sum_{j=0}^{N} a_j s^{jq}}, \quad M, N \in \mathbb{N}_0 \wedge q, a_j, b_i \in \mathbb{R} \tag{3.2}$$

[38] It is always good to recall that fractional calculus is a misname and that irrational numbers are handled exactly as fractions are.

Transfer functions such as (3.2) are called *commensurate*. And although any real q suffices for the function being commensurate, q is usually the inverse of a natural number:

$$F(s)=\frac{\sum_{i=0}^{M} b_i s^{i/Q}}{\sum_{j=0}^{N} a_j s^{j/Q}}, \quad M,N\in\mathbb{N}_0 \wedge Q\in\mathbb{N} \wedge a_j, b_i \in\mathbb{R} \tag{3.3}$$

Fractional transfer functions appearing in practice need not be so. But those that are so are far more manageable (as shall be seen below) and show up very often. Usual, integer-order transfer functions are particular cases of (3.3), when $Q = 1$.

3.1.2. Fractional transfer function matrixes

A transfer function such as (3.3) is fit for describing the relation between the input and the output of a SISO plant. If a fractional plant is MIMO, its n inputs and m outputs shall be related by means of a $n\times m$ transfer function matrix—a matrix with elements similar to (3.3).

3.1.3. Fractional state-space representations

Fractional systems (with a rational commensurate order) also allow for a state-space representation:

$$\begin{aligned} D^{1/Q}\mathbf{x} &= \mathbf{Ax}+\mathbf{Bu} \\ \mathbf{y} &= \mathbf{Cx}+\mathbf{Du} \end{aligned} \tag{3.4}$$

In (3.4), **u** are the inputs, **x** the states, and **y** the outputs of the system. Usual integer state-space representations of dynamical systems are particular cases of (3.4), with $Q = 1$.

Remark: Just as with integer state-space representations, fractional state-space representations are not unique; in other words, there are many fractional state-space representations corresponding to the same fractional transfer function.

Theorem 3.1: State-space representation (3.4) corresponds to transfer function matrix

$$\mathbf{y}=\left[\mathbf{C}\left(s^{1/Q}\mathbf{I}-\mathbf{A}\right)^{-1}\mathbf{B}+\mathbf{D}\right]\mathbf{u} \tag{3.5}$$

provided that no non-null initial conditions appear when Laplace transforms are applied to the first equality.

Proof: Thanks to the last hypothesis, we may write

$$s^{1/Q}\mathbf{x} = \mathbf{Ax} + \mathbf{Bu} \Leftrightarrow \left(s^{1/Q}\mathbf{I} - \mathbf{A}\right)\mathbf{x} = \mathbf{Bu} \Rightarrow \mathbf{x} = \left(s^{1/Q}\mathbf{I} - \mathbf{A}\right)^{-1}\mathbf{Bu} \tag{3.6}$$

which, replaced in the second equality of (3.4), gives (3.5). **Q.E.D.**

Remark: It really should be stressed that a transfer function may be obtained from a state-space representation only if initial conditions allow for it—and if the Riemann-Liouville definition of fractional derivatives is used, these initial conditions may include fractional derivatives of functions. For that reason the Caputo and Grünwald-Letnikoff definitions are often used instead.

The following results are easily proved from Theorem 3.1, just like its integer counterparts[39].

Theorem 3.2: Transfer function

$$F(s) = \frac{\sum_{i=0}^{N} b_i s^{i/Q}}{s^{N/Q} + \sum_{i=0}^{N-1} a_i s^{i/Q}}, \quad N, Q \in \mathbb{N} \wedge a_i, b_i \in \mathbb{R} \tag{3.7}$$

is equivalent to the following state-space representation, known as controllable canonical form, provided that the conditions for applying Theorem 3.1 are verified:

$$D^{1/Q}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_{N-1} \\ x_N \end{bmatrix} = \begin{bmatrix} -a_{N-1} & -a_{N-2} & -a_{N-3} & \cdots & -a_1 & -a_0 \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_{N-1} \\ x_N \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} u$$

$$y = \begin{bmatrix} b_{N-1} - a_{N-1}b_N & | & b_{N-2} - a_{N-2}b_N & | & \cdots & | & b_0 - a_0 b_N \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} + b_N u \tag{3.8}$$

Theorem 3.3: Transfer function (3.7) is equivalent to the following state-space representation, known as observable canonical form, provided that the conditions for applying Theorem 3.1 are verified:

[39] These are given for instance by Ogata (1997, p. 711-713).

$$\begin{aligned}
&D^{1/Q}\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{N-1} \\ x_N \end{bmatrix} = \begin{bmatrix} 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & \cdots & 0 & -a_1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & -a_{N-2} \\ 0 & 0 & \cdots & 1 & -a_{N-1} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{N-1} \\ x_N \end{bmatrix} + \begin{bmatrix} b_0 - a_0 b_N \\ b_1 - a_1 b_N \\ \vdots \\ b_{N-2} - a_{N-2} b_N \\ b_{N-1} - a_{N-1} b_N \end{bmatrix} u \\
&y = \begin{bmatrix} 0 & | & \cdots & | & 0 & | & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_{N-1} \\ x_N \end{bmatrix} + b_N u
\end{aligned} \tag{3.9}$$

Theorem 3.4: If transfer function (3.7) has no repeated poles, it is possible to rewrite it, using a partial fraction expansion, thus:

$$F(s) = b_N + \sum_{i=1}^{N} \frac{c_i}{s - p_i} \tag{3.10}$$

and this is equivalent to the following state-space representation, known as diagonal canonical form, provided that the conditions for applying Theorem 3.1 are verified:

$$\begin{aligned}
&D^{1/Q}\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} = \begin{bmatrix} p_1 & 0 & \cdots & 0 \\ 0 & p_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & p_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} u \\
&y = \begin{bmatrix} c_1 & | & c_2 & | & \cdots & | & c_N \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} + b_N u
\end{aligned} \tag{3.11}$$

Theorem 3.5: If transfer function (3.7) has a pole of order n (which we will assume to be the first, without loss of generality), it is possible to rewrite it, using a partial fraction expansion, thus:

$$F(s) = b_N + \frac{c_1}{(s-p_1)^n} + \frac{c_2}{(s-p_1)^{n-1}} + \ldots + \frac{c_{n-1}}{(s-p_1)^2} + \frac{c_n}{s-p_1} + \sum_{i=n+1}^{N} \frac{c_i}{s-p_i} \tag{3.12}$$

and this is equivalent to the following state-space representation, known as Jordan canonical form, provided that the conditions for applying Theorem 3.1 are verified:

$$D^{1/Q}\left[\begin{array}{c} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_{n-2} \\ x_{n-1} \\ x_n \\ \hline x_{n+1} \\ \vdots \\ x_N \end{array}\right]=\left[\begin{array}{ccccccc|ccc} p_1 & 1 & 0 & \cdots & 0 & 0 & 0 & & & \\ 0 & p_1 & 1 & \cdots & 0 & 0 & 0 & & & \\ 0 & 0 & p_1 & \cdots & 0 & 0 & 0 & & & \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & & 0 & \\ 0 & 0 & 0 & \cdots & p_1 & 1 & 0 & & & \\ 0 & 0 & 0 & \cdots & 0 & p_1 & 1 & & & \\ 0 & 0 & 0 & \cdots & 0 & 0 & p_1 & & & \\ \hline & & & & & & & p_{n+1} & \cdots & 0 \\ & & & 0 & & & & \vdots & \ddots & \vdots \\ & & & & & & & 0 & \cdots & p_N \end{array}\right]\left[\begin{array}{c} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_{n-2} \\ x_{n-1} \\ x_n \\ \hline x_{n+1} \\ \vdots \\ x_N \end{array}\right]+\left[\begin{array}{c} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 1 \\ \hline 1 \\ \vdots \\ 1 \end{array}\right]u$$

$$y=\left[\begin{array}{c|c|c|c} c_1 & c_2 & \cdots & c_N \end{array}\right]\left[\begin{array}{c} x_1 \\ x_2 \\ \vdots \\ x_N \end{array}\right]+b_N u \tag{3.13}$$

Other multiple poles are handled likewise.

Remark: Notice that (3.7) is equivalent to (3.3) with the only restriction that $M \leq N$. It need not be $M = N$ since the last *b* coefficients may well be zero.

3.1.4. Observability, controllability and stability of fractional plants

The two following results may be proved in a manner similar to that of their integer-order counterparts[40].

Theorem 3.6: Fractional system (3.4) is observable if and only if the observability matrix

$$\mathscr{O}=\begin{bmatrix} \mathbf{C} \\ \mathbf{CA} \\ \vdots \\ \mathbf{CA}^{N-1} \end{bmatrix} \tag{3.14}$$

(where *N* is the number of states) is a full-rank matrix.

Theorem 3.7: Fractional system (3.4) is controllable if and only if the controllability matrix

$$\mathscr{C}=\begin{bmatrix} \mathbf{B} & \mathbf{AB} & \cdots & \mathbf{A}^{N-1}\mathbf{B} \end{bmatrix} \tag{3.15}$$

[40] Vinagre (2001, p. 59-64). The major difference for the integer case is that, instead of the exponential function, function E_t shows up. See the Appendix A, section A.2.5, on function E_t.

(where N is the number of states) is a full-rank matrix.

The following result is quoted without proof.

Theorem 3.8 (Matignon's theorem[41]): System (3.3) is stable (in the sense that bounded inputs will always lead to bounded outputs) if and only if all (complex) roots r of its denominator verify

$$\left|\arg(r)\right| > \frac{\pi}{2Q} \tag{3.16}$$

Remark 1: The branch of function argument that has $\left]-\pi;+\pi\right]$ rad as counter-domain is assumed.

Remark 2: The usual stability condition for integer systems is obtained from (3.16) if $Q = 1$.

3.1.5. Other fractional plants

It is possible to conceive fractional order plants where the non-integer exponents do not affect the Laplace transform variable s only. Among such plants the simplest will correspond to transfer functions given by

$$F(s) = \left[G(s)\right]^{\nu} \tag{3.17}$$

where $G(s)$ is some integer order transfer function (other than s). Linear combinations of terms given by (3.17) and products and ratios thereof are also possible.

Such models appear seldom and will not be considered in what follows.

3.2. Function s^{ν}

Function

$$F(s) = s^{\nu} \tag{3.18}$$

is not only the simplest fractional order transfer function that may appear but is also very important for applications, as shall be seen subsequently. For that reason, this section analyses its time and frequency responses.

3.2.1. Time responses of s^{ν}

So as to find the time responses of (3.18), some auxiliary results are needed.

Theorem 3.9: The derivatives of the exponential function are given by

[41] Malti *et al.* (2003, p. 5).

$$ {}_0D_t^{\nu} e^{at} = E_t(-\nu, a), \quad t > 0 \tag{3.19}$$

Proof: For negative orders, we have, from definition (2.41),

$$ {}_0D_t^{-\nu} e^{at} = \frac{1}{\Gamma(\nu)} \int_0^t (t-\xi)^{\nu-1} e^{a\xi} d\xi, \ \nu \in \mathbb{R}^+ \tag{3.20}$$

By means of the substitution $x = t - \xi$, in the first place, and of the substitution $ax = y$, in the second place, we obtain

$$ {}_0D_t^{-\nu} e^{at} = -\frac{1}{\Gamma(\nu)} \int_t^0 x^{\nu-1} e^{a(t-x)} dx = \frac{e^{at}}{\Gamma(\nu)} \int_0^t x^{\nu-1} e^{-ax} dx =$$
$$= \frac{e^{at}}{\Gamma(\nu)} \int_0^{at} \left(\frac{y}{a}\right)^{\nu-1} e^{-y} \frac{dy}{a} = \frac{e^{at}}{\Gamma(\nu) a^{\nu}} \int_0^{at} y^{\nu-1} e^{-y} dy = E_t(\nu, a) \tag{3.21}$$

For positive orders, the same definition gives

$$ {}_0D_t^{\nu} e^{at} = D^n \, {}_0D_t^{\nu-n} e^{at} = \frac{d^n}{dt^n} E_t(n-\nu, a) = E_t(-\nu, a),$$
$$\nu \in \mathbb{R}^+ \wedge n = \min\{k \in \mathbb{N} : k > \nu\} \tag{3.22}$$

(The last equality stems from (A.67).)

Finally, if $\nu = 0$, (A.68) shows that

$$E_t(0, a) = \sum_{k=0}^{+\infty} \frac{(at)^k}{\Gamma(k+1)} \tag{3.23}$$

which is the series development of e^{at}. **Q.E.D.**

Theorem 3.10: The Laplace transform of E_t is

$$\mathscr{L}[E_t(\nu, a)] = \frac{1}{s^{\nu}(s-a)} \tag{3.24}$$

Proof: For negative orders, applying the convolution theorem (2.46) to the right-hand member of (3.20) we obtain

$$\mathscr{L}[E_t(\nu, a)] = \frac{1}{\Gamma(\nu)} \mathscr{L}\left[t^{\nu-1}\right] \mathscr{L}\left[e^{at}\right] = \frac{1}{s^{\nu}(s-a)} \tag{3.25}$$

For positive orders, applying the Laplace transform to the last equality of (3.22),

$$\mathscr{L}[E_t(-\nu, a)] = \mathscr{L}\left[\frac{d^n}{dt^n} E_t(n-\nu, a)\right] = s^n \frac{1}{s^{n-\nu}(s-a)} = \frac{1}{s^{-\nu}(s-a)} \tag{3.26}$$

And when $\nu = 0$, (3.24) also yields the correct result:

$$\mathscr{L}\left[E_t\left(0,a\right)\right] = \mathscr{L}\left[e^{at}\right] = \frac{1}{s-a} \tag{3.27}$$

Q.E.D.

Theorem 3.11: The impulse response of (3.18) is

$$\frac{t^{-\nu-1}}{\Gamma\left(-\nu\right)}, \quad t > 0 \tag{3.28}$$

provided that $\nu \notin \mathbb{N}$ (otherwise the denominator would be infinite).

Proof[42]: By letting $a = 0$ in (3.24) we get

$$\mathscr{L}\left[E_t\left(\nu,0\right)\right] = \frac{1}{s^{\nu+1}} \tag{3.29}$$

Hence

$$\mathscr{L}^{-1}\left[s^{\nu}\right] = E_t\left(-\nu-1,0\right) \tag{3.30}$$

which (A.64) shows to be (3.28). **Q.E.D.**

Corollary: The step response of (3.18) is given by

$$\frac{t^{-\nu}}{\Gamma\left(-\nu+1\right)}, \quad t > 0 \tag{3.31}$$

provided that $\nu - 1 \notin \mathbb{N}$.

Proof: The step response of (3.18) may be easily reckoned from (3.28) whenever it may be applied:

$$\mathscr{L}^{-1}\left\{G\left(s\right)\mathscr{L}\left[1\right]\right\} = \mathscr{L}^{-1}\left\{G\left(s\right)\times\frac{1}{s}\right\} = \mathscr{L}^{-1}\left\{s^{\nu-1}\right\} = \frac{t^{-\nu}}{\Gamma\left(-\nu+1\right)} \tag{3.32}$$

Q.E.D.

3.2.2. Bode and Nichols plots of s^{ν} for real orders

The frequency response of (3.18) is[43]

---

[42] Oustaloup (1991, p. 368-369) gives an alternative proof.
[43] Oustaloup (1991, p. 366-367).

$$F(j\omega) = (j\omega)^{\nu} \tag{3.33}$$

$$\left|F(j\omega)\right| = \left|j^{\nu}\omega^{\nu}\right| = \left|\omega^{\nu}\right| = \omega^{\nu} \tag{3.34}$$

$$\arg\left[F(j\omega)\right] = \arg\left(j^{\nu}\omega^{\nu}\right) = \arg\left(j^{\nu}\right) \tag{3.35}$$

Now there are several complex numbers z with different arguments such that $z = j^{\nu}$; by choosing the one with a lower argument in interval $[0; 2\pi[$ we will obtain

$$\arg\left[F(j\omega)\right] = \nu\frac{\pi}{2} \tag{3.36}$$

The gain in decibel shall be

$$\left|F(j\omega)\right| = 20\log_{10}\omega^{\nu} = 20\nu\log_{10}\omega \quad (\text{dB}) \tag{3.37}$$

Thus the Bode and Nichols plots of $F(s) = s^{\nu}$ are those shown in Figure 3.1 and Figure 3.2.

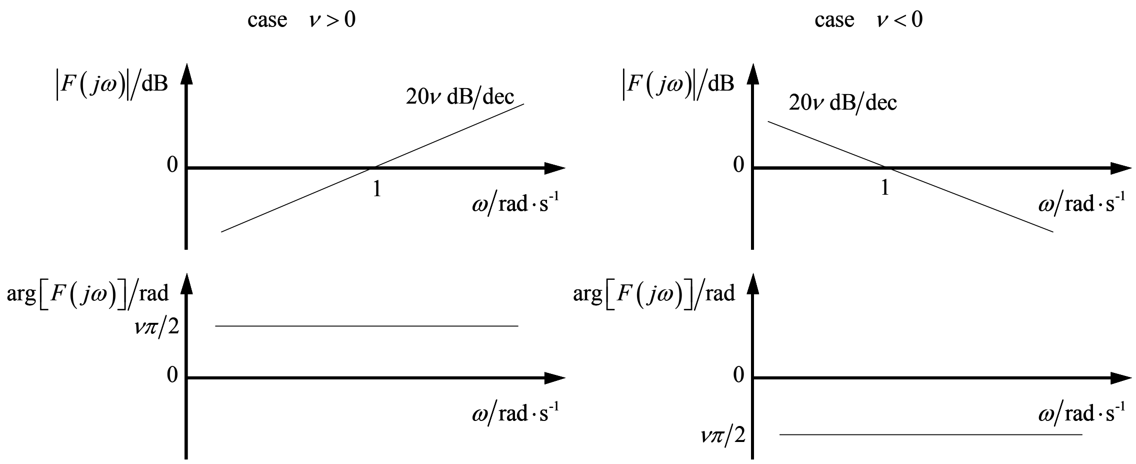


Figure 3.1 — Bode diagrams of (3.18)

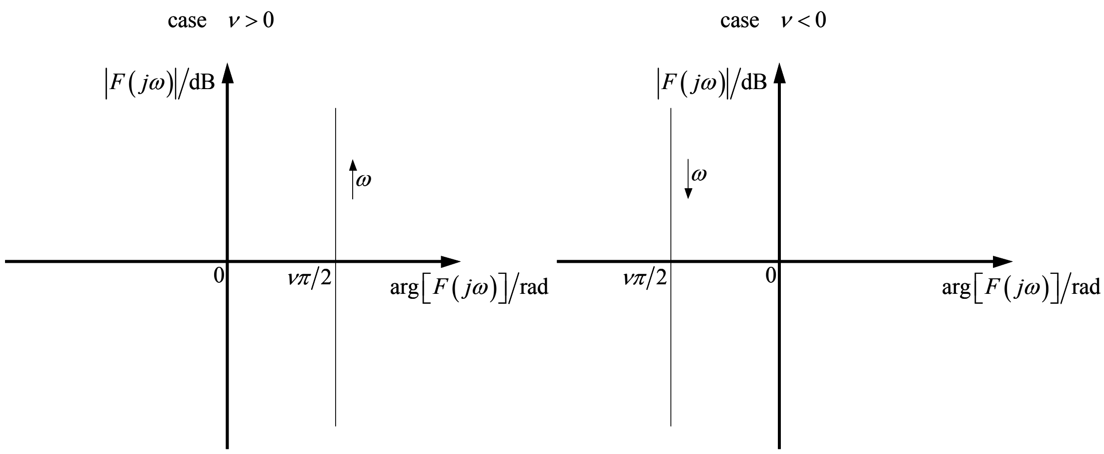


Figure 3.2 — Nichols diagrams of (3.18)

3.2.3. Bode and Nichols plots of s^ν for complex orders

Let[44]

$$G(s) = s^{a+jb} \tag{3.38}$$

Using (A.18) we get

$$G(s) = s^a \cos(b\log s) + js^a \sin(b\log s) \tag{3.39}$$

Let us define

$$G_r(s) = \mathrm{Re}\left[G(s)\right] = s^a \cos(b\log s) \tag{3.40}$$

$$G_i(s) = \mathrm{Im}\left[G(s)\right] = s^a \sin(b\log s) \tag{3.41}$$

These are the two functions whose Bode and Nichols plots we are interested in[45].

❖ First case: transfer function (3.40)

$$\begin{aligned} G_r(j\omega) &= j^a\omega^a \cos(b\log j\omega) = \\ &= j^a\omega^a \cos(b\log j + b\log\omega) = j^a\omega^a \cos\left(bj\frac{\pi}{2} + b\log\omega\right) = \\ &= j^a\omega^a \left[\cos\left(jb\frac{\pi}{2}\right)\cos(b\log\omega) - \sin\left(jb\frac{\pi}{2}\right)\sin(b\log\omega)\right] = \\ &= j^a\omega^a \left[\cosh\left(b\frac{\pi}{2}\right)\cos(b\log\omega) - j\sinh\left(b\frac{\pi}{2}\right)\sin(b\log\omega)\right] \end{aligned} \tag{3.42}$$

Thus the gain will be

$$\left|G_r(j\omega)\right| = \omega^a \left[\cosh^2\left(b\frac{\pi}{2}\right)\cos^2(b\log\omega) + \sinh^2\left(b\frac{\pi}{2}\right)\sin^2(b\log\omega)\right]^{\frac{1}{2}} \tag{3.43}$$

It is now expedient to make the following substitutions:

$$\cosh^2 x = \left(\frac{e^x + e^{-x}}{2}\right)^2 = \frac{e^{2x} + 2e^0 + e^{-2x}}{4} = \frac{\cosh(2x)+1}{2} \tag{3.44}$$

$$\sinh^2 x = \left(\frac{e^x - e^{-x}}{2}\right)^2 = \frac{e^{2x} - 2e^0 + e^{-2x}}{4} = \frac{\cosh(2x)-1}{2} \tag{3.45}$$

44 In all this subsection expressions from Appendix A, section A.1, will be repeatedly used. References will often not be given for they would be too numerous.

45 Oustaloup (1991), p. 293-300.

Hence

$$\left|G_r(j\omega)\right| = \omega^a \left[\frac{\cosh(b\pi)+1}{2}\cos^2(b\log\omega) + \frac{\cosh(b\pi)-1}{2}\sin^2(b\log\omega)\right]^{\frac{1}{2}} =$$

$$= \omega^a \left[\frac{\cosh(b\pi)+\cos^2(b\log\omega)-\sin^2(b\log\omega)}{2}\right]^{\frac{1}{2}} =$$

$$= \omega^a \left[\frac{\cosh(b\pi)+\cos(2b\log\omega)}{2}\right]^{\frac{1}{2}} \tag{3.46}$$

The phase will be

$$\arg\left[G_r(j\omega)\right] =$$

$$= \arg\left(j^a\omega^a\right) + \arg\left[\cosh\left(b\frac{\pi}{2}\right)\cos(b\log\omega) - j\sinh\left(b\frac{\pi}{2}\right)\sin(b\log\omega)\right] =$$

$$= a\frac{\pi}{2} + \operatorname{arctg}\frac{-\sinh\left(b\frac{\pi}{2}\right)\sin(b\log\omega)}{\cosh\left(b\frac{\pi}{2}\right)\cos(b\log\omega)} = a\frac{\pi}{2} - \operatorname{arctg}\left[\operatorname{tgh}\left(b\frac{\pi}{2}\right)\operatorname{tg}(b\log\omega)\right] \tag{3.47}$$

Even though neither gain nor phase are linear, expressions (3.46) and (3.47) correspond to a Bode diagram very close from linearity, especially if $|b|>1$. Let us linearise around $\omega=1$. We will have

$$\left|G_r(j\omega)\right|\Big|_{\omega=1} = \left[\frac{\cosh(b\pi)+\cos 0}{2}\right]^{\frac{1}{2}} = \left[\cosh^2\left(b\frac{\pi}{2}\right)\right]^{\frac{1}{2}} = \cosh\left(b\frac{\pi}{2}\right) \tag{3.48}$$

$$\arg\left[G_r(j\omega)\right]\Big|_{\omega=1} = a\frac{\pi}{2} - \operatorname{arctg}\left[\operatorname{tgh}\left(b\frac{\pi}{2}\right)\operatorname{tg} 0\right] = a\frac{\pi}{2} \tag{3.49}$$

In what concerns the slope of the gain, it is expedient to reckon, in the first place,

$$\frac{d\left|G_r(j\omega)\right|}{d\log\omega} = ae^{a\log\omega}\left[\frac{\cosh(b\pi)+\cos(2b\log\omega)}{2}\right]^{\frac{1}{2}} +$$

$$+ e^{a\log\omega}\frac{1}{2}\left[\frac{\cosh(b\pi)+\cos(2b\log\omega)}{2}\right]^{-\frac{1}{2}}\frac{-2b\sin(2b\log\omega)}{2} \tag{3.50}$$

$$\left.\frac{d\left|G_r(j\omega)\right|}{d\log\omega}\right|_{\omega=1} = a\left[\frac{\cosh(b\pi)+\cos 0}{2}\right]^{\frac{1}{2}} - \frac{1}{2}\left[\frac{\cosh(b\pi)+\cos 0}{2}\right]^{-\frac{1}{2}} b\sin 0 =$$

$$= a\left[\cosh^2\left(b\frac{\pi}{2}\right)\right]^{\frac{1}{2}} = a\cosh\left(b\frac{\pi}{2}\right) \tag{3.51}$$

And since

$$\log_a b = \frac{\log b}{\log a} \Rightarrow d\log_a b = \frac{d\log b}{\log a} \tag{3.52}$$

we will have

$$\left.\frac{d\left|G_r\left(j\omega\right)\right|}{d\log_{10}\omega}\right|_{\omega=1} = \left.\frac{d\left|G_r\left(j\omega\right)\right|}{d\log\omega}\right|_{\omega=1} \log 10 = a\log 10\cosh\left(b\frac{\pi}{2}\right) \tag{3.53}$$

Finally, by the chain rule of derivatives,

$$\frac{d\left[20\log_{10}\left|G_r\left(j\omega\right)\right|\right]}{d\log_{10}\omega} = \frac{20}{\log 10}\frac{\frac{d\left|G_r\left(j\omega\right)\right|}{d\log_{10}\omega}}{\left|G_r\left(j\omega\right)\right|} \tag{3.54}$$

and so

$$\left.\frac{d\left[20\log_{10}\left|G_r\left(j\omega\right)\right|\right]}{d\log_{10}\omega}\right|_{\omega=1} = \frac{20}{\log 10}\frac{a\log 10\cosh\left(b\frac{\pi}{2}\right)}{\cosh\left(b\frac{\pi}{2}\right)} = 20a \tag{3.55}$$

In what concerns the slope of the phase, it is expedient to reckon, in the first place,

$$\frac{d\arg\left[G_r\left(j\omega\right)\right]}{d\log\omega} = -\frac{1}{1+\operatorname{tgh}^2\left(b\frac{\pi}{2}\right)\operatorname{tg}^2\left(b\log\omega\right)}\operatorname{tgh}\left(b\frac{\pi}{2}\right)\frac{b}{\cos^2\left(b\log\omega\right)} \tag{3.56}$$

$$\left.\frac{d\arg\left[G_r\left(j\omega\right)\right]}{d\log\omega}\right|_{\omega=1} = -\frac{1}{1+\operatorname{tgh}^2\left(b\frac{\pi}{2}\right)\operatorname{tg}^2 0}\operatorname{tgh}\left(b\frac{\pi}{2}\right)\frac{b}{\cos^2 0} = -b\operatorname{tgh}\left(b\frac{\pi}{2}\right) \tag{3.57}$$

Applying (3.52)

$$\left.\frac{d\arg\left[G_r\left(j\omega\right)\right]}{d\log_{10}\omega}\right|_{\omega=1} = -b\log 10\operatorname{tgh}\left(b\frac{\pi}{2}\right) \tag{3.58}$$

The Bode and Nichols plots will be as in Figure 3.3 and Figure 3.4.

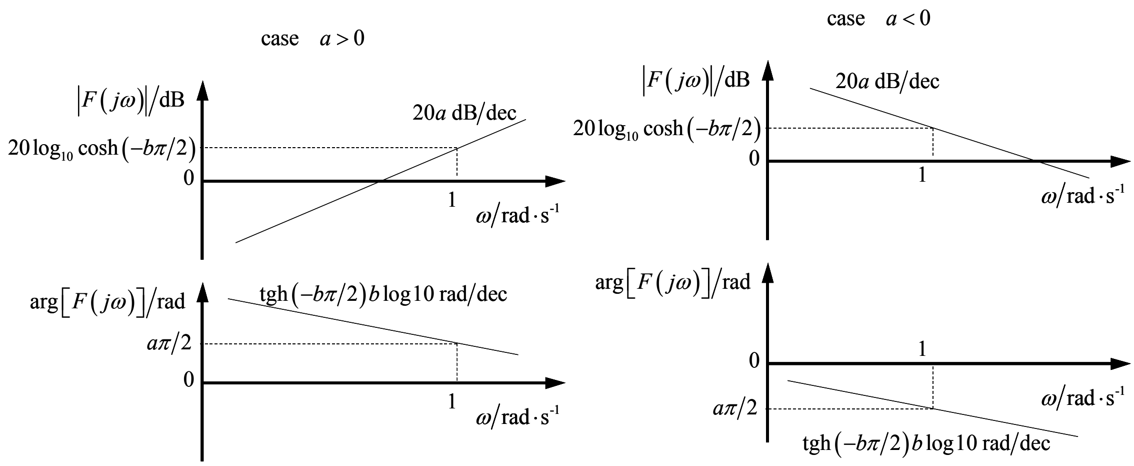


Figure 3.3 — Linearised Bode diagrams of (3.40)

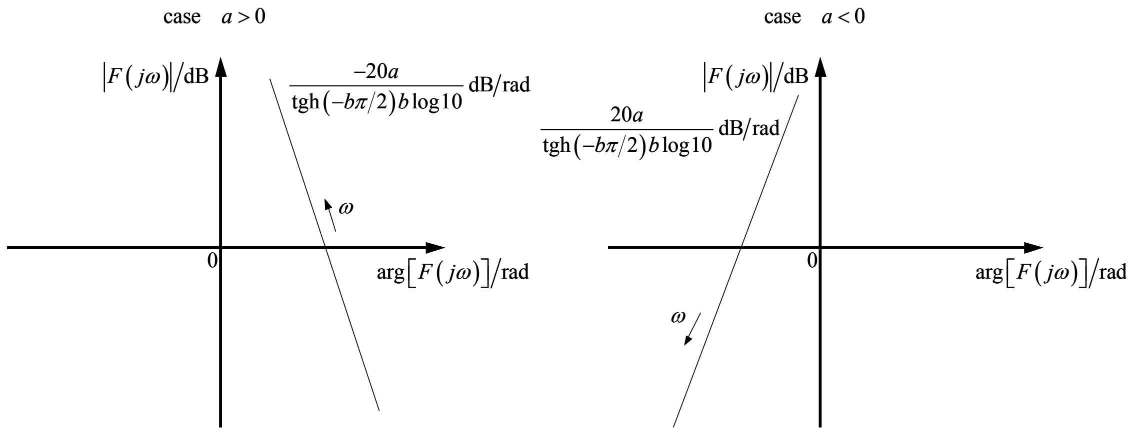


Figure 3.4 — Linearised Nichols diagrams of (3.40)

❖ Second case: transfer function (3.41)

$$G_i(j\omega)=j^a\omega^a\sin(b\log j\omega)=$$
$$=j^a\omega^a\sin(b\log j+b\log\omega)=j^a\omega^a\sin\left(bj\frac{\pi}{2}+b\log\omega\right)=$$
$$=j^a\omega^a\left[\sin\left(jb\frac{\pi}{2}\right)\cos(b\log\omega)+\cos\left(jb\frac{\pi}{2}\right)\sin(b\log\omega)\right]=$$
$$=j^a\omega^a\left[j\sinh\left(b\frac{\pi}{2}\right)\cos(b\log\omega)+\cosh\left(b\frac{\pi}{2}\right)\sin(b\log\omega)\right] \tag{3.59}$$

Thus the gain will be

$$|G_i(j\omega)|=\omega^a\left[\cosh^2\left(b\frac{\pi}{2}\right)\sin^2(b\log\omega)+\sinh^2\left(b\frac{\pi}{2}\right)\cos^2(b\log\omega)\right]^{\frac{1}{2}}=$$
$$=\omega^a\left[\frac{\cosh(b\pi)+1}{2}\sin^2(b\log\omega)+\frac{\cosh(b\pi)-1}{2}\cos^2(b\log\omega)\right]^{\frac{1}{2}}=$$

$$= \omega^a \left[\frac{\cosh(b\pi) + \sin^2(b \log \omega) - \cos^2(b \log \omega)}{2} \right]^{\frac{1}{2}} =$$

$$= \omega^a \left[\frac{\cosh(b\pi) - \cos(2b \log \omega)}{2} \right]^{\frac{1}{2}} \tag{3.60}$$

The phase will be

$$\arg\left[G_i(j\omega)\right] =$$

$$= \arg\left(j^a \omega^a\right) + \arg\left[\cosh\left(b\frac{\pi}{2}\right)\sin(b \log \omega) + j \sinh\left(b\frac{\pi}{2}\right)\cos(b \log \omega)\right] =$$

$$= a\frac{\pi}{2} + \operatorname{arctg}\frac{\sinh\left(b\frac{\pi}{2}\right)\cos(b \log \omega)}{\cosh\left(b\frac{\pi}{2}\right)\sin(b \log \omega)} =$$

$$= a\frac{\pi}{2} + \operatorname{arctg}\left[\operatorname{tgh}\left(b\frac{\pi}{2}\right)\operatorname{cotg}(b \log \omega)\right] \tag{3.61}$$

Even though the gain is not linear, expression (3.60) corresponds to a Bode diagram with a gain plot very close from linearity, especially if $|b| > 1$. Expression (3.61), beyond being non-linear, is not continuous, as will be seen in a moment. Apart from the discontinuity, the phase plot of the Bode diagram is also very close from linearity, especially if $|b| > 1$. Let us linearise around $\omega = 1$. We will have

$$\left|G_i(j\omega)\right|\Big|_{\omega=1} = \left[\frac{\cosh(b\pi) - \cos 0}{2}\right]^{\frac{1}{2}} = \left[\sinh^2\left(b\frac{\pi}{2}\right)\right]^{\frac{1}{2}} = \sinh\left(b\frac{\pi}{2}\right) \tag{3.62}$$

$$\arg\left[G_i(j\omega)\right]\Big|_{\omega=1} = a\frac{\pi}{2} + \operatorname{arctg}\left[\operatorname{tgh}\left(b\frac{\pi}{2}\right)\operatorname{cotg} 0\right] = a\frac{\pi}{2} \pm \frac{\pi}{2} \tag{3.63}$$

Both these values will be met, depending on the way we approach frequency $\omega = 1$: if $\omega \to 1^+$, then $\log \omega$ will approach 0 by positive values, and the cotangent will be evaluated in the first quadrant, resulting in a positive arc of tangent; if $\omega \to 1^-$, then $\log \omega$ will approach 0 by negative values, and the cotangent will be evaluated in the fourth quadrant, resulting in a negative arc of tangent.

In what concerns the slope of the gain, it is expedient to reckon, in the first place,

$$\frac{d\left|G_i(j\omega)\right|}{d \log \omega} = a e^{a \log \omega}\left[\frac{\cosh(b\pi) - \cos(2b \log \omega)}{2}\right]^{\frac{1}{2}} +$$

$$+ e^{a \log \omega} \frac{1}{2}\left[\frac{\cosh(b\pi) - \cos(2b \log \omega)}{2}\right]^{-\frac{1}{2}} \frac{2b \sin(2b \log \omega)}{2} \tag{3.64}$$

$$\left.\frac{d\left|G_i(j\omega)\right|}{d\log\omega}\right|_{\omega=1} = a\left[\frac{\cosh(b\pi)-\cos 0}{2}\right]^{\frac{1}{2}} + \frac{1}{2}\left[\frac{\cosh(b\pi)-\cos 0}{2}\right]^{-\frac{1}{2}} b\sin 0 =$$

$$= a\left[\sinh^2\left(b\frac{\pi}{2}\right)\right]^{\frac{1}{2}} = a\sinh\left(b\frac{\pi}{2}\right) \tag{3.65}$$

Applying (3.52), we will have

$$\left.\frac{d\left|G_i(j\omega)\right|}{d\log_{10}\omega}\right|_{\omega=1} = \left.\frac{d\left|G_i(j\omega)\right|}{d\log\omega}\right|_{\omega=1}\log 10 = a\log 10\sinh\left(b\frac{\pi}{2}\right) \tag{3.66}$$

Finally

$$\frac{d\left[20\log_{10}\left|G_i(j\omega)\right|\right]}{d\log_{10}\omega} = \frac{20}{\log 10}\frac{\dfrac{d\left|G_i(j\omega)\right|}{d\log_{10}\omega}}{\left|G_i(j\omega)\right|} \tag{3.67}$$

and so

$$\left.\frac{d\left[20\log_{10}\left|G_i(j\omega)\right|\right]}{d\log_{10}\omega}\right|_{\omega=1} = \frac{20}{\log 10}\frac{a\log 10\sinh\left(b\frac{\pi}{2}\right)}{\sinh\left(b\frac{\pi}{2}\right)} = 20a \tag{3.68}$$

In what concerns the slope of the phase, it is expedient to reckon, in the first place,

$$\frac{d\arg\left[G_i(j\omega)\right]}{d\log\omega} = \frac{1}{1+\operatorname{tgh}^2\left(b\frac{\pi}{2}\right)\operatorname{tg}^2(b\log\omega)}\operatorname{tgh}\left(b\frac{\pi}{2}\right)\frac{-b}{\sin^2(b\log\omega)} =$$

$$= \frac{-b\operatorname{tgh}\left(b\frac{\pi}{2}\right)}{\sin^2(b\log\omega)+\operatorname{tgh}^2\left(b\frac{\pi}{2}\right)\cos^2(b\log\omega)} \tag{3.69}$$

$$\left.\frac{d\arg\left[G_i(j\omega)\right]}{d\log\omega}\right|_{\omega=1} = \frac{-b\operatorname{tgh}\left(b\frac{\pi}{2}\right)}{\sin^2 0+\operatorname{tgh}^2\left(b\frac{\pi}{2}\right)\cos^2 0} = -b\operatorname{cotgh}\left(b\frac{\pi}{2}\right) \tag{3.70}$$

Applying (3.52)

$$\left.\frac{d\arg\left[G_i\left(j\omega\right)\right]}{d\log_{10}\omega}\right|_{\omega=1}=-b\log 10\,\mathrm{cotgh}\left(b\frac{\pi}{2}\right) \tag{3.71}$$

The Bode plots will be as in Figure 3.5. It should be noticed that in both these Figures the gain at unit frequency might be negative, since hyperbolic sinus function has the entire $\mathbb{R}$ set as counter-domain. Nichols plots are not shown since they are irrelevant for what follows.

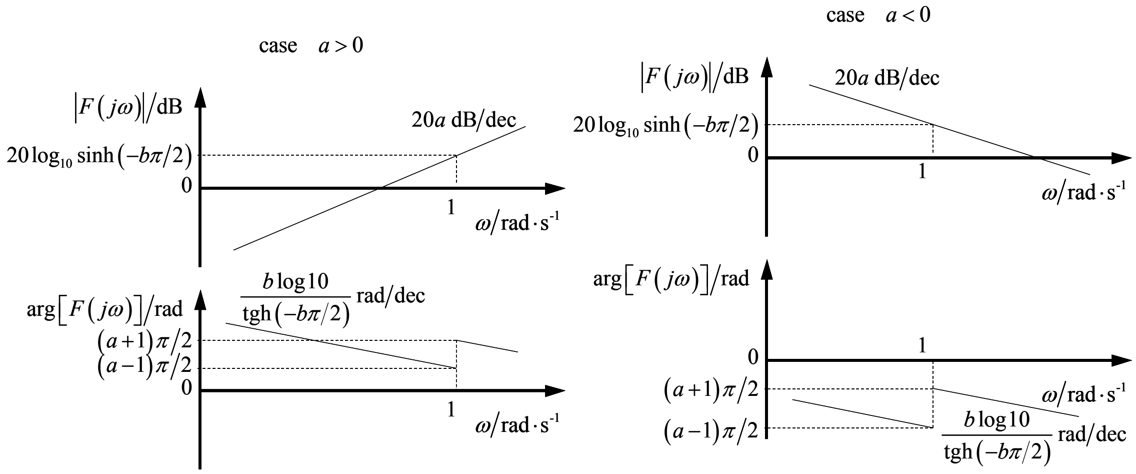


Figure 3.5 — Linearised Bode diagrams of (3.41)

Remark 1: Linearised expressions (3.48), (3.49), (3.55) and (3.58) for the case of $G_r(s)$ show that this transfer function has a Bode diagram nearly equal to that of function s^{ν}, the major difference being that the phase now has a slope. Additionally the gain is shifted up (since the hyperbolic cosinus is always positive).

Remark 2: Linearised expressions (3.62), (3.63), (3.68) and (3.71) for the case of $G_i(s)$ show a rather similar picture, but the discontinuity of the phase given by (3.63) prevents this transfer function from having much interest.

Remark 3: On the contrary, transfer function $G_r(s)$, as given by (3.40), will be of much use in what follows, because of its rather linear phase. It should be noticed, however, that the slope given by (3.58) is always negative. Should we need a positive slope for the phase, we may use function

$$G_r^{-1}(s)=s^{-a}\sec\left(b\log s\right) \tag{3.72}$$

whose phase slope is always positive. Oustaloup (1991, p. 305) combines both cases as follows:

$$G_r^*(s;a,b)=\begin{cases}s^a\cos\left(b\log s\right), & \text{if } b<0 \text{ (negative slope phase)}\\ s^{-a}\sec\left(b\log s\right), & \text{if } b>0 \text{ (positive slope phase)}\end{cases} \tag{3.73}$$

3.3. Approximations of fractional derivatives

Quantity is either discrete or continuous.

Aristotle, *Organon* I, 6

Transfer functions such as (3.3) are not easy to implement for computational purposes. Simulations are usually carried out with software prepared to deal with integer powers of s only. Controllers are often implemented in laboratories by means of the very same software. And hardware implementations of controllers are nowadays usually achieved with electronic components allowing implementation of usual integer transfer functions easily, while how fractional transfer functions can be achieved with them is not clear at all.

This makes the task of finding integer order approximations of fractional transfer functions a most important one. What is meant by this is that when simulations are to be performed or controllers are to be implemented, fractional transfer functions are usually replaced by integer transfer functions, with a behaviour close enough to the one desired, but much easier to handle.

There are many different ways of finding such approximations but unfortunately it is not possible to say that one of them is the best, because even though some are better than others in regard to certain characteristics, the relative merits of each approximation depend on the differentiation order, on whether one is more interested in an accurate frequency behaviour or in accurate time responses, on how large admissible transfer functions may be, and other factors such like these. For that reason this section shall present several alternatives and conclude with a comparison of them.

Approximations are available both in the s-domain and in the z-domain. The former shall henceforth be called continuous approximations, or approximations in the frequency domain; the latter, discrete approximations, or approximations in the time domain.

3.3.1. The Crone approximation

The Crone[46] methodology provides a continuous approximation of (3.18) based on a recursive distribution of zeros and poles[47]. Such a distribution, alternating zeros and poles at well-chosen intervals, allows building a transfer function with a gain nearly linear on the logarithm of the frequency and a phase nearly constant—being possible for the values of the slope of the gain and of the phase to correspond to those of (3.18) for any value of ν.

The structure of the approximation $\hat{F}$ is

$$\hat{F}(s) = C_0 \prod_{n=1}^{N} \frac{1+\frac{s}{\omega_{zn}}}{1+\frac{s}{\omega_{pn}}} \tag{3.74}$$

and has thus N zeros and N poles. It is clear that they will exist in an interval of

[46] Crone is the (French) acronym of *Commande Robuste d'Ordre Non-Entier*.
[47] Oustaloup (1991, p. 154-174).

frequencies only, and that consequently no fractional-like behaviour is to found at low or high frequencies. Let $\left[\omega_l;\omega_h\right]$ be the interval of frequencies we are interested in. So as to have *N* zeros and poles therein recursively placed, their frequencies will be given by

$$\begin{aligned}
&\omega_{z1} = \omega_l\sqrt{\eta} \\
&\omega_{pn} = \omega_{zn}\alpha, \quad n = 1\ldots N \\
&\omega_{zn+1} = \omega_{pn}\eta, \quad n = 1\ldots N-1 \\
&\omega_h = \omega_{pN}\sqrt{\eta}
\end{aligned} \tag{3.75}$$

if we begin with a zero or

$$\begin{aligned}
&\omega_{p1} = \omega_l\sqrt{\eta} \\
&\omega_{zn} = \omega_{pn}\alpha, \quad n = 1\ldots N \\
&\omega_{pn+1} = \omega_{zn}\eta, \quad n = 1\ldots N-1 \\
&\omega_h = \omega_{zN}\sqrt{\eta}
\end{aligned} \tag{3.76}$$

if we begin with a pole. Actually (3.75) results in an increasing gain and a positive phase, that is to say, corresponds to $\nu > 0$; while (3.76) results in a decreasing gain and a negative phase, that is to say, corresponds to $\nu < 0$. Both (3.75) and (3.76) lead to

$$\omega_h = \omega_l\left(\alpha\eta\right)^N \Rightarrow \alpha\eta = \left(\frac{\omega_h}{\omega_l}\right)^{1/N} \tag{3.77}$$

which is decomposed as

$$\alpha = \left(\alpha\eta\right)^{|\nu|} = \left(\frac{\omega_h}{\omega_l}\right)^{|\nu|/N} \tag{3.78}$$

$$\eta = \left(\alpha\eta\right)^{1-|\nu|} = \left(\frac{\omega_h}{\omega_l}\right)^{(1-|\nu|)/N} \tag{3.79}$$

The sign of ν is not considered in the expressions above since it is handled by the use of recursive relations (3.75) or (3.76).

To use this approximation, the frequency range of interest $\left[\omega_l;\omega_h\right]$ and the number of zeros and poles *N* must be fixed in advance. It is a good idea to try to see what happens with several values of *N*: low values will result in approximations with significant ripples in the gain and in the phase, while high values, although having a nearly linear gain and a nearly constant phase, may be computationally too heavy.

Remark: Unlike all approximations that follow, the Crone approximation may easily be extended to the case when the power of *s* is complex, that is to say, to approximate (3.38). Such an approximation is not under discussion here and is used only in a very specific circumstance. For that reason such extension is given below in subsection 5.3.5.

3.3.2. The Carlson approximation

The Carlson methodology also provides a continuous approximation of (3.18) which stems from rewriting that equality as[48]

$$F^{1/\nu}(s) = s \tag{3.80}$$

and then solving this equation for transfer function $F(s)$ using Newton's iterative method. Unfortunately this is only possible if $1/\nu \in \mathbb{Z}$; in other words, ν can only assume values ±1/2, ±1/3, ±1/4, ±1/5, and so on.

Newton's iterative method will lead to a sequence of approximations $F_i(s),\quad i \in \mathbb{N}$, given by

$$F_i(s) = F_{i-1}(s)\frac{\left(\frac{1}{\nu}-1\right)F_{i-1}^{1/\nu}(s)+\left(\frac{1}{\nu}+1\right)s}{\left(\frac{1}{\nu}+1\right)F_{i-1}^{1/\nu}(s)+\left(\frac{1}{\nu}-1\right)s} \tag{3.81}$$

beginning with

$$F_0(s) = 1 \tag{3.82}$$

Expression (3.81) clearly shows why it is that $1/\nu$ must be integer: if it were not, fractional powers of *s* would still appear in the approximations $F_i(s)$.

Each iteration decreases the ripples in the gain and the phase of the approximation and widens the frequency range where it is valid. Iterations should be performed until these characteristics are satisfactory, but it should be taken into account that the number of zeros and poles grows very fast. The interval of frequencies where the approximation is valid is always centred on 1. Should a different value be sought, the zeros and poles of (3.81) must be determined and multiplied by the new value desired for the centre of the interval of frequencies.

3.3.3. The Matsuda approximation

The Matsuda methodology provides continuous approximations of fractional plants obtained by identifying a model from its gain[49]. Suppose that we want to approximate (3.18), whose gain is given by (3.34). The gain must be found at several frequencies, the number thereof determining how many zeros and poles the approximation will have. For *M* zeros and *M* poles, $2M+1$ frequencies must be used. It is good to always use an odd number of frequencies; if an even number is used, the number of zeros will be the number of poles plus one, and thus the model is not proper.

---

[48] Carlson *et al.* (1964); Vinagre (2001, p. 153).

[49] Matsuda *et al.* (1993); Vinagre (1991, p. 153-154).

Let the frequencies chosen be $\omega_0, \omega_1, \omega_2, \ldots, \omega_N$. This method requires defining functions

$$\begin{aligned}
&d_0(\omega) = |F(j\omega)| \\
&d_1(\omega) = \frac{\omega - \omega_0}{d_0(\omega) - d_0(\omega_0)} \\
&d_2(\omega) = \frac{\omega - \omega_1}{d_1(\omega) - d_1(\omega_1)} \\
&\vdots \\
&d_k(\omega) = \frac{\omega - \omega_{k-1}}{d_{k-1}(\omega) - d_{k-1}(\omega_{k-1})} \\
&\vdots \\
&d_N(\omega) = \frac{\omega - \omega_{N-1}}{d_{N-1}(\omega) - d_{N-1}(\omega_{N-1})}
\end{aligned} \quad (3.83)$$

It is then expedient to build the $(N+1)\times(N+1)$ superior triangular matrix

$$\mathbf{D} = \begin{bmatrix} d_0(\omega_0) & d_0(\omega_1) & d_0(\omega_2) & \cdots & d_0(\omega_N) \\ & d_1(\omega_1) & d_1(\omega_2) & \cdots & d_1(\omega_N) \\ & & d_2(\omega_2) & \cdots & d_2(\omega_N) \\ & & & \ddots & \vdots \\ & & & & d_N(\omega_N) \end{bmatrix} \quad (3.84)$$

wherefrom a set of coefficients is defined as

$$\alpha_k = D_{kk} = \begin{cases} |F(j\omega_0)|, & \text{if } k = 0 \\ \dfrac{\omega_k - \omega_{k-1}}{d_{k-1}(\omega_k) - d_{k-1}(\omega_{k-1})}, & \text{if } k = 1, 2, \ldots, N \end{cases} \quad (3.85)$$

The desired approximation is then given by the continued fraction[50]

$$\hat{F}(s) = \alpha_0 + \cfrac{s - \omega_0}{\alpha_1 + \cfrac{s - \omega_1}{\alpha_2 + \cfrac{s - \omega_2}{\alpha_3 + \ldots}}} = \alpha_0 + \frac{s - \omega_0}{\alpha_1 +}\frac{s - \omega_1}{\alpha_2 +}\frac{s - \omega_2}{\alpha_3 +}\ldots = \left[\alpha_0; \frac{s - \omega_{k-1}}{\alpha_k}\right]_{k=1}^{+\infty} \quad (3.86)$$

Just like the Crone method, this method provides an approximation which will only be useful inside an interval of frequencies, that this time is $[\omega_0; \omega_N]$, but now yet another problem arises, that of choosing frequencies therein. The function to

[50] Continued fractions, which will appear again in this chapter, are mathematical entities dealt with in section A.4 of Appendix A, where all that has to be known about them for what follows is summed up.

approximate being rather regular, it is a good idea just to fix in advance $\left[\omega_0;\omega_N\right]$ and *N*, and then distribute frequencies logarithmically inside that range.

3.3.4. Continuous continued fraction approximations

Two other continuous approximations of (3.18) involving continued fractions may be obtained taking into account that[51]

$$s^\nu \approx \begin{cases} \left(1+s\right)^\nu, & \text{if } \omega \gg 1 \\ \left(1+\dfrac{1}{s}\right)^{-\nu}, & \text{if } \omega \ll 1 \end{cases} \tag{3.87}$$

The idea is to expand the powers as continued fractions. Obviously the first resulting approximation will be valid for frequencies higher than 1 rad/s only while the second will be valid for frequencies below 1 rad/s only. How broad the range of frequencies where the approximation is valid will be is something dependent on the number of terms of the continued fraction that is kept. However, different validity frequency ranges may be obtained rewriting (3.87) as

$$\left(\frac{s}{\lambda}\right)^\nu \approx \begin{cases} \left(1+\dfrac{s}{\lambda}\right)^\nu, & \text{if } \dfrac{\omega}{\lambda} \gg 1 \\ \left(1+\dfrac{\lambda}{s}\right)^{-\nu}, & \text{if } \dfrac{\omega}{\lambda} \ll 1 \end{cases} \Leftrightarrow s^\nu \approx \begin{cases} \lambda^\nu \left(1+\dfrac{s}{\lambda}\right)^\nu, & \text{if } \omega \gg \lambda \\ \lambda^\nu \left(1+\dfrac{\lambda}{s}\right)^{-\nu}, & \text{if } \omega \ll \lambda \end{cases} \tag{3.88}$$

Obvious variations of expansion (A.138) turn this into

$$\hat{F}(s) = \lambda^\nu \left[0; \frac{1}{1}, \frac{-\nu\frac{s}{\lambda}}{1}, \left\{ \frac{\frac{i(i+\nu)}{(2i-1)2i}\frac{s}{\lambda}}{1}, \frac{\frac{i(i-\nu)}{2i(2i+1)}\frac{s}{\lambda}}{1} \right\} \right]_{i=1}^{+\infty}, \quad \text{for } \omega > \lambda \tag{3.89}$$

$$\hat{F}(s) = \lambda^\nu \left[0; \frac{1}{1}, \frac{\nu\frac{\lambda}{s}}{1}, \left\{ \frac{\frac{i(i-\nu)}{(2i-1)2i}\frac{\lambda}{s}}{1}, \frac{\frac{i(i+\nu)}{2i(2i+1)}\frac{\lambda}{s}}{1} \right\} \right]_{i=1}^{+\infty}, \quad \text{for } \omega < \lambda \tag{3.90}$$

Truncating (3.89) after *N* terms results in a transfer function with *N* poles and *N* – 1 zeros. This approximation will be called high-frequency continued fraction approximation. Truncating (3.90) after *N* terms results in a transfer function with *N* poles and *N* zeros. This approximation will be called low-frequency continued fraction approximation.

[51] Vinagre (2001, p. 153).

3.3.5. The Grünwald-Letnikoff approximation

The digital Grünwald-Letnikoff approximation stems from noticing that transfer function (3.18) may appear applying the Laplace transform to the fractional differential equation

$$y(t) = D^{\nu} u(t) \tag{3.91}$$

provided that no initial conditions appear, or that those appearing be zero. So it is reasonable to approximate it using (2.71). That is exactly what we have when we use a digital transfer function, and so the vanishing time interval *h* is to be replaced by the sampling time *T*:

$$\hat{F}\left(z^{-1}\right) = \frac{\sum_{k=0}^{N} (-1)^k \binom{\nu}{k} z^{-k}}{T^{\nu}} = \frac{1}{T^{\nu}} \sum_{k=0}^{N} \frac{(-1)^k \Gamma(\nu+1)}{\Gamma(k+1)\Gamma(\nu-k+1)} z^{-k} \tag{3.92}$$

According to (A.44), this last equality, *k* being non-negative, holds if ν is not a negative integer (a reasonable assumption since *fractional* derivatives are what we are interested in). The upper limit of integration, here called *N*, should vary with integration limits:

$$N = \frac{t-c}{T} \tag{3.93}$$

But if $\nu > -1$ the fraction in the right hand side of (3.92) will decrease as *k* increases and thus, if *u* is bounded, it is possible to neglect the contributions of the earlier time instants. Theorem 2.8 shows this can be done even if $\nu \le -1$, since it is much more practical to have a time-invariant controller, something bound not to happen if *n* were to vary with *t*; so a constant number of terms *N* is always kept. The larger *N* is, the better the approximation is expected to be, as will be seen in detail below.

3.3.6. MacLaurin series based approximations

There are several well-known ways of approximating a first-order derivative with a *z*-domain transfer function. For example, one of the simplest is a first-order backwards finite difference[52]:

$$s \approx \frac{1-z^{-1}}{T} \tag{3.94}$$

This allows approximating (3.18) when $\nu = 1$. For other values a solution is raising both sides of (3.94) to ν. The problem with that is that the result is *not* the ratio of two polynomials in z^{-1}, but that issue can be overcome expanding it into a MacLaurin

[52] Centred or forwards finite differences would never do for control purposes since resulting transfer functions are not causal.

series[53]. Such a development has an infinite number of terms, but a finite transfer function can be obtained truncating the series somewhere, and neglecting further terms[54]. Summing up, the steps mentioned are:

- ❖ choosing a z-domain transfer function that approximates transfer function s;
- ❖ raising it to the desired fractional power ν;
- ❖ expanding the result into a MacLaurin series;
- ❖ truncating the series after a reasonable number of terms.

Theorem 3.12: The result of applying the steps above to a first-order backwards finite difference (3.94) is given by (3.92).

Proof: This is an obvious consequence of (A.77). **Q.E.D.**

Because of this, approximating formula (3.92) will be henceforth mentioned as first-order backwards finite difference formula.

By using other z-domain transfer functions approximating s as a departure point, other approximations of (3.18) can be obtained following the same four steps.

Theorem 3.13: The result of applying the steps above to a second-order backwards finite difference[55]

$$s \approx \frac{3 - 4z^{-1} + z^{-2}}{2T} \tag{3.95}$$

is given by

$$\boxed{\begin{aligned}\hat{F}\left(z^{-1}\right) &= \frac{\left[\Gamma(\nu+1)\right]^2}{(2T)^\nu} \times \\ &\times \sum_{k=0}^{N} z^{-k} \sum_{j=0}^{k} \frac{3^{\nu-j}(-1)^k}{\Gamma(j+1)\Gamma(\nu-j+1)\Gamma(k-j+1)\Gamma(\nu-k+j+1)}\end{aligned}} \tag{3.96}$$

Proof: This is an obvious consequence of (A.90). **Q.E.D.**

Theorem 3.14: The result of applying the steps above to a third-order backwards finite difference [56]

$$s \approx \frac{11 - 18z^{-1} + 9z^{-2} - 2z^{-3}}{6T} \tag{3.97}$$

is given by

[53] Actually any Taylor series would do, but since sampling times are expected to be small, $z^{-1} = 0$ is the best point to develop the series around.

[54] Bibliography about digital approximations is quoted in section 1.3.

[55] See Pina (1995, p. 103) on second order backward finite differences.

[56] See Pina (1995, p. 50-51) on third order backward finite differences.

$$\hat{F}\left(z^{-1}\right)=\frac{\left[\Gamma\left(\nu+1\right)\right]^{3}}{\left(3T\right)^{\nu}}\sum_{k=0}^{N}z^{-k}\left[\sum_{t=0}^{k}\sum_{p=0}^{t}\frac{\left(-1\right)^{p}}{\Gamma\left(p+1\right)\Gamma\left(\nu-p+1\right)}\right.$$
$$\left.\frac{\left(-\frac{7}{4}-\frac{\sqrt{39}}{4}j\right)^{\nu-t+p}\left(-\frac{7}{4}+\frac{\sqrt{39}}{4}j\right)^{\nu-k+t}}{\Gamma\left(t-p+1\right)\Gamma\left(\nu-t+p+1\right)\Gamma\left(k-t+1\right)\Gamma\left(\nu-k+t+1\right)}\right] \tag{3.98}$$

Proof: This is an obvious consequence of (A.95). **Q.E.D.**

In what follows finite differences of order higher than three are not considered. Lubich (1986, p. 708) shows that for orders above six the result does not converge to the desired fractional derivative. This means that orders four, five and six could be considered here, but are not; actually, as shall be seen below in subsections 3.3.13 and following, the three orders addressed suffice to show that increasing the order of the finite difference does not improve the quality of the approximation. Of course explicit formulas similar to (3.92), (3.96) and (3.98) could be found in a similar way for orders four to six.

Theorem 3.15: The result of applying the steps above to Tustin formula

$$s\approx\frac{2}{T}\frac{1-z^{-1}}{1+z^{-1}} \tag{3.99}$$

is given by

$$\hat{F}\left(z^{-1}\right)=\left(\frac{2}{T}\right)^{\nu}\Gamma\left(\nu+1\right)\Gamma\left(-\nu+1\right)\times$$
$$\times\sum_{k=0}^{N}z^{-k}\left[\sum_{j=0}^{k}\frac{\left(-1\right)^{j}}{\Gamma\left(\nu-j+1\right)\Gamma\left(j+1\right)\Gamma\left(k-j+1\right)\Gamma\left(-\nu+j-k+1\right)}\right] \tag{3.100}$$

Proof: This is an obvious consequence of (A.78). **Q.E.D.**

Theorem 3.16: The result of applying the steps above to Simpson formula

$$s\approx\frac{3}{T}\frac{\left(1+z^{-1}\right)\left(1-z^{-1}\right)}{1+4z^{-1}+z^{-2}}=\frac{3}{T}\frac{1-z^{-2}}{1+4z^{-1}+z^{-2}} \tag{3.101}$$

is given by

$$\hat{F}(s)=\left(\frac{3}{T}\right)^{\nu}\Gamma(\nu+1)\left[\Gamma(-\nu+1)\right]^{2}\sum_{q=0}^{N}z^{-q}\left[\sum_{n=0}^{\mathscr{E}\left(\frac{q}{2}\right)}\sum_{p=2n}^{q}\frac{(-1)^{n}}{\Gamma(n+1)\Gamma(\nu-n+1)}\right.$$
$$\left.\frac{\left(2-\sqrt{3}\right)^{-\nu-p+2n}\left(2+\sqrt{3}\right)^{-\nu-q+p}}{\Gamma(p-2n+1)\Gamma(-\nu-p+2n+1)\Gamma(q-p+1)\Gamma(-\nu-q+p+1)}\right] \tag{3.102}$$

Proof: This is an obvious consequence of (A.82). **Q.E.D.**

Remark: All approximations in this subsection are finite impulse response (FIR) filters[57].

3.3.7. Time response based approximations

We may find an approximation of (3.18) in the time domain building a transfer function with an impulse response equal to the impulse response of (3.18) at sampling times. As an alternative we may build a transfer function with a step response equal to the step response of (3.18) at sampling times.

A filter $\hat{F}\left(z^{-1}\right)=\sum_{k=0}^{+\infty}a_k z^{-k}$ with sampling time T has an impulse response given by

$$y(kT)=a_k,\quad k\in\mathbb{N}_0 \tag{3.103}$$

and a step response given by

$$y(kT)=\sum_{i=0}^{k}a_i,\quad k\in\mathbb{N}_0 \tag{3.104}$$

So, if we require the impulse response (3.103) to follow (3.28) at sampling times, we get

$$a_k=\frac{(kT)^{-\nu-1}}{\Gamma(-\nu)},\quad k\in\mathbb{N} \tag{3.105}$$

Should this expression hold also for $k=0$, step responses would often be shifted up or down far from the values they should assume. Thus it is reasonable to set a_0 so that the step response at time T has the correct value. From (3.104) and (3.31) it is seen that this means

[57] Chen and Vinagre (2003) suggest performing the operations above using a linear combination of formulas for approximating a derivative, instead of a single formula. This method will not be developed in what follows.

Diethelm *et al.* (2003, p. 6-7) present a recursive expression that allows finding the coefficients of the approximations in this subsection.

$$a_0 + a_1 = \frac{T^{-\nu}}{\Gamma(-\nu+1)} \Leftrightarrow a_0 = \frac{T^{-\nu}}{\Gamma(-\nu+1)} - \frac{T^{-\nu-1}}{\Gamma(-\nu)} \tag{3.106}$$

and so, joining together (3.105) and (3.106), and truncating after *N* terms, we get

$$\hat{F}\left(z^{-1}\right) = \frac{T^{-\nu}}{\Gamma(-\nu+1)} - \frac{T^{-\nu-1}}{\Gamma(-\nu)} + \sum_{i=1}^{N} \frac{(iT)^{-\nu-1}}{\Gamma(-\nu)} z^{-i} \tag{3.107}$$

Alternatively, if we require step response (3.104) to follow (3.31) at sampling times, we get the following system of linear equations:

$$\sum_{i=0}^{k} a_i = \frac{(kT)^{-\nu}}{\Gamma(-\nu+1)}, \quad k \in \mathbb{N} \tag{3.108}$$

Should (3.108) hold also for $k = 0$, impulse responses would often be shifted up or down far from the values they should assume. Thus it is reasonable to set a_0 so that the impulse response at time *T* has the correct value. This is imposing the same conditions that led to (3.106). So, joining it together with (3.108), and truncating after *N* terms, we get

$$\hat{F}\left(z^{-1}\right) = \sum_{i=0}^{N} a_i z^{-i} \tag{3.109}$$

$$a_0 = \frac{T^{-\nu}}{\Gamma(-\nu+1)} - \frac{T^{-\nu-1}}{\Gamma(-\nu)} \tag{3.110}$$

$$a_k = -\sum_{i=0}^{k-1} a_i + \frac{(kT)^{-\nu}}{\Gamma(-\nu+1)}, \quad k = 1, 2, \ldots, N \tag{3.111}$$

Remark: It turned out to be impossible to build an approximation following *both* responses at sampling times. This is a reminder that no approximation is perfect.

3.3.8. Discrete continued fraction approximations

In subsection 3.3.6 five approximations of (3.18) involving truncated MacLaurin series expansions were given. The expansion was needed for providing a polynomial approximation of a non-rational quantity.

Instead of a MacLaurin series expansion a continued fraction expansion can be used. The four steps for building the approximation become:

- choosing a *z*-domain transfer function that approximates transfer function *s*;
- raising it to the desired fractional power ν;
- expanding the result into a continued fraction;
- truncating the series after a reasonable number of terms.

Theorem 3.17: The result of applying the steps above to a first-order backwards finite difference (3.94) is given by

$$\hat{F}\left(z^{-1}\right)=T^{-\nu}\left[0;\frac{1}{1},\frac{\nu z^{-1}}{1},\left\{\frac{-\frac{i(i+\nu)}{(2i-1)2i}z^{-1}}{1},\frac{-\frac{i(i-\nu)}{2i(2i+1)}z^{-1}}{1}\right\}\right]_{i=1}^{N} \tag{3.112}$$

(In this last expression, curly brackets are meant to show that each value of dumb variable i adds two terms to the continued fraction.)

Proof: This is an obvious consequence of (A.138). **Q.E.D.**

Theorem 3.18: The result of applying the steps above to Tustin formula (3.99) is given by

$$\hat{F}\left(z^{-1}\right)=\left(\frac{2}{T}\right)^{\nu}\left[1;\frac{2\nu}{-\frac{1}{z^{-1}}-\nu},\left\{\frac{\nu^2-i^2}{-\frac{2i+1}{z^{-1}}}\right\}\right]_{i=1}^{N} \tag{3.113}$$

(Once more, the curly brackets intend to show that only the last term is to be repeated with each iteration of i.)

Proof: This is an obvious consequence of (A.151). **Q.E.D.**

No explicit expressions were found for the cases of second and third order backwards finite differences, nor for the case of Simpson formula. But in those cases power series expansions (3.96), (3.98) and (3.102) may be expanded into continued fractions by means of (A.115) and (A.116).

Remark 1: Power series expansions (3.92) and (3.100), corresponding to a first-order finite difference and to Tustin formula, may also be expanded into continued fractions by means of (A.115) and (A.116). Results obtained are equal to those found with (A.115) and (A.116).

Remark 2: Impulse and time response approximations (3.107) and (3.109) are not the result of power series expansions. Nevertheless, it also is possible to expand them into continued fractions with (A.115) and (A.116).

3.3.9. Inverted approximations

Function (3.18) may be rewritten as

$$F(s)=\frac{1}{s^{-\nu}} \tag{3.114}$$

Then the denominator may be approximated with any of the approximations (continuous or discrete) from the previous subsections, and the result inverted.

Remark 1: This should *not* be interpreted as

$$D^{\nu} f(x) \approx \frac{1}{D^{-\nu} f(x)} \tag{3.115}$$

Clearly (3.115) would not hold if it were an equality, neither are the Laplace transforms of the two sides of (3.115) equal.

Remark 2: The Crone approximation with zeros and poles given by (3.75) and the Crone approximation with zeros and poles given by (3.76) are also the inverse of each other.

Remark 3: If the approximation in the denominator of the right-hand side of (3.115) has more poles than zeros (that is for instance the case of the continuous high-frequency continued fraction approximation) the result will not be a causal system, which is likely to be undesirable.

Remark 4: Approximations based upon (3.114) are especially important since (as will be seen below) some formulas have better performances for $\nu > 0$ than for $\nu < 0$, or vice-versa. So (3.114) allows choosing the sign with which results will be better.

3.3.10. Switching between continuous and discrete approximations

Instead of using any of the discrete approximations of (3.18) given above, a continuous approximation may be built and then discretised using any of the suitable discretisation methods available (Tustin, Simpson, and so forth).

Likewise, instead of using any of the continuous approximations of (3.18) given above, a discrete approximation may be built and then turned into a continuous one by the inverse of some discretisation method.

3.3.11. Summing up the approximations of (3.18)

Table 3.1 and Table 3.2 list all discrete formulas presented so far for approximating (3.18).

			High-frequency	Low-frequency
Crone	Carlson	Matsuda	continued fraction	
(3.74)	(3.81)	(3.86)	(3.89)	(3.90)

Table 3.1 — List of continuous approximations of (3.18)

		First	Second	Third		Impulse	Step
	Tustin	order finite difference			Simpson	response	response
truncated MacLaurin expansion of a power of the formula	(3.100)	(3.92)	(3.96)	(3.98)	(3.102)	—	—
truncated interpolation of the ideal response at sampling times	—	—	—	—	—	(3.107)	(3.109)
truncated continued fraction expansion of a power of the formula	(3.113)	(3.112)	apply (A.115) and (A.116) to the formulas above				

Table 3.2 — List of discrete approximations of (3.18)

3.3.12. Approximations of transfer functions other than (3.18)

Fractional transfer functions that appear in applications are more often than not more complex than (3.18). There are two ways to deal with such a complexity.

❖ *A piecewise approximation.* Each derivative is approximated with a suitable method and the several approximations are then linearly combined.

Suppose, for example, that it is desired to implement a controller with a transfer function

$$C(s) = 0.6021 + \frac{0.6187}{s^{1.3646}} + 0.3105 s^{1.0618} \tag{3.116}$$

(This controller will appear below in section 6.3.3.) A piecewise approximation of (3.116) is found developing two integer-order approximations

$$\hat{F}_1(s) \approx s^{1.3646} \tag{3.117}$$

$$\hat{F}_2(s) \approx s^{1.0618} \tag{3.118}$$

and then building an approximation of C given by

$$\hat{C}(s) = 0.6021 + \frac{0.6187}{\hat{F}_1(s)} + 0.3105 \hat{F}_2(s) \tag{3.119}$$

❖ *A lumped approximation.* The transfer function is approximated as a whole.

This last hypothesis means the following for the several continuous approximation methods above presented:

❖ *Matsuda approximation.* For the Matsuda method, a lumped approximation means nothing more than feeding the gain of the entire transfer function to approximate to build functions (3.83). (If the gain changes abruptly at some frequency, it may be good to concentrate sampling frequencies $\omega_i, i = 1, 2, \ldots, N$ around that point.)

❖ *Continued fraction approximations.* Continuous lumped continued fraction approximations may be found when some analytical expression for the continued fraction expansion of the transfer function to approximate—or some good approximation thereof—is known. This is likely to happen seldom (explicit analytical formulas for expanding functions into continued fractions are something hard to come by). So, as an alternative, a power series expansion may be performed, and then (A.115) and (A.116) applied to the result.

❖ *Carlson approximation.* A lumped Carlson approximation can only be found if the transfer function to approximate F satisfies

$$F^{1/v}(s) = G(s), \quad \frac{1}{v} \in \mathbb{Z} \tag{3.120}$$

where G is some integer order transfer function.

❖ *Crone approximation.* The Crone approximation is not suitable for a lumped approach.

For the several discrete approximation methods above presented, a lumped approximation means this:

❖ *McLaurin series approximations.* When a power series expansion is desired, the four steps listed in subsection 3.3.6 have to be performed for the function as a whole. To give an example, suppose that a lumped discrete approximation of (3.116) is desired and that Tustin's formula (3.99) is chosen to approximate *s*. Then function

$$C(s) = 0.6021 + \frac{0.6187}{\left(\frac{2}{T}\frac{1-z^{-1}}{1+z^{-1}}\right)^{1.3646}} + 0.3105\left(\frac{2}{T}\frac{1-z^{-1}}{1+z^{-1}}\right)^{1.0618} \quad (3.121)$$

must be expanded into a MacLaurin series, and the result truncated after some suitable number of terms *N*.

❖ *Time response approximations.* If a time response based approximation is desired, the impulse response and step responses of the transfer function to approximate must be found. Then steps similar to those described in subsection 3.3.7 must be taken.

❖ *Continued fraction approximations.* When a discrete continued fraction expansion is desired, the four steps listed in subsection 3.3.8 have to be performed for the function as a whole. Since it is expectable that no analytical continued fraction be available, an alternative is to perform a power series expansion, and then apply (A.115) and (A.116) to the result.

Of course, when a lumped approximation is performed, it is possible to find a lumped continuous approximation and discretise it, or to find a lumped discrete approximation and find its continuous counterpart, just as suggested in subsection 3.3.10.

Lumped approximations are the most obvious method to deal with transfer functions such as (3.17) (or linear combinations or products and ratios thereof), where the power does not affect *s* only, and which the piecewise approximation mentioned fails to cover[58].

---

[58] An example makes this clearer. Consider plant

$$L(s) = \left(\frac{1+as}{1-as}\right)^{\nu}$$

which is a particular case of the fractional lead-lag.

To find a Matsuda approximation of *L*, gain

$$\left|L(j\omega)\right| = \left|\left(\frac{1+ja\omega}{1-ja\omega}\right)^{\nu}\right|$$

has to be evaluated for several values of ω. Then the result is fed to (3.83) and the method proceeds as given in subsection 3.3.3.

A continuous continued fraction approximation of *L* may be found truncating an obvious variation of (A.151):

$$\hat{L}(s) = \left[1; \frac{2\nu}{\frac{1}{as}-\nu}, \left\{\frac{\nu^2-i^2}{\frac{2i+1}{as}}\right\}\right]_{i=1}^{N}$$

The same formula may also provide a discrete continued fraction approximation of *L*, after using a

Choosing the best method for a lumped approximation is likely to be harder than for a piecewise one. In the following subsections a comparison of approximations of (3.18) is given. But it is of course impossible to present such a comparison taking into account *all* possible transfer functions; so the adequacy of the solution must be checked for each case. Furthermore, lumped approximations require complicated programming and lengthy computations.

A piecewise approximation only requires the formulas listed in Table 3.1 and Table 3.2. It is likely easier to program. However, an approximation made up of several others suffers from all their approximation errors.

3.3.13. Performance comparison of approximations of (3.18)

Since so many approximations of (3.18) are available, it is reasonable to ask what the merits and drawbacks of each are. The following subsections address this problem, comparing:

❖ the placement of zeros and poles of the approximations;

❖ the frequency responses of the approximations, and how they relate to the ideal one;

❖ the impulse and step responses of the approximations, and how they relate to the ideal ones.

To deal with these issues analytically would be far too difficult given the complexity of approximating formulas. The discussion that follows dwells upon results observed to hold for several particular cases, of which a few will be shown in Figures (in which very poor results are not shown; only good and reasonable ones) by the way of illustration[59].

3.3.14. Placement of zeros and poles

Where zeros and poles of the approximations are placed is important since unstable poles are certainly not desired, and non-minimum phase zeros may not be wanted.

With the approximations of Table 3.1, zeros and poles are placed as follows:

❖ Both the Crone and the Matsuda formula lead to alternating, stable, real zeros and poles. Symmetrical values of ν correspond to exchanging the roles of zeros and poles. Examples are given in Figure 3.6.

---

first order backward finite difference, with sampling time T, to approximate s:

$$\hat{L}\left(z^{-1}\right)=\left[1;\frac{2\nu}{\frac{T}{a\left(1-z^{-1}\right)}-\nu},\left\{\frac{\nu^2-i^2}{\frac{(2i+1)T}{a\left(1-z^{-1}\right)}}\right\}_{i=1}^{N}\right]$$

Other approximations would be found similarly.

[59] All these approximations have been implemented as part of toolbox Ninteger for Matlab. Results that follow were obtained with that implementation. See the Appendix B.

❖ The Carlson approximation always leads to stable poles and zeros, many of which are usually non-real. Examples are given in Figure 3.7.

❖ The high-frequency continued fraction approximation leads to stable poles and zeros for $-2 < \nu < 0$. For $\nu < -2$ non-minimum phase zeros appear. For $\nu > 0$ unstable poles appear.

❖ The high-frequency continued fraction approximation leads to stable poles and zeros for $0 < \nu < 2$. For $\nu > 2$ non-minimum phase zeros appear. For $\nu < 0$ unstable poles appear.

These last two cases are exemplified in Figure 3.8.

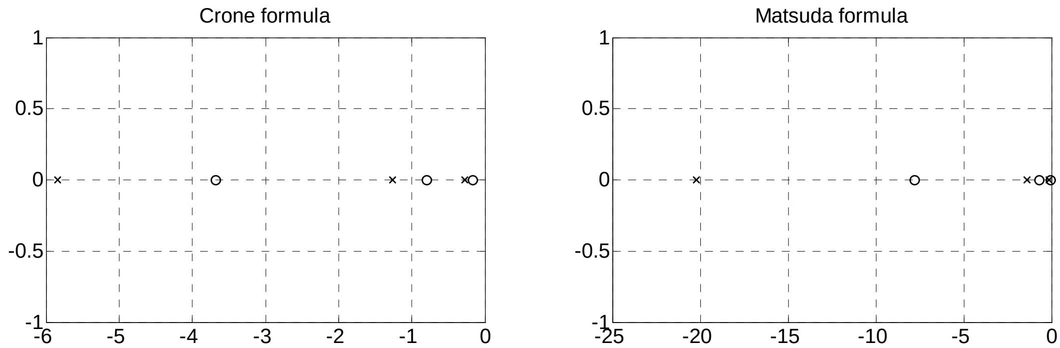


Figure 3.6 — Placement of zeros and poles for approximations of (3.18) (N set to have 3 zeros and poles; approximations were devised for the [0.1; 10] rad/s range; ○ are the zeros and × the poles if $\nu = 0.3$; ○ are the poles and × the zeros if $\nu = -0.3$)

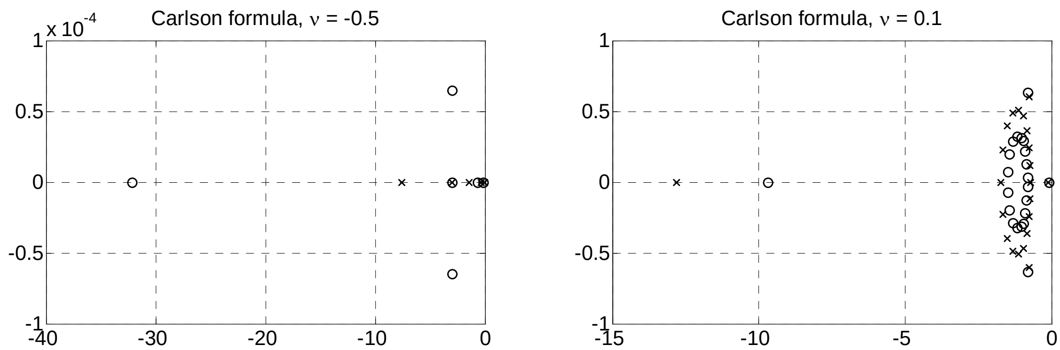


Figure 3.7 — Placement of zeros and poles for approximations of (3.18) (2 iterations performed; development about 1 rad/s; ○ are the zeros and × the poles)

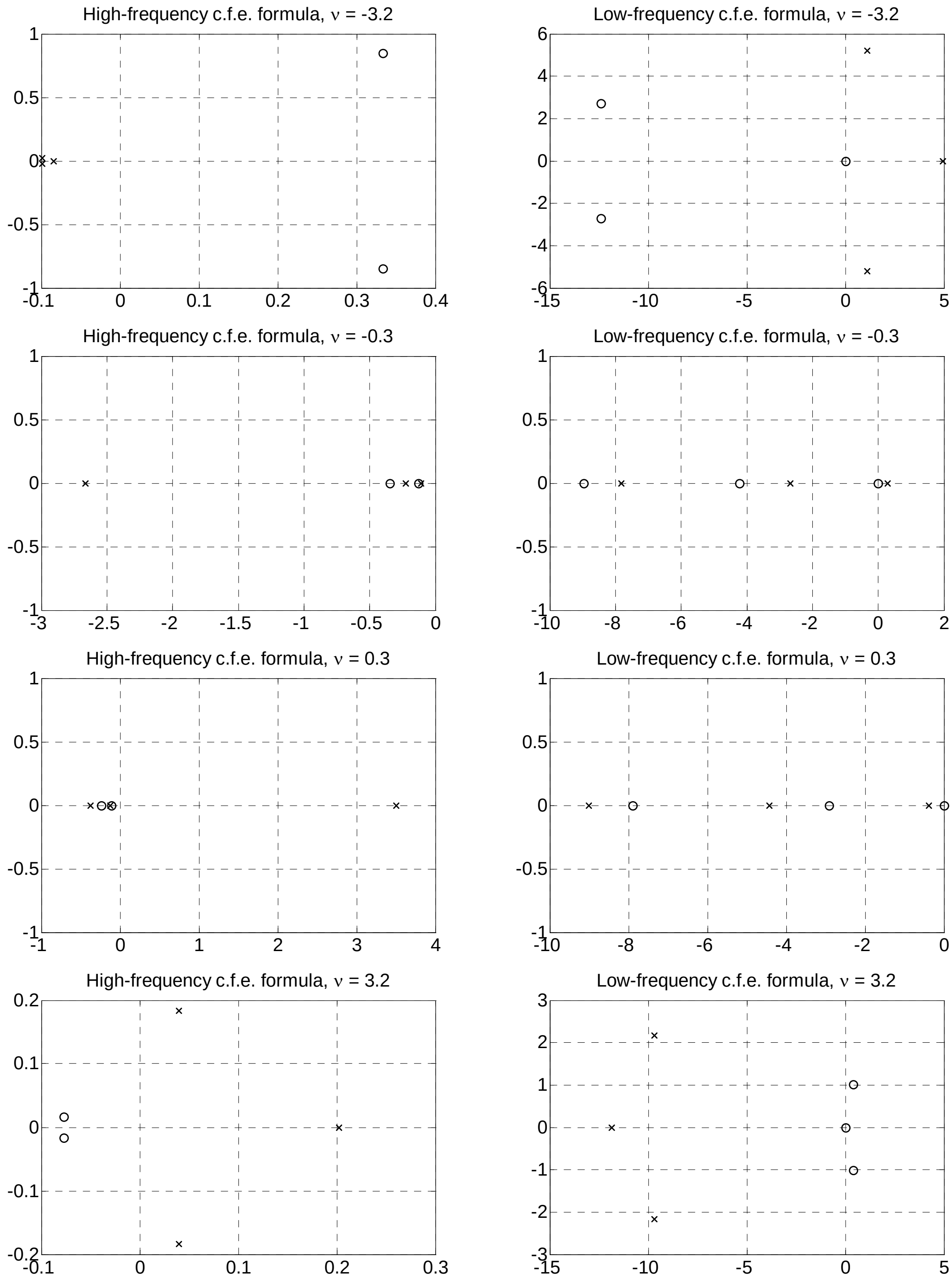


Figure 3.8 — Placement of zeros and poles for approximations of (3.18) (continued fractions truncated after 3 terms; approximations devised for the [0.1; 10] rad/s range; ○ are the poles and × the zeros)

In what concerns discrete approximations, it must be noticed that zero and pole placement depends on the differentiation order ν and on the number of terms retained after truncation N; the sampling time T, on the other hand, is but a scale factor affecting all terms likewise, as is clear from all formulas presented, and does not change the roots of the polynomials of transfer functions. Hence a unit sampling time will be assumed in this subsection.

With the first seven formulas of Table 3.2, zeros are placed as follows:

❖ Whatever the formula, zeros are more or less arranged around the origin.

❖ When finite difference formulas are used, all zeros are all non-minimum phase if $\nu < -1$. If $-1 < \nu < 1$, all zeros are minimum phase. If $\nu > 1$, some are minimum phase and some are non-minimum phase.

❖ When Tustin formula is used, zeros are minimum phase if $-0.5 < \nu < 0.5$. If $0.5 < |\nu| < 1$, some zeros are minimum phase and some are non-minimum phase. If $|\nu| > 1$, all are non-minimum phase. Furthermore, symmetrical values of ν correspond to zeros symmetrical with respect to the imaginary axis.

❖ When Simpson formula is used, there are always non-minimum phase zeros. Only for small values of *n* is it possible to find some minimum phase ones.

❖ When impulse or step responses formulas are used, all zeros are minimum phase if $\nu > 0$. If $\nu < -1$, all zeros are non-minimum phase. If $-1 < \nu < 0$, some may be non-minimum phase, and the more zeros there are, the more likely is that some will be non-minimum phase.

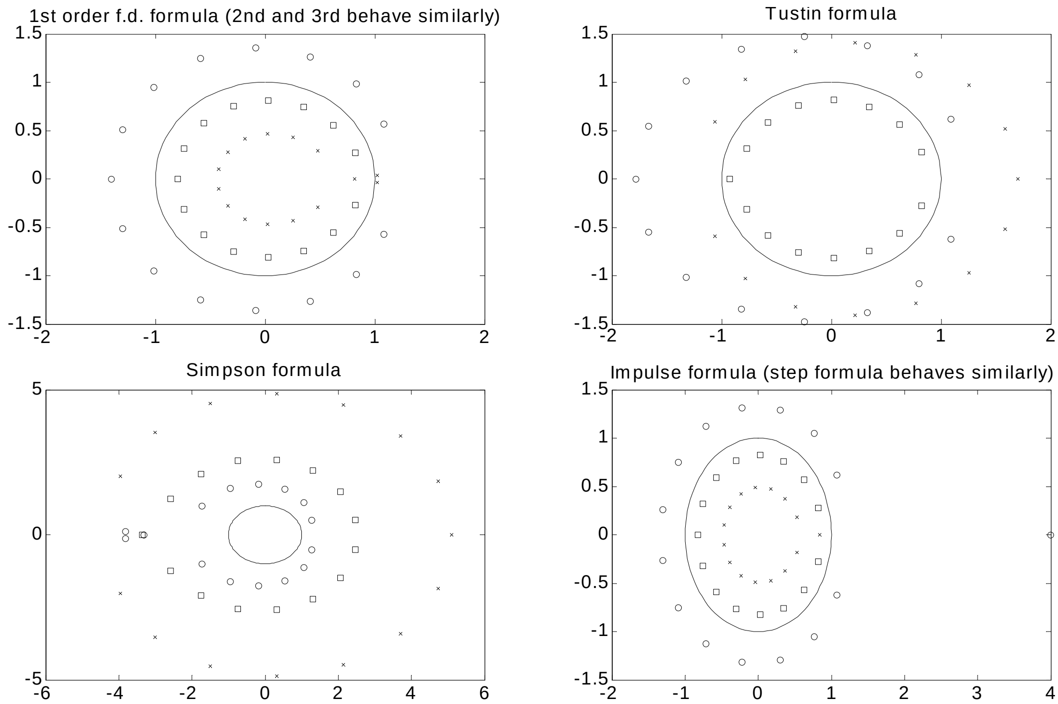


Figure 3.9 — Placement of zeros for approximations of (3.18) (MacLaurin series and time response formulas; $N = 15$; ν variable as follows: ○ $\nu = -2.75$; □ $\nu = -0.3$; × $\nu = 2.5$)

Some examples are shown in Figure 3.9 for some values of ν.

When continued fractions are used, zeros and poles are placed as follows:

❖ When finite difference formulas are used, some poles and sometimes also some zeros appear outside the unit radius circle if $\nu < -1$. If $-1 < \nu < 1$ poles are stable and zeros are minimum phase. If $\nu > 1$ some zeros are non-minimum phase and sometimes also some poles are unstable.

❖ When Tustin formula is used, symmetrical values of ν correspond to zeros and poles symmetrical in respect to the imaginary axis, so changing the sign of ν is the same as inverting the transfer function. If $-1 < \nu < 1$ poles and zeros are real and are inside the unit radius circle; if not, there will always be unstable poles and non-minimum phase zeros.

❖ When Simpson formula is used there are always unstable poles and non-minimum phase zeros.

❖ When impulse or step responses formulas are used, all zeros and poles are inside the unit radius circle if $\nu > 0$. If $\nu < -1$, there are always some unstable poles and non-minimum phase zeros. If $-1 < \nu < 0$, some zeros and poles may be outside, but if ν approaches zero all appear inside.

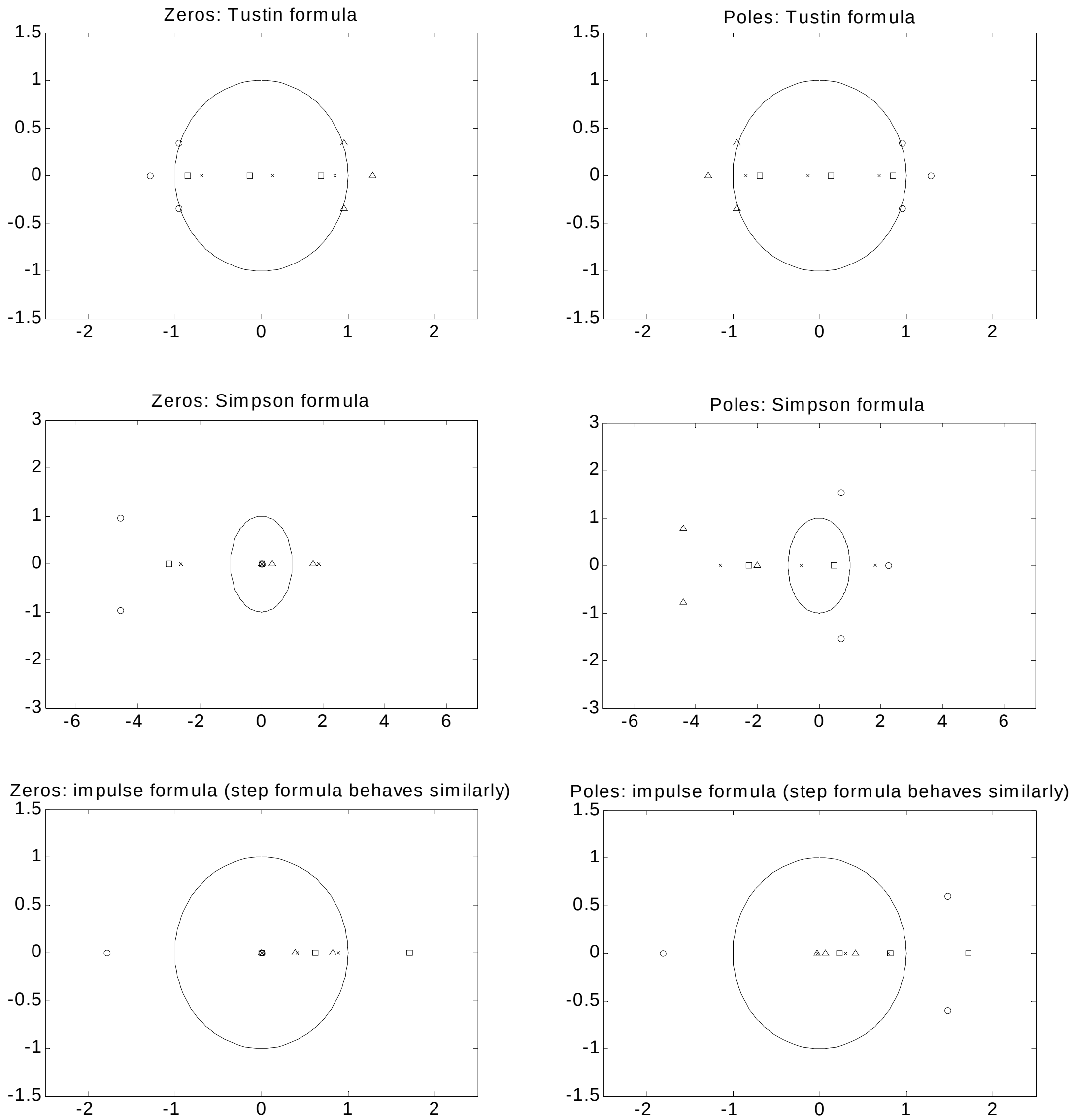


Figure 3.10 — Placement of zeros and poles for approximations of (3.18) (discrete continued fraction expansion formulas; N set to have 3 zeros and poles; ν variable as follows: ○ $\nu = -3.2$; □ $\nu = -0.3$; × $\nu = 0.3$; Δ $\nu = 3.2$)

Some examples are shown in Figure 3.10.

Of course, when approximating formulas (either continuous or discrete) are inverted using (3.115) the above situations still hold, the only difference being that zeros become poles and vice-versa.

3.3.15. Frequency responses

Frequency responses of continuous approximations of (3.18) are as follows:

❖ The high-frequency continued fraction approximation is valid in a rather short range of frequencies above λ only.

❖ The low-frequency continued fraction approximation is valid in a rather short range of frequencies below λ only.

❖ The Crone approximation is rather good well within the frequency range it is devised for, but its performance deteriorates near its limits.

❖ The Matsuda approximation is uniformly good over all the frequency range it is devised for, but the ripple of its phase behaviour is more significant.

❖ The Carlson approximation is in a certain sense the best, but comparison with other approximating methods is difficult for two reasons. Firstly, as pointed above in subsection 3.3.2, few values of ν may be used with this approximation. Secondly, each iteration of Newton's method adds several zeros and poles to the approximation. This means the number of zeros and poles of the approximation cannot be chosen as freely as with the other approximations.

Some examples are given in Figure 3.11[60].

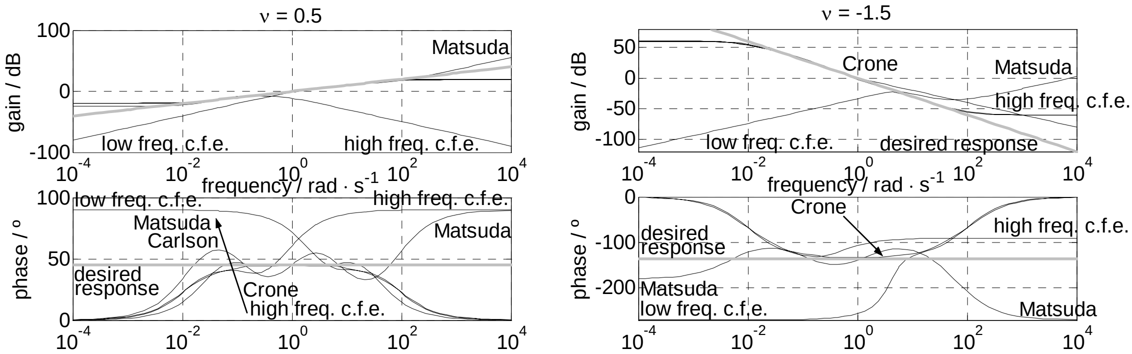


Figure 3.11 — Frequency responses of continuous approximations of (3.18)

In what concerns discrete approximations, the effect of the sampling is that of a scale factor, as stated in subsection 3.3.14. Thus increasing the sampling time will cause the gain plot to shift downwards (since it appears in the denominator of formulas) but will shift the phase plot neither up nor down. This happens together with a left shift that appears whenever the sampling time of a digital controller is increased (since the bandwidth was shifted for lower frequencies). The gain, which is linear in a certain zone, undergoes both effects, so that the linear zone remains aligned; the phase, which is constant in a certain zone, is simply shifted left, and so that constant value is not affected. Both effects can be seen in the example of Figure 3.12. In what follows a unit sampling time is used.

[60] In all remaining figures of this chapter, the desired response is shown as a thick grey line.

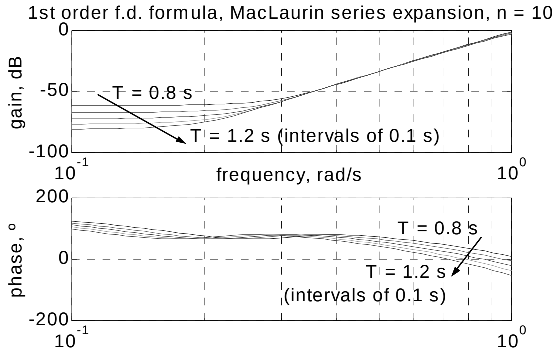


Figure 3.12 — Frequency responses of time-domain approximations of $s^{5.5}$

The effect of the length of the series used for implementing the controller is much more critical. Figure 3.13 shows that increasing the number of terms improves the frequency behaviour, especially in low frequencies (high frequencies stabilise with less terms). This is true for all discrete approximations.

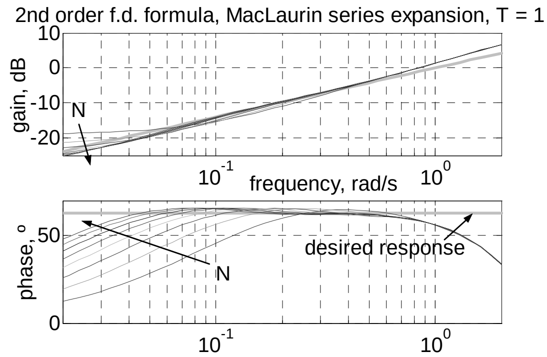


Figure 3.13 — Frequency responses of time-domain approximations of $s^{0.7}$ when N is 4, 7, 10, 13, 16, 19, 22 and 25

Different discrete approximations approximate the ideal frequency behaviour with different success. If a MacLaurin series expansion is used:

❖ Only finite differences formulas achieve acceptable results when $\nu > 1$ (the highest the order of the formula, the better the approximation).
❖ Only Simpson formula does not have an acceptable result for $-1 < \nu < 1$.
❖ No formula has good results for $\nu < -1$.

When continued fraction expansions are used:

❖ All formulas achieve good results for $\nu < 1$ (save Simpson, good for $-1 < \nu < 1$ only).
❖ The best approximations are those of first and second order backwards finite differences formulas.
❖ Step and impulse response formulas and Simpson formula do not achieve good results for $\nu > 1$.

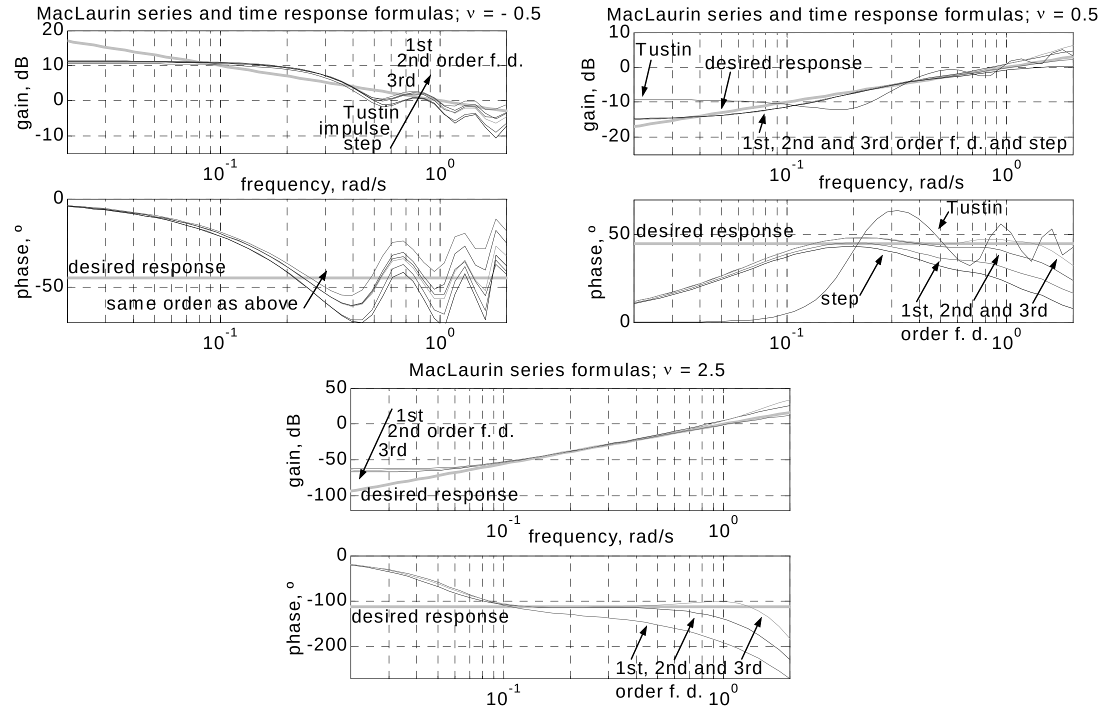


Figure 3.14 — Frequency responses of time-domain approximations of (3.18) (n = 10; T = 1 s)

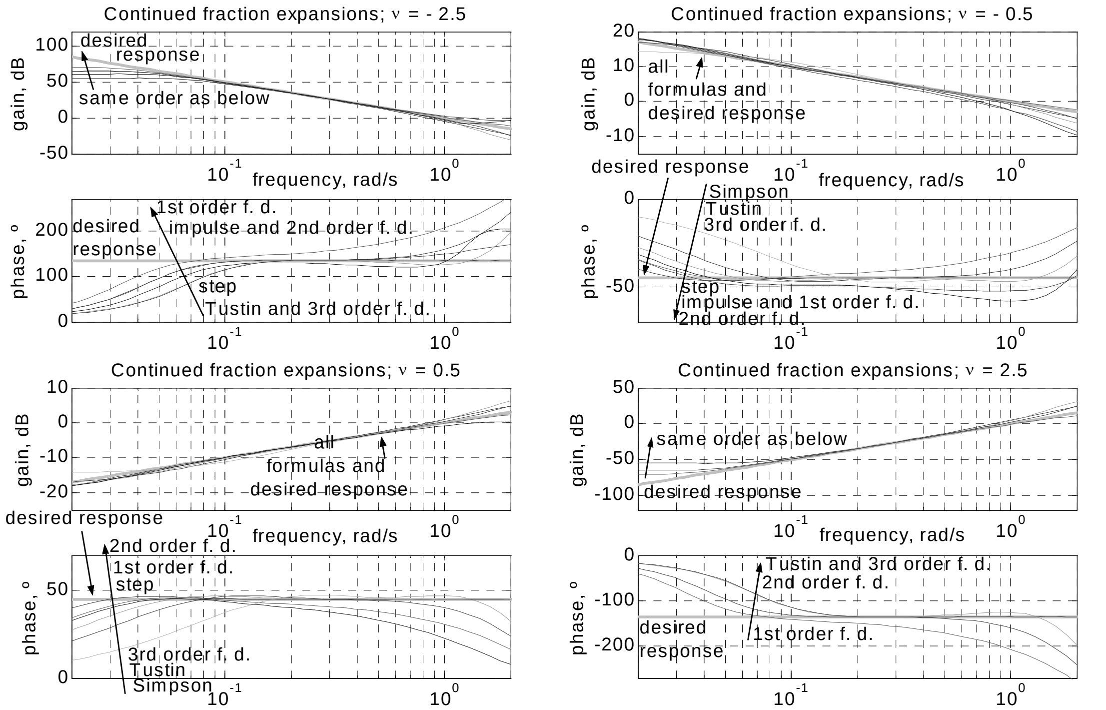


Figure 3.15 — Frequency responses of time-domain approximations of (3.18) (n set to have 5 zeros and poles; T = 1 s)

When these formulas are inverted using (3.115), the behaviour for negative orders becomes the behaviour for positive ones and vice-versa. Some examples are shown in Figure 3.14 and Figure 3.15.

3.3.16. Impulse responses

In what concerns approximations of Table 3.1:

❖ The Crone approximation has an impulse response nearly identical to the ideal one if $-1<\nu<1$. Even otherwise the response is close to the ideal one.

❖ The Carlson approximation (when it exists) also has an impulse response nearly identical to the ideal one.

❖ The Matsuda approximation has an impulse response close to the ideal one if $-1<\nu<1$. If $\nu<-1$, it follows the general trend of the ideal response, but not so closely. For $\nu>-1$, responses no longer resemble the ideal ones.

❖ The high-frequency continued fraction approximation has an impulse response that follows the general trend of the ideal response only if $\nu<0$.

❖ The low-frequency continued fraction approximation has an impulse response close to the ideal one only if $0<\nu<1$. Inverting it with (3.114) never leads to good results.

Examples are shown in Figure 3.16.

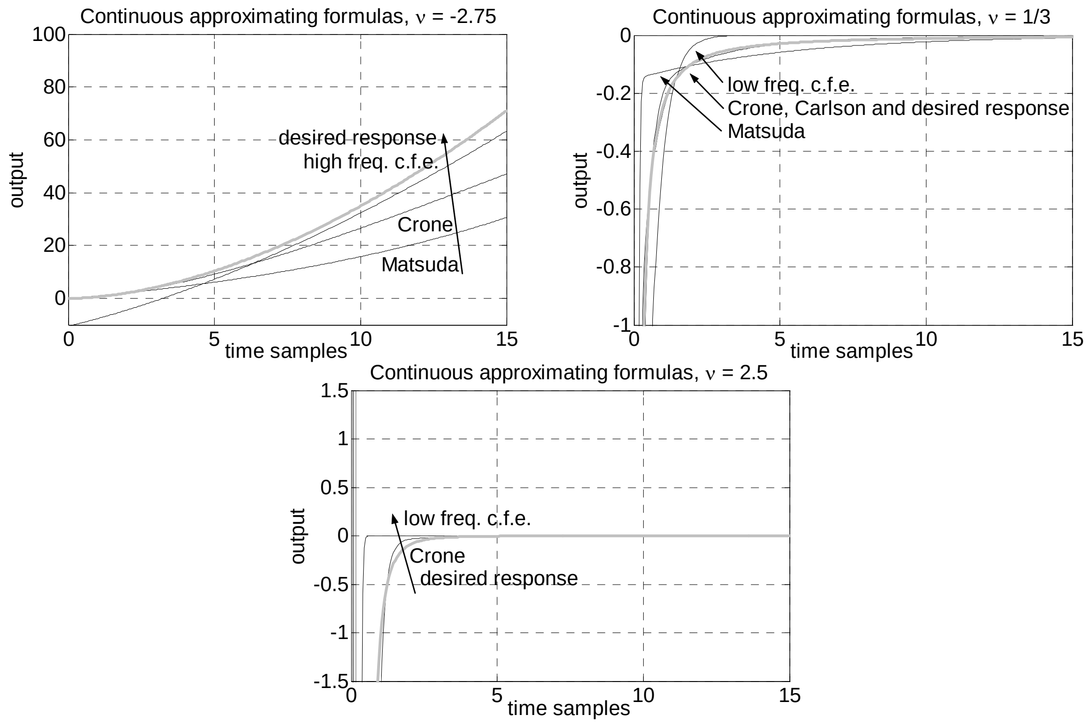


Figure 3.16 — Impulse responses of continuous approximations of (3.18) (*n* set to have 4 zeros and poles)

Impulse responses of formulas that use a MacLaurin series expansion are as follows:

❖ Simpson formula's impulse responses undergo undesired oscillations. For $\nu>-1$, responses no longer resemble the ideal ones.

❖ Tustin formula's impulse responses are fine for $\nu<-1$. If $-1<\nu<1$ oscillations appear. For $\nu>-1$, responses no longer resemble the ideal ones.

❖ Finite differences formulas' impulse responses are good for $\nu<0$, and the higher the order of the finite difference, the closer the response gets to the ideal one. If

$\nu > 0$, responses are not very good, and the higher the order of the finite difference, the worst the response is, especially because oscillations appear for $\nu > -1$.

❖ Impulse response formula's responses are correct at sampling times (as they should), and step response formula's responses are good with only very little oscillations.

❖ Whatever the case, for $t = 0$ all responses are somewhat far from the ideal one—also because the ideal response would be infinite if $\nu > -1$.

❖ Furthermore, all responses become zero after a number of sampling times equal to the number of terms kept in the series, as seen in plots. This is not very relevant for $\nu > -1$, since the response goes to zero anyway, but for $\nu < -1$ responses follow the ideal one up to a certain instant only.

If continued fraction expansions are used, results are very similar. The main difference is that responses do not become zero after a finite number of sampling times. If the theoretical response goes to infinity, so do these formulas' responses, oscillating.

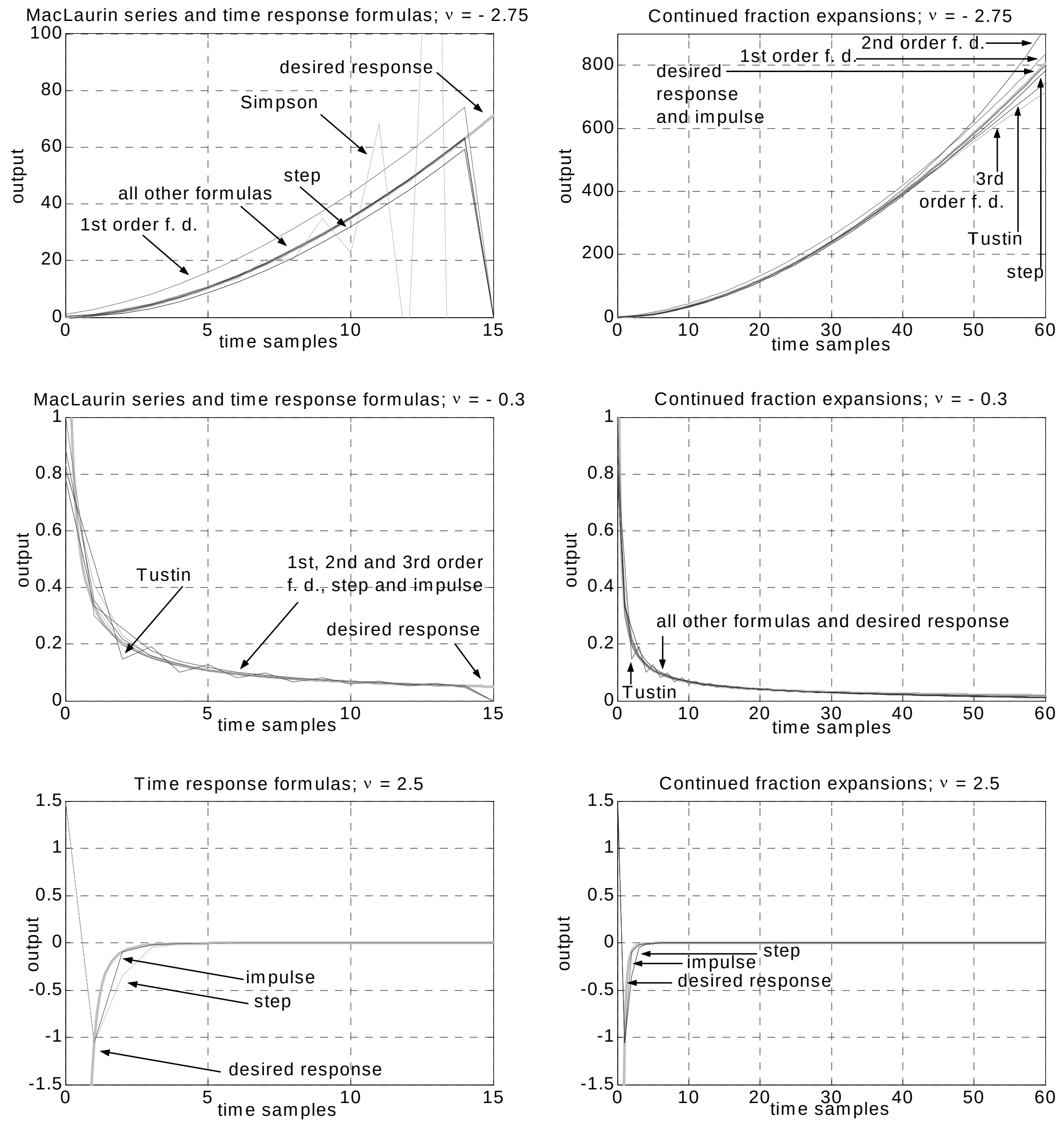


Figure 3.17 — Impulse responses of time-domain approximations of (3.18), without inversion (n set to have 14 zeros and poles)

Some illustrative examples are shown in Figure 3.17 for some values of ν. Formulas with poor results are not shown; acceptable ones are.

When inversion with (3.115) is used with MacLaurin series expansion formulas:

❖ No good results ever appear for $\nu > 1$.

❖ Step and impulse response formulas do not perform well for $\nu < -1$.

❖ The best results are obtained with finite differences formulas.

❖ Responses go to infinity, without oscillations, if the theoretical response goes to infinity.

When inversion with (3.115) is used with continued fraction expansion formulas:

❖ Finite differences formulas, Tustin formula and Simpson formula have the same behaviour observed when inversion is not used.

❖ Step and impulse response formulas behave differently, and achieve reasonable results only when $-1 < \nu < 0$.

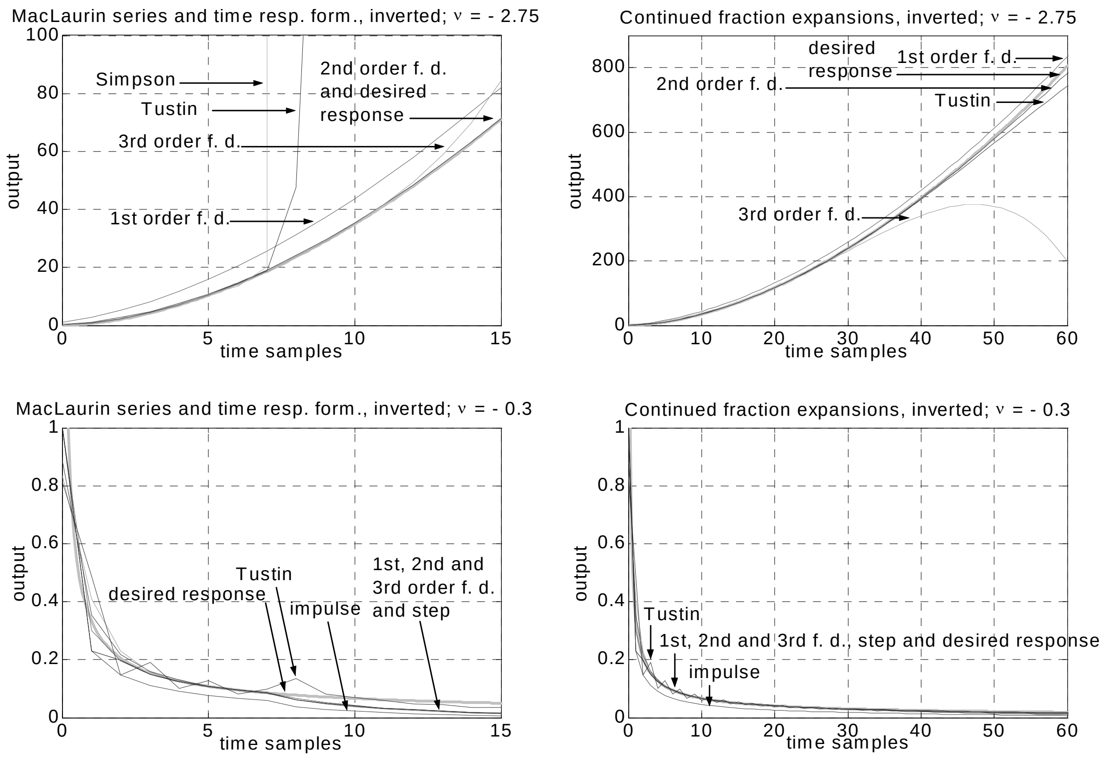


Figure 3.18 — Impulse responses of time-domain approximations of (3.18), with inversion (n set to have 14 zeros and poles)

Examples are given in Figure 3.18.

It should be noticed that continued fraction expansions follow the desired response for a longer time. MacLaurin series without inversion follow it for less time than MacLaurin series with inversion. This is seen in Figure 3.17 and Figure 3.18, where the total number of poles and zeros is always the same and cannot thus be responsible for the different performance.

3.3.17. Step responses

Comments concerning impulse responses might be repeated about step responses. (Similarities are so striking that no Figures with examples are deemed necessary.) The only differences are:

❖ The low-frequency continuous continued fraction approximating formula never leads to goods results.
❖ When Tustin formula is used, oscillations appear for $\nu > 1$ only.
❖ Truncated MacLaurin series being finite impulse response filters, their step responses remain constant when their impulse responses go to zero.

3.3.18. Summing up comparison results

In short, results presented show that:

❖ Whatever the approximation used, impulse and step responses never follow the theoretical response forever. Depending on the case, and after a variable number of time samples, they may go to zero, remain constant, or diverge to infinity.
❖ Whatever the approximation used, frequency responses follow the ideal one for a limited frequency range only.
❖ For $|\nu| > 1$, only the Crone approximation has acceptable frequency and time responses and stable poles and zeros.
❖ For $-1 < \nu < 1$, the Crone, Carlson and Matsuda formulas, finite differences formulas (whatever the order and the expansion) and Tustin formula using a continued fraction expansion (inverted or not) also have those desirable characteristics.
❖ For $-0.5 < \nu < 0.5$, Tustin formula using a MacLaurin series expansion (inverted or not) also behaves so.
❖ Some formulas (continuous continued fractions, Simpson) are never a good choice.

Additionally, the actual formula chosen should depend on what is more important (frequency response, time response, pole or zero placement…).

From all this it is clear that the case when $|\nu| > 1$ is to be avoided as much as possible. For that reason, when piecewise integer approximations of fractional transfer functions are built, it is usual not only to approximate each derivative separately but also to split derivatives into integer and fractional parts. That is to say that (3.18) is turned into

$$F(s) = s^\nu = s^n s^\alpha, \quad n \in \mathbb{Z} \wedge \alpha \in \left]-1;1\right[\wedge n + \alpha = \nu \qquad (3.122)$$

The fractional part is approximated with some of the methods considered above. The integer part is left as it is if the approximation is continuous; if it is discrete, some usual discretisation formula is used.

Remark: There are additional reasons why one should want to do this. The Riemann-Liouville and the Caputo definitions also split (at least for positive values of ν) the fractional derivative into an integer and a fractional part. If the approximation is intended for control purposes, it may be useful (depending on the type of reference expected) that it include one (or more) integer integrators to avoid or reduce steady-state errors[61].

[61] Actually Oustaloup (1991, p. 158-161), Vinagre (2001, p. 151) and other authors go even farther and make $\alpha \in \left]0;1\right[$ in (3.122). In this manner an (integer) integration always appears whenever $\nu < 0$.

4. Identification for fractional control

> Io lavoro in una casa editrice e in una casa editrice vengono savi e matti. Il mestiere del redattore è riconoscere a colpo d'occhio i matti.
>
> Umberto Eco, *Il pendolo di Foucault* III, 10

This chapter presents methods for identifying models from experimental (or simulation) data useful when working with fractional plants:

❖ Section 4.1 addresses Levy's identification method. This is a well-known method with which an integer model is found from frequency data. In this section it is generalised so as to provide fractional models as well. Improvements on this method (Vinagre's weights, Sanathanan and Koerner's iterative method, Lawrence and Rogers's iterative method) are also expounded.

❖ Section 4.2 addresses the Crone identification method. This is an identification method for finding an integer model from frequency data that is often used together with the second generation Crone methodology for devising fractional controllers, described below in section 5.2. Both the original method, resulting in an *s*-domain model, and modifications thereof for finding models in the *z*-domain, are presented.

❖ Section 4.3 addresses the question of how a fractional model may be identified from a time response.

4.1. Levy's identification method

Levy's method is a well-established method for finding the coefficients of an integer transfer function that models a plant with a known frequency behaviour[62].

4.1.1. Levy's method extended for fractional orders

Let us suppose we have a plant G with a known frequency behaviour. Let us suppose we want to model it using a commensurate transfer function[63]

$$\hat{G}(s)=\frac{b_0+b_1s^q+b_2s^{2q}+\ldots+b_ms^{mq}}{1+a_1s^q+a_2s^{2q}+\ldots+a_ns^{nq}}=\frac{\sum_{k=0}^{m}b_ks^{kq}}{1+\sum_{k=1}^{n}a_ks^{kq}} \tag{4.1}$$

Remark: Levy's original method requires setting in advance orders m and n. With this extension for fractional models the commensurate order q is also needed in advance. Some comments concerning that will be found below in subsection 1.2.14.

---

[62] Levy (1959).

[63] Hartley *et al.* (2003), addressing the identification of fractional models in a way rather close to Levy's method, also restrict themselves to commensurate transfer functions. But they only make use of simpler forms of $\hat{G}$, namely requesting the numerator to be 1.

The frequency response of (4.1) is given by

$$\hat{G}(j\omega) = \frac{\sum_{k=0}^{m} b_k (j\omega)^{kq}}{1+\sum_{k=1}^{n} a_k (j\omega)^{kq}} = \frac{N(\omega)}{D(\omega)} = \frac{\alpha(\omega) + j\beta(\omega)}{\sigma(\omega) + j\tau(\omega)} \tag{4.2}$$

where N and D are complex-valued and α, β, σ and τ, the real and imaginary parts thereof, are real-valued. The error between model and plant, for a given frequency ω, will be

$$\varepsilon(\omega) = G(j\omega) - \frac{N(\omega)}{D(\omega)} \tag{4.3}$$

It might be possible to adjust the parameters of (4.1) by minimising the norm (or the square of the norm) of this error. Levy's method, however, tries to minimise the square of the norm of

$$\varepsilon(\omega) D(\omega) = G(j\omega) D(\omega) - N(\omega) \tag{4.4}$$

instead. So as to alleviate notation let us define

$$E(\omega) = \varepsilon(\omega) D(\omega) \tag{4.5}$$

and drop the dependency on ω; we will have

$$\begin{aligned} &E = GD - N = \\ &= \left[\mathrm{Re}(G) + j\,\mathrm{Im}(G)\right](\sigma + j\tau) - (\alpha + j\beta) = \\ &= \left[\mathrm{Re}(G)\sigma - \mathrm{Im}(G)\tau - \alpha\right] + j\left[\mathrm{Re}(G)\tau + \mathrm{Im}(G)\sigma - \beta\right] \end{aligned} \tag{4.6}$$

The square of the norm of E is

$$|E|^2 = \left[\mathrm{Re}(G)\sigma - \mathrm{Im}(G)\tau - \alpha\right]^2 + \left[\mathrm{Re}(G)\tau + \mathrm{Im}(G)\sigma - \beta\right]^2 \tag{4.7}$$

From (4.2) we see that

$$\alpha(\omega) = \sum_{k=0}^{m} b_k \,\mathrm{Re}\left[(j\omega)^{kq}\right] \tag{4.8}$$

$$\beta(\omega) = \sum_{k=0}^{m} b_k \,\mathrm{Im}\left[(j\omega)^{kq}\right] \tag{4.9}$$

$$\sigma(\omega) = 1 + \sum_{k=1}^{n} a_k \,\mathrm{Re}\left[(j\omega)^{kq}\right] \tag{4.10}$$

$$\tau(\omega) = \sum_{k=1}^{n} a_k \,\mathrm{Im}\left[(j\omega)^{kq}\right] \tag{4.11}$$

Thus, if we differentiate[64] $|E|^2$ with respect to one of the coefficients b_k, we shall have

$$\begin{aligned}&\frac{\partial|E|^2}{\partial b_k}=\\&=-2\left[\operatorname{Re}(G)\sigma-\operatorname{Im}(G)\tau-\alpha\right]\operatorname{Re}\left[(j\omega)^{kq}\right]-\\&-2\left[\operatorname{Re}(G)\tau+\operatorname{Im}(G)\sigma-\beta\right]\operatorname{Im}\left[(j\omega)^{kq}\right]\end{aligned} \tag{4.12}$$

Equalling the derivative to zero,

$$\begin{aligned}&\frac{\partial|E|^2}{\partial b_k}=0\Leftrightarrow\\&\Leftrightarrow\left[\operatorname{Re}(G)\sigma-\operatorname{Im}(G)\tau-\alpha\right]\operatorname{Re}\left[(j\omega)^{kq}\right]+\\&+\left[\operatorname{Re}(G)\tau+\operatorname{Im}(G)\sigma-\beta\right]\operatorname{Im}\left[(j\omega)^{kq}\right]=0\end{aligned} \tag{4.13}$$

And if we differentiate $|E|^2$ with respect to one of the coefficients a_k we shall have

$$\begin{aligned}&\frac{\partial|E|^2}{\partial a_k}=\\&=2\left[\operatorname{Re}(G)\sigma-\operatorname{Im}(G)\tau-\alpha\right]\operatorname{Re}(G)\operatorname{Re}\left[(j\omega)^{kq}\right]+\\&+2\left[\operatorname{Re}(G)\tau+\operatorname{Im}(G)\sigma-\beta\right]\operatorname{Im}(G)\operatorname{Re}\left[(j\omega)^{kq}\right]-\\&-2\left[\operatorname{Re}(G)\sigma-\operatorname{Im}(G)\tau-\alpha\right]\operatorname{Im}(G)\operatorname{Im}\left[(j\omega)^{kq}\right]+\\&+2\left[\operatorname{Re}(G)\tau+\operatorname{Im}(G)\sigma-\beta\right]\operatorname{Re}(G)\operatorname{Im}\left[(j\omega)^{kq}\right]\end{aligned} \tag{4.14}$$

Equalling the derivative to zero,

$$\frac{\partial|E|^2}{\partial a_k}=0\Leftrightarrow$$

[64] See Appendix A, subsection A.1.3, on derivatives.

$$\begin{aligned}
&\Leftrightarrow \sigma\left[\mathrm{Re}(G)\right]^2 \mathrm{Re}\left[(j\omega)^{kq}\right] - \tau\,\mathrm{Im}(G)\mathrm{Re}(G)\mathrm{Re}\left[(j\omega)^{kq}\right] - \\
&-\alpha\,\mathrm{Re}(G)\mathrm{Re}\left[(j\omega)^{kq}\right] + \tau\,\mathrm{Im}(G)\mathrm{Re}(G)\mathrm{Re}\left[(j\omega)^{kq}\right] + \\
&+\sigma\left[\mathrm{Im}(G)\right]^2 \mathrm{Re}\left[(j\omega)^{kq}\right] - \beta\,\mathrm{Im}(G)\mathrm{Re}\left[(j\omega)^{kq}\right] - \\
&-\sigma\,\mathrm{Im}(G)\mathrm{Re}(G)\mathrm{Im}\left[(j\omega)^{kq}\right] + \tau\left[\mathrm{Im}(G)\right]^2 \mathrm{Im}\left[(j\omega)^{kq}\right] + \\
&+\alpha\,\mathrm{Im}(G)\mathrm{Im}\left[(j\omega)^{kq}\right] + \tau\left[\mathrm{Re}(G)\right]^2 \mathrm{Im}\left[(j\omega)^{kq}\right] + \\
&+\sigma\,\mathrm{Im}(G)\mathrm{Re}(G)\mathrm{Im}\left[(j\omega)^{kq}\right] - \beta\,\mathrm{Re}(G)\mathrm{Im}\left[(j\omega)^{kq}\right] = 0 \Leftrightarrow \\
&\Leftrightarrow \sigma\left\{\left[\mathrm{Im}(G)\right]^2 + \left[\mathrm{Re}(G)\right]^2\right\}\mathrm{Re}\left[(j\omega)^{kq}\right] + \\
&+\tau\left\{\left[\mathrm{Im}(G)\right]^2 + \left[\mathrm{Re}(G)\right]^2\right\}\mathrm{Im}\left[(j\omega)^{kq}\right] + \\
&+\alpha\left\{\mathrm{Im}(G)\mathrm{Im}\left[(j\omega)^{kq}\right] - \mathrm{Re}(G)\mathrm{Re}\left[(j\omega)^{kq}\right]\right\} + \\
&+\beta\left\{-\mathrm{Im}(G)\mathrm{Re}\left[(j\omega)^{kq}\right] - \mathrm{Re}(G)\mathrm{Im}\left[(j\omega)^{kq}\right]\right\} = 0
\end{aligned} \tag{4.15}$$

The $m+1$ equations given by (4.13) and the n equations given by (4.15) make up a linear system that may be solved so as to find the coefficients of (4.1). Usually the frequency behaviour of the plant is known in more than one frequency (otherwise it is likely that the model will be rather poor). Let us suppose that it is known at f frequencies. Then the system to solve, given by equations (4.13) and (4.15) written explicitly on coefficients a and b, is

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}\begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} e \\ g \end{bmatrix} \tag{4.16}$$

where

$$\begin{aligned}
&\mathbf{A}_{l,c} = \sum_{p=1}^{f}\left\{-\mathrm{Re}\left[(j\omega_p)^{lq}\right]\mathrm{Re}\left[(j\omega_p)^{cq}\right] - \mathrm{Im}\left[(j\omega_p)^{lq}\right]\mathrm{Im}\left[(j\omega_p)^{cq}\right]\right\}, \\
&\quad l = 0\ldots m \wedge c = 0\ldots m
\end{aligned} \tag{4.17}$$

$$\begin{aligned}
&\mathbf{B}_{l,c} = \sum_{p=1}^{f}\left\{\mathrm{Re}\left[(j\omega_p)^{lq}\right]\mathrm{Re}\left[(j\omega_p)^{cq}\right]\mathrm{Re}\left[G(j\omega_p)\right] + \right. \\
&+\mathrm{Im}\left[(j\omega_p)^{lq}\right]\mathrm{Re}\left[(j\omega_p)^{cq}\right]\mathrm{Im}\left[G(j\omega_p)\right] - \\
&-\mathrm{Re}\left[(j\omega_p)^{lq}\right]\mathrm{Im}\left[(j\omega_p)^{cq}\right]\mathrm{Im}\left[G(j\omega_p)\right] + \\
&\left.+\mathrm{Im}\left[(j\omega_p)^{lq}\right]\mathrm{Im}\left[(j\omega_p)^{cq}\right]\mathrm{Re}\left[G(j\omega_p)\right]\right\}, \\
&\quad l = 0\ldots m \wedge c = 1\ldots n
\end{aligned} \tag{4.18}$$

$$\mathbf{C}_{l,c} = \sum_{p=1}^{f} \Big\{ -\mathrm{Re}\Big[\big(j\omega_p\big)^{lq}\Big]\mathrm{Re}\Big[\big(j\omega_p\big)^{cq}\Big]\mathrm{Re}\Big[G\big(j\omega_p\big)\Big] +$$
$$+\mathrm{Im}\Big[\big(j\omega_p\big)^{lq}\Big]\mathrm{Re}\Big[\big(j\omega_p\big)^{cq}\Big]\mathrm{Im}\Big[G\big(j\omega_p\big)\Big] -$$
$$-\mathrm{Re}\Big[\big(j\omega_p\big)^{lq}\Big]\mathrm{Im}\Big[\big(j\omega_p\big)^{cq}\Big]\mathrm{Im}\Big[G\big(j\omega_p\big)\Big] - \tag{4.19}$$
$$-\mathrm{Im}\Big[\big(j\omega_p\big)^{lq}\Big]\mathrm{Im}\Big[\big(j\omega_p\big)^{cq}\Big]\mathrm{Re}\Big[G\big(j\omega_p\big)\Big]\Big\},$$
$$l = 1\ldots n \wedge c = 0\ldots m$$

$$\mathbf{D}_{l,c} = \sum_{p=1}^{f} \Big(\Big\{\mathrm{Re}\Big[G\big(j\omega_p\big)\Big]\Big\}^2 + \Big\{\mathrm{Im}\Big[G\big(j\omega_p\big)\Big]\Big\}^2 \Big)$$
$$\Big\{\mathrm{Re}\Big[\big(j\omega_p\big)^{lq}\Big]\mathrm{Re}\Big[\big(j\omega_p\big)^{cq}\Big] + \mathrm{Im}\Big[\big(j\omega_p\big)^{lq}\Big]\mathrm{Im}\Big[\big(j\omega_p\big)^{cq}\Big]\Big\}, \tag{4.20}$$
$$l = 1\ldots n \wedge c = 1\ldots n$$

$$b = \begin{bmatrix} b_0 \\ \vdots \\ b_m \end{bmatrix} \tag{4.21}$$

$$a = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \tag{4.22}$$

$$e_{l,1} = \sum_{p=1}^{f} \Big\{ -\mathrm{Re}\Big[\big(j\omega_p\big)^{lq}\Big]\mathrm{Re}\Big[G\big(j\omega_p\big)\Big] - \mathrm{Im}\Big[\big(j\omega_p\big)^{lq}\Big]\mathrm{Im}\Big[G\big(j\omega_p\big)\Big]\Big\}, \tag{4.23}$$
$$l = 0\ldots m$$

$$g_{l,1} = \sum_{p=1}^{f} -\mathrm{Re}\Big[\big(j\omega_p\big)^{lq}\Big]\Big(\Big\{\mathrm{Re}\Big[G\big(j\omega_p\big)\Big]\Big\}^2 + \Big\{\mathrm{Im}\Big[G\big(j\omega_p\big)\Big]\Big\}^2 \Big) \tag{4.24}$$
$$l = 1\ldots n$$

If q is 1, the real and imaginary parts of $(j\omega)^k$ reduce (k being a natural) to either $\pm\omega^k$ or $\pm j\omega^k$ and matrices **A**, **B**, **C** and **D** and vectors e and g assume the usual structures of Levy's identification method.

4.1.2. First improvement: Vinagre's weights

Levy's method's drawbacks are well known, one of them being that low frequency data has little influence in (4.16) and the resulting fit is poor for such frequencies. Using well-chosen weights for increasing the influence of low frequency data is a means of dealing with this. Vinagre (2001, p. 140-141) notes that if $g(t)$ is the step response of our system then

$$\int_0^{+\infty} \left| g(t) - \hat{g}(t) \right|^2 dt = \int_0^{+\infty} \left| \mathscr{L}^{-1}\left[G(s)\frac{1}{s} \right] - \mathscr{L}^{-1}\left[\hat{G}(s)\frac{1}{s} \right] \right|^2 dt \tag{4.25}$$

and that Parseval's theorem turns this into

$$\int_{-\infty}^{+\infty}\left|G(j\omega)\frac{1}{j\omega}-\hat{G}(j\omega)\frac{1}{j\omega}\right|^2 d\omega=\int_{-\infty}^{+\infty}\frac{\left|\varepsilon(\omega)\right|^2}{\omega^2}d\omega \tag{4.26}$$

Using the trapezoidal numerical integration rule this can be approximated by

$$\sum_{p=1}^{f-1}\left[\left(\omega_{p+1}-\omega_p\right)\frac{\frac{\left|\varepsilon\left(\omega_{p+1}\right)\right|^2}{\omega_{p+1}^2}+\frac{\left|\varepsilon\left(\omega_p\right)\right|^2}{\omega_p^2}}{2}\right]=\sum_{p=1}^{f}\frac{\left|\varepsilon\left(\omega_p\right)\right|^2}{\omega_p^2}\varphi_p \tag{4.27}$$

where

$$\varphi_p=\begin{cases}\frac{\omega_2-\omega_1}{2}, & \text{if } p=1\\ \frac{\omega_{p+1}-\omega_{p-1}}{2}, & \text{if } 1<p<f\\ \frac{\omega_f-\omega_{f-1}}{2}, & \text{if } p=f\end{cases} \tag{4.28}$$

are the coefficients of the trapezoidal integration rule. Just as Levy's method minimises $\sum_{p=1}^{f}\left|E(\omega)\right|^2$ instead of $\sum_{p=1}^{f}\left|\varepsilon(\omega)\right|^2$, so this time, instead of the right-hand member of (4.26), the quantity

$$\sum_{p=1}^{f}\left|E\left(\omega_p\right)\right|^2\frac{\varphi_p}{\omega_p^2} \tag{4.29}$$

will be minimised instead. The fraction multiplying the square of the norm is the weight[65],

$$w_p=\frac{\varphi_p}{\omega_p^2} \tag{4.30}$$

that clearly increases the influence of low frequencies. Since the weight does not depend on coefficients *a* and *b*, it will not change the values of derivatives (4.12) and (4.14). The only difference in the method is that matrixes and vectors in (4.16) will now be given by

[65] Based upon energetic considerations, Vinagre (2001, p. 140-141) adds yet another term to this weight, that depends neither on *p* nor on coefficients *a* or *b*, and thus may be neglected by the minimisation.

$$\mathbf{A}_{l,c} = \sum_{p=1}^{f} \Big\{ -\mathrm{Re}\Big[\big(j\omega_p\big)^{lq}\Big]\mathrm{Re}\Big[\big(j\omega_p\big)^{cq}\Big] - \\ -\mathrm{Im}\Big[\big(j\omega_p\big)^{lq}\Big]\mathrm{Im}\Big[\big(j\omega_p\big)^{cq}\Big]\Big\} w_p, \\ l = 0\ldots m \wedge c = 0\ldots m \tag{4.31}$$

$$\mathbf{B}_{l,c} = \sum_{p=1}^{f} \Big\{ \mathrm{Re}\Big[\big(j\omega_p\big)^{lq}\Big]\mathrm{Re}\Big[\big(j\omega_p\big)^{cq}\Big]\mathrm{Re}\Big[G\big(j\omega_p\big)\Big] + \\ +\mathrm{Im}\Big[\big(j\omega_p\big)^{lq}\Big]\mathrm{Re}\Big[\big(j\omega_p\big)^{cq}\Big]\mathrm{Im}\Big[G\big(j\omega_p\big)\Big] - \\ -\mathrm{Re}\Big[\big(j\omega_p\big)^{lq}\Big]\mathrm{Im}\Big[\big(j\omega_p\big)^{cq}\Big]\mathrm{Im}\Big[G\big(j\omega_p\big)\Big] + \\ +\mathrm{Im}\Big[\big(j\omega_p\big)^{lq}\Big]\mathrm{Im}\Big[\big(j\omega_p\big)^{cq}\Big]\mathrm{Re}\Big[G\big(j\omega_p\big)\Big]\Big\} w_p, \\ l = 0\ldots m \wedge c = 1\ldots n \tag{4.32}$$

$$\mathbf{C}_{l,c} = \sum_{p=1}^{f} \Big\{ -\mathrm{Re}\Big[\big(j\omega_p\big)^{lq}\Big]\mathrm{Re}\Big[\big(j\omega_p\big)^{cq}\Big]\mathrm{Re}\Big[G\big(j\omega_p\big)\Big] + \\ +\mathrm{Im}\Big[\big(j\omega_p\big)^{lq}\Big]\mathrm{Re}\Big[\big(j\omega_p\big)^{cq}\Big]\mathrm{Im}\Big[G\big(j\omega_p\big)\Big] - \\ -\mathrm{Re}\Big[\big(j\omega_p\big)^{lq}\Big]\mathrm{Im}\Big[\big(j\omega_p\big)^{cq}\Big]\mathrm{Im}\Big[G\big(j\omega_p\big)\Big] - \\ -\mathrm{Im}\Big[\big(j\omega_p\big)^{lq}\Big]\mathrm{Im}\Big[\big(j\omega_p\big)^{cq}\Big]\mathrm{Re}\Big[G\big(j\omega_p\big)\Big]\Big\} w_p, \\ l = 1\ldots n \wedge c = 0\ldots m \tag{4.33}$$

$$\mathbf{D}_{l,c} = \sum_{p=1}^{f} \Big(\Big\{\mathrm{Re}\Big[G\big(j\omega_p\big)\Big]\Big\}^2 + + \Big\{\mathrm{Im}\Big[G\big(j\omega_p\big)\Big]\Big\}^2 \Big) \\ \Big\{\mathrm{Re}\Big[\big(j\omega_p\big)^{lq}\Big]\mathrm{Re}\Big[\big(j\omega_p\big)^{cq}\Big] + \mathrm{Im}\Big[\big(j\omega_p\big)^{lq}\Big]\mathrm{Im}\Big[\big(j\omega_p\big)^{cq}\Big]\Big\} w_p, \\ l = 1\ldots n \wedge c = 1\ldots n \tag{4.34}$$

$$e_{l,1} = \sum_{p=1}^{f} \Big\{ -\mathrm{Re}\Big[\big(j\omega_p\big)^{lq}\Big]\mathrm{Re}\Big[G\big(j\omega_p\big)\Big] - \mathrm{Im}\Big[\big(j\omega_p\big)^{lq}\Big]\mathrm{Im}\Big[G\big(j\omega_p\big)\Big]\Big\} w_p, \\ l = 0\ldots m \tag{4.35}$$

$$g_{l,1} = \sum_{p=1}^{f} -\mathrm{Re}\Big[\big(j\omega_p\big)^{lq}\Big]\Big(\Big\{\mathrm{Re}\Big[G\big(j\omega_p\big)\Big]\Big\}^2 + \Big\{\mathrm{Im}\Big[G\big(j\omega_p\big)\Big]\Big\}^2\Big) w_p \\ l = 1\ldots n \tag{4.36}$$

4.1.3. Second improvement: the iterative method of Sanathanan and Koerner

Another way of improving Levy's method was proposed by Sanathanan and Koerner (1963) and results from performing several iterations where variable E is replaced by

$$E_L = \frac{GD - N}{D_{L-1}} \tag{4.37}$$

where L is the iteration number and D_{L-1} is the denominator found in the previous iteration. In the first iteration this is assumed to be 1 and the result is that of Levy's method. If convergence exists, subsequent iterations will see E_L converge to ε. This time the variable minimised is

$$\sum_{p=1}^{f} \left|E\left(\omega_p\right)\right|^2 \frac{1}{\left|D_{L-1}\left(\omega_p\right)\right|^2} \tag{4.38}$$

and the fraction is the weight:

$$w_p = \frac{1}{\left|D_{L-1}\left(\omega_p\right)\right|^2} \tag{4.39}$$

It depends on coefficients known from the last iteration, not the current one, and so derivatives (4.12) and (4.14) are again not affected. Thus (4.31) to (4.36) remain valid (save that w_p is given by a different expression), and these are the values with which (4.16) is to be solved in each iteration. The resulting values of a will be used to find the new weights for the next iteration. The process may be stopped when no significant change in parameters is achieved or after some pre-set number of iterations (which is sometimes advisable because too many iterations may cause numerical errors to accumulate causing the result to diverge).

4.1.4. Third improvement: the iterative method of Lawrence and Rogers

All possibilities addressed this far involve solving a linear set of equations, and with all of them, if new data from new frequencies appear, the system will have to be solved again. Lawrence and Rogers (1979) developed an iterative method to avoid solving the system again if new data is obtained; this method deals with each frequency at one time. (This is not only for saving time. As will be seen in the subsections that follow, equation systems that show up with the methods of previous subsections may cause numerical problems to arise. Avoiding such systems may thus be numerically favourable.) It stems from writing (4.6) in the following form:

$$E = GD - N = G\left(1 + a^T s\right) - b^T t = G + a^T G s - b^T t \tag{4.40}$$

where

$$s = \begin{bmatrix} \left(j\omega\right)^q \\ \vdots \\ \left(j\omega\right)^{nq} \end{bmatrix} \tag{4.41}$$

$$t = \begin{bmatrix} 1 \\ (j\omega)^q \\ \vdots \\ (j\omega)^{mq} \end{bmatrix} \quad (4.42)$$

If we let

$$v = \begin{bmatrix} b \\ a \end{bmatrix} \quad (4.43)$$

$$u = \begin{bmatrix} t \\ -Gs \end{bmatrix} \quad (4.44)$$

then (4.40) becomes

$$E = G - v^T u \quad (4.45)$$

Now instead of (4.7) we may alternatively write

$$|E|^2 = \left(G - v^T u\right)\overline{\left(G - v^T u\right)^T} = G\overline{G} - G\overline{u^T}v - \overline{G}v^T u + v^T u\overline{u^T}v \quad (4.46)$$

where it has been taken into account that G is a scalar (whereas u and v are vectors) and that v is real-valued (whereas G and u are complex-valued). Differentiating (4.46) in order to v gives

$$\frac{\partial |E|^2}{\partial v} = -G\overline{u} - \overline{G}u + u\overline{u^T}v + \overline{u}u^T v \quad (4.47)$$

and equalling (4.47) to zero gives

$$\left(u\overline{u^T} + \overline{u}u^T\right)v = G\overline{u} + \overline{G}u \quad (4.48)$$

It should be noticed that both the matrix in the left-hand side multiplying v and the vector in the right-hand side are real-valued. And since we usually deal not with only one but with f frequencies, this becomes

$$\sum_{k=1}^{f}\left(u_k\overline{u_k^T} + \overline{u_k}u_k^T\right)v = \sum_{k=1}^{f}\left(G_k\overline{u_k} + \overline{G_k}u_k\right) \quad (4.49)$$

And, if weights are included, we shall want to minimise

$$\frac{\partial\left(|E|^2 w^2\right)}{\partial v} = w^2\left(-G\overline{u} - \overline{G}u + u\overline{u^T}v + \overline{u}u^T v\right) \quad (4.50)$$

and (4.49) becomes

$$\sum_{k=1}^{f} w_k^2 \left(u_k \overline{u_k^T} + \overline{u_k} u_k^T \right) v = \sum_{k=1}^{f} w_k^2 \left(G_k \overline{u_k} + \overline{G_k} u_k \right) \tag{4.51}$$

Until now this is solely putting (4.16) under an equivalent, more compact form (the resulting system of equations is, of course, equivalent; the dimension of the matrix and the size of the vector in (4.51) are the same as those in (4.16)). Yet (4.51) allows for the developments that follow. Let

$$\mathbf{H}_f^{-1} = \sum_{k=1}^{f} w_k^2 \left(u_k \overline{u_k^T} + \overline{u_k} u_k^T \right) = \mathbf{H}_{f-1}^{-1} + w_f^2 \left(u_f \overline{u_f^T} + \overline{u_f} u_f^T \right) \tag{4.52}$$

Then (4.51) becomes

$$\mathbf{H}_f^{-1} v_f = \sum_{k=1}^{f} w_k^2 \left(G_k \overline{u_k} + \overline{G_k} u_k \right) \tag{4.53}$$

where the subscript on v has been added to show that the solution is obtained from data concerning f frequencies. Additionally,

$$\begin{aligned} &\sum_{k=1}^{f} w_k^2 \left(G_k \overline{u_k} + \overline{G_k} u_k \right) = w_f^2 \left(G_f \overline{u_f} + \overline{G_f} u_f \right) + \sum_{k=1}^{f-1} w_k^2 \left(G_k \overline{u_k} + \overline{G_k} u_k \right) = \\ &= w_f^2 \left(G_f \overline{u_f} + \overline{G_f} u_f \right) + \mathbf{H}_{f-1}^{-1} v_{f-1} \end{aligned} \tag{4.54}$$

Hence

$$\begin{aligned} &\mathbf{H}_f^{-1} v_f = w_f^2 \left(G_f \overline{u_f} + \overline{G_f} u_f \right) + \mathbf{H}_{f-1}^{-1} v_{f-1} = \\ &= w_f^2 \left(G_f \overline{u_f} + \overline{G_f} u_f \right) + \left[\mathbf{H}_f^{-1} - w_f^2 \left(u_f \overline{u_f^T} + \overline{u_f} u_f^T \right) \right] v_{f-1} = \\ &= \mathbf{H}_f^{-1} v_{f-1} - w_f^2 \left(u_f \overline{u_f^T} + \overline{u_f} u_f^T \right) v_{f-1} + w_f^2 \left(G_f \overline{u_f} + \overline{G_f} u_f \right) \Rightarrow \\ &\Rightarrow v_f = v_{f-1} - \mathbf{H}_f w_f^2 \left(u_f \overline{u_f^T} + \overline{u_f} u_f^T \right) v_{f-1} + \mathbf{H}_f w_f^2 \left(G_f \overline{u_f} + \overline{G_f} u_f \right) = \\ &= v_{f-1} + \mathbf{H}_f w_f^2 \left[u_f \left(\overline{G_f} - \overline{u_f^T} v_{f-1} \right) + \overline{u_f} \left(G_f - u_f^T v_{f-1} \right) \right] \end{aligned} \tag{4.55}$$

This last equality means that once a vector v with parameters for the model is obtained from data concerning $f-1$ frequencies, it is possible to improve it taking into account data from another frequency. It is even possible to find an expression for $\mathbf{H}$ that does not require inverting $\mathbf{H}^{-1}$, developing (4.52) as follows:

$$\begin{aligned} &\mathbf{H}_f^{-1} = \mathbf{H}_{f-1}^{-1} + w_f^2 \left(\mathrm{Re}\left[u_f \right] + j \, \mathrm{Im}\left[u_f \right] \right) \left(\mathrm{Re}\left[u_f^T \right] - j \, \mathrm{Im}\left[u_f^T \right] \right) + \\ &+ w_f^2 \left(\mathrm{Re}\left[u_f \right] - j \, \mathrm{Im}\left[u_f \right] \right) \left(\mathrm{Re}\left[u_f^T \right] + j \, \mathrm{Im}\left[u_f^T \right] \right) = \end{aligned}$$

$$
\begin{aligned}
&= \mathbf{H}_{f-1}^{-1} + w_f^2 \Big(\mathrm{Re}\big[u_f\big]\mathrm{Re}\big[u_f^T\big] + \mathrm{Im}\big[u_f\big]\mathrm{Im}\big[u_f^T\big] - \\
&- j\,\mathrm{Re}\big[u_f\big]\mathrm{Im}\big[u_f^T\big] + j\,\mathrm{Im}\big[u_f\big]\mathrm{Re}\big[u_f^T\big] + \\
&+ \mathrm{Re}\big[u_f\big]\mathrm{Re}\big[u_f^T\big] + \mathrm{Im}\big[u_f\big]\mathrm{Im}\big[u_f^T\big] + \\
&+ j\,\mathrm{Re}\big[u_f\big]\mathrm{Im}\big[u_f^T\big] - j\,\mathrm{Im}\big[u_f\big]\mathrm{Re}\big[u_f^T\big] \Big) = \\
&= \mathbf{H}_{f-1}^{-1} + 2w_f^2 \Big(\mathrm{Re}\big[u_f\big]\mathrm{Re}\big[u_f^T\big] + \mathrm{Im}\big[u_f\big]\mathrm{Im}\big[u_f^T\big] \Big)
\end{aligned}
\tag{4.56}
$$

Let

$$
\mathbf{Z}_f^{-1} = \mathbf{H}_{f-1}^{-1} + 2w_f^2\,\mathrm{Re}\big[u_f\big]\mathrm{Re}\big[u_f^T\big] \tag{4.57}
$$

Multiplying this by $\mathbf{Z}_f$, by $\mathbf{H}_{f-1}$ and by $\mathrm{Re}\big[u_f\big]$

$$
\mathbf{I} = \mathbf{Z}_f\mathbf{H}_{f-1}^{-1} + 2w_f^2\mathbf{Z}_f\,\mathrm{Re}\big[u_f\big]\mathrm{Re}\big[u_f^T\big] \Leftrightarrow \tag{4.58}
$$

$$
\Leftrightarrow \mathbf{H}_{f-1} = \mathbf{Z}_f + 2w_f^2\mathbf{Z}_f\,\mathrm{Re}\big[u_f\big]\mathrm{Re}\big[u_f^T\big]\mathbf{H}_{f-1} \Leftrightarrow \tag{4.59}
$$

$$
\begin{aligned}
&\Leftrightarrow \mathbf{H}_{f-1}\mathrm{Re}\big[u_f\big] = \mathbf{Z}_f\,\mathrm{Re}\big[u_f\big] + 2w_f^2\mathbf{Z}_f\,\mathrm{Re}\big[u_f\big]\mathrm{Re}\big[u_f^T\big]\mathbf{H}_{f-1}\mathrm{Re}\big[u_f\big] = \\
&= \mathbf{Z}_f\,\mathrm{Re}\big[u_f\big]\Big(1 + 2w_f^2\,\mathrm{Re}\big[u_f^T\big]\mathbf{H}_{f-1}\mathrm{Re}\big[u_f\big]\Big)
\end{aligned}
\tag{4.60}
$$

It should be noticed that the term within parenthesis is scalar. Rearranging and then multiplying by $\mathrm{Re}\big[u_f^T\big]\mathbf{H}_{f-1}$,

$$
\begin{aligned}
&\mathbf{Z}_f\,\mathrm{Re}\big[u_f\big] = \frac{\mathbf{H}_{f-1}\mathrm{Re}\big[u_f\big]}{1 + 2w_f^2\,\mathrm{Re}\big[u_f^T\big]\mathbf{H}_{f-1}\mathrm{Re}\big[u_f\big]} \Leftrightarrow \\
&\Leftrightarrow \mathbf{Z}_f\,\mathrm{Re}\big[u_f\big]\mathrm{Re}\big[u_f^T\big]\mathbf{H}_{f-1} = \frac{\mathbf{H}_{f-1}\mathrm{Re}\big[u_f\big]\mathrm{Re}\big[u_f^T\big]\mathbf{H}_{f-1}}{1 + 2w_f^2\,\mathrm{Re}\big[u_f^T\big]\mathbf{H}_{f-1}\mathrm{Re}\big[u_f\big]}
\end{aligned}
\tag{4.61}
$$

Recall that the denominator is a scalar. Now (4.59) shows that

$$
\begin{aligned}
&\mathbf{H}_{f-1} = \mathbf{Z}_f + 2w_f^2\mathbf{Z}_f\,\mathrm{Re}\big[u_f\big]\mathrm{Re}\big[u_f^T\big]\mathbf{H}_{f-1} \Leftrightarrow \\
&\Leftrightarrow \mathbf{Z}_f\,\mathrm{Re}\big[u_f\big]\mathrm{Re}\big[u_f^T\big]\mathbf{H}_{f-1} = \frac{\mathbf{H}_{f-1} - \mathbf{Z}_f}{2w_f^2}
\end{aligned}
\tag{4.62}
$$

From (4.61) and (4.62)

$$
\frac{\mathbf{H}_{f-1}\mathrm{Re}\big[u_f\big]\mathrm{Re}\big[u_f^T\big]\mathbf{H}_{f-1}}{1 + 2w_f^2\,\mathrm{Re}\big[u_f^T\big]\mathbf{H}_{f-1}\mathrm{Re}\big[u_f\big]} = \frac{\mathbf{H}_{f-1} - \mathbf{Z}_f}{2w_f^2} \Leftrightarrow
$$

$$\Leftrightarrow \mathbf{Z}_f = \mathbf{H}_{f-1} - \frac{\mathbf{H}_{f-1}\operatorname{Re}\left[u_f\right]\operatorname{Re}\left[u_f^T\right]\mathbf{H}_{f-1}}{\frac{1}{2w_f^2} + \operatorname{Re}\left[u_f^T\right]\mathbf{H}_{f-1}\operatorname{Re}\left[u_f\right]} =$$
$$= \mathbf{H}_{f-1}\left(\mathbf{I} - \frac{\operatorname{Re}\left[u_f\right]\operatorname{Re}\left[u_f^T\right]\mathbf{H}_{f-1}}{\frac{1}{2w_f^2} + \operatorname{Re}\left[u_f^T\right]\mathbf{H}_{f-1}\operatorname{Re}\left[u_f\right]}\right) \tag{4.63}$$

The identity matrix above has the same size of $\mathbf{H}_{f-1}$, which is also the size of matrix $\operatorname{Re}\left[u_f\right]\operatorname{Re}\left[u_f^T\right]$.

The steps that follow are close parallels of those from (4.57) to (4.63). From (4.56) and (4.57) we know that

$$\mathbf{H}_f^{-1} = \mathbf{Z}_f^{-1} + 2w_f^2\operatorname{Im}\left[u_f\right]\operatorname{Im}\left[u_f^T\right] \tag{4.64}$$

Multiplying this by $\mathbf{H}_f$, by $\mathbf{Z}_f$ and by $\operatorname{Im}\left[u_f\right]$

$$\mathbf{I} = \mathbf{H}_f\mathbf{Z}_f^{-1} + 2w_f^2\mathbf{H}_f\operatorname{Im}\left[u_f\right]\operatorname{Im}\left[u_f^T\right] \Leftrightarrow \tag{4.65}$$

$$\Leftrightarrow \mathbf{Z}_f = \mathbf{H}_f + 2w_f^2\mathbf{H}_f\operatorname{Im}\left[u_f\right]\operatorname{Im}\left[u_f^T\right]\mathbf{Z}_f \Leftrightarrow \tag{4.66}$$

$$\Leftrightarrow \mathbf{Z}_f\operatorname{Im}\left[u_f\right] = \mathbf{H}_f\operatorname{Im}\left[u_f\right] + 2w_f^2\mathbf{H}_f\operatorname{Im}\left[u_f\right]\operatorname{Im}\left[u_f^T\right]\mathbf{Z}_f\operatorname{Im}\left[u_f\right] =$$
$$= \mathbf{H}_f\operatorname{Im}\left[u_f\right]\left(1 + 2w_f^2\operatorname{Im}\left[u_f^T\right]\mathbf{Z}_f\operatorname{Im}\left[u_f\right]\right) \tag{4.67}$$

Rearranging and then multiplying by $\operatorname{Im}\left[u_f^T\right]\mathbf{Z}_f$,

$$\mathbf{H}_f\operatorname{Im}\left[u_f\right] = \frac{\mathbf{Z}_f\operatorname{Im}\left[u_f\right]}{1 + 2w_f^2\operatorname{Im}\left[u_f^T\right]\mathbf{Z}_f\operatorname{Im}\left[u_f\right]} \Leftrightarrow$$
$$\Leftrightarrow \mathbf{H}_f\operatorname{Im}\left[u_f\right]\operatorname{Im}\left[u_f^T\right]\mathbf{Z}_f = \frac{\mathbf{Z}_f\operatorname{Im}\left[u_f\right]\operatorname{Im}\left[u_f^T\right]\mathbf{Z}_f}{1 + 2w_f^2\operatorname{Im}\left[u_f^T\right]\mathbf{Z}_f\operatorname{Im}\left[u_f\right]} \tag{4.68}$$

Now (4.66) shows that

$$\mathbf{Z}_f = \mathbf{H}_f + 2w_f^2\mathbf{H}_f\operatorname{Im}\left[u_f\right]\operatorname{Im}\left[u_f^T\right]\mathbf{Z}_f \Leftrightarrow$$
$$\Leftrightarrow \mathbf{H}_f\operatorname{Im}\left[u_f\right]\operatorname{Im}\left[u_f^T\right]\mathbf{Z}_f = \frac{\mathbf{Z}_f - \mathbf{H}_f}{2w_f^2} \tag{4.69}$$

From (4.68) and (4.69)

$$\frac{\mathbf{Z}_f \operatorname{Im}\left[u_f\right]\operatorname{Im}\left[u_f^T\right]\mathbf{Z}_f}{1+2w_f^2 \operatorname{Im}\left[u_f^T\right]\mathbf{Z}_f \operatorname{Im}\left[u_f\right]} = \frac{\mathbf{Z}_f - \mathbf{H}_f}{2w_f^2} \Leftrightarrow$$

$$\mathbf{H}_f = \mathbf{Z}_f - \frac{\mathbf{Z}_f \operatorname{Im}\left[u_f\right]\operatorname{Im}\left[u_f^T\right]\mathbf{Z}_f}{\frac{1}{2w_f^2} + \operatorname{Im}\left[u_f^T\right]\mathbf{Z}_f \operatorname{Im}\left[u_f\right]} = \mathbf{Z}_f\left(\mathbf{I} - \frac{\operatorname{Im}\left[u_f\right]\operatorname{Im}\left[u_f^T\right]\mathbf{Z}_f}{\frac{1}{2w_f^2} + \operatorname{Im}\left[u_f^T\right]\mathbf{Z}_f \operatorname{Im}\left[u_f\right]}\right) \quad (4.70)$$

The identity matrix above has the same size of $\mathbf{Z}_f$, which is also the size of matrix $\operatorname{Im}\left[u_f\right]\operatorname{Im}\left[u_f^T\right]$.

The best way to use this method is to begin with some values for **H** and v (which is made up of parameters a and b), obtained applying (4.52) and (4.53) with a few frequencies. Data from each of the further frequencies is then taken into account using (4.63) and (4.70), with which it is possible to obtain a value for $\mathbf{H}_f$,, from the value of $\mathbf{H}_{f-1}$, inverting only a scalar. Then (4.55) is used to update vector of parameters v.

Actually it is possible to begin with no estimate at all, making

$$\begin{aligned} &v_0 = 0 \\ &\mathbf{H}_0^{-1} = 0 \Rightarrow \mathbf{H}_0 = \mathbf{I}\times\infty \end{aligned} \quad (4.71)$$

Since infinity is not an available numerical value, some positive real number x is used instead and

$$\begin{aligned} &v_0 = 0 \\ &\mathbf{H}_0 = \mathbf{I}\times x \end{aligned} \quad (4.72)$$

However, it is rather hard to tell in advance which number to use; large real numbers, close to the floating-point limit, are good approximations of infinity but are likely to cause overflow errors; furthermore, there are cases when a moderate choice performs better than a very large one.

The specificity of the fractional case consists solely in the definition of s and t, in (1.41) and (1.42).

4.1.5. First example: numerical results with exact data

The exact frequency responses of some transfer functions were reckoned[66] and the methods of the previous sections were used to reconstruct the function. As Table 4.1 shows[67], this is usually possible, provided that a compatible structure is offered. Column J of that Table presents an index showing how close the frequency response of the identified model $\hat{G}(j\omega)$ is to the data from which the model was obtained $G(j\omega)$. It is given by

[66] The identification methods above were implemented as part of toolbox Ninteger for Matlab. Results that follow were obtained with that implementation. See the Appendix B.

[67] This Table is found in page 113 for reasons of layout.

$$J = \frac{1}{f}\sum_{i=1}^{f}\left[G(j\omega) - \hat{G}(j\omega)\right]^2 \tag{4.73}$$

where f is the number of sampling frequencies. Insignificant values of J appear when only slight numerical discrepancies exist; higher values reflect the lack of quality of the model identified.

Results of the iterative method of Lawrence and Rogers are not shown because, if the initial conditions in (4.72) are assumed, it is necessary to have data from many frequencies to get any acceptable results. Actually the best way of using that iterative method is to combine it with the weights of Sanathanan and Koerner's method. (For this method the last column shows the number of iterations n needed to find the result given.)

4.1.6. Second example: numerical results with corrupted data

Noise is usually present in experimental data. To check how it may affect this identification method, frequency responses were added a Gaussian distributed, zero mean noise, with a 1 dB or 1 degree variance. As Table 4.2 shows[68], this does not necessarily prevent a reasonable approximation of the original transfer function to be found, but the structure offered needs to be closer to the correct one.

4.1.7. Comments

In what concerns numerical problems, it should be noticed that all poor results of Table 4.1 were returned by Matlab together with a warning about a bad condition number of the matrix in (4.16). Choosing structures that avoid such condition numbers may prove to be a good option. The number of sampling frequencies also affects the conditioning of the matrix. Increasing the number of sampling frequencies usually helps improving the results, though not always, there being a few exceptions in which results even deteriorate.

It is also worth of notice that index J is not always a sufficient indicator of the goodness of an approximation. Sometimes poorer models achieve a lower value of this index. Above all, the improvements of Levy's method do sometimes lead to better models, but, again, this is not always so. It seems that the best policy is to use all methods—Levy's method, Levy's method with weights such as those devised by Vinagre, Sanathanan and Koerner's iterative method—and then check which one has a more reasonable result, even using Lawrence and Roger's iteration to assist in the reckonings, avoiding inverting ill-conditioned matrixes, when good estimates for **H** and v are already available.

So, in short, this extension of Levy's method appears to enjoy the same merits and suffer from the same drawbacks of the original integer-order version. It namely requires providing in advance the orders of the numerator and the denominator, n and m; and this extension also requires q, the commensurate order. Of course, numerical problems usually arise when excessively high values for n and m or excessively low values for q are provided. There are two possible solutions for dealing with this requirement: a visual inspection of frequency data may suggest the appropriate orders; or several possible combinations of values may be tried, and the best retained. This last option is

[68] This Table is found in page 114 for reasons of layout.

possible because the algorithm runs fast enough in modern computers. The same comments may be made on the improvements of Levy's method also extended in this paper.

4.2. Crone identification method

This section presents the Crone identification method for finding the coefficients of an integer transfer function that models a plant with a known frequency behaviour. A variation of this method for finding a digital model is also presented.

4.2.1. Identification of a model in the *s*-domain

Let us suppose[69] we know the frequency behaviour of plant G at f frequencies $\omega_i, i = 1 \ldots f$, where its gain is g_i and its phase is ϕ_i:

$$g_i = \left|G\left(j\omega_i\right)\right| \tag{4.74}$$

$$\phi_i = \arg\left[G\left(j\omega_i\right)\right] \tag{4.75}$$

Let us also suppose we want to fit an s-domain model $\hat{G}$, with a total of M zeros and poles, to this data:

$$\hat{G}\left(s\right) = C_0 \frac{\prod_{k=1}^{m}\left(1+\frac{s}{b_k}\right)}{\prod_{k=1}^{n}\left(1+\frac{s}{a_k}\right)}, \quad C_0 > 0, m+n = M \tag{4.76}$$

The phase of its frequency response is

$$\arg\left[\hat{G}\left(j\omega\right)\right] = \arg\left[C_0 \frac{\prod_{k=1}^{m}\left(1+\frac{j\omega}{b_k}\right)}{\prod_{k=1}^{n}\left(1+\frac{j\omega}{a_k}\right)}\right] = \sum_{k=1}^{m} \operatorname{arctg}\frac{\omega}{b_k} - \sum_{k=1}^{n} \operatorname{arctg}\frac{\omega}{a_k} \tag{4.77}$$

Since the arc of tangent is an odd function, this is the same as the phase of

$$\tilde{G}\left(s\right) = C_0 \frac{1}{\prod_{k=1}^{M}\left(1+\frac{s}{c_k}\right)} \tag{4.78}$$

where

[69] Oustaloup (1991, p. 204-208).

$$c_k = \begin{cases} a_k, & \text{if } k \le n \\ -b_{k-n}, & \text{if } k \ge n+1 \end{cases} \tag{4.79}$$

The expression giving the phase (4.77) is now

$$\arg\left[\tilde{G}(j\omega)\right] = -\sum_{k=1}^{M} \operatorname{arctg}\frac{\omega}{c_k} \tag{4.80}$$

We want that

$$\arg\left[\tilde{G}(j\omega_i)\right] = \phi_i,\ i = 1\ldots f \tag{4.81}$$

The next step is solving the non-linear set of equations above.

4.2.2. Solving equations (4.81)

First we make

$$y_{i,k} = \frac{\omega_i}{c_k} \tag{4.82}$$

and rewrite (4.81) as

$$\sum_{k=1}^{M} \operatorname{arctg} y_{i,k} = -\phi_i,\ i = 1\ldots f \tag{4.83}$$

Then we apply Theorem A. 9 and rewrite (4.83), after taking the tangent of its both sides, as

$$\frac{\sum_{k\in I_M} (-1)^{\mathscr{C}(k/2)} \mathscr{S}_{i,k}}{1+\sum_{k\in P_M} (-1)^{\mathscr{C}(k/2)} \mathscr{S}_{i,k}} = -\operatorname{tg}(\phi_i),\ i = 1\ldots f \tag{4.84}$$

where I_a and P_a are the sets of odd and even natural numbers smaller than or equal to a

$$I_a = \left\{n \in \mathbb{N} : n \le a \wedge \exists_{p\in\mathbb{N}} n = 2p-1\right\} \tag{4.85}$$

$$P_a = \left\{n \in \mathbb{N} : n \le a \wedge \exists_{p\in\mathbb{N}} n = 2p\right\} \tag{4.86}$$

and

$$\mathscr{S}_{i,1} = y_{i,1} + y_{i,2} + \ldots + y_{i,a} \tag{4.87}$$

$$\begin{aligned}\mathscr{S}_{i,2} &= y_{i,1}y_{i,2} + y_{i,1}y_{i,3} + \ldots + y_{i,1}y_{i,a} + \\ &+ y_{i,2}y_{i,3} + \ldots + y_{i,2}y_{i,a} + \\ &+ \ldots + \\ &+ y_{i,a-1}y_{i,a}\end{aligned} \tag{4.88}$$

$$\begin{aligned}\mathscr{S}_{i,3} &= y_{i,1}y_{i,2}y_{i,3} + y_{i,1}y_{i,2}y_{i,4} + \ldots + y_{i,1}y_{i,2}y_{i,a} + \\ &+ y_{i,1}y_{i,3}y_{i,4} + \ldots + y_{i,1}y_{i,3}y_{i,a} + \\ &+ \ldots + \\ &+ y_{i,1}y_{i,a-1}y_{i,a} + \\ &+ y_{i,2}y_{i,3}y_{i,4} + \ldots + y_{i,2}y_{i,3}y_{i,a} + \\ &+ \ldots + \\ &+ y_{i,2}y_{i,a-1}y_{i,a} + \\ &\vdots \\ &+ y_{i,a-2}y_{i,a-1}y_{i,a}\end{aligned} \tag{4.89}$$

or, generally,

$$\mathscr{S}_{i,k} = \sum \prod_{j \in \mathscr{B}_{M,k}} y_{i,j} \tag{4.90}$$

$$\begin{cases} \#\left(\mathscr{B}_{a,k}\right) = k \\ \mathscr{B}_{a,k} \subset \{n \in \mathbb{N} : n \le a\} \end{cases} \tag{4.91}$$

Equations (4.84) may further be put as

$$\sum_{k=1}^{M} (-1)^{\mathscr{C}(k/2)} \Phi_{i,k} \mathscr{S}_{i,k} = -\Phi_{i,0}, \; i = 1 \ldots f \tag{4.92}$$

where

$$\Phi_{i,k} = \begin{cases} \operatorname{tg}(\phi_i), & \exists_{p \in \mathbb{N}_0} k = 2p \\ 1, & \exists_{p \in \mathbb{N}} k = 2p - 1 \end{cases} \tag{4.93}$$

The set of equations (4.92) is finally a linear one, but variables are not the same over all equations. Luckily, they may all be written as linear combinations of those of the first equation. This is because, as is clear from definition (4.82),

$$y_{i,k} = y_{1,k} A_i, \quad i = 1 \ldots f \tag{4.94}$$

$$A_i = \frac{\omega_i}{\omega_1}, \quad i = 1 \ldots f \tag{4.95}$$

Thus

$$\mathscr{S}_{i,1} = \mathscr{S}_{1,1} A_i \tag{4.96}$$

$$\mathscr{S}_{i,2} = \mathscr{S}_{1,2} A_i^2 \tag{4.97}$$

$$\mathscr{S}_{i,3} = \mathscr{S}_{1,3} A_i^3 \tag{4.98}$$

and so on; or, generally,

$$\mathscr{S}_{i,k} = \mathscr{S}_{1,k} A_i^k \tag{4.99}$$

Now, we may use (4.99) in (4.92) to get

$$\sum_{k=1}^{M} (-1)^{\mathscr{C}(k/2)} \Phi_{i,k} \mathscr{S}_{1,k} A_i^k = -\Phi_{i,0},\, i = 1 \ldots f \tag{4.100}$$

4.2.3. Summing up the *s*-domain case

To find the coefficients of (4.76), we solve the set of equations (4.100), which is linear on variables $\mathscr{S}$. Symbols used in (4.100) are defined in (4.95) and (4.93). Notice that the system will be square only if $M = f$. Otherwise a solution in least-squares sense is to be found.

Then the values of variables y_1 are found solving the polynomial equation

$$y_1^M + \sum_{k=1}^{M} (-1)^k y_1^{M-k} \mathscr{S}_{1,k} = 0 \tag{4.101}$$

that has M roots. When the y_1 are known, coefficients c are obtained from them inverting (4.82):

$$c_k = \frac{\omega_1}{y_{1,k}},\, k = 1 \ldots M \tag{4.102}$$

These are the coefficients of (4.76), as (4.79) shows.

At this point it is still not known which of the values of c correspond to zeros and which correspond to poles. So, all possibilities (which are 2^M in number) are checked, and the transfer function with a gain closer to g is retained.

Since it is possible that spurious zeros and poles appear when using this identification method (e.g. zeros or poles at frequencies higher than those we are interested in, or with effects in the phase that cancel each other, or resulting from the periodicity of circular functions), especially when M is unnecessarily high, a common-sense judgement must be passed on the relevance of every c for the model.

4.2.4. Identification of a digital model

The following subsections develop an identification method similar to that of the previous section but for digital models. This must be done separately for three different cases:

- the model has M zeros and no poles;
- the model has M poles and no zeros;
- the model has both poles and zeros, M in all.

4.2.5. Identification of a FIR filter

Let us suppose we know the frequency behaviour of a plant G at f (normalised) frequencies $\Omega_i, i=1\dots f$, where its gain is g_i and its phase is ϕ_i. Let us also suppose that we want to fit a digital model $\hat{G}$ with M zeros (that is to say, a finite impulse response filter, or FIR filter) to this data:

$$\hat{G}\left(z^{-1}\right)=C_0\prod_{k=1}^{M}\left(1-b_k z^{-1}\right), C_0>0 \tag{4.103}$$

The phase of its frequency response is

$$\arg\left[\hat{G}\left(e^{j\Omega}\right)\right]=\arg\left[C_0\prod_{k=1}^{M}\left(1-b_k e^{-j\Omega}\right)\right]=\arg\left[C_0\prod_{k=1}^{M}\left(1-b_k\cos\Omega+jb_k\sin\Omega\right)\right]=$$
$$=\sum_{k=1}^{M}\operatorname{arctg}\frac{b_k\sin\Omega}{1-b_k\cos\Omega} \tag{4.104}$$

We want that

$$\arg\left[\hat{G}\left(e^{j\Omega_i}\right)\right]=\phi_i, i=1\dots f \tag{4.105}$$

The next step is solving the non-linear set of equations above.

4.2.6. Solving equations (4.105)

The strategy used for solving (4.105) closely follows that for the s-domain case (the major difference being that now calculations will be more tedious). First we make

$$y_{i,k}=\frac{1}{b_k}\operatorname{cosec}\Omega_i-\operatorname{cotg}\Omega_i \tag{4.106}$$

and rewrite (4.105) as

$$\sum_{k=1}^{M}\operatorname{arctg}\frac{1}{y_{i,k}}=\phi_i\Leftrightarrow\sum_{k=1}^{M}\left(\frac{\pi}{2}-\operatorname{arctg}y_{i,k}\right)=\phi_i\Leftrightarrow$$
$$\Leftrightarrow\sum_{k=1}^{M}\operatorname{arctg}y_{i,k}=M\frac{\pi}{2}-\phi_i, i=1\dots f \tag{4.107}$$

Theorem A. 9 allows rewriting (4.107), after taking the tangent of its both sides, as

$$\frac{\sum_{k\in I_M}(-1)^{\mathscr{C}(k/2)}\mathscr{S}_{i,k}}{1+\sum_{k\in P_M}(-1)^{\mathscr{C}(k/2)}\mathscr{S}_{i,k}}=-\operatorname{tg}\left(\phi_i-M\frac{\pi}{2}\right),\quad i=1\ldots N \tag{4.108}$$

where the $\mathscr{S}$ are defined as in (4.90) and sets I, P and $\mathscr{B}$ are defined as in (4.85), (4.86) and (4.91). Equations (4.108) may further be put into the form

$$\sum_{k=1}^{M}(-1)^{\mathscr{C}(k/2)}\Phi_{i,k}\mathscr{S}_{i,k}=-\Phi_{i,0},\, i=1\ldots f \tag{4.109}$$

(which is the same as (4.92)), if we define

$$\Phi_{i,k}=\begin{cases}\operatorname{tg}\left(\phi_i-M\frac{\pi}{2}\right), & \exists_{p\in\mathbb{N}_0}k=2p\\ 1, \quad \exists_{p\in\mathbb{N}}k=2p-1\end{cases} \tag{4.110}$$

The set of equations (4.109) is finally a linear one, but variables are not the same over all equations. Luckily, they may all be written as linear combinations of those of the first equation. This is because, as is clear from definition (4.106),

$$y_{i,k}=y_{1,k}A_i+B_i,\quad i=1\ldots f \tag{4.111}$$

$$A_i=\frac{\sin\Omega_1}{\sin\Omega_i},\quad i=1\ldots f \tag{4.112}$$

$$B_i=\frac{\cos\Omega_1}{\sin\Omega_i}-\operatorname{cotg}\Omega_i,\quad i=1\ldots f \tag{4.113}$$

Thus

$$\mathscr{S}_{i,1}=\mathscr{S}_{1,1}A_i+MB_i \tag{4.114}$$

$$\mathscr{S}_{i,2}=\mathscr{S}_{1,2}A_i^2+(M-1)\mathscr{S}_{1,1}A_iB_i+\left(\sum_{i=1}^{M-1}i\right)B_i^2 \tag{4.115}$$

$$\mathscr{S}_{i,3}=\mathscr{S}_{1,3}A_i^3+(M-2)\mathscr{S}_{1,2}A_i^2B_i+\left(\sum_{i=1}^{M-2}i\right)\mathscr{S}_{1,1}A_iB_i^2+\left(\sum_{j=1}^{M-2}\sum_{i=1}^{j}i\right)B_i^3 \tag{4.116}$$

and so on; or, generally,

$$\mathscr{S}_{i,k}=\sum_{r=0}^{k}\left(\mathscr{T}_{k-r-1,M-k+1}\mathscr{S}_{1,r}A_i^rB_i^{k-r}\right) \tag{4.117}$$

where

$$\mathscr{S}_{1,0}=1 \tag{4.118}$$

$$\mathscr{T}_{-1,a}=1 \tag{4.119}$$

$$\mathscr{T}_{0,a} = a \tag{4.120}$$

$$\mathscr{T}_{1,a} = \sum_{i=1}^{a} i \tag{4.121}$$

$$\mathscr{T}_{2,a} = \sum_{j=1}^{a}\sum_{i=1}^{j} i \tag{4.122}$$

and so on; or, generally,

$$\mathscr{T}_{b,a} = \sum_{i=1}^{a} \mathscr{T}_{b-1,i}, \quad b \in \mathbb{N}_0 \tag{4.123}$$

Now, we may use (4.117) in (4.109) to get

$$\begin{aligned}
&\sum_{k=1}^{M}\left[(-1)^{\mathscr{C}(k/2)}\Phi_{i,k}\sum_{r=0}^{k}\left(\mathscr{T}_{k-r-1,M-k+1}\mathscr{S}_{1,r}A_i^r B_i^{k-r}\right)\right] = -\Phi_{i,0} \Leftrightarrow \\
&\Leftrightarrow \sum_{k=1}^{M}\sum_{r=1}^{k}\left[(-1)^{\mathscr{C}(k/2)}\Phi_{i,k}\mathscr{T}_{k-r-1,M-k+1}\mathscr{S}_{1,r}A_i^r B_i^{k-r}\right] = \\
&\quad = -\Phi_{i,0} - \sum_{k=1}^{M}\left[(-1)^{\mathscr{C}(k/2)}\Phi_{i,k}\mathscr{T}_{k-1,M-k+1}B_i^k\right], \quad i = 1\ldots f
\end{aligned} \tag{4.124}$$

4.2.7. Summing up the FIR filter case

To find the coefficients of (4.103), we solve the set of equations (4.124), which is linear on variables $\mathscr{S}$. Symbols used in (4.124) are defined in (4.123), (4.112), (4.113) and (4.110). Notice that the system will be square only if $M = f$. Otherwise a solution in least-squares sense is to be found.

Then the values of variables y_1 are found solving the polynomial equation (4.101) (the same of the *s*-domain case). When the y_1 are known, coefficients *b* are obtained from them inverting (4.106):

$$b_k = \frac{\operatorname{cosec}\Omega_1}{y_{1,k} + \operatorname{cotg}\Omega_1} \tag{4.125}$$

4.2.8. Identification of a filter without zeros

Let us now suppose we want to fit a digital model $\hat{G}$ with *M* poles to the data available about plant *G*:

$$\hat{G}\left(z^{-1}\right) = C_0 \frac{1}{\prod_{k=1}^{M}\left(1 - a_k z^{-1}\right)}, C_0 > 0 \tag{4.126}$$

The phase of its frequency response is

$$\arg\left[\hat{G}\left(e^{j\Omega}\right)\right]=\arg\left[C_0\frac{1}{\prod_{k=1}^{M}\left(1-a_k e^{-j\Omega}\right)}\right]=\arg\left[C_0\frac{1}{\prod_{k=1}^{M}\left(1-a_k\cos\Omega+ja_k\sin\Omega\right)}\right]=$$
$$=-\sum_{k=1}^{M}\operatorname{arctg}\frac{a_k\sin\Omega}{1-a_k\cos\Omega} \tag{4.127}$$

Again what we want is

$$\arg\left[\hat{G}\left(e^{j\Omega_i}\right)\right]=\phi_i,\ i=1\dots f \tag{4.128}$$

(this is the same as (1.108)). The next step is solving that non-linear set of equations.

4.2.9. Solving the set of equations (4.128)

The exact solution for the set of equations (4.128) is nearly the same as that developed in subsection 4.2.6. First we make

$$y_{i,k}=\frac{1}{a_k}\operatorname{cosec}\Omega_i-\operatorname{cotg}\Omega_i \tag{4.129}$$

and rewrite (4.128) as

$$\sum_{k=1}^{M}\operatorname{arctg}\frac{1}{y_{i,k}}=-\phi_i\Leftrightarrow\sum_{k=1}^{M}\left(\frac{\pi}{2}-\operatorname{arctg}y_{i,k}\right)=-\phi_i\Leftrightarrow$$
$$\Leftrightarrow\sum_{k=1}^{M}\operatorname{arctg}y_{i,k}=M\frac{\pi}{2}+\phi_i,\quad i=1\dots f \tag{4.130}$$

From now on, the very same steps of subsection 1.2.6 may be followed, save that, instead of (4.110), we will have

$$\Phi_{i,k}=\begin{cases}\operatorname{tg}\left(-\phi_i-M\frac{\pi}{2}\right), & \exists_{p\in\mathbb{N}_0}k=2p\\ 1, & \exists_{p\in\mathbb{N}}k=2p-1\end{cases} \tag{4.131}$$

4.2.10. Summing up the case of a filter without zeros

To find the coefficients of (4.126), we solve the set of equations (4.124), which is linear on the variables $\mathscr{S}$. Symbols used in (4.124) are defined in (4.123), (4.112), (4.113) and (4.131). Notice that the system will be square only if $M=f$. Otherwise a solution in least-squares sense is to be found.

Then the values of variables y_1 are found solving the polynomial equation (4.101). When the y_1 are known, coefficients a are obtained from them inverting (4.129):

$$a_k = \frac{\operatorname{cosec}\Omega_1}{y_{1,k} + \operatorname{cotg}\Omega_1} \tag{4.132}$$

4.2.11. Identification of a general digital model

Let us finally suppose we want to fit a digital model $\hat{G}$ with *m* zeros and *n* poles, in a total of $M = m + n$, to the data available about plant *G*:

$$\hat{G}\left(z^{-1}\right) = C_0 \frac{\prod_{k=1}^{m}\left(1 - b_k z^{-1}\right)}{\prod_{k=1}^{n}\left(1 - a_k z^{-1}\right)}, \quad C_0 > 0, m, n \in \mathbb{N} \tag{4.133}$$

The phase of its frequency response is

$$\arg\left[\hat{G}\left(e^{j\Omega}\right)\right] = \arg\left[C_0 \frac{\prod_{k=1}^{m}\left(1 - b_k e^{-j\Omega}\right)}{\prod_{k=1}^{n}\left(1 - a_k e^{-j\Omega}\right)}\right] = \arg\left[C_0 \frac{\prod_{k=1}^{m}\left(1 - b_k \cos\Omega + j b_k \sin\Omega\right)}{\prod_{k=1}^{n}\left(1 - a_k \cos\Omega + j a_k \sin\Omega\right)}\right] =$$

$$= \sum_{k=1}^{m} \operatorname{arctg}\frac{b_k \sin\Omega}{1 - b_k \cos\Omega} - \sum_{k=1}^{n} \operatorname{arctg}\frac{a_k \sin\Omega}{1 - a_k \cos\Omega} \tag{4.134}$$

If we let

$$c_k = \begin{cases} a_k, & k \le n \\ b_{k-m}, & k \ge n+1 \end{cases} \tag{4.135}$$

$$\omega_k = \begin{cases} -\Omega, & k \le n \\ +\Omega, & k \ge n+1 \end{cases} \tag{4.136}$$

expression (4.134) becomes

$$\arg\left[\hat{G}\left(e^{j\Omega}\right)\right] = \sum_{k=1}^{M} \operatorname{arctg}\frac{c_k \sin\omega_k}{1 - c_k \cos\omega_k} \tag{4.137}$$

Again what we want is

$$\arg\left[\hat{G}\left(e^{j\Omega_i}\right)\right] = \phi_i, i = 1 \ldots f \tag{4.138}$$

(this is the same as (4.105) and (4.128)). The next step is solving that non-linear set of equations.

4.2.12. Solving equations (4.138)

Unfortunately, this time the strategy used so far fails[70]. Parameters c have to be found employing a numerical method to solve (4.138) or to fit them in some other wise to experimental data.

4.2.13. Summing up the general digital case

So as to find the coefficients of (4.133), we solve numerically the set of equations (4.138), where the argument is given by (4.137). The coefficients found are those of (4.133), as (4.135) and (4.136) show.

4.2.14. Final comments on the digital case

It is time to notice a difference between the digital case and that of the s-domain: in the former, the number of zeros m and the number of poles n of the model must be specified beforehand, while in the latter only the total number of poles and zeros must be specified in advance.

Of course, in the digital case the s-domain method should be imitated, requiring in advance only the total M, and trying all the $M+1$ possible combinations to see which one results in a gain behaviour close to g. This brings, however, a problem: equations (4.105) (or (4.128) or (4.138), which are all the same) must be solved $M+1$ times. In the s-domain it was not so: equations (4.81) had to be solved only once, and then the resulting frequencies—always the same—were checked to see if they should be poles or zeros. This means that in the digital case the computational effort is clearly more significant, for, even though there are fewer cases (2^M for the continuous case against $M+1$ for the digital case), solving the system (be it (4.81), (4.105), (4.128) or (4.138)) is the major computational task. It also means that, while in the s-domain case all possible solutions would have the same phase behaviour, the only differences being those of the gain behaviour, in the digital case the several candidate models will differ in both gain and phase.

It is of course possible to perform the identification so as to always obtain an s-domain model, and then discretise it if necessary. But this does not usually lead to results better than those obtained identifying a digital model directly.

In both the continuous and the digital cases, care must be taken to see, after obtaining the result, if all poles and zeros are really relevant.

[70] We could still make

$$y_{i,k} = \frac{\operatorname{cosec} \omega_{i,k}}{c_k} - \operatorname{cotg} \omega_{i,k}$$

and rewrite (4.138) in a form similar to (4.107) or (4.130), but it would *not* be possible to find all values of matrix $\mathscr{S}$ as linear combinations of the values in its first line. Hence no linear system of equations would be obtained.

4.2.15. Example

The methods of this section were used[71] to identify a model from the frequency response of

$$G\left(z^{-1}\right)=1+0.6z^{-1}-0.16z^{-2} \tag{4.139}$$

$$\Omega=\begin{bmatrix}0.1 & 0.5 & 1.0 & 2.5 & 3.0\end{bmatrix} \tag{4.140}$$

The numerical method used for models with both poles and zeros was Nelder-Mead's multidimensional minimisation method[72]. All values of M up to 10 were tried; the exact result was obtained for 2, and the second best results for 5:

$$\hat{G}\left(z^{-1}\right)=1+0.6z^{-1}-0.16z^{-2}+10^{-15}z^{-3}-10^{-16}z^{-4}-10^{-15}z^{-5} \tag{4.141}$$

$$\hat{G}\left(z^{-1}\right)=\frac{1+0.063z^{-1}-0.48z^{-2}+0.086z^{-3}-10^{-8}z^{-4}}{1-0.54z^{-1}} \tag{4.142}$$

This last model (4.142) has a pole and a zero that nearly cancel each other and are thus useless. Realising this requires, as explained in subsection 4.2.14, human supervision of the identification process.

4.3. Time domain identification

If the experimental or simulation data to which a model is to be fit consists in a time response, the model's parameters are best found by means of some non-linear minimisation technique. There are scores of them and many may be used with good results. In what follows two will be used: the Nelder-Mead simplex agorithm and genetic algorithms. They are both described in section A.6 of Appendix A.

Each of these two non-linear minimisation algorithms—or any other—may be used for identifying a fractional model by trying to minimise the error between the data available and the appropriate time response of the model (responses to impulses, steps, or even to some other input signal may be used). The time response of the model may be reckoned exactly, or the time response of an approximation thereof may be used instead. This last option has the advantage of taking immediately into account the deterioration of the time response every approximation involves.

Whatever the case, this strategy makes it necessary to choose a structure beforehand. The structure may include a commensurate order, but the commensurate order may also be considered a parameter to optimise, or even not exist at all, the non-integer orders being free to vary without being multiples of some q. As is usual, if there is no good reason to cling to some specified number of zeros and poles and to some commensurate order, it is a good idea to test several possibilities and choose the one providing the best fit.

Another issue to take into account is that minimisation algorithms may be trapped

[71] The Crone identification method (in all its versions given above) was implemented as part of toolbox Ninteger for Matlab. Results that follow were obtained with that implementation. See the Appendix B.

[72] See subsection A.6.1 of Appendix A below.

in local minima[73]. Running them more than once with the same structure but with different initial guesses may prove to be a wise move.

It is possible that some restrictions on what the model may be are known beforehand—that it is stable, or that poles and zeros are within a given frequency range, for instance. Such restrictions may be embedded in the optimisation method, preventing undesired solutions from being found, or the optimisation method may be run without restrictions and undesired solutions discarded at the end.

Of course, all this brings into the picture the question of how fast optimisation algorithms perform their task. The simplex algorithm is more favourable from this point of view: genetic algorithms take more time, but, on the other hand, because of their use of randomness, it is likely that they will have to be run fewer times.

[73] In what concerns the two methods from section A.6, genetic algorithms try to cope with that possibility by employing randomness; the Nelder-Mead algorithm, on the other hand, is, from the moment the initial simplex is obtained, deterministic. But in both cases local minima may be found.

Table 4.1 — Identification results

Transfer function	**Sampling frequencies**	***q***	***n***	***m***	**Levy's method**		**Levy's method with Vinagre's weights**		**Sanathanan and Koerner's method**		
					Identified transfer function	***J***	**Identified transfer function**	***J***	**Identified transfer function**	***J***	***n***
$\frac{1}{1+s}$	0.1, 1, 10	1	1	1	$\frac{1.0000}{1+1.0000s}$	4.7506×10^{-33}	$\frac{1.0000}{1+1.0000s}$	7.1420×10^{-33}	$\frac{1.0000}{1+1.0000s}$	6.4238×10^{-33}	2
		1	2	2	$\frac{1.0000-0.0042s}{1+0.9958s-0.0042s^2}$	8.8683×10^{-33}	$\frac{1.0000}{1+1.0000s}$	7.0547×10^{-33}	$\frac{1.0000}{1+1.0000s}$	6.4198×10^{-34}	3
		0.5	2	2	$\frac{1.0000}{1+1.0000s}$	1.9033×10^{-33}	$\frac{1.0000}{1+1.0000s}$	4.2560×10^{-31}	$\frac{1.0000}{1+1.0000s}$	5.7478×10^{-31}	3
	0.01, 0.1, 1, 10, 100	0.25	4	4	$\frac{1.0000}{1+1.0000s}$	1.2623×10^{-26}	$\frac{1.0000}{1+1.0000s}$	1.1257×10^{-24}	$\frac{1.0000}{1+1.0000s}$	7.3572×10^{-26}	3
		0.25	5	5	$\frac{1.000+0.1606s^{0.25}}{1+0.1606s^{0.25}+1.000s+0.1606s^{1.25}}$	8.0421×10^{-24}	$\frac{1.000-42.6087s^{0.25}}{1-42.6087s^{0.25}+1.000s-42.6087s^{1.25}}$	1.3533×10^{-26}	$\frac{1.000+0.3824s^{0.25}}{1+0.3824s^{0.25}+1.000s+0.3824s^{1.25}}$	2.0285×10^{-26}	100
$\frac{1}{1+s^{0.5}}$	0.1, 1, 10	0.5	1	1	$\frac{1.0000}{1+1.0000s^{0.5}}$	1.3225×10^{-31}	$\frac{1.0000}{1+1.0000s^{0.5}}$	5.5731×10^{-30}	$\frac{1.0000}{1+1.0000s^{0.5}}$	5.6494×10^{-33}	2
		0.25	2	2	$\frac{1.0000}{1+1.0000s^{0.5}}$	7.7515×10^{-29}	$\frac{1.0000}{1+1.0000s^{0.5}}$	3.6635×10^{-28}	$\frac{1.0000}{1+1.0000s^{0.5}}$	9.9147×10^{-30}	2
	0.01, 0.03, 0.1, 0.3, 1, 3, 10, 30, 100	0.25	3	3	$\frac{1.0000+0.0104s^{0.25}}{1+0.0104s^{0.25}+1.0000s^{0.5}+0.0104s^{0.75}}$	4.7447×10^{-30}	$\frac{1.0000+1.1260s^{0.25}}{1+1.1260s^{0.25}+1.0000s^{0.5}+1.1260s^{0.75}}$	8.5356×10^{-30}	$\frac{1.0000-2.5156s^{0.25}}{1-2.5156s^{0.25}+1.0000s^{0.5}-2.5156s^{0.75}}$	5.0961×10^{-29}	100
$\frac{4+5s^{0.5}+6s}{1+2s^{0.5}+3s}$	0.1, 1, 10	0.5	2	2	$\frac{4.0000+5.0000s^{0.5}+6.0000s}{1+2.0000s^{0.5}+3.0000s}$	2.7242×10^{-25}	$\frac{4.0000+5.0000s^{0.5}+6.0000s}{1+2.0000s^{0.5}+3.0000s}$	2.4901×10^{-28}	$\frac{4.0000+5.0000s^{0.5}+6.0000s}{1+2.0000s^{0.5}+3.0000s}$	2.4993×10^{-28}	2
	50 logarithmically spaced points between 10^{-3} and 10^{3}	0.5	3	3	$\frac{4.0000+5.0060s^{0.5}+6.0075s+0.0090s^{1.5}}{1+2.0015s^{0.5}+3.0030s+0.0045s^{1.5}}$	1.5720×10^{-20}	$\frac{4.0000+5.1137s^{0.5}+6.1421s+0.1705s^{1.5}}{1+2.0284s^{0.5}+3.0568s+0.0852s^{1.5}}$	7.7583×10^{-28}	$\frac{4.0000+4.5038s^{0.5}+5.3798s-0.7443s^{1.5}}{1+1.8760s^{0.5}+2.7519s-0.3721s^{1.5}}$	1.5516×10^{-28}	100

Table 4.2 — Identification results when noise is present.

Transfer function	**Sampling frequencies**	***q***	***n***	***m***	**Levy's method**		**Levy's method with Vinagre's weights**		**Sanathanan and Koerner's method**		
					Identified transfer function	***J***	**Identified transfer function**	***J***	**Identified transfer function**	***J***	***n***
$\frac{1}{1+s}$	0.1, 1, 10	1	1	1	$\frac{0.9683+0.0025s}{1+1.0457s}$	1.1557×10^{-3}	$\frac{0.9553+0.0188s}{1+0.9682s}$	7.9640×10^{-4}	$\frac{0.9636-0.0071s}{1+0.9535s}$	6.2058×10^{-4}	4
		0.5	2	2	$\frac{0.9731-0.0956s^{0.5}+0.0227s}{1-0.0519s^{0.5}+0.8812s}$	4.1194×10^{-4}	$\frac{1.0119-0.3437s^{0.5}+0.0609s}{1-0.0539s^{0.5}+0.6127s}$	9.9986×10^{-4}	$\frac{0.9877-0.1884s^{0.5}+0.0370s}{1-0.0522s^{0.5}+0.7805s}$	2.9884×10^{-4}	7
$\frac{1}{1+s^{0.5}}$	0.1, 1, 10	0.5	1	1	$\frac{0.9940+0.0091s^{0.5}}{1+0.9384s^{0.5}}$	4.2694×10^{-4}	$\frac{0.9949+0.0355s^{0.5}}{1+0.9656s^{0.5}}$	9.4836×10^{-4}	$\frac{0.9951+0.0082s^{0.5}}{1+0.9726s^{0.5}}$	3.4494×10^{-4}	4
	50 logarithmically spaced points between 10^{-3} and 10^{3}	0.25	2	2	$\frac{0.9294-0.0517s^{0.25}+0.0070s^{0.5}}{1-0.1255s^{0.25}+0.8591s^{0.5}}$	2.5614×10^{-3}	$\frac{1.0096-2.8410s^{0.25}+0.7753s^{0.5}}{1-2.2890s^{0.25}-0.5489s^{0.5}}$	2.5935×10^{-2}	$\frac{0.9547-0.3434s^{0.25}+0.0686s^{0.5}}{1-0.2973s^{0.25}+0.6253s^{0.5}}$	2.8507×10^{-3}	8
		0.25	3	3	$\frac{0.9255-0.3572s^{0.25}+0.0223s^{0.5}-0.0022s^{0.75}}{1-0.4995s^{0.25}+0.9439s^{0.5}-0.2930s^{0.75}}$	2.4983×10^{-3}	$\frac{0.9458-3.3266s^{0.25}+0.7870s^{0.5}-0.1249s^{0.75}}{1-3.4428s^{0.25}+1.3446s^{0.5}-2.2052s^{0.75}}$	2.5448×10^{-3}	$\frac{0.9537-1.1239s^{0.25}+0.3482s^{0.5}-0.0560s^{0.75}}{1-1.1235s^{0.25}+0.8792s^{0.5}-0.5235s^{0.75}}$	2.8170×10^{-3}	11

5. Synthesis of fractional controllers

Fractional controllers are those involving fractional order systems in their conception. Fractional controllers may be devised for both integer and fractional plants. Controllers may have themselves a fractional behaviour, or cause the controlled system to behave as though it were fractional. In this chapter five techniques for devising fractional controllers are addressed.

- ❖ Section 5.1 concerns first generation Crone controllers.
- ❖ Section 5.2 concerns second generation Crone controllers.
- ❖ Section 5.3 concerns third generation Crone controllers.
- ❖ Section 5.4 concerns fractional PID controllers. Three tuning methods are considered: internal model control, numerical minimisation, and use of tuning rules.
- ❖ Section 5.5 concerns H_2 and H_∞ fractional controllers.

5.1. First generation Crone controllers

First generation Crone controllers consist solely in[74]

$$C(s) = s^{\nu} \tag{5.1}$$

Their purpose is to increase the phase margin of the plant by some suitable value. This is obviously easier if fractional orders are contemplated; otherwise, only variations multiple of 90º would be admissible with a structure as simple as (5.1).

In its original formulation, (5.1) was to be approximated by an integer controller given by the Crone approximation, but of course any approximating formula from those given in section 3.3 may be used.

This type of controller is useful when the plant to control already has a constant phase, at least in a frequency range around the gain crossover frequency. In that case, the loop will be robust in face of plant gain variations, since even though the gain crossover frequency may change, the plant's phase margin will not—and neither will the controller's.

5.2. Second generation Crone controllers

If a plant does not have a constant phase, first generation Crone controllers will not be robust in face of plant gain variations. But it is possible to devise a controller that is[75]. Such a controller must ensure a constant open-loop phase; in other words, if the controller is C and the plant is F,

$$\arg\left[F(j\omega)C(j\omega)\right] = \arg\left[F(j\omega)\right] + \arg\left[C(j\omega)\right] = \varphi \tag{5.2}$$

Of course the equality needs to hold only in some suitable limited frequency range only. Now this is what we get if

[74] Oustaloup (1991, p. 141-153).
[75] Oustaloup (1991, p. 180-183).

$$F(s)C(s) = s^{2\varphi/\pi} \tag{5.3}$$

in that suitable frequency range.

Second generation Crone controllers result from solving (5.3) in order to C and then identifying it from its frequency behaviour. In its original formulation, the Crone identification method was to be used. Though not being the sole possibility, this has the advantage—if a continuous controller is desired—of taking into account the frequency behaviour (the sole really important) only. From (5.2)

$$\arg\left[C(j\omega)\right] = \varphi - \arg\left[F(j\omega)\right] \tag{5.4}$$

Since the continuous Crone identification method finds frequencies of zeros and poles altogether, and only in the end is the decision taken about which are what (see subsection 4.2.3), it is possible to distribute such frequencies as zeros and poles so that all will be stable. This advantage does not exist if a discrete controller is desired (see subsection 4.2.14).

5.3. Third generation Crone controllers

Both first and second generation Crone controllers aim at being robust in face of plant gain variations—but other types of model uncertainty, including pole and zero misplacement, are not covered. Third generation Crone controllers try to make good this shortcoming. To see how it is necessary to consider first two important sets of curves in the Nichols plane.

5.3.1. Curves of constant closed-loop gain

Let us consider an open-loop frequency response $G(j\omega)$ given (for a certain frequency) by

$$\Theta \overset{\text{def}}{=} \left|G(j\omega)\right| \tag{5.5}$$

$$\theta \overset{\text{def}}{=} \arg\left[G(j\omega)\right] \tag{5.6}$$

The closed-loop shall be

$$L(j\omega) = \frac{G(j\omega)}{1+G(j\omega)} = \frac{\Theta e^{j\theta}}{1+\Theta e^{j\theta}} \tag{5.7}$$

and thus

$$\left|L(j\omega)\right| = \left|\frac{\Theta e^{j\theta}}{1+\Theta e^{j\theta}}\right| = \frac{\Theta}{\left|1+\Theta\cos\theta + j\Theta\sin\theta\right|} = \frac{\Theta}{\sqrt{\left(1+\Theta\cos\theta\right)^2 + \left(\Theta\sin\theta\right)^2}} =$$

$$= \frac{\Theta}{\sqrt{1+\Theta^2\cos^2\theta+2\Theta\cos\theta+\Theta^2\sin^2\theta}} = \frac{\Theta}{\sqrt{1+\Theta^2+2\Theta\cos\theta}} \tag{5.8}$$

Closed-loop resonance gain shall be

$$\left|L(j\omega)\right|_R = \max_{\omega} \frac{\left|G(j\omega)\right|}{\sqrt{1+\left|G(j\omega)\right|^2+2\left|G(j\omega)\right|\cos\left[\arg G(j\omega)\right]}} \tag{5.9}$$

Curves of constant values of $\left|L(j\omega)\right|$ are represented in a Nichols chart in Figure 5.1. They have a periodicity of 2π rad in the phase axis and are symmetric in respect to all vertical straight lines given by $x = n\pi$ rad, $n \in \mathbb{Z}$ (including the phase axis), and for that reason it suffices to show them in the $[-\pi;0]$ rad interval[76].

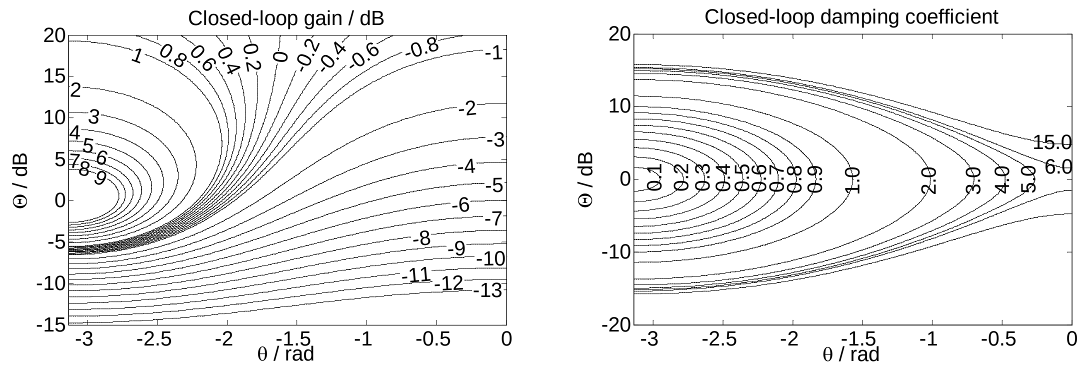


Figure 5.1 — Nichols charts with curves of constant values of closed-loop gain and damping coefficient

5.3.2. Curves of approximately constant damping coefficient

Oustaloup (1991, p. 395-398) shows that the damping coefficient is approximately given by

$$\xi \approx -\cos\frac{\pi^2}{2\arccos\left(\dfrac{1+\Theta^2+2\Theta\cos\theta}{2\Theta}-1\right)} \tag{5.10}$$

(even though the exact relationship depends on the plant). The smaller value of the damping coefficient will be

$$\min_{\omega}\xi \approx \min_{\omega}\left\{-\cos\frac{\pi^2}{2\arccos\left(\dfrac{1+\left|G(j\omega)\right|^2+2\left|G(j\omega)\right|\cos\left[\arg G(j\omega)\right]}{2\left|G(j\omega)\right|}-1\right)}\right\} \tag{5.11}$$

[76] Ogata (1997, p. 562); Oustaloup (1991, p. 398).

Curves of constant values of ξ are represented in a Nichols chart in Figure 5.1. Comments on their periodicity similar to those of the previous subsection could be made.

5.3.3. Objective of a third generation Crone controller

The objective of a third generation Crone controller[77] is to ensure that the closed-loop gain will never get beyond a certain value—even if some parameters of the plant vary within a known range. Or the objective may be to ensure that the closed-loop damping coefficient will never get below a certain value—even if some parameters of the plant vary within a known range. Both these objectives correspond to prevent the Nichols plot of the open-loop from approaching the zones where the closed-loop gain is high and the damping coefficient is low. These are the zones around points $\left(0\text{ dB}; 2n\pi+\pi\text{ rad}\right),\quad n\in\mathbb{Z}$.

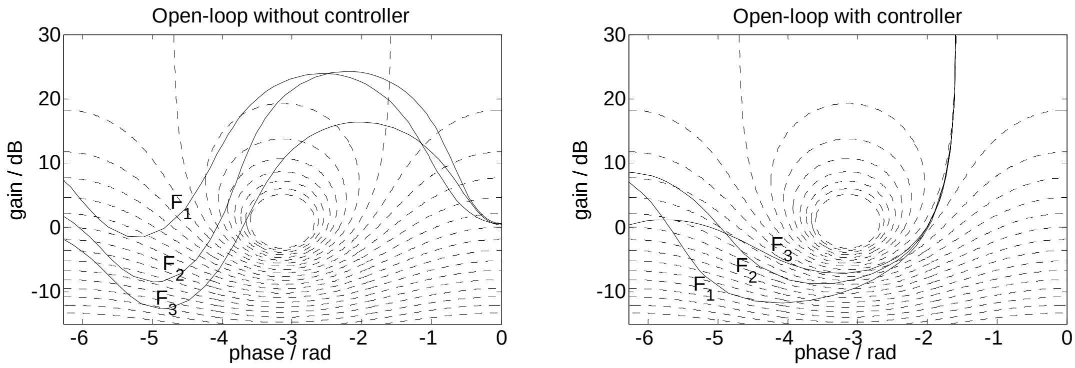


Figure 5.2 — Objective of a third generation Crone controller

Figure 5.2 presents an example of how that objective is to be applied[78]. A plant may assume three different configurations that correspond to the three following models:

$$F_1\left(z^{-1}\right)=\frac{0.50666z^{-4}+0.28261z^{-3}}{0.88642z^{-4}-1.31608z^{-3}+1.58939z^{-2}-1.41833z^{-1}+1} \tag{5.12}$$

$$F_2\left(z^{-1}\right)=\frac{0.18123z^{-4}+0.10276z^{-3}}{0.89413z^{-4}-1.84083z^{-3}+2.20265z^{-2}-1.99185z^{-1}+1} \tag{5.13}$$

$$F_3\left(z^{-1}\right)=\frac{0.10407z^{-4}+0.06408z^{-3}}{0.87129z^{-4}-1.93353z^{-3}+2.31962z^{-2}-2.09679z^{-1}+1} \tag{5.14}$$

The sampling time is T = 0.05 s. The Nichols plots of such models come dangerously close to points where the closed-loop gain is rather high. A controller is wanted for this plant that can deal with all three configurations ensuring that the closed-loop gain is never higher than 1 dB. This is the case of

[77] Oustaloup (1991, p. 283-284 and 290-292); Oustaloup *et al.* (2000).

[78] The example is a benchmark for digital robust control presented in Landau *et al.* (1995). The controllers for this benchmark are given by Oustaloup *et al.* (1995).

$$C\left(z^{-1}\right)=\frac{N\left(z^{-1}\right)}{D\left(z^{-1}\right)} \tag{5.15}$$

$$\begin{aligned}N\left(z^{-1}\right)=&-0.1461419891z^{-7}+0.6076390925z^{-6}-0.5801126047z^{-5}\\&-0.8034704435z^{-4}+1.603182672z^{-3}-0.2081961813z^{-2}\\&-0.8704306516z^{-1}+0.4105249593\end{aligned} \tag{5.16}$$

$$\begin{aligned}D\left(z^{-1}\right)=&-0.0532773097z^{-7}+0.1032290834z^{-6}-0.5418714841z^{-5}\\&+0.5887788033z^{-4}+1.324106971z^{-3}-1.686052320z^{-2}\\&-0.7349137439z^{-1}+1\end{aligned} \tag{5.17}$$

as the Figure shows.

For reasons of simplicity, the desired nominal open-loop is usually assumed to correspond to a straight line on a Nichols plot[79]. Such a plot is obtained with (3.73) (usually restricted to negative values of a only, so as to ensure that the gain decays for high frequencies, as is bound to happen in reality).

In order to find the most convenient slope, a quadratic penalty function is minimised—hence the name of optimal Crone control used for this type of controllers. Let $S(0)$ be the steady-state plant gain, let ω_{zi} and ω_{pi} be the plant's zeros and poles, and let there be n_z zeros and n_p poles. The steady-state gain, poles and zeros may undergo variations, $\Delta S(0)$, $\Delta\omega_{zi}$ and $\Delta\omega_{pi}$ respectively, limited between maximal and minimal values[80].

It is clear that, if the open-loop transfer function $G(s)$ is affected by plant variations $\Delta S(0)$, $\Delta\omega_{zi}$ and $\Delta\omega_{pi}$, then $\left|L(j\omega)\right|_R$ and $\min_{\omega}\xi$ will also vary. Let the variations be $\Delta\left|L(j\omega)\right|_{R,S(0)_{\min}}$, $\Delta\left|L(j\omega)\right|_{R,S(0)_{\max}}$, $\Delta\left|L(j\omega)\right|_{R,\omega_{zi\min}}$ and so forth; and $\Delta\xi_{S(0)_{\min}}$, $\Delta\xi_{S(0)_{\max}}$, $\Delta\xi_{\omega_{zi\min}}$ and so forth. Thus, each open-loop dynamic $G(s)$, corresponding to a different phase slope, is assigned a penalty function given by

$$\begin{aligned}J\left(\Delta\left|L(j\omega)\right|_R\right)=&\left(\Delta\left|L(j\omega)\right|_{R,S(0)_{\min}}\right)^2+\left(\Delta\left|L(j\omega)\right|_{R,S(0)_{\max}}\right)^2+\\&+\sum_{i=1}^{n_z}\left[\left(\Delta\left|L(j\omega)\right|_{R,\omega_{zi\min}}\right)^2+\left(\Delta\left|L(j\omega)\right|_{R,\omega_{zi\max}}\right)^2\right]+\\&+\sum_{i=1}^{n_p}\left[\left(\Delta\left|L(j\omega)\right|_{R,\omega_{pi\min}}\right)^2+\left(\Delta\left|L(j\omega)\right|_{R,\omega_{pi\max}}\right)^2\right]\end{aligned} \tag{5.18}$$

or

[79] That of Figure 5.2 is not so for frequencies higher than the Nyquist frequency since what happens then is irrelevant.

[80] Uncertainties about the plant's structure, that is to say, situations when the number of zeros and poles is not surely known, are dealt with assuming that all possible poles and zeros do exist, but allowing frequencies of zeros and poles that may not exist to have uncertainties placing them clearly above the frequency range where control is intended. If they are placed at such high frequencies, it will be as though they did not exist.

$$J(\Delta\xi) = \left(\Delta\xi_{S(0)_{\min}}\right)^2 + \left(\Delta\xi_{S(0)_{\max}}\right)^2 + \\ + \sum_{i=1}^{n_z}\left[\left(\Delta\xi_{\omega_{zi\min}}\right)^2 + \left(\Delta\xi_{\omega_{zi\max}}\right)^2\right] + \sum_{i=1}^{n_p}\left[\left(\Delta\xi_{\omega_{pi\min}}\right)^2 + \left(\Delta\xi_{\omega_{pi\max}}\right)^2\right] \tag{5.19}$$

and the slope corresponding to the lowest value of the penalty function is chosen.

5.3.4. Implementation

Once the desired open-loop frequency behaviour is known, the controller behaviour is obtained as

$$C(j\omega) = \frac{G(j\omega)}{F(j\omega)} \tag{5.20}$$

Then the controller may be obtained using some identification method, just as second generation Crone controllers.

Third generation Crone controllers may be improved by including in the open-loop terms aiming at attenuating the open-loop response or the output sensitivity frequency response at certain frequencies. This is likely to be important for resonant plants[81].

5.3.5. Crone implementation for the logarithmic phase case

Should (5.20) lead to a controller having a logarithmic phase (as is the case when the nominal plant to control has a constant phase), it is possible to implement it using a variation of the Crone approximation[82], in which the number of poles and zeros is *not* the same:

$$\hat{C}(s) = C_0 \frac{\prod_{n=1}^{M}\left(1+\frac{s}{\omega_{zn}}\right)}{\prod_{n=1}^{N}\left(1+\frac{s}{\omega_{pn}}\right)} \tag{5.21}$$

$$\omega_{zj+1} = \zeta\omega_{zj}, \quad j = 1, 2, \ldots, M-1 \tag{5.22}$$

$$\omega_{pi+1} = \beta\omega_{pi}, \quad i = 1, 2, \ldots, N-1 \tag{5.23}$$

In this manner poles and zeros will be equidistant when represented in logarithmic scale.

Since each pole lowers the phase $\pi/2$ rad, if there is a pole per decade (which will be the case for $\beta = 10$) and there are no zeros, its effects shall overlap, and the transfer function phase will be nearly linear with a negative slope of $-\pi/2$ rad per

[81] An application of this composite methodology is given by Oustaloup *et al.* (1995).
[82] Oustaloup (1991, p. 336-353).

decade. If there are two poles per decade (which will be the case for $\beta = \sqrt{10}$), slope will be $-\pi$ rad per decade, and so on. The general formula for slope (3.58), as function of the recursive parameter β, is

$$-b \log 10 \operatorname{tgh}\left(b\frac{\pi}{2}\right) = -\frac{\pi}{2\log_{10}\beta} \tag{5.24}$$

If there are zeros but no poles the phase slope will be positive:

$$-b \log 10 \operatorname{tgh}\left(b\frac{\pi}{2}\right) = \frac{\pi}{2\log_{10}\zeta} \tag{5.25}$$

If there are both poles and zeros,

$$-b \log 10 \operatorname{tgh}\left(b\frac{\pi}{2}\right) = \frac{\pi}{2\log_{10}\zeta} - \frac{\pi}{2\log_{10}\beta} \tag{5.26}$$

Poles and zeros should cover the entire frequency range where control is intended.

Remark 1: If M and N are the same, the phase will have no slope. This corresponds to the real case (3.74), when $\zeta = \beta = \alpha\eta$.

Remark 2: Approximation (5.21) may be further adjusted by the inclusion of an integer part, as discussed in subsection 3.3.18.

5.4. Fractional PID controllers

PID controllers are those whose output is a linear combination of the input, the derivative of the input and the integral of the input (hence their name). Fractional PID controllers are generalisations of PIDs: their output is a linear combination of the input, a fractional derivative of the input and a fractional integral of the input[83]:

$$C(s) = P + \frac{I}{s^\lambda} + Ds^\mu \tag{5.27}$$

For this reason fractional PIDs are also known as $\mathrm{PI}^\lambda\mathrm{D}^\mu$ controllers. If both λ and μ are 1, the result is a usual PID (henceforth called *integer* PID as opposed to a fractional PID).

The following subsections expose three different ways of tuning the five parameters P, I, λ, D and μ of (5.27).

5.4.1. Fractional PID tuning by internal model control

The internal model control (IMC) methodology may, in some cases, be used to obtain PID or fractional PID controllers. It makes use of the control scheme[84] of Figure

[83] Podlubny (1999, p. 249-250).
[84] Hägglund *et al.* (1996, p. 821-822).

5.3 (left), where F is the plant to control, F^* is an inverse of F or at least a plant as close as possible to the inverse of F, F' is a model of F and B is some judiciously chosen filter. If F' were exact, the error e would be equal to disturbance d. If, additionally, F^* were the exact inverse of F and B were unity, control would be perfect. Since no models are perfect, e will not be exactly the disturbance. That is also exactly why B exists and is usually a low-pass filter: to reduce the influence of high-frequency modelling errors. It also helps ensuring that product BF^* is realisable. These interconnections are equivalent to those of Figure 5.3 (right) if

$$C(s)=\frac{B(s)F^*(s)}{1-B(s)F^*(s)F'(s)} \tag{5.28}$$

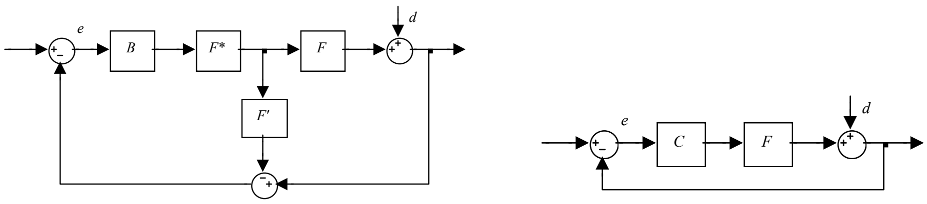


Figure 5.3 — Block diagram for internal model control (left) and its equivalent (right)

Controller C is not, in the general case, a PID or a fractional PID, but it will if

$$F(s)=\frac{K}{1+Ts^{\mu}}e^{-Ls} \tag{5.29}$$

Let

$$B(s)=\frac{1}{1+T_B s} \tag{5.30}$$

$$F^*(s)=\frac{1+Ts^{\mu}}{K} \tag{5.31}$$

$$F'(s)=\frac{K}{1+Ts^{\mu}}(1-Ls) \tag{5.32}$$

Notice that the delay of F was neglected in F^* but not in F'. Then (5.28) becomes

$$C_1(s)=\frac{\overline{\dfrac{K}{K(T_B+L)}}}{s}+\frac{\overline{\dfrac{T}{K(T_B+L)}}}{s^{1-\mu}} \tag{5.33}$$

that can be viewed as a fractional PID controller with the proportional part equal to zero. And if the model of F' in (5.32) is replaced with the improved approximation

$$F'(s)=\frac{K}{1+Ts^{\mu}}\frac{1-\frac{Ls}{2}}{1+\frac{Ls}{2}} \tag{5.34}$$

then (5.28) becomes

$$C_2(s)=\frac{L}{2K(T_P+L)}+\frac{\frac{1}{K(T_B+L)}}{s}+\frac{\frac{T}{K(T_B+L)}}{s^{1-\mu}}+\frac{TL}{2K(T_P+L)}s^{\mu} \tag{5.35}$$

If one of the two integral parts is neglectable, (5.35) will be a fractional PID controller. Obviously, should $\mu\in\mathbb{Z}$, both (5.33) and (5.35) become usual PIDs.

5.4.2. Fractional PID tuning by minimisation

Fractional PIDs may be tuned by finding parameters that satisfy the following conditions[85] (*C* being the controller and *G* the plant):

❖ The gain-crossover frequency ω_{cg} is to have some specified value:

$$\left|C(\omega_{cg})F(\omega_{cg})\right|=0\text{ dB} \tag{5.36}$$

❖ The phase margin φ_m is to have some specified value:

$$-\pi+\varphi_m=\arg\left[C(\omega_{cg})F(\omega_{cg})\right] \tag{5.37}$$

❖ So as to reject high-frequency noise, the closed-loop transfer function must have a small magnitude at high frequencies; thus it is required that at some specified frequency ω_h its magnitude be less than some specified gain:

$$\left|\frac{C(\omega_h)F(\omega_h)}{1+C(\omega_h)F(\omega_h)}\right|<H \tag{5.38}$$

❖ So as to reject output disturbances and closely follow references, the sensitivity function must have a small magnitude at low frequencies; thus it is required that at some specified frequency ω_l its magnitude be less than some specified gain:

$$\left|\frac{1}{1+C(\omega_l)F(\omega_l)}\right|<N \tag{5.39}$$

❖ So as to be robust in face of gain variations of the plant, the phase of the open-

[85] Monje *et al.* (2004).

loop transfer function must be (at least roughly) constant around the gain-crossover frequency:

$$\frac{d}{d\omega}\arg\left[C(\omega)F(\omega)\right]\Big|_{\omega=\omega_{cg}}=0 \tag{5.40}$$

Conditions are five because five are the parameters to tune[86]. To satisfy them all numerical optimisation algorithms may be used, such as Nelder-Mead's simplex algorithm[87]. This is effective but allows local minima to be obtained. In practice most solutions found with this optimisation method are good enough, but they strongly depend on initial estimates of the parameters provided. Some may be discarded because they are unfeasible or lead to unstable loops, but in many cases it is possible to find more than one acceptable fractional PID; in others, only well-chosen initial estimates of the parameters allow finding a solution.

5.4.3. Fractional PID tuning by the use of tuning rules

Integer PID controllers are widely used and enjoy significant popularity not only because they are simple, effective and robust, but also because of the existence of tuning rules for finding suitable parameters, rules that do not require any model of the plant to control. All that is needed to apply such rules is to have a certain time response of the plant. Examples of such sets of rules are those due to Ziegler and Nichols, those due to Cohen and Coon, and the Kappa-Tau rules[88]. It is true that PIDs tuned with such rules often perform in a non-optimal way. But even though further fine-tuning be possible and sometimes necessary, rules provide a good starting point. Their usefulness is obvious when no model of the plant is available, and thus no analytic means of tuning a controller exists, but rules may also be used when a model is known.

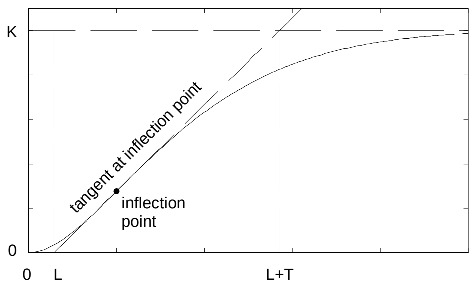


Figure 5.4 — S-shaped unit step response

The two methods of the two previous subsections require a model of the plant to control, and the same happens with other alternatives suggested in the literature[89]. On

[86] A similar minimisation method may be applied to the tuning of integer PIDs as well, but since there are only three parameters it is likely that such minimisation will not work properly with more than three conditions. The conditions may be chosen so as to try to have a closed-loop behaving as similarly as possible as a fractional system. This approach is followed by Barbosa *et al.* (2003, 2004a, 2004b) and by Chen *et al.* (2004) (these authors differ significantly on how that is to be done).

[87] This is the proposal of Monje *et al.* (2004). These authors namely suggest the implementation in Matlab's function *fmincon*, that allows for constraints in the minimisation: condition (5.36) is to be assumed as the condition to minimise, and conditions (5.37) to (5.40) as constraints.

[88] Ziegler *et al.* (1942); Hägglund *et al.* (1996, p. 819-820).

[89] Caponetto (2002, 2004); Monje *et al.* (2004).

the contrary, the tuning rules of the two following subsections try to emulate what happens with existing tuning rules for integer PIDs. They bear similarities to the first rule proposed by Ziegler and Nichols for integer PIDs, making use of the same plant time response data: that rule assumes the plant to have an S-shaped unit-step response, as that of Figure 5.4, where *L* is an apparent delay and *T* may be interpreted as a pole. The method cannot be applied if the unit-step response is shaped otherwise. The simplest plant with such a response is

$$F(s) = \frac{K}{1+sT} e^{-Ls} \tag{5.41}$$

So the tuning rules given below were obtained applying the minimisation method of subsection 5.4.2 to plants given by (5.41) for several values of *L* and *T*, with $K = 1$. The parameters of fractional PIDs thus obtained vary in a regular manner. Using the least-squares method it is possible to translate that regularity into formulas to find values of the parameters from *L* and *T* which are acceptable even if the plant is not given by (5.41)—it suffices that its step-response be like that in Figure 5.4.

Remark 1: The optimisation method employed was Nelder-Mead's simplex method[90] of subsection A.6.1. The last condition was verified numerically, evaluating argument in (5.40) at two frequencies, equal to $\omega_{cg}/1.122$ and $1.122\omega_{cg}$ (this corresponds to 1/20 of a decade). It is of course possible to evaluate the argument at other frequencies around ω_{cg}; actually, the larger the interval where the argument is constant (or nearly so) the better, and thus using more than two points might ensure that. However, it was verified that such stronger requirements often prevent a solution from being found.

Remark 2: Least-squares method-adjusted formulas cannot exactly reproduce every change in parameters. This means that fractional PIDs tuned with the rules presented below never behave as well as those tuned analytically, neither are they so robust. Conditions (5.36) to (5.40) will only approximately be verified.

Remark 3: Just as the two sets of rules given below were obtained, other sets of rules might have been found in the same way for different specifications.

5.4.4. First set of tuning rules

A first set of rules is given in Table 5.1. This is to be read as

$$P = -0.0048 + 0.2664L + 0.4982T + 0.0232L^2 - 0.0720T^2 - 0.0348TL \tag{5.42}$$

and so on. They may be used if

$$0.1 \le T \le 50 \wedge L \le 2 \tag{5.43}$$

and were designed for the following specifications:

$$\omega_{cg} = 0.5 \text{ rad/s} \tag{5.44}$$

$$\varphi_m = 2/3 \text{ rad} \approx 38^\circ \tag{5.45}$$

[90] The implementation allowing for constraints mentioned above in footnote 87 was used.

$$\omega_h = 10 \text{ rad/s} \quad (5.46)$$
$$\omega_l = 0.01 \text{ rad/s} \quad (5.47)$$
$$H = -10 \text{ dB} \quad (5.48)$$
$$N = -20 \text{ dB} \quad (5.49)$$

Recall that specifications are only approximately verified.

	Parameters to use when $0.1 \le T \le 5$					Parameters to use when $5 \le T \le 50$				
	P	I	λ	D	μ	P	I	λ	D	μ
1	–0.0048	0.3254	1.5766	0.0662	0.8736	2.1187	–0.5201	1.0645	1.1421	1.2902
L	0.2664	0.2478	–0.2098	–0.2528	0.2746	–3.5207	2.6643	–0.3268	–1.3707	–0.5371
T	0.4982	0.1429	–0.1313	0.1081	0.1489	–0.1563	0.3453	–0.0229	0.0357	–0.0381
L^2	0.0232	–0.1330	0.0713	0.0702	–0.1557	1.5827	–1.0944	0.2018	0.5552	0.2208
T^2	–0.0720	0.0258	0.0016	0.0328	–0.0250	0.0025	0.0002	0.0003	–0.0002	0.0007
LT	–0.0348	–0.0171	0.0114	0.2202	–0.0323	0.1824	–0.1054	0.0028	0.2630	–0.0014

Table 5.1 — Parameters for the first set of tuning rules

5.4.5. Second set of tuning rules

A second set of rules is given in Table 5.2. These may be applied if

$$0.1 \le T \le 50 \wedge L \le 0.5 \quad (5.50)$$

Only one set of parameters is needed in this case because the range of values of L these rules cope with is more reduced. They were designed for the following specifications:

$$\omega_{cg} = 0.5 \text{ rad/s} \quad (5.51)$$
$$\varphi_m = 1 \text{ rad} \approx 57^\circ \quad (5.52)$$
$$\omega_h = 10 \text{ rad/s} \quad (5.53)$$
$$\omega_l = 0.01 \text{ rad/s} \quad (5.54)$$
$$H = -20 \text{ dB} \quad (5.55)$$
$$N = -20 \text{ dB} \quad (5.56)$$

	P	I	λ	D	μ
1	–1.0574	0.6014	1.1851	0.8793	0.2778
L	24.5420	0.4025	–0.3464	–15.0846	–2.1522
T	0.3544	0.7921	–0.0492	–0.0771	0.0675
L^2	–46.7325	–0.4508	1.7317	28.0388	2.4387
T^2	–0.0021	0.0018	0.0006	0.0000	–0.0013
LT	–0.3106	–1.2050	0.0380	1.6711	0.0021

Table 5.2 — Parameters for the second set of tuning rules

5.4.6. Comments on the tuning rules above

The new rules are clearly more complicated than those of Ziegler-Nichols: they have to be quadratic, approximations of lower order being unsatisfactory. And the broader the application range of the rules is to be, the more complicated they become (the first rule needs two tables of parameters while the second, good for a narrower interval of values of L only, needs only one). The usefulness of these rules is that of all sets of rules: they may be applied even if no model of the plant is available, provided a suitable time response is; they may be used as a departing point for fine-tuning (this is relevant if the optimisation tuning method is used, since its results depend significantly from the initial estimate provided); they are easier and faster to apply than analytic methods. Their drawbacks are also those of all sets of rules: their performance is often inferior to the one sought, fine-tuning being often needed; they perform worse than controllers tuned analytically; they cannot be applied to all types of plants, but only to those with a particular sort of time response.

5.5. H_2 and H_∞ controllers

Optimal controllers are so called because they minimise a given performance criterion. This means they achieve the best possible result in what that criterion is concerned. Of course, should the criterion be poorly chosen the controller's performance would probably be unsatisfactory, even though it would still be optimal in the sense above. Finding an optimal controller requires having a model of the plant to control.

Third generation Crone controllers described above in section 5.3 are optimal controllers in this sense. In what follows two additional performance criteria usually employed for devising optimal controllers are described. These may be used for both SISO and MIMO plants.

5.5.1. H_2 norm

The H_2 norm of a transfer function matrix **G** is defined as

$$\left\|\mathbf{G}(s)\right\|_2 = \sqrt{\frac{1}{2\pi}\int_{-\infty}^{+\infty} \operatorname{tr}\left[\mathbf{G}(j\omega)\overline{\mathbf{G}(j\omega)^T}\right] d\omega} \tag{5.57}$$

In simple words, the H_2 norm of a transfer function reflects how much it amplifies (or attenuates) its input over all frequencies.

We now make use of the following result:

Theorem 5.1: Let **A** be a matrix with m lines and n columns. Then

$$\operatorname{tr}\left[\mathbf{A}\overline{\mathbf{A}^T}\right] = \sum_{i=1}^{m}\sum_{j=1}^{n}\left|\mathbf{A}_{ij}\right|^2 \tag{5.58}$$

Proof: The product $\mathbf{A}\overline{\mathbf{A}^T}$ is a square matrix with m lines and m columns. Its elements are given by

$$\left[\mathbf{A}\overline{\mathbf{A}^T}\right]_{ij} = \sum_{k=1}^{n}\mathbf{A}_{ik}\overline{\left[\mathbf{A}^T\right]_{kj}} = \sum_{k=1}^{n}\mathbf{A}_{ik}\overline{\mathbf{A}_{jk}} \tag{5.59}$$

Elements in the main diagonal are given by

$$\left[\mathbf{A}\overline{\mathbf{A}^T}\right]_{ii} = \sum_{k=1}^{n} \mathbf{A}_{ik}\overline{\mathbf{A}_{ik}} = \sum_{k=1}^{n}\left(\mathrm{Re}\left[\mathbf{A}_{ik}\right] + j\,\mathrm{Im}\left[\mathbf{A}_{ik}\right]\right)\left(\mathrm{Re}\left[\mathbf{A}_{ik}\right] - j\,\mathrm{Im}\left[\mathbf{A}_{ik}\right]\right) = \\ = \sum_{k=1}^{n} \mathrm{Re}^2\left[\mathbf{A}_{ik}\right] + \mathrm{Im}^2\left[\mathbf{A}_{ik}\right] = \sum_{k=1}^{n}\left|\mathbf{A}_{ik}\right|^2 \tag{5.60}$$

The hypothesis follows immediately from the definition of trace. **Q.E.D.**

This means that, for a MIMO system with *n* lines and *m* columns,

$$\left\|\mathbf{G}\right\|_2 = \sqrt{\sum_{j=1}^{n}\sum_{i=1}^{m}\left\|\mathbf{G}_{ij}\right\|_2^2} \tag{5.61}$$

This reduces the problem to finding the H_2 norm of SISO systems.

5.5.2. H_2 norm of integer systems

The following result is quoted without proof.

Theorem 5.2[91]**:** If **G** is an integer order system (SISO or MIMO) given by (3.4) with $Q = 1$, its H_2 norm may be found solving one of the equations

$$\mathbf{A}\mathbf{L}_C + \mathbf{L}_C\mathbf{A}^T + \mathbf{B}\mathbf{B}^T = \mathbf{0} \tag{5.62}$$

$$\mathbf{A}^T\mathbf{L}_O + \mathbf{L}_O\mathbf{A} + \mathbf{C}^T\mathbf{C} = \mathbf{0} \tag{5.63}$$

for $\mathbf{L}_C$ (the controllability gramian matrix) or $\mathbf{L}_O$ (the observability gramian matrix). (These two linear matrix equations belong to a class of equations known as Riccatti type equations.) Then

$$\left\|\mathbf{G}\right\|_2 = \sqrt{\mathrm{tr}\left(\mathbf{C}\mathbf{L}_C\mathbf{C}^T\right)} \tag{5.64}$$

$$\left\|\mathbf{G}\right\|_2 = \sqrt{\mathrm{tr}\left(\mathbf{B}^T\mathbf{L}_C\mathbf{B}\right)} \tag{5.65}$$

5.5.3. H_2 norm of fractional systems

No results similar to (5.64) and (5.65) have been found for fractional systems, but the result may be found as follows[92]. Let $1\,/\,Q$ be the commensurate order of the system, and

$$\left.\begin{matrix} q = \mathscr{C}\left(Q\right) \\ p = Q - q \end{matrix}\right\} \Rightarrow p + q = Q \tag{5.66}$$

$$x = \omega^{1/Q} \Leftrightarrow \omega = x^Q \Rightarrow d\omega = Qx^{Q-1}dx \tag{5.67}$$

[91] Lublin *et al.* (1996, p. 636).
[92] Malti *et al.* (2003).

Since the complex conjugate may be obtained changing the sign of the imaginary part, we will have

$$\|G\|_2^2 = \frac{1}{\pi}\int_0^{+\infty} Qx^{Q-1}G(jx)G(-jx)\,dx \tag{5.68}$$

Let *A* and *B* be the polynomials in the numerator and denominator of product

$$G(jx)G(-jx) = \frac{A(x)}{B(x)} \tag{5.69}$$

Notice that the imaginary unit *j* has been considered as part of polynomials *A* and *B*. If we now let

$$\frac{x^q A(x)}{B(x)} = \frac{R(x)}{B(x)} + \sum_{k=0}^{q+\deg(A)-\deg(B)} a_k x^k \tag{5.70}$$

(where deg(*P*) represents the degree of polynomial *P*), then (5.68) becomes

$$\|G\|_2^2 = \frac{Q}{\pi}\int_0^{+\infty} x^{p-1}x^q \frac{A(x)}{B(x)}dx = \frac{Q}{\pi}\int_0^{+\infty} x^{p-1}\left(\frac{R(x)}{B(x)} + \sum_{k=0}^{q+\deg(A)-\deg(B)} a_k x^k\right)dx \tag{5.71}$$

Three cases are to be distinguished when reckoning (5.71).

❖ Case $q+\deg(A)-\deg(B)>0$

In this case the summation in (5.71) is not identically zero. Its integral will be

$$\int_0^{+\infty}\sum_{k=0}^{q+\deg(A)-\deg(B)} a_k x^{k+p-1}dx = \sum_{k=0}^{q+\deg(A)-\deg(B)}\left[\frac{a_k}{k+p}x^{k+p}\right]_0^{+\infty} \tag{5.72}$$

and, *k* being 1 or higher and *p* being positive,

$$x^{k+p}\Big|_{x=+\infty} = +\infty \tag{5.73}$$

Hence

$$\|G\|_2 = \infty \tag{5.74}$$

❖ Case $q+\deg(A)-\deg(B)\le 0 \wedge p\ne 0$

In this case, let $B(x)$ have *b* different poles, $s_1, s_2, \ldots s_b$, and let m_k be the multiplicity of pole s_k. Then we may perform a partial fraction expansion of

$$\frac{x^q A(x)}{B(x)} = \sum_{k=1}^{b} \sum_{n=1}^{m_k} \frac{a_{k,n}}{(x+s_k)^n} \tag{5.75}$$

(Recall that a pole of multiplicity *m* will appear *m* times in the expansion.) Then (5.71) becomes

$$\|G\|_2^2 = \frac{Q}{\pi} \int_0^{+\infty} \sum_{k=1}^{b} \sum_{n=1}^{m_k} \frac{a_{k,n} x^{p-1}}{(x+s_k)^n} dx = \frac{Q}{\pi} \sum_{k=1}^{b} \sum_{n=1}^{m_k} \frac{a_{k,n}}{s_k^n} \int_0^{+\infty} \frac{x^{p-1}}{\left(\frac{x}{s_k}+1\right)^n} dx \tag{5.76}$$

This becomes[93]

$$\|G\|_2^2 = \frac{Q}{\pi} \sum_{k=1}^{b} \sum_{n=1}^{m_k} \frac{a_k}{s_k^n} (-1)^{n-1} \frac{\pi}{s_k^{-p}} \binom{p-1}{n-1} \operatorname{cosec}(p\pi) = \sum_{k=1}^{b} \sum_{n=1}^{m_k} \frac{(-1)^{n-1} Q a_k s_k^{p-n} \binom{p-1}{n-1}}{\sin(p\pi)} \tag{5.77}$$

❖ Case $q + \deg(A) - \deg(B) \le 0 \wedge p = 0$

In this case, the integration rule of footnote 93 cannot be applied, so an alternative expansion is carried out. Let s_1 be one of the poles of (5.75), chosen arbitrarily[94]. Then we may write

$$\frac{x^{q-1} A(x)}{B(x)} = \sum_{k=2}^{b} \frac{c_k}{(x+s_1)(x+s_k)} + \sum_{k=1}^{b} \sum_{n=2}^{m_k} \frac{d_{k,n}}{(x+s_k)^n} \tag{5.78}$$

Notice that poles with multiplicity 1 do not appear in the second summation. Expression (5.71) becomes

$$\|G\|_2^2 = \frac{Q}{\pi} \left(\sum_{k=2}^{b} \int_0^{+\infty} \frac{c_k}{(x+s_1)(x+s_k)} dx + \sum_{k=1}^{b} \sum_{n=2}^{m_k} \int_0^{+\infty} \frac{d_{k,n}}{(x+s_k)^n} dx \right) =$$

$$= \sum_{k=2}^{b} \left[\frac{Q}{\pi} \int_0^{+\infty} \left(\frac{1}{x+s_1} + \frac{-1}{x+s_k} \right) \frac{c_k}{s_k - s_1} dx \right] + \sum_{k=1}^{b} \sum_{n=2}^{m_k} \left[\frac{Q}{\pi} \int_0^{+\infty} \frac{d_{k,n}}{(x+s_k)^n} dx \right] =$$

$$= \sum_{k=2}^{b} \left\{ \frac{Q c_k}{\pi (s_k - s_1)} \left[\ln \frac{x+s_1}{x+s_k} \right]_0^{+\infty} \right\} + \sum_{k=1}^{b} \sum_{n=2}^{m_k} \left\{ \frac{Q d_{k,n}}{\pi(-n+1)} \left[(x+s_k)^{-n+1} \right]_0^{+\infty} \right\} =$$

---

[93] Gradshteyn *et al.* (1980, p. 285) gives the following result (the original notation was preserved):

$$\int_0^{\infty} \frac{x^{\mu-1} dx}{(1+\beta x)^{n+1}} = (-1)^n \frac{\pi}{\beta^\mu} \binom{\mu-1}{n} \operatorname{cosec}(\mu\pi)$$

$$|\arg \beta| < \pi, \quad 0 < \operatorname{Re} \nu < n+1$$

In this last condition, ν appears to be a typographical mistake for μ.

[94] Actually the better choice would be the one minimising numerical errors, but it is difficult to know beforehand which one is.

$$= \sum_{k=2}^{b} \left[\frac{Qc_k}{\pi(s_k - s_1)} \ln \frac{s_k}{s_1} \right] + \sum_{k=1}^{b} \sum_{n=2}^{m_k} \frac{Qd_{k,n} s_k^{1-n}}{\pi(-n+1)} \tag{5.79}$$

5.5.4. Summing up the computation of the H_2 norm

The algorithm for finding the H_2 norm of a matrix transfer function **G** may be summed up as follows:

> If **G** is of integer order ($Q = 1$), then apply (5.62) and (5.64) or (5.63) and (5.65).
> If **G** is of fractional order ($Q \neq 1$), then:
> - If **G** is MIMO, find the H_2 norm of its components and then apply (5.61).
> - If G is SISO, then:
>   - If $q + \deg(A) - \deg(B) > 0$, the norm is ∞.
>   - If $q + \deg(A) - \deg(B) \leq 0$, then:
>   - If $p \neq 0$, apply (5.77).
>   - If $p = 0$, apply (5.79).

It should be noticed that, even though several different formulas are to be applied depending on the value of Q, the norm is a continuous function thereof.

5.5.5. H_∞ norm

The H_∞ norm of a transfer function matrix **G** is defined as

$$\|\mathbf{G}\|_\infty = \sup_\omega \max \sigma\left[\mathbf{G}(j\omega)\right] \tag{5.80}$$

where $\sigma(A)$ represents the set of singular values of matrix A. (This set has a finite number of values, and thus has a maximum; on the other hand, the set resulting of sweeping all frequencies may have no maximum, but only a supreme value.) If G is SISO, (5.80) becomes simply

$$\|G\|_\infty = \sup_\omega \max |G(j\omega)| \tag{5.81}$$

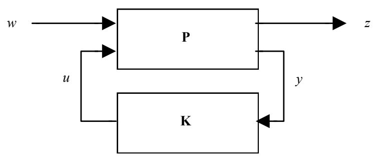
In simple words, the H_∞ norm reflects how much it amplifies (or attenuates) its input at the frequency at which the amplification is maximal.

This norm may be found by direct evaluation at several frequencies. Frequencies clearly above or below all the frequencies of poles and zeros need not be searched. The result is, of course, equal to or below the exact result—it can never be above.

5.5.6. The control paradigm of H_2 and H_∞ controllers

Controllers minimising the H_2 or the H_∞ norm of a suitable loop transfer function involving the plant to control are in use over the last decades, in what integer order plants are concerned. Roughly speaking, the idea is to minimise (at least over a certain range of frequencies we are interested in) one of those norms, ensuring that the input is never amplified to such an extent that instability will arise. It is usual to include judiciously chosen weights, that is to say, shaping transfer functions, in the control loop so that control efforts be exerted at those frequencies desired by the control designer. Should the weights be adequately chosen, it is possible to find, by minimising one of the two norms, a control-loop that is stable and robust to plant variations. These are expected to cause a worse performance but not instability[95].

H_2 and H_∞ controllers make use of the control structures of the block diagrams in Figure 5.5, where **K** is the controller[96], **A**, **B**, **C** and **D** are the matrixes of the state space representation of the plant **F**, and **L** models how noise affects the states (or the inputs, should $\mathbf{L} = \mathbf{B}$). Vector *w* collects all inputs (references, noise, disturbances…) save the control actions *u*. Vector *z* collects all variables showing the performance of the control system, namely outputs and control actions (whose magnitude may have to be limited). Weights W_1 to W_4 may be transfer functions, and usually are. They let the control designer shape the result by telling the loop in what frequencies control actions, outputs, etc., have to be large or small.



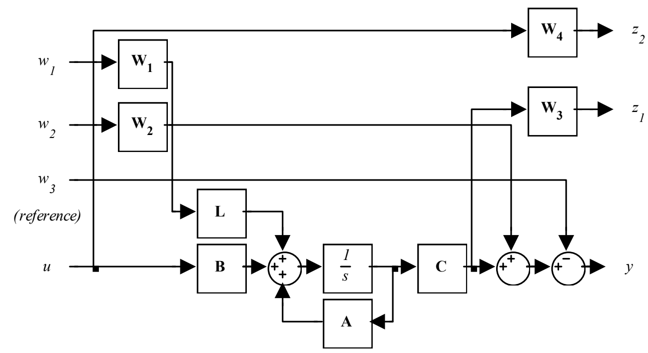


Figure 5.5 — Block diagram for H_2 and H_∞ controllers (left); sub-block diagram of **P** (right)

The above interconnections give rise to this transfer function:

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} W_3\mathbf{S}\mathbf{G}_1W_1 & W_3\mathbf{S}\mathbf{G}_2\mathbf{K}W_2 \\ W_4\mathbf{K}\mathbf{S}\mathbf{G}_1W_1 & W_4\mathbf{K}\mathbf{S}W_2 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \tag{5.82}$$

$$\mathbf{G}_1 = \mathbf{C}\left(s\mathbf{I} - \mathbf{A}\right)^{-1}\mathbf{L} \tag{5.83}$$

$$\mathbf{G}_2 = \mathbf{C}\left(s\mathbf{I} - \mathbf{A}\right)^{-1}\mathbf{B} \tag{5.84}$$

$$\mathbf{S} = \left(\mathbf{I} - \mathbf{G}_2\mathbf{K}\right)^{-1} \tag{5.85}$$

It is a norm of the matrix transfer function in (5.82) that we want to minimise.

5.5.7. Finding controllers

For integer order plants it is possible to find a H_2 or H_∞ controllers by analytical

[95] This is of course an oversimplified description. See for instance Lublin *et al.* (1996) for details.

[96] Though *C* has been used so far to denote the controller, this is no longer practical, under penalty of confusion with matrix **C** of the state-space representation.

means[97]. Unfortunately no such relations have yet been found for fractional-order plants; those for integer-order plants are derived from mathematical formulations for the norms very different from the formulations available for the fractional case. But it is possible to use numerical methods such as those of subsections A.6.1 and A.6.2 to minimise such norms; genetic algorithms, in particular, appear to be particularly suited[98].

[97] See for instance Lublin *et al.* (1996, p. 653-654) for details.

[98] Actually genetic algorithms ensure reasonable results for nearly all minimisation problems. Other methods may perform better in particular cases, but knowing in advance which ones might is guesswork. On the other hand, genetic algorithms have fairly good results in almost all cases. See Silva *et al.* (2005).

6. Applications of fractional control

— Oh! avec les chiffres on prouve tout ce qu'on veut!
— Et avec les faits, mon garçon, en est-il de même ?

Jules Verne, *Voyage au centre de la terre*, 6

Applications of fractional control are by now becoming current, as is documented in section 1.3. In this chapter a few of the author's are presented.

❖ In section 6.1 the tuning of fractional controllers for the hybrid force / position control of a rigid robot is addressed. Tuning was performed by optimisation. Validation was performed by simulation only.

❖ In section 6.2 the tuning of fractional PIDs for the position control of a flexible robot is addressed. Tuning was performed by optimisation. Validation was performed both by simulation and laboratory implementation.

❖ In section 6.3 the tuning of fractional PIDs for some benchmark plants is addressed. Tuning was performed by means of tuning rules and internal model control. Validation was performed by simulation only.

❖ In section 6.4 the tuning of fractional H_2 and H_∞ controllers for a thermal plant is addressed. Tuning was performed by optimisation. Validation was performed by simulation only.

6.1. Fractional control of a rigid robot

This section describes the optimal control of a two-link robotic arm, for which hybrid position / force control had been developed and implemented using fractional order derivatives by Machado *et al.* (1998). Parameter optimisation was performed using a genetic algorithm. This method is appropriate here since the plant is highly non-linear and complex.

6.1.1. The robotic arm

The robot is an arm with two degrees of motion consisting of two rigid links, connected by joints, leaning on a surface, that makes a slope ϕ with the horizontal, as seen in Figure 6.1. Angular joint coordinates are θ_1 and θ_2; they consist of the angle made by the first link with the horizontal and the angle between the two links, respectively. The torques applied to each of the two joints of the robot are τ_1 and τ_2. The force exerted by the environment on the robot's tip is assumed as having the direction of the x-axis (this means that friction with the wall is neglected) and modelled by means of a second-order differential equation (such as those involving a mass, a spring and a dashpot).

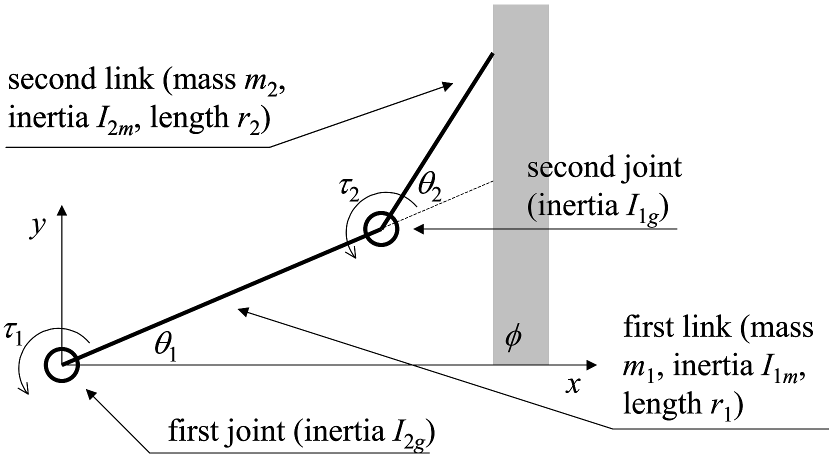


Figure 6.1 — Robotic arm

A complete list of symbols used in the model follows.

c	vector of centrifugal and Coriolis terms
F	force exerted by the environment on the robot's tip
g	vector of gravitational terms
H	inertia matrix, dimension 2×2
I_{1g}, I_{2g}	moments of inertia of the joints
I_{1m}, I_{2m}	moments of inertia of the links
J	Jacobian matrix of the robot, dimension 2×2
$\mathbf{J}_c$	Jacobian matrix of kinematic equations Λ_c, dimension 2×2
m_1	mass of the first link
m_2	mass of the second link
q	vector consisting of coordinates θ_1 and θ_2
r_1	length of the first link
r_2	length of the second link
S	selection matrix for choosing which variable is controlled
x_c, y_c	Cartesian coordinates of the robot's tip found using Λ_c
Λ_c	kinematic equations for finding x_c and y_c from θ_1 and θ_2
θ_1, θ_2	joint coordinates
τ_1, τ_2	actuator torques on the joints

The robot's and the environment's dynamic behaviours are given by[99]

$$\tau = \mathbf{H}(q)\ddot{q} + c(q,\dot{q}) + g(q) - \mathbf{J}^T(q)\mathbf{F} \tag{6.1}$$

$$\mathbf{H}_{11} = (m_1 + m_2)r_1^2 + m_2 r_2^2 + 2m_2 r_1 r_2 \cos(\theta_2) + I_{1m} + I_{1g} \tag{6.2}$$

$$\mathbf{H}_{12} = m_2 r_2^2 + m_2 r_1 r_2 \cos(\theta_2) \tag{6.3}$$

$$\mathbf{H}_{21} = m_2 r_2^2 + m_2 r_1 r_2 \cos(\theta_2) \tag{6.4}$$

$$\mathbf{H}_{22} = m_2 r_2^2 + I_{2m} + I_{2g} \tag{6.5}$$

$$c_1 = -m_2 r_1 r_2 \sin(\theta_2)\dot{\theta}_2^2 - 2m_2 r_1 r_2 \sin(\theta_2)\dot{\theta}_1\dot{\theta}_2 \tag{6.6}$$

$$c_2 = m_2 r_1 r_2 \sin(\theta_2)\dot{\theta}_1^2 \tag{6.7}$$

[99] Machado *et al.* (1998).

$$g_1 = g\left[m_1 r_1 \cos(\theta_1) + m_2 r_1 \cos(\theta_1) + m_2 r_2 \cos(\theta_1 + \theta_2)\right] \tag{6.8}$$

$$g_2 = g m_2 r_2 \cos(\theta_1 + \theta_2) \tag{6.9}$$

$$\mathbf{J}_{11} = -r_1 \sin(\theta_1) + r_2 \sin(\theta_1 + \theta_2) \tag{6.10}$$

$$\mathbf{J}_{21} = r_1 \cos(\theta_1) + r_2 \cos(\theta_1 + \theta_2) \tag{6.11}$$

$$\mathbf{J}_{12} = -r_2 \sin(\theta_1 + \theta_2) \tag{6.12}$$

$$\mathbf{J}_{22} = r_2 \cos(\theta_1 + \theta_2) \tag{6.13}$$

$$F_1 = M\ddot{x}_c + B\dot{x}_c + Kx_c \tag{6.14}$$

$$F_2 = 0 \tag{6.15}$$

where

$$q = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} \tag{6.16}$$

$$\tau = \begin{bmatrix} \tau_1 \\ \tau_2 \end{bmatrix} \tag{6.17}$$

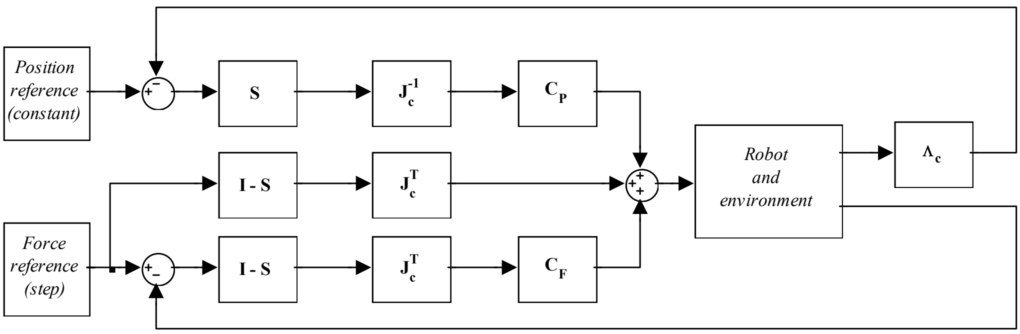


Figure 6.2 — Hybrid control scheme

Coordinate x is force-controlled and coordinate y (which is vertical) position-controlled. The hybrid control scheme[100] is found in Figure 6.2. For this control loop, we have:

$$\mathbf{S} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \tag{6.18}$$

$$\Lambda_c : \begin{cases} x_c = r_1 \sin(\phi - \theta_1) + r_2 \sin(\phi - \theta_1 - \theta_2) \\ y_c = r_1 \cos(\phi - \theta_1) + r_2 \cos(\phi - \theta_1 - \theta_2) \end{cases} \tag{6.19}$$

$$\mathbf{J}_{c11} = -r_1 \cos(\phi - \theta_1) - r_2 \cos(\phi - \theta_1 - \theta_2) \tag{6.20}$$

$$\mathbf{J}_{c12} = -r_2 \cos(\phi - \theta_1 - \theta_2) \tag{6.21}$$

$$\mathbf{J}_{c21} = r_1 \sin(\phi - \theta_1) + r_2 \sin(\phi - \theta_1 - \theta_2) \tag{6.22}$$

$$\mathbf{J}_{c22} = r_2 \sin(\phi - \theta_1 - \theta_2) \tag{6.23}$$

[100] Khatib (1987); Machado *et al.* (1998).

The parameters of the robot and of the environment are the following:

$$m_1 = 0.5 \text{ kg} \tag{6.24}$$
$$m_2 = 6.25 \text{ kg} \tag{6.25}$$
$$r_1 = 1.0 \text{ m} \tag{6.26}$$
$$r_2 = 0.8 \text{ m} \tag{6.27}$$
$$I_{1m} = 1.0 \text{ kg m}^2 \tag{6.28}$$
$$I_{2m} = 1.0 \text{ kg m}^2 \tag{6.29}$$
$$I_{1g} = 4.0 \text{ kg m}^2 \tag{6.30}$$
$$I_{2g} = 4.0 \text{ kg m}^2 \tag{6.31}$$
$$M = 0.03 \text{ kg} \tag{6.32}$$
$$B = 0.03 \text{ N s/m} \tag{6.33}$$
$$K = 400 \text{ N/m} \tag{6.34}$$
$$\phi = \frac{\pi}{2} \text{ rad} \tag{6.35}$$

The reference test is that of Machado *et al.* (1998). The robot begins at

$$\theta_1 = \theta_2 = \frac{15\pi}{36} \text{ rad} \tag{6.36}$$

without exerting force on the environment; then a step at the force reference, with magnitude $F_{\text{ref}} = 1 \text{ N}$, is applied when t = 0; the vertical position y_c should remain constant, equal to its original position $y_{c\text{ref}}$.

6.1.2. Controllers used

Non-integer order controllers given by Machado *et al.* (1998) for this robot are

$$C_F = \begin{bmatrix} G_{F1} s^{\nu_{F1}} \\ G_{F2} s^{\nu_{F2}} \end{bmatrix} \tag{6.37}$$
$$C_P = \begin{bmatrix} G_{P1} s^{\nu_{P1}} \\ G_{P2} s^{\nu_{P2}} \end{bmatrix} \tag{6.38}$$

implemented with the parameters given in the first row of Table 6.1 (found by trial and error) and using the first order backwards finite difference approximation based upon a MacLaurin series, (3.92), with a sampling time T = 1 ms. In what follows, a lower number of parameters in the transfer function, N = 8, was kept, because a large value for N results in a larger steady-state error for the vertical position, in slightly larger overshoots for the force and in larger settling times: the value of N used here is a compromise with a smaller steady-state error for the force.

6.1.3. The genetic algorithm used for parameter tuning

An optimal controller for this plant was found by means of a genetic algorithm. The controllers of Machado *et al.* (1998) were used as a starting point for the optimisation process. The algorithm used was as follows:

❖ *Initialisation.* Create a population consisting of 50 possible solutions. The first has the parameters of the controllers given by Machado *et al.* (1998); others are randomly obtained as follows.

❖ Differentiation orders ν_{F1} and ν_{F2} are normally distributed with mean $-1/5$ and variance 5.

❖ Differentiation orders ν_{P1} and ν_{P2} are normally distributed with mean $1/2$ and variance 5.

❖ Gain G_{F1} is normally distributed with mean $10^3 T^{\nu_{F1}+\frac{1}{5}}$ (where T is the sampling time) and variance 250.

❖ Gain G_{F2} is normally distributed with mean $10^3 T^{\nu_{F2}+\frac{1}{5}}$ (where T is the sampling time) and variance 250.

❖ Gain G_{P1} is normally distributed with mean $10^3 T^{\nu_{P1}-\frac{1}{2}}$ (where T is the sampling time) and variance 25000.

❖ Gain G_{P2} is normally distributed with mean $10^3 T^{\nu_{P2}-\frac{1}{2}}$ (where T is the sampling time) and variance 25000.

❖ Those values are codified into a binary number as follows.

❖ Differentiation order ν_{F1} is codified using 15 bits: 1 bit represents the sign, the other 14 the absolute value, multiplied by 1000. As a consequence, the resolution obtained is 0.001.

❖ Gain G_{F1} is codified using 18 bits. Only positive integer values are possible (or, in other words, the resolution is 1).

❖ Differentiation order ν_{F2} is codified using 15 bits, just as ν_{F1} above.

❖ Gain G_{F2} is codified using 18 bits, just as G_{F1} above.

❖ Differentiation order ν_{P1} is codified using 15 bits, just as ν_{F1} above.

❖ Gain G_{P1} is codified using 18 bits, just as G_{F1} above.

❖ Differentiation order ν_{P2} is codified using 15 bits, just as ν_{F1} above.

❖ Gain G_{P2} is codified using 18 bits, just as G_{F1} above.

❖ *Iterations.* Repeat the following steps until the exit condition is satisfied.

❖ Control is simulated for all individuals in the population. Fractional powers of s were implemented using one of four digital approximations: those based upon truncating MacLaurin series and continuous fraction expansions of a first order backwards finite difference ((3.92) and (3.112)) and those based upon truncating MacLaurin series and continuous fraction expansions of Tustin formula ((3.100) and (3.113)). Maximum force overshoot M_F, maximum position error M_{y_c}, steady-state force error S_F and stationary position error S_{y_c} are determined:

$$M_F = \frac{\max(F) - F_{\text{ref}}}{F_{\text{ref}}} \tag{6.39}$$

$$M_{y_c} = \frac{\max(|y_c - y_{c\text{ref}}|)}{y_{c\text{ref}}} \tag{6.40}$$

$$S_F = \frac{F(1\text{ s}) - F_{\text{ref}}}{F_{\text{ref}}} \tag{6.41}$$

$$S_{y_c} = \frac{y_c(1\text{ s}) - y_{c\text{ref}}}{y_{c\text{ref}}} \tag{6.42}$$

These values are then combined into a performance index. Several have been experimented; the one retained was

$$J = S_F^4 + 6M_F^2 + 1000S_{y_c}^2 + 1000M_{y_c}^2 \tag{6.43}$$

❖ *Algorithm termination.* If this is the 100th iteration, or if the best value of (6.43) did not change in the previous 10 iterations, stop the algorithm.

❖ *Elitism.* Eliminate all individuals in the population save those with the 6 best indexes of performance. Eliminated individuals are replaced as follows.

❖ *Mutation.* One-half of the eliminated individuals are replaced by mutations of the six remaining individuals. A random individual is chosen, and each binary digit thereof may undergo a change, with a probability of 4 %, so that the overall probability of no mutation is 0.5 %.

❖ *Three-point crossover.* One-fourth of the eliminated individuals are replaced by descendants of the six remaining individuals. Two random individuals are chosen. Each of the controllers $G_{F1}s^{\nu_{F1}}$, $G_{F2}s^{\nu_{F2}}$, $G_{P1}s^{\nu_{P1}}$ and $G_{P2}s^{\nu_{P2}}$ may be taken from one of those two, with equal probability.

❖ *Spontaneous generation.* One-fourth of the eliminated individuals are replaced by a population created exactly as in the beginning (save that the total of created individuals is no longer 49, but one-fourth of the missing ones, as mentioned).

This algorithm deserves two remarks.

Firstly, building a significant performance index is always hard, and the result is always questionable—quantifying *exactly* how much we prefer overshoot to steady-state error is not always straightforward. Regarding this issue, it should be stressed that values of S_{y_c} and M_{y_c} are systematically very low. In spite of the 1000 weights in (6.43), the performance of the force control loop is always more important. This is quite reasonable since y_c never deviates more than 2 mm from its reference. And it should also be stressed that performance indexes as the one chosen take no account of settling times or (even worst) high-frequency rippled responses. But settling times are no issue, since simulations are 1 s long and this is short enough to ensure that a good steady-state response also has a good settling time; and solutions presenting a high-frequency ripple may be rejected after the end of the algorithm.

Secondly, the accuracy of approximating formulas being variable, the optimisation led, depending on the formula used, to different parameters for the controllers.

	G_{F1}	v_{F1}	G_{F2}	v_{F2}	G_{F3}	v_{F3}	G_{F4}	v_{F4}	N	Approx.
Original	10^3	–0.200	10^3	–0.200	10^5	0.500	10^5	0.500	17	(3.92)
Set A	4.570×10^4	–0.750	1.741×10^3	–0.155	9.009×10^5	0.180	1.381×10^5	0.484	8	(3.92)
Set B	1.232×10^5	–0.894	0.704×10^3	–0.187	1.128×10^8	–0.496	9.479×10^4	0.465	8	(3.112)
Set C	1.191×10^4	–0.556	2.075×10^4	–0.639	1.789×10^6	0.113	8.146×10^4	0.470	8	(3.100)
Set D	2.281×10^4	–0.650	2.024×10^3	–0.292	2.628×10^6	0.119	4.326×10^4	0.465	8	(3.113)

Table 6.1 — Control parameters

6.1.4. Results and conclusions

Figure 6.3 and Table 6.2 show the performance of the controller found by Machado *et al.* (1998) (implemented, as explained above, with $N = 8$) together with that of the best controllers found with the genetic algorithm, given in Table 6.1. Four sets of controllers are given, one for each of the four approximations mentioned above. Significant performance improvements are clear over steady-state errors, maximum overshoots and settling times, and hence over the performance index as well.

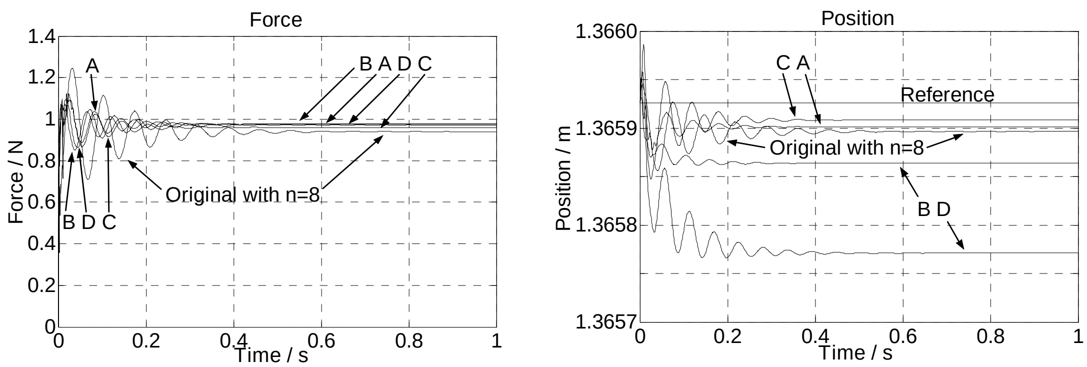


Figure 6.3 — Control results

	Original (with $N = 8$)	Set A	Variation	Set B	Variation	Set C	Variation	Set D	Variation
M_F	24.5 %	12.3 %	–50 %	7.0 %	–71 %	9.4 %	–62 %	6.8 %	–72 %
$1000\ M_{y_c}$	5.1 %	3.6 %	–29 %	4.7 %	–8 %	5.7 %	12 %	11.7 %	129 %
S_F	–6.0 %	–2.5 %	–58 %	–2.4 %	–60 %	–4.0 %	–33 %	–2.9 %	–52 %
$1000\ S_{y_c}$	–2.2 %	–1.8 %	–18 %	–4.5 %	105 %	–1.3 %	–41 %	–11.3 %	414 %
J	4959	959	–81 %	370	–93 %	815	–84 %	620	–87 %
F settling time	362 ms	177 ms	–51 %	178 ms	–51 %	181 ms	–50 %	217 ms	–40 %
y_c settling time	576 ms	334 ms	–42 %	223 ms	–61 %	429 ms	–26 %	343 ms	–40 %

Table 6.2 — Control results (settling times were computed with respect to the steady-state value, using the 2 % criterion)

A few remarks ought to be made. Firstly, every attempt to reduce the force overshoot further, by increasing its weight in the performance index or by increasing its exponent, had as consequence a significant increase of the steady-state error. Secondly, the higher importance of the force control loop performance is clear from results, since position control, in certain cases, became *worst*, and nevertheless index (6.43) undergoes a clear decrease (this happened with controllers implemented with approximations using continued fractions). Thirdly, using parameters tuned with one formula for implementing non-integer order derivatives together with another formula was tried, but this does not (with very few exceptions) result in an improved behaviour. It seems clear that this optimisation is formula-dependent: as expected, the chosen formula conditions the control parameters obtained.

Finally, these controllers show robustness in face of changes in initial conditions. Figure 6.4 shows how the performance index changes when θ_1 and θ_2 vary (the position of the wall has been accordingly modified so that the simulation still begins with the robot at the desired position without exerting force on the environment). It was reckoned with the controllers of set B, but similar results might have been obtained with other sets. It shows that changing the initial conditions results in a poorer behaviour, but there is still a reasonably broad margin for which the control still remains stable. If variations of θ_1 and θ_2 are so large that behaviour becomes unacceptably poor or even unstable, other controllers will have to be obtained for the new conditions.

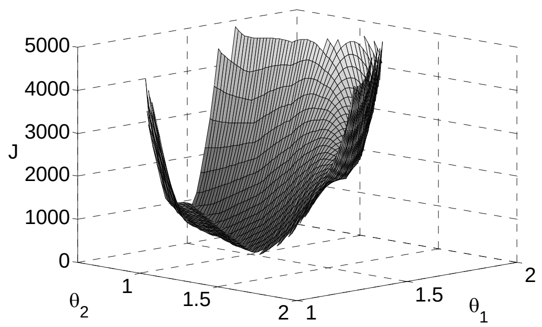


Figure 6.4 — Control results: performance index as a function of initial conditions

6.2. Fractional control of a flexible robot

In this section fractional order control for a flexible robot is presented. A robot is said to be flexible if designed in such a manner that its structure will undergo deformation (and consequently vibration) under normal operation conditions to such a degree that rigid body models become too poor for providing suitable approximations for control. Low robot mass / carried mass ratios make flexible robots rather attractive from the energy saving point of view, but difficulties arising from its control are significant, especially for tasks demanding a high accuracy.

6.2.1. The flexible robot

The flexible robot addressed is a two-degree of freedom planar horizontal robot extant at the Control, Automation and Robotics Laboratory of the Technical University of Lisbon[101]. It consists of a rigid link and a flexible link, connected by rigid hubs, as seen in Figure 6.5 and in Figure 6.6. In each hub there are a motor, a tachometer and an encoder; the vibration of the flexible link is measured by means of two extensometer bridges. Control is performed using Matlab's xPC target toolbox. The robot's tip position can be reckoned from the two angles θ_1 and θ_2 and the elastic displacement of the flexible link v.

[101] The robot's model is taken from Martins (2000, p. 67-69) and from Nabais (2002, p. 34-41).

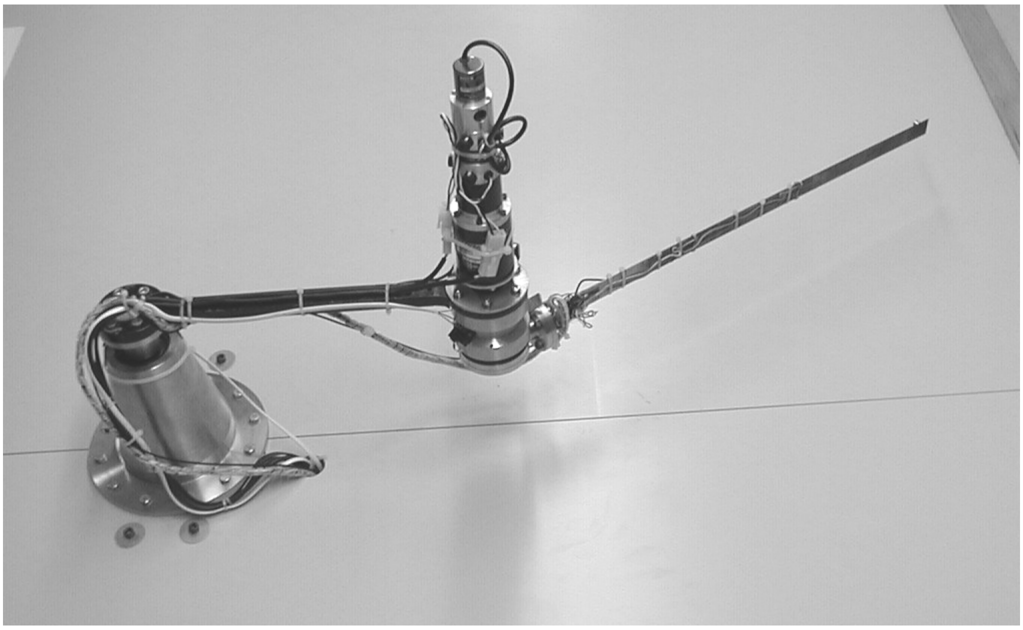


Figure 6.5 — Flexible robotic arm

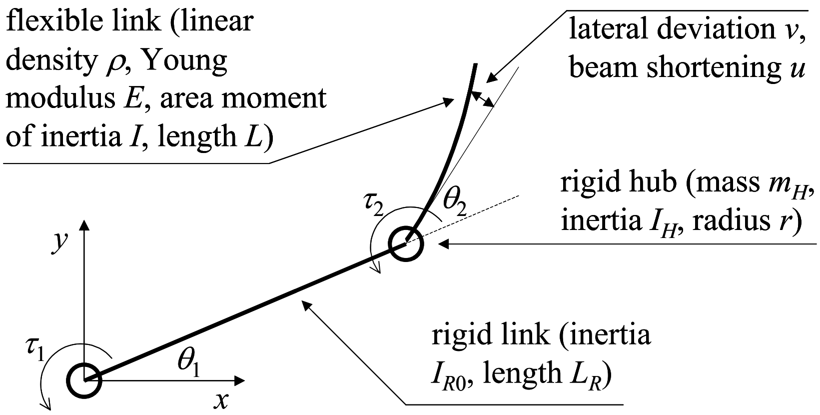


Figure 6.6 — Scheme of the robot in Figure 6.5

A complete list of symbols used in the model follows.

C	centrifugal and Coriolis vector
E	Young modulus of the robot's flexible link
F	viscous friction coefficients vector
$\mathbf{H}$	mass matrix
I_b	moment of inertia of the robot's flexible link
I_{m1}	moment of inertia of the motor of the rigid link's joint
I_{m2}	moment of inertia of the motor of the joint between the two links
I_H	moment of inertia of the hub connecting the links
I_{R0}	moment of inertia of the robot's rigid link
$\mathbf{J}_c$	Jacobian of kinematic equations Λ_c
$\mathbf{K}$	stiffness matrix
L	length of the robot's flexible link
L_R	length of the robot's rigid link
$\mathbf{M}$	auxiliary matrix
M	auxiliary function
m_b	mass of the robot's flexible link, equal to ρL

m_H mass of the hub connecting the links
N auxiliary matrix
q time-varying vector of the coordinates that define the robot's state
r radius of the hub connecting the links
r_1 transmission relation of the rigid link's joint
r_2 transmission relation of the joint between the two links
S auxiliary function
T vector with the torques applied
u beam shortening in the flexible link measured longitudinally
v elastic displacement (lateral deviation) of the flexible link
x_c, y_c Cartesian coordinates of the robot's tip found using Λ_c
α_{ij} auxiliary quantity
β_i eigenvalues corresponding to the free vibration modes of the robot's flexible link
γ_{ij} auxiliary quantity
$\eta_i(t)$ weighing coefficients called elastic coordinates
Λ_c kinematic equations for finding x_c and y_c from θ_1, θ_2, u and v
θ_1, θ_2 angles of the links
ρ linear density of the robot's flexible link
τ_1, τ_2 torques applied at the joints by motors
Φ auxiliary quantity
χ_i normalised clamped-free vibration modes

The following auxiliary quantities will be used:

$$\mathbf{M}_{i,j} = \int_r^{r+L} M(x) \frac{d\chi_i(x)}{dx} \frac{d\chi_j(x)}{dx} dx \tag{6.44}$$

$$\mathbf{N}_{i,j} = \int_r^{r+L} S(x) \frac{d\chi_i(x)}{dx} \frac{d\chi_j(x)}{dx} dx \tag{6.45}$$

$$M(x) = \int_x^{r+L} \rho d\xi \tag{6.46}$$

$$S(x) = \int_x^{r+L} \rho\xi d\xi \tag{6.47}$$

$$\gamma_{ij} = \int_r^{r+L} S(x) \chi_i(x) \frac{d^2\chi_j(x)}{dx^2} dx \tag{6.48}$$

$$\begin{aligned}\mathbf{H}_{1j} = \Big(\rho\int_r^{r+L} x\chi_{j-2}(x)dx + \\ +L_R\cos\theta_2\rho\int_r^{r+L} \chi_{j-2}(x)dx\Big)\Big/ r_1, \quad 3 \le j \le n\end{aligned} \tag{6.49}$$

$$\alpha_{ij} = \int_r^{r+L} M(x) \chi_i(x) \frac{d^2\chi_j(x)}{dx^2} dx \tag{6.50}$$

$$\alpha'_{ij} = \int_r^{r+L} \frac{dM(x)}{dx} \chi_i(x) \frac{d\chi_j(x)}{dx} dx \tag{6.51}$$

$$\Phi(i) = \int_r^{r+L} \chi_i(x) dx \tag{6.52}$$

$$\begin{aligned}\chi_i(x) &= \cosh\left[\beta_i(x-r)\right]-\cos\left[\beta_i(x-r)\right]- \\ &-\frac{\cosh(\beta_i L)+\cos(\beta_i L)}{\sinh(\beta_i L)+\sin(\beta_i L)}\left\{\sinh\left[\beta_i(x-r)\right]-\sin\left[\beta_i(x-r)\right]\right\}\end{aligned} \tag{6.53}$$

Beam-shortening u was neglected. Elastic lateral deviation v was approximated by a finite series

$$v(x,t)=\sum_{i=1}^{n}\chi_i(x)\eta_i(t) \tag{6.54}$$

This allows for a discrete model of the robot which is

$$\mathbf{H}(q)\ddot{q}+F\dot{q}+\mathbf{K}(q,\dot{q})q+C(q,\dot{q})=T \tag{6.55}$$

where

$$T=\begin{bmatrix}\tau_1\\ \tau_2\\ 0\\ \vdots\end{bmatrix} \tag{6.56}$$

$$\eta=\begin{bmatrix}\eta_1\\ \vdots\\ \eta_n\end{bmatrix} \tag{6.57}$$

$$q=\begin{bmatrix}\theta_1\\ \theta_2\\ \eta\end{bmatrix} \tag{6.58}$$

Matrixes **H** and **K** are symmetric. The non-zero elements of **H**, **K** and C are

$$\begin{aligned}\mathbf{H}_{11}=\frac{1}{r_1^2}\Big[&I_{m1}+I_{R0}+L_R^2(m_H+m_b)+I_H+I_b+\rho L\eta^T\eta-2\rho L_R\sin\theta_2\Phi^T\eta+ \\ &+2\rho L_R\cos\theta_2\int_r^{r+L}xdx-\eta^T\mathbf{N}\eta-L_R\cos\theta_2\eta^T\mathbf{M}\eta\Big]\end{aligned} \tag{6.59}$$

$$\begin{aligned}\mathbf{H}_{12}=\mathbf{H}_{21}=\frac{1}{r_1r_2}\Big[&I_H+I_b+\rho L\eta^T\eta-\rho L_R\sin\theta_2\Phi^T\eta+ \\ &+\rho L_R\cos\theta_2\int_r^{r+L}xdx+\frac{L_R\cos\theta_2}{2}\eta^T\mathbf{M}\eta-\eta^T\mathbf{N}\eta\Big]\end{aligned} \tag{6.60}$$

$$\mathbf{H}_{22}=\frac{I_{m2}+I_H+I_b+\rho L\eta^T\eta-\eta^T\mathbf{N}\eta}{r_2^2} \tag{6.61}$$

$$\mathbf{H}_{1j}=\mathbf{H}_{j1}=\frac{\rho\int_r^{r+L}x\chi_{j-2}(x)dx+L_R\cos\theta_2\rho\int_r^{r+L}\chi_{j-2}(x)dx}{r_1},\quad 3\le j\le n \tag{6.62}$$

$$\mathbf{H}_{2j}=\mathbf{H}_{j2}=\frac{\rho\int_r^{r+L}x\chi_{j-2}(x)\,dx}{r_2},\quad 3\le j\le n \tag{6.63}$$

$$\mathbf{H}_{3j}=\mathbf{H}_{j3}=\rho L,\quad 3\le j\le n \tag{6.64}$$

$$\begin{aligned}\mathbf{K}_{i+2,i+2}&=EIL\beta_i^4-\dot\theta_1^2\left[\rho L+\gamma_{ii}+\gamma'_{ii}+L_R\cos\theta_2\left(\alpha_{ii}+\alpha'_{ii}\right)\right]-\\&-\dot\theta_2^2\left(\rho L+\gamma_{ii}+\gamma'_{ii}\right)-2\dot\theta_1\dot\theta_2\left[\rho L+\gamma_{ii}+\gamma'_{ii}+\frac{L_R\cos\theta_2}{2}\left(\alpha_{ii}+\alpha'_{ii}\right)\right],\\&1\le i\le n\end{aligned} \tag{6.65}$$

$$\begin{aligned}\mathbf{K}_{i+2,j+2}&=\mathbf{K}_{j+2,i+2}=-\dot\theta_1^2\left[\gamma_{ii}+\gamma'_{ii}+L_R\cos\theta_2\left(\alpha_{ii}+\alpha'_{ii}\right)\right]-\dot\theta_2^2\left(\gamma_{ii}+\gamma'_{ii}\right)-\\&-2\dot\theta_1\dot\theta_2\left[\gamma_{ii}+\gamma'_{ii}+\frac{L_R\cos\theta_2}{2}\left(\alpha_{ii}+\alpha'_{ii}\right)\right],\quad 1\le i\le n\wedge 1\le j\le n\wedge i\ne j\end{aligned} \tag{6.66}$$

$$\begin{aligned}C(1)=\frac{1}{r_1}\Big[&\left(\rho L\eta^T-\rho L_R\sin\theta_2\Phi^T-\eta^T\mathbf{N}-L_R\cos\theta_2\eta^T\mathbf{M}\right)2\dot\theta_1\dot\eta+\\&+\left(\rho L\eta^T-\rho L_R\sin\theta_2\Phi^T-\eta^T\mathbf{N}+\frac{L_R\cos\theta_2}{2}\eta^T\mathbf{M}\right)2\dot\theta_2\dot\eta+\\&+\left(-\rho L_R\sin\theta_2\int_r^{r+L}xdx-\frac{L_R\sin\theta_2}{2}\eta^T\mathbf{M}\eta-\rho L_R\cos\theta_2\Phi^T\eta\right)\dot\theta_2^2+\\&+\left(-\rho L_R\sin\theta_2\int_r^{r+L}xdx+\frac{L_R\sin\theta_2}{2}\eta^T\mathbf{M}\eta-\rho L_R\cos\theta_2\Phi^T\eta\right)2\dot\theta_1\dot\theta_2\Big]\end{aligned} \tag{6.67}$$

$$\begin{aligned}C(2)=\frac{1}{r_2}\Big[&\left(\rho L\eta^T+\frac{L_R\cos\theta_2}{2}\eta^T\mathbf{M}-\eta^T\mathbf{N}\right)2\dot\theta_1\dot\eta+\\&+\left(\rho L_R\sin\theta_2\int_r^{r+L}xdx+\rho L_R\cos\theta_2\Phi^T\eta-\frac{L_R\sin\theta_2}{2}\eta^T\mathbf{M}\eta\right)\dot\theta_1^2+\\&+\left(\rho L\eta^T-\eta^T\mathbf{N}\right)2\dot\eta\dot\theta_2-\dot\theta_1\dot\theta_2L_R\sin\theta_2\eta^T\mathbf{M}\eta\Big]\end{aligned} \tag{6.68}$$

$$C(i)=\dot\theta_1L_R\sin\theta_2\rho\int_r^{r+L}\chi_{i-2}(x)\,dx,\quad 3\le i\le n \tag{6.69}$$

Beyond vector F, friction in the joints was modelled by means of a dead zone and of Coulomb and viscous friction affecting the input of each.

$$\tau_{i.\text{effective}}=\text{sign}(\tau_i)\left(\mu_i|\tau_i|+\tau_{0i}\right),\quad i=1,2 \tag{6.70}$$

Friction parameters, given in Table 6.3, were found by minimising, using a genetic algorithm, the quadratic error of the model responses to steps and impulses.

Robot parameters are as follows:

$$L_R=0.32\text{ m} \tag{6.71}$$

$$I_{R0}=0.25\text{ kg m}^2 \tag{6.72}$$

$$I_H=13.22\times10^{-4}\text{ kg m}^2 \tag{6.73}$$

$$m_H=0.47\text{ kg} \tag{6.74}$$

$$r = 0.075\ \text{m} \tag{6.75}$$
$$L = 0.5\ \text{m} \tag{6.76}$$
$$I_b = 99\times10^{-4}\ \text{kg}\,\text{m}^2 \tag{6.77}$$
$$\rho = 0.157\ \text{kg/m} \tag{6.78}$$
$$m_b = 0.0785\ \text{kg} \tag{6.79}$$
$$EI = 0.349\ \text{N}\,\text{m}^2 \tag{6.80}$$
$$n = 3 \tag{6.81}$$
$$\beta_1 = \frac{1.8751}{L} \tag{6.82}$$
$$\beta_2 = \frac{4.6941}{L} \tag{6.83}$$
$$\beta_3 = \frac{7.8548}{L} \tag{6.84}$$

Viscous friction vector	$F = \left[2.0\times10^{-4} \quad 2.3\times10^{-4} \quad 0 \quad 0 \quad 0\right]$
First joint's dead zone	$\left[-3.2379 \quad 2.9129\right]$ N m
Second joint's dead zone	$\left[-0.5543 \quad 0.6608\right]$ N m
First joint's Coulomb friction τ_{01}	4.56 N m
Second joint's Coulomb friction τ_{02}	0 N m
First joint's viscous friction μ_1	0.2585
Second joint's viscous friction μ_2	1.0231

Table 6.3 — Friction parameters

The kinematic equations for finding the Cartesian coordinates x and y of the robot's tip from the joint coordinates and variables u and v reflecting the second link's flexibility are

$$\Lambda_c : \begin{cases} x = L_R \cos\theta_1 + (L + r + u)\cos(\theta_1 + \theta_2) - v\sin(\theta_1 + \theta_2) \\ y = L_R \sin\theta_1 + (L + r + u)\sin(\theta_1 + \theta_2) + v\cos(\theta_1 + \theta_2) \end{cases} \tag{6.85}$$

The resulting Jacobian is

$$\mathbf{J}_{c11} = -L_R \sin\theta_1 - (L + r + u)\sin(\theta_1 + \theta_2) - v\cos(\theta_1 + \theta_2) \tag{6.86}$$
$$\mathbf{J}_{c12} = -(L + r + u)\sin(\theta_1 + \theta_2) - v\cos(\theta_1 + \theta_2) \tag{6.87}$$
$$\mathbf{J}_{c21} = L_R \cos\theta_1 + (L + r + u)\cos(\theta_1 + \theta_2) - v\sin(\theta_1 + \theta_2) \tag{6.88}$$
$$\mathbf{J}_{c22} = (L + r + u)\cos(\theta_1 + \theta_2) - v\sin(\theta_1 + \theta_2) \tag{6.89}$$

6.2.2. Controllers used

Both PIDs and fractional PIDs were used to control of the position of the tip of the robot. Fractional derivatives were implemented using Tustin's FIR approximation (3.100), with a sampling time of 1 ms. The use of other digital approximations was attempted but the best results were obtained with this one.

The control structure employed is that of Figure 6.7. Notice that v enters the control loop with its signal changed. This is a way of dealing with the non-minimum phase behaviour of the plant that results from the flexibility of the second link: when it turns to one side, its tip bends to the other side in a first moment[102].

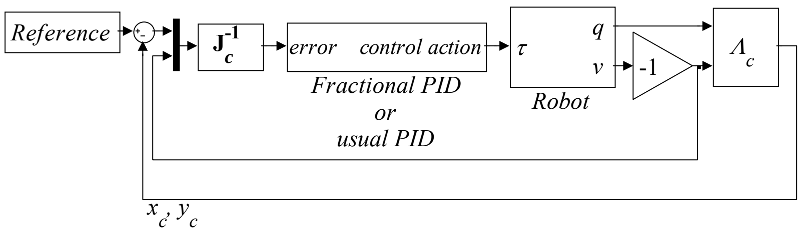


Figure 6.7 — Control loop

6.2.3. The genetic algorithm used for parameter tuning

Both PIDs and fractional PIDs were tuned using a genetic algorithm. That for fractional PIDs, implemented in Matlab, was as follows:

❖ *Initialisation.* A population of 50 elements is created. Each corresponds to a set of two fractional PIDs, the first (C_1) for the first joint of the robot and the second (C_2) for the second joint, given by

$$C_1 = k_1 D^{\nu_1} + k_2 D^{\nu_2} + k_3 \quad (6.90)$$

$$C_2 = k_4 D^{\nu_4} + k_5 D^{\nu_5} + k_6 \quad (6.91)$$

Parameters, stored as real numbers, are randomly chosen with normal distributions with the means and standard deviations of Table 6.4, which are based on previous trial-and-error tuning.

	k_1	ν_1	k_2	ν_2	k_3	k_4	ν_4	k_5	ν_5	k_6
$\bar{x}$	10^4	0	1	0	10^{-2}	10^2	0	1	0	10^{-2}
σ_x	2000	2	0.1	0.1	0.1	20	2	0.1	0.1	0.1

Table 6.4 — Parameters' distributions

❖ *Algorithm termination.* Following steps are performed until 100 iterations are reached, or until no improvement arises over 10 iterations.

❖ *Iterations.* Control is simulated for all individuals. The number of terms retained in (3.100) was set to $N = 10$. A performance index is reckoned including the sum of the integrals of the squares of the errors in both coordinates, and a term for penalising large vibrations of the tip of the robot, since they are hard to deal with and a potential source of inaccuracy. Vibration at the end of the simulation should be as small as possible, so that the robot may come to rest; thus, as vibration cannot be completely avoided, it is penalised increasingly with time:

[102] Benosman *et al.*, (2002).

$$J=\int_0^{t_{end}}\left[\left(x-x_{ref}\right)^2+\left(y-y_{ref}\right)^2+t^2v^4\right]dt \tag{6.92}$$

❖ *Elitism.* Eliminate all individuals save those with the 6 best performance indexes. These are allowed into the next generation, and are the only ones that may reproduce or mutate.

❖ *Mutation.* One-half of the eliminated individuals are replaced by mutations of the surviving ones. Mutants begin as copies of a randomly chosen survivor. The number of parameters that mutate is randomly chosen; the probability decreases exponentially with the number of parameters, so that it will be more likely that few parameters will change. The change consists in adding a Gaussianly distributed random number with zero-mean and variance equal to one-tenth of the mutating parameter.

❖ *Nine-point crossover.* One-fourth of the eliminated individuals are replaced by descendents of the surviving ones. Each descendant is the offspring of two randomly chosen survivors; each of the ten parameters listed in Table 2 is drawn, randomly, from one of the parents.

❖ *Spontaneous generation.* One-fourth of the eliminated individuals are replaced by randomly generated individuals, such as those of the original population—save that they are fewer in number.

A similar algorithm was used for fitting PIDs to the robot. The distributions given in Table 6.4 for parameters k_1, k_2, k_3, k_4, k_5 and k_6 were used with this algorithm as well; ν_1 and ν_4 were forced to be 1; ν_2 and ν_5 were forced to be –1.

6.2.4. Results and conclusions

Three different trajectories were considered and different fractional PIDs were tuned for each. Trajectories correspond to movements usually required from robots. Figure 6.8, Figure 6.9 and Figure 6.10 show, for each trajectory, how the position of the tip of the robot evolves with time, the trajectory it describes in space, and the vibration it undergoes. Both simulation and experimental results are shown. Controllers' parameters are given in Table 6.5; since only once the two available fractional derivatives were needed, the structure of the fractional PID seems to have been a reasonable choice.

	Figure 6.8	Figure 6.9	Figure 6.10
k_1	4.81×10^{14}	3.89×10^{14}	59.4
ν_1	–3.63	–3.54	0.715
k_2	0	0	0
ν_2	—	—	—
k_3	0.0265	0.318	0.117
k_4	11.8	3.91×10^{3}	304
ν_4	0.298	–0.498	–0.114
k_5	0	1.6106	0
ν_5	—	0.0101	—
k_6	0.0929	0.400	0.0121

Table 6.5 — Controllers' parameters

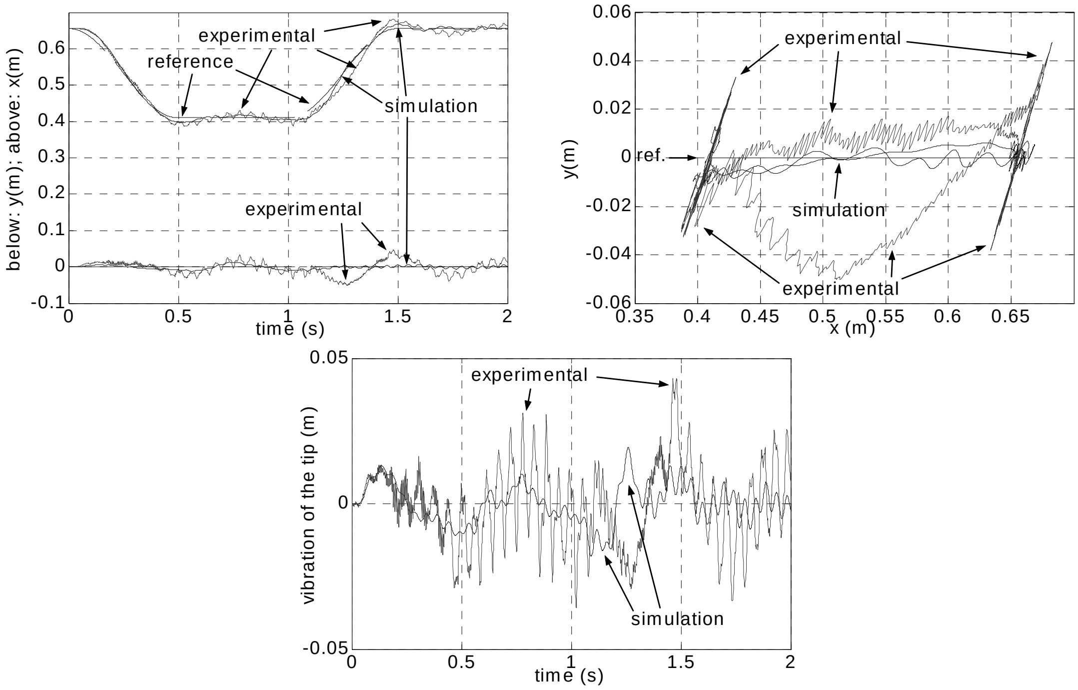


Figure 6.8 — Control results for a linear trajectory (reference trajectory dashed)

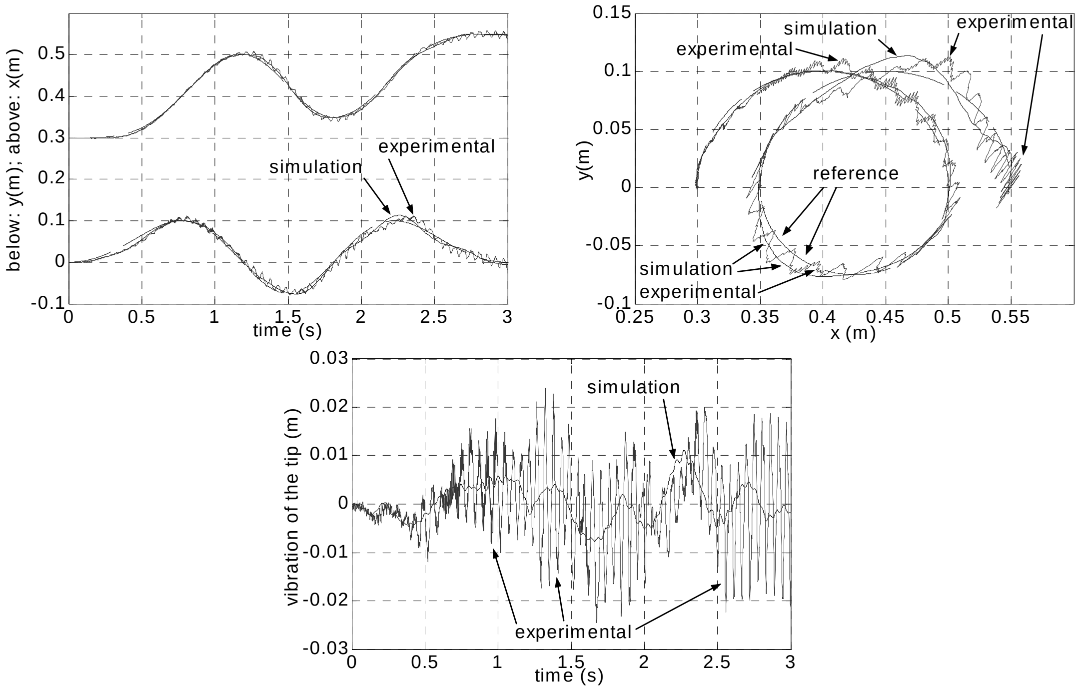


Figure 6.9 — Control results for a two-loop trajectory (reference trajectory dashed).

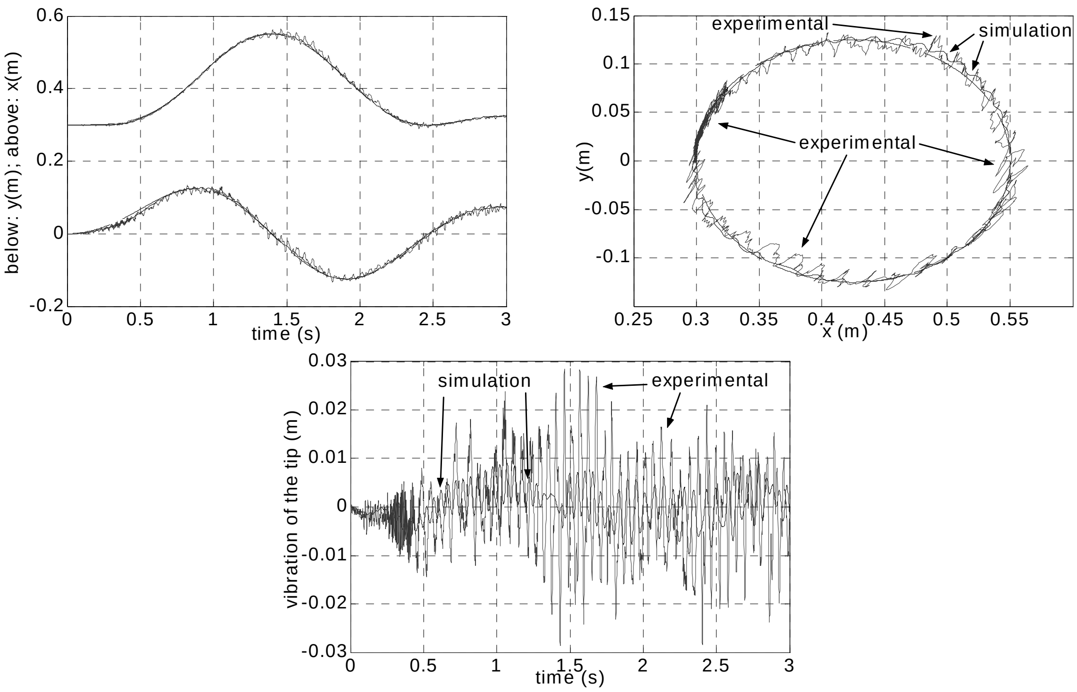


Figure 6.10 — Control results for a circular trajectory (reference trajectory dashed).

Whatever the trajectory, no integer PIDs were found that could stabilise the control-loop. This does not prove, of course, that the plant cannot be controlled with integer PIDs. But the fact that no such controllers were found, while acceptable fractional PIDs were, shows that the latter are clearly superior for this application. Furthermore, controllers developed for one trajectory also work for other references as well. An example (among others possible) is shown in Figure 6.11, showing that controllers work for trajectories different from that they were devised for.

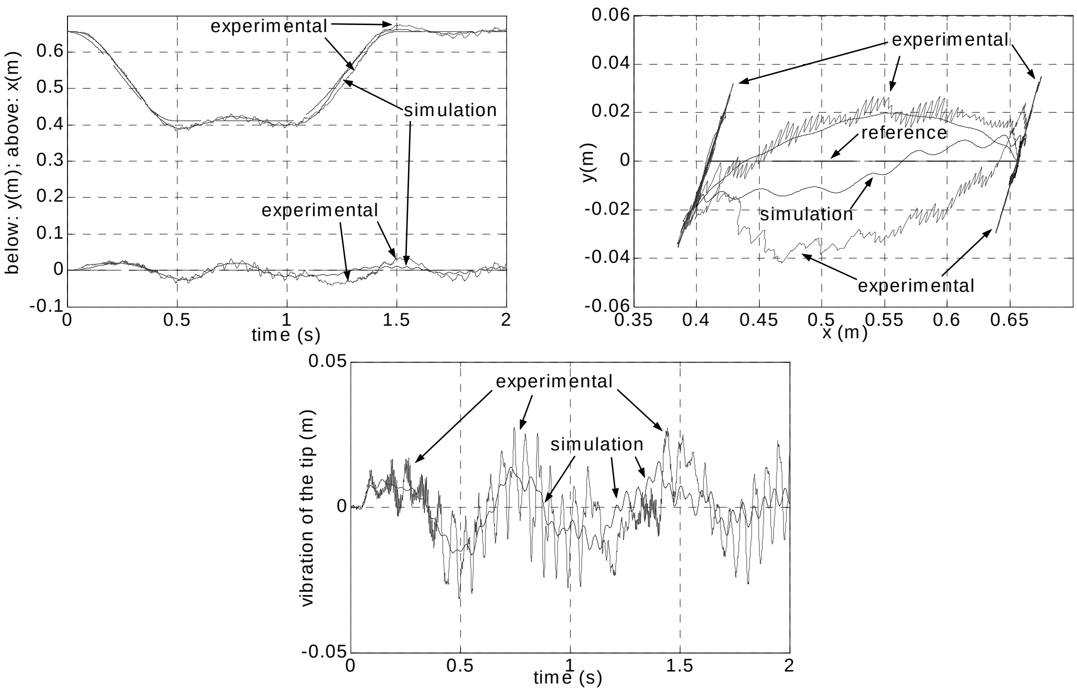


Figure 6.11 — Control results for the trajectory of Figure 6.8 when the controller devised for the trajectory of Figure 6.9 is used

However, it is also clear from these results that the experimental performance is clearly worse than what was expected from simulations. Table 6.6 gives the half-width of an envelope around the reference trajectory that fully covers each simulation or experimental result. These envelopes are a set of circles centred on the successive points of the reference; their radius is the envelope half-width. Simulation values, though not good, may still be bearable, but experimental ones are clearly excessive for most applications. This is especially due to higher vibrations in experimental data and may happen because of insufficiencies in the simulation model or because of inaccurate measurements of the vibration, which is thus not properly dealt with and gets larger than it should. The model may be insufficient either in describing the vibration of the flexible link or in describing the friction in the joints. It is hard to improve the former, because taking into account more vibration modes leads to numerical problems when the kinematic equations are inverted; the latter is also hard to improve, since friction depends on temperature and on the current configuration of the robot.

	Simulation	Experimental
Figure 6.8	0.016 m	0.053 m
Figure 6.9	0.014 m	0.021 m
Figure 6.10	0.006 m	0.018 m
Figure 6.11	0.033 m	0.042 m

Table 6.6 — Half-width of the envelope of the reference containing simulated and experimental control results

It should also be noticed that fractional controllers stabilise the loop by coping with vibrations, not by suppressing them.

6.3. Applications of fractional PIDs

In this section three different benchmark plants, that may be used to model several usual plants[103], are controlled by means of fractional PIDs, with parameters found either with the tuning rules of subsections 5.4.3 to 5.4.6 or, when that is possible, with IMC, as presented in subsection 5.4.1. To assert the goodness of the results these are compared to those obtained with integer PIDs tuned with the first Ziegler-Nichols rule.

Bode diagrams presented are exact; all time-responses involving fractional derivatives and integrals were obtained with simulations making use of the Crone approximation (3.74), with

$$\omega_l = 10^{-3} \text{ rad/s} \tag{6.93}$$

$$\omega_h = 10^{3} \text{ rad/s} \tag{6.94}$$

$$N = 7 \tag{6.95}$$

6.3.1. First-order plant with delay

The plant considered was

[103] Morari *et al.* (1989, p. 114-117).

$$F(s)=\frac{K}{1+s}e^{-0.1s} \tag{6.96}$$

The nominal value of K is 1. Controllers obtained with the two tuning rules from subsections 5.4.4 and 5.4.5 and with the first Ziegler-Nichols rule are

$$C_1(s)=0.4448+\frac{0.5158}{s^{1.4277}}+0.2045s^{1.0202} \tag{6.97}$$

$$C_2(s)=1.2507+\frac{1.3106}{s^{1.1230}}-0.2589s^{0.1533} \tag{6.98}$$

$$C_{ZN}(s)=12.0000+\frac{60.0000}{s}+0.6000s \tag{6.99}$$

(Notice that due to the approximations involved one of the gains is negative; this will not, however, affect results.) Corresponding step responses are given in Figure 6.12, Figure 6.13 and Figure 6.14. These show what happens for several values of K, the plant's gain, assumed as known with uncertainty. It is to be noticed that fractional PIDs can deal with a clearly broader range of values of K. This is likely because the specifications the integer PID tries to achieve are different: that is why responses are all faster, at the cost of greater overshoots. More important is that the overshoot is fairly constant with fractional PIDs, at least for those values closer to 1. This is because fractional PIDs attempt to verify (5.40), which the integer PID does not. Data on these responses is summed up in Table 6.7. (In this and the following tables, the rise time is reckoned according to the 10 %—90 % rule and the settling time is reckoned according to the ±5 % rule.)

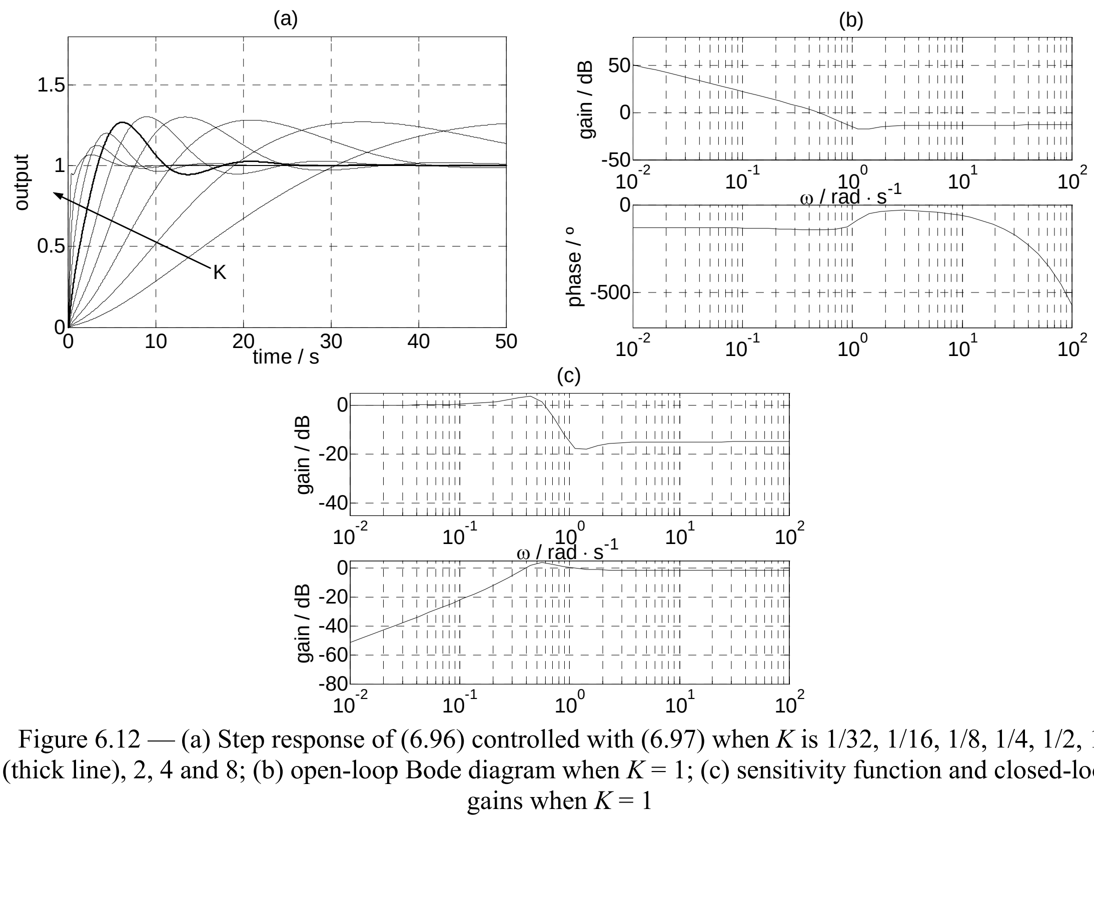


Figure 6.12 — (a) Step response of (6.96) controlled with (6.97) when K is 1/32, 1/16, 1/8, 1/4, 1/2, 1 (thick line), 2, 4 and 8; (b) open-loop Bode diagram when K = 1; (c) sensitivity function and closed-loop gains when K = 1

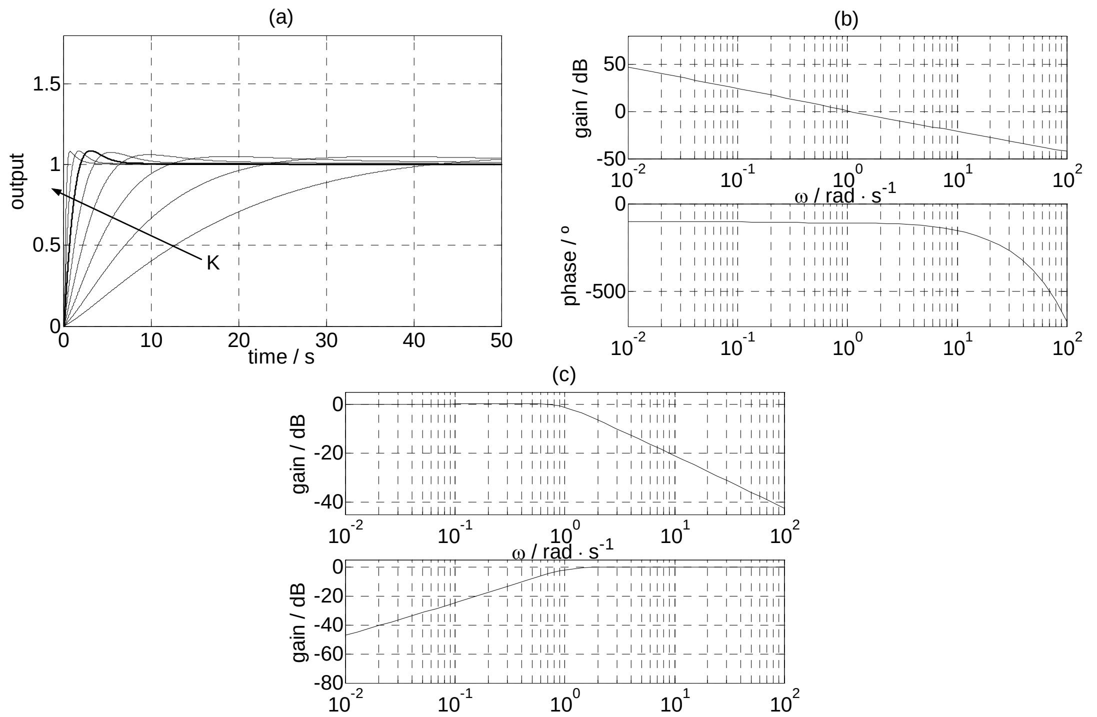


Figure 6.13 — (a) Step response of (6.96) controlled with (6.98) when *K* is 1/32, 1/16, 1/8, 1/4, 1/2, 1 (thick line), 2 and 4; (b) open-loop Bode diagram when *K* = 1; (c) sensitivity function and closed-loop gains when *K* = 1

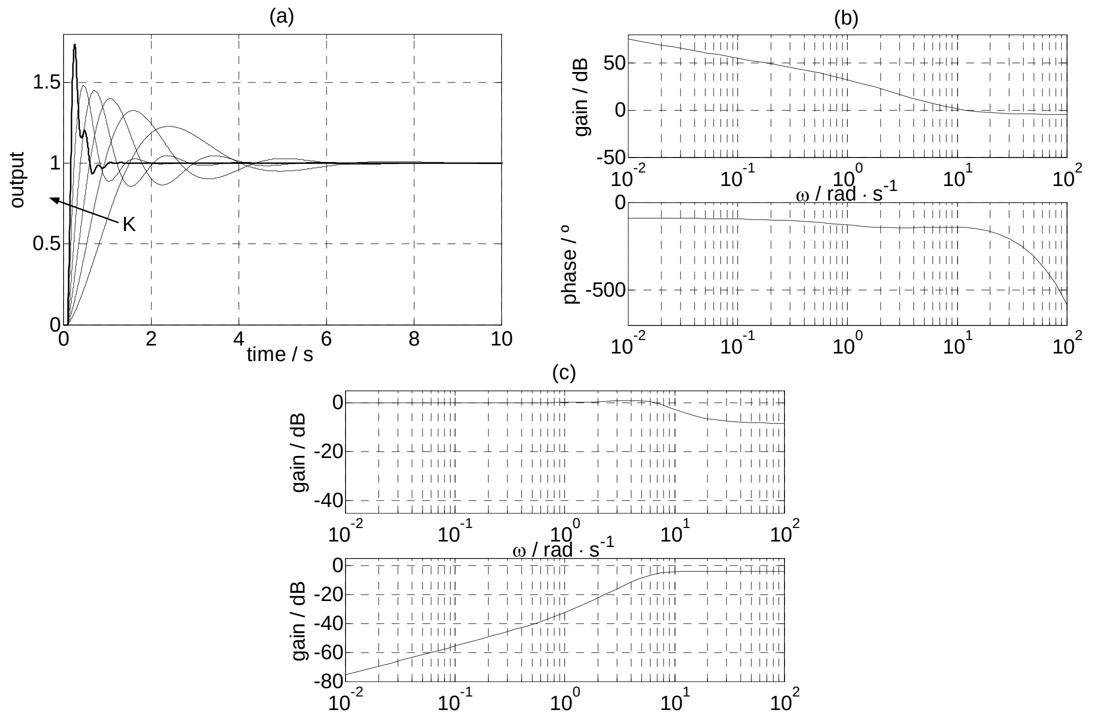


Figure 6.14 — (a) Step response of (6.96) controlled with (6.99) when *K* is 1/32, 1/16, 1/8, 1/4, 1/2 and 1 (thick line); (b) open-loop Bode diagram when *K* = 1; (c) sensitivity function and closed-loop gains when *K* = 1

Those figures also give the corresponding open-loop Bode diagrams and the gains of sensitivity and closed-loop functions. They show that the desired conditions—given in subsection 5.4.2—are reasonably—though not exactly—followed. Differences are due to the approximations incurred by the least-squares fit.

	C_1 given by (6.97)			C_2 given by (6.98)			C_{ZN} given by (6.99)		
K	rise time	overshoot	settling time	rise time	overshoot	settling time	rise time	overshoot	settling time
1/32	22.1 s	26 %	94.6 s	28.1 s	5 %	78.2 s	1.0 s	23 %	3.7 s
1/16	13.8 s	27 %	59.1 s	15.1 s	5 %	19.2 s	0.6 s	33 %	3.9 s
1/8	8.9 s	28 %	36.5 s	8.1 s	5 %	10.2 s	0.4 s	40 %	2.8 s
1/4	5.9 s	30 %	22.6 s	4.4 s	6 %	12.4 s	0.2 s	45 %	1.9 s
1/2	4.0 s	30 %	19.8 s	2.4 s	8 %	7.7 s	0.1 s	48 %	1.3 s
1	2.6 s	27 %	14.6 s	1.3 s	9 %	4.7 s	0.1 s	74 %	0.7 s
2	1.7 s	20 %	7.4 s	0.7 s	8 %	2.8 s	—	—	—
4	0.9 s	12 %	5.5 s	0.3 s	8 %	1.4 s	—	—	—
8	0.2 s	7 %	3.9 s	—	—	—	—	—	—

Table 6.7 — Data on step responses of Figure 6.12, Figure 6.13 and Figure 6.14

6.3.2. Second-order plant

The plant considered was

$$F(s) = \frac{K}{4.3200s^2 + 19.1801s + 1} \approx \frac{K}{1 + 20s} e^{-0.2s} \tag{6.100}$$

with a nominal value of K of 1. The approximation stems from the values of L and T obtained from its step-response. Controllers obtained with the two tuning rules from subsections 5.4.4 and 5.4.5 and with the first Ziegler-Nichols rule are

$$C_1(s) = 0.0880 + \frac{6.5185}{s^{0.6751}} + 2.5881s^{0.6957} \tag{6.101}$$

$$C_2(s) = 6.9928 + \frac{12.4044}{s^{0.6000}} - 4.1066s^{0.7805} \tag{6.102}$$

$$C_{ZN}(s) = 120.0000 + \frac{300.0000}{s} + 12.0000s \tag{6.103}$$

The step-responses obtained (together with open-loop Bode diagrams and sensitivity and closed-loop functions' gains) are given in Figure 6.15, Figure 6.16 and Figure 6.17. This time, since there is no delay, the plant is easier to control and a wider variation of K is supported by all controllers. But fractional PIDs still achieve an overshoot that is more constant, in spite of the different structure of the plant.

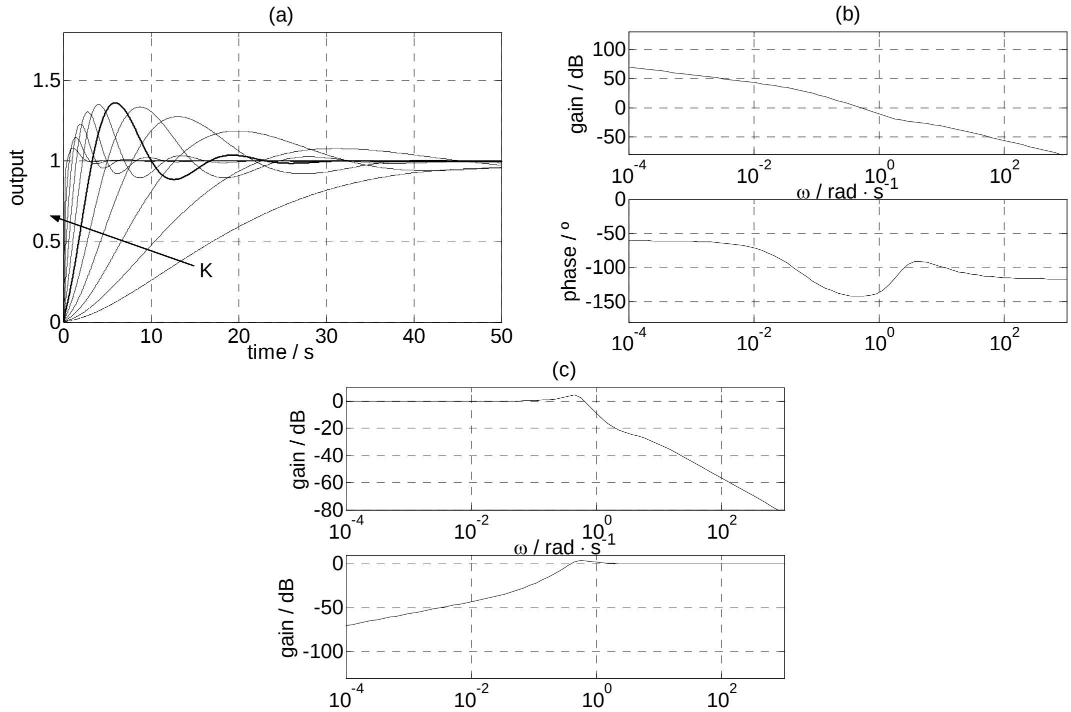


Figure 6.15 — (a) Step response of (6.100) controlled with (6.101) when K is 1/32, 1/16, 1/8, 1/4, 1/2, 1 (thick line), 2, 4, 8, 16 and 32; (b) open-loop Bode diagram when $K = 1$; (c) sensitivity function and closed-loop gains when $K = 1$

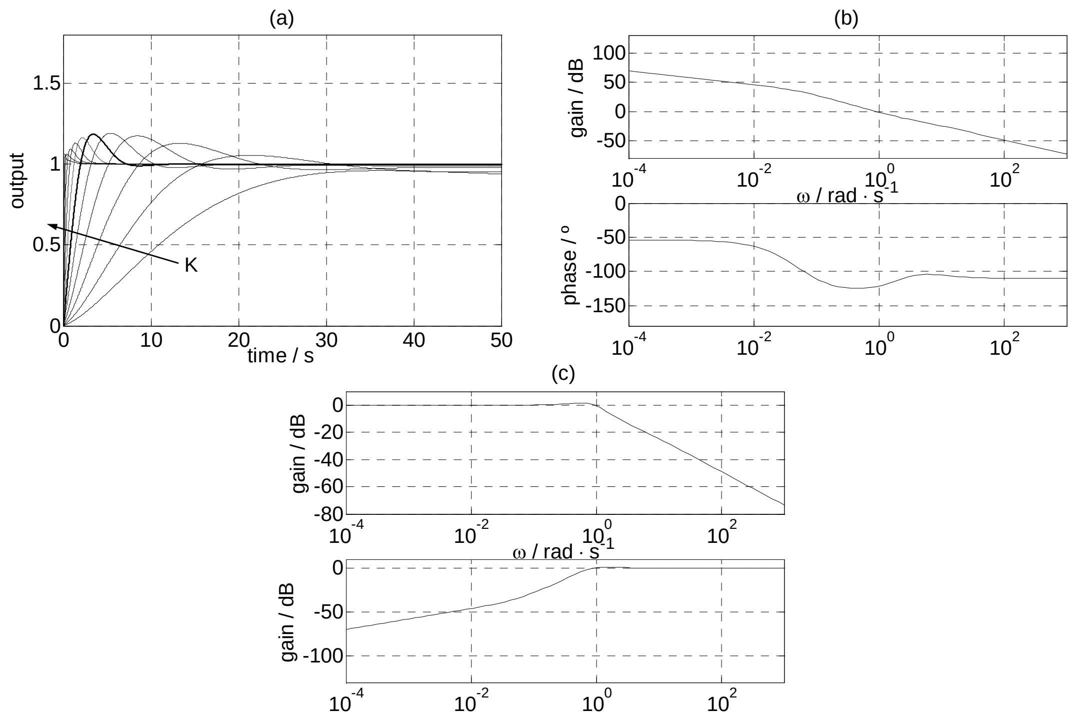


Figure 6.16 — (a) Step response of (6.100) controlled with (6.102) when K is 1/32, 1/16, 1/8, 1/4, 1/2, 1 (thick line), 2, 4, 8, 16 and 32; (b) open-loop Bode diagram when $K = 1$; (c) sensitivity function and closed-loop gains when $K = 1$

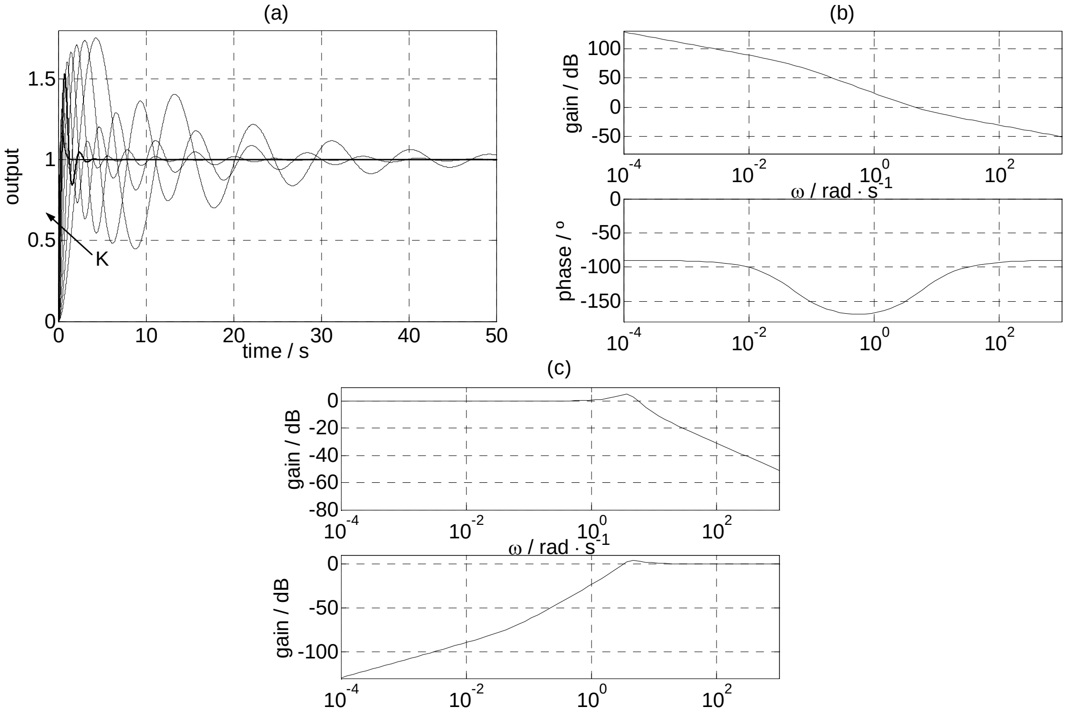


Figure 6.17 — (a) Step response of (6.100) controlled with (6.103) when K is 1/32, 1/16, 1/8, 1/4, 1/2, 1 (thick line), 2, 4, 8, 16 and 32; (b) open-loop Bode diagram when K = 1; (c) sensitivity function and closed-loop gains when K = 1

	C_1 given by (6.101)			C_2 given by (6.102)			C_{ZN} given by (6.103)		
K	rise time	overshoot	settling time	rise time	overshoot	settling time	rise time	overshoot	settling time
1/32	31.4 s	—	45.5 s	21.7 s	—	148.0 s	1.5 s	75 %	41.2 s
1/16	15.6 s	8 %	38.2 s	10.9 s	5 %	23.3 s	1.0 s	74 %	25.9 s
1/8	9.1 s	19 %	47.1 s	6.3 s	13 %	19.2 s	0.7 s	71 %	14.0 s
1/4	5.7 s	28 %	32.1 s	3.8 s	17 %	13.3 s	0.5 s	66 %	8.2 s
1/2	3.8 s	34 %	22.1 s	2.4 s	19 %	9.0 s	0.4 s	61 %	4.6 s
1	2.6 s	36 %	15.3 s	1.5 s	19 %	5.9 s	0.3 s	53 %	2.0 s
2	1.7 s	35 %	10.5 s	0.9 s	16 %	3.8 s	0.2 s	43 %	1.3 s
4	1.2 s	31 %	7.2 s	0.5 s	13 %	2.5 s	0.2 s	31 %	0.8 s
8	0.8 s	23 %	3.3 s	0.3 s	9 %	1.5 s	0.1 s	22 %	0.8 s
16	0.5 s	15 %	2.5 s	0.2 s	6 %	0.7 s	0.1 s	17 %	0.8 s
32	0.2 s	8 %	1.8 s	0.1 s	6 %	0.3 s	0.1 s	15 %	0.8 s

Table 6.8 — Data on step-responses of Figure 6.15, Figure 6.16 and Figure 6.17

6.3.3. Fractional-order plant with delay

The plant considered was

$$F(s) = \frac{K}{1+\sqrt{s}} e^{-0.5s} \approx \frac{K}{1+1.5s} e^{-0.1s} \tag{6.104}$$

with a nominal value of K of 1. The approximation is derived from the plant's step-response at t = 0.92 s (the step response at t = 0.5 s cannot be used since it has an infinite derivative). Controllers obtained with the two tuning rules from subsections 5.4.4 and 5.4.5 and with the first Ziegler-Nichols rule are

$$C_1(s) = 0.6021 + \frac{0.6187}{s^{1.3646}} + 0.3105s^{1.0618} \quad (6.105)$$

$$C_2(s) = 1.4098 + \frac{1.6486}{s^{1.1011}} - 0.2139s^{0.1855} \quad (6.106)$$

$$C_{ZN}(s) = 18.0000 + \frac{90.0000}{s} + 0.9000s \quad (6.107)$$

The step-responses obtained (together with open-loop Bode diagrams and sensitivity and closed-loop functions' gains) are given in Figure 6.18, Figure 6.19 and Figure 6.20. The PID performs poorly because it tries to obtain a fast response and thus employs higher gains, but what is relevant here is that fractional PIDs still achieve practically constant overshoots, since, in spite of the different plant structure, the conditions they were expected to verify are still verified to a reasonable degree, as the frequency-response plots show.

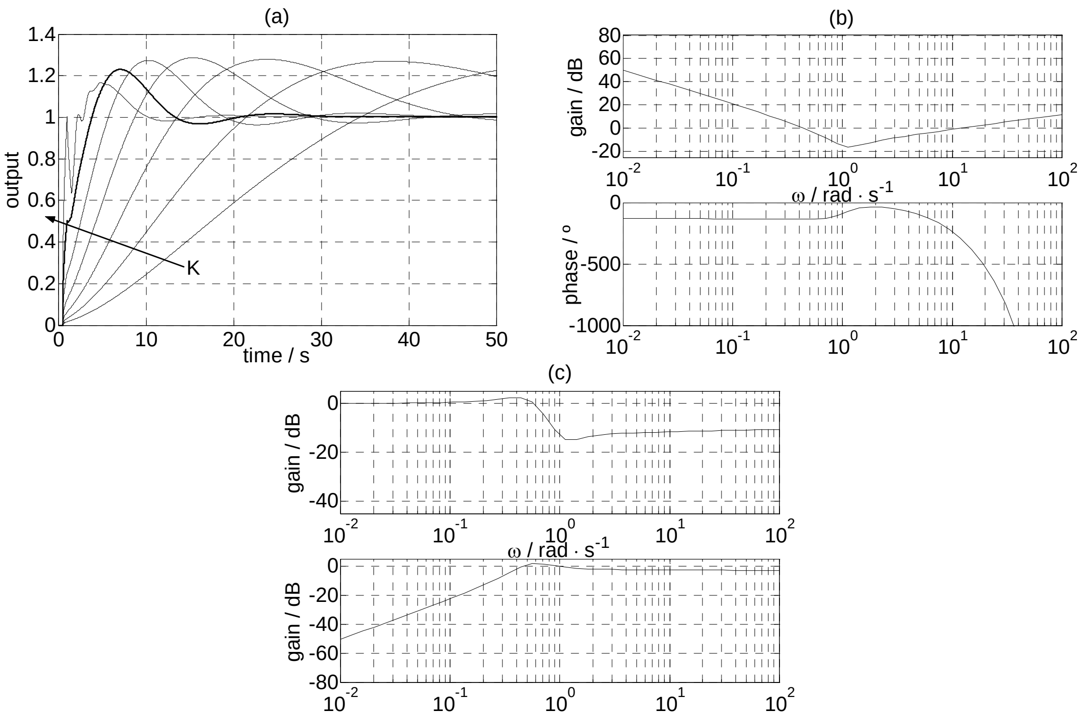


Figure 6.18 — (a) Step response of (6.104) controlled with (6.105) when K is 1/32, 1/16, 1/8, 1/4, 1/2, 1 (thick line) and 2; (b) open-loop Bode diagram when $K = 1$; (c) sensitivity function and closed-loop gains when $K = 1$

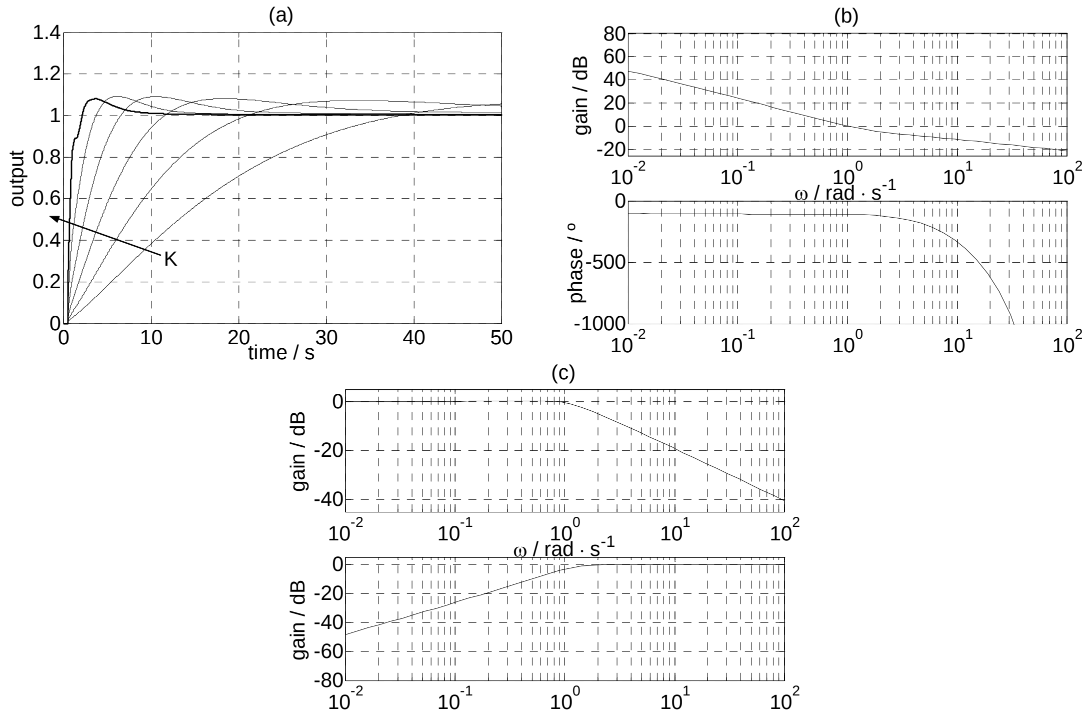


Figure 6.19 — (a) Step response of (6.104) controlled with (6.106) when K is 1/32, 1/16, 1/8, 1/4, 1/2 and 1 (thick line); (b) open-loop Bode diagram when $K = 1$; (c) sensitivity function and closed-loop gains when $K = 1$

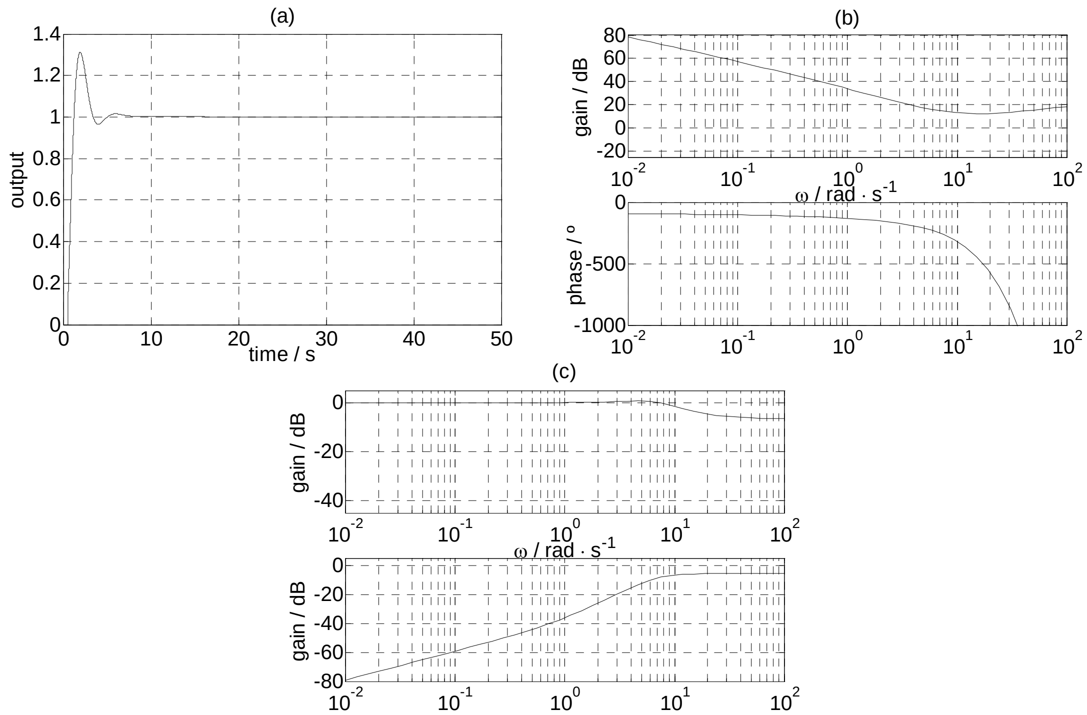


Figure 6.20 — (a) Step response of (6.104) controlled with (6.107) when K is 1/32; (b) open-loop Bode diagram when $K = 1$; (c) sensitivity function and closed-loop gains when $K = 1$

K	C_1 given by (6.105)			C_2 given by (6.106)			C_{ZN} given by (6.107)		
	rise time	overshoot	settling time	rise time	overshoot	settling time	rise time	overshoot	settling time
1/32	25.1 s	26 %	105.5 s	26.5 s	7 %	86.6 s	0.6 s	31 %	3.1 s
1/16	15.8 s	27 %	65.8 s	14.7 s	7 %	47.7 s	—	—	—
1/8	10.3 s	28 %	41.2 s	8.2 s	8 %	27.1 s	—	—	—
1/4	6.9 s	29 %	26.1 s	4.6 s	9 %	16.0 s	—	—	—
1/2	4.4 s	27 %	17.2 s	2.4 s	9 %	9.6 s	—	—	—
1	2.7 s	23 %	11.9 s	1.1 s	8 %	5.6 s	—	—	—
2	1.5 s	17 %	8.6 s	—	—	—	—	—	—

Table 6.9 — Data on step-responses of Figure 6.18, Figure 6.19 and Figure 6.20

Comparing these results with those obtained with IMC-tuned fractional PIDs shows that rule-tuned fractional PIDs perform nearly as well as those found with this analytical method. In this case, by letting $T_B = \frac{1}{2}$, we get, from (5.35), the following controller:

$$C_{IMC}(s) = \frac{1}{4} + \frac{1}{s} + \frac{1}{s^{1/2}} + \frac{1}{4}s \qquad (6.108)$$

It is not clear which of the two integral terms is better to discard; thus it is better to try both cases:

$$C_{IMC1}(s) = \frac{1}{4} + \frac{1}{s} + \frac{1}{4}s \qquad (6.109)$$

$$C_{IMC2}(s) = \frac{1}{4} + \frac{1}{s^{1/2}} + \frac{1}{4}s \qquad (6.110)$$

Step-responses obtained are shown in Figure 6.21 and compare well with those of Figure 6.18 and Figure 6.19.

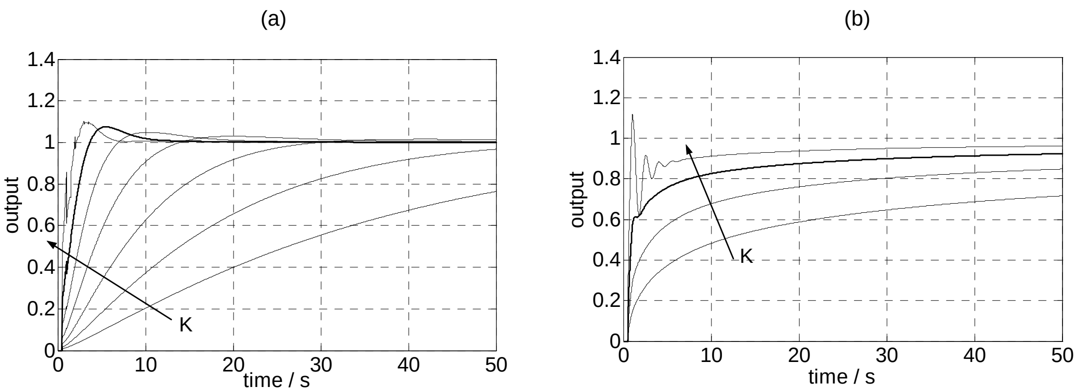


Figure 6.21 — (a) Step response of (6.104) controlled with (6.109) without the fractional integration when K is 1/32, 1/16, 1/8, 1/4, 1/2, 1 (thick line) and 2; (b) step response of (6.104) controlled with (6.110) without the fractional integration when K is 1/4, 1/2, 1 (thick line) and 2

6.3.4. Final comments

A few comments.

Firstly, as stated in subsection 5.4.6, rules usually lead to results poorer than those they were devised to achieve. (The same happens with Ziegler-Nichols rules: they are expected to result in an overshoot around 25 %, but it is not hard to find plants with which the overshoot is 100 % or even more.)

Secondly, Ziegler-Nichols rules make no attempt to reach always the same gain-crossover frequency, or the same phase margin. Actually, these two performance indicators vary widely as L and T vary. This adds some flexibility to Ziegler-Nichols rules: they can be applied for wide ranges of L and T and still achieve a controller that stabilises the plant. Rules from subsections 5.4.4 and 5.4.5 always aim at fulfilling the same specifications, and that is why their application range is never so broad as that of Ziegler-Nichols rules.

Lastly, fractional PIDs tuned with these rules compare well with integer PIDs tuned according to the first Ziegler-Nichols rule, even though the comparison is difficult because Ziegler-Nichols rules achieve different specifications for different values of T and L while rules developed for fractional PIDs attempt to keep always a uniform result. (It is of course likely that carefully tuned integer PIDs perform better than rule-tuned fractional PIDs.)

6.4. Fractional control of a thermal plant

In this section fractional H_2 and H_∞ controllers, implemented as summed up above in subsection 5.5.6, are developed for a fractional plant modelling a thermal system.

6.4.1. The thermal plant

Let

$$G(s) = \frac{1}{39.69 s^{1.26} + 0.598} \tag{6.111}$$

This transfer function describes a thermal system[104] heated by an electrical radiator (the input being a voltage) with the temperature measured by a pyrometer (the output being a voltage as well).

The parameters of (6.111) have been identified by numerically fitting its step response to experimental values. So there are no reasons why the much more tractable commensurate (Q = 4) transfer function

$$G(s) = \frac{1}{39.69 s^{5/4} + 0.598} \tag{6.112}$$

should not be used instead, its step and frequency responses being indistinguishable from those of (6.111). Suppose that we model (white, 0.01 V^2 intensity) noise as affecting both input ($\mathbf{L} = \mathbf{B}$ in Figure 5.5) and output. We want the output to remain unchanged in spite of noise, with the transfer function from w_1 to z_1 smaller than –6 dB over all frequencies and the (not weighted) transfer function from w_2 to z_1 decaying significantly (say, at –40 dB/decade at least) for high frequencies (say, above 1 rad/s,

[104] Vinagre *et al.* (2001); Vinagre (2001, p. 252-253).

given the nature of the plant). The transfer functions mentioned are those without weights, SG_1 and SG_2K.

After some trial and error, the following weights have been selected:

$$W_1(s) = \frac{1}{0.01}\frac{0.1s^{1/4}+10}{s^{1/4}+1.25} \tag{6.113}$$

$$W_2(s) = \frac{1}{0.01}\frac{1429s^{1/4}+5000}{s^{1/4}+1000} \tag{6.114}$$

$$W_3 = 1 \tag{6.115}$$

$$W_4 = 1 \tag{6.116}$$

These weights are fractional-order transfer functions, but integer-order weights might have been used instead. For this particular problem, fractional-order weights allowed attaining the control objectives more easily, but this is not always necessarily so. Integer-order weights have the additional advantage of having frequency responses easier to obtain. Furthermore, even though in this particular case a single set of weights sufficed for both the H_2 and the H_∞ controllers, this is not always necessarily so: different weights might have been necessary.

6.4.2. The genetic algorithm used for parameter tuning

A genetic algorithm was used, as suggested in subsection 5.5.7, to minimise the desired norm. It was as follows:

❖ *Initialisation.* A population with fifty individuals is created. Each individual is a transfer function matrix with a dimension compatible with the dimensions of the plant (in this case, controllers are SISO). The orders of the numerators and the denominators are those of the plant or the ones immediately above or below. Parameters are stored as real numbers.

❖ *Iterations.* The H_2 or H_∞ norm of the matrix transfer function in (5.82) is evaluated for all individuals (in this case, this is a 2×2 matrix). The smaller the norm, the fitter the individual is.

❖ A new population is created with 90 % of the size of the original one. Individuals are selected for this group according to their fitness. These will be the parents in the next step.

❖ *Crossover.* The parents are recombined and replaced by their offspring. In other words, parents are matched in pairs; each pair is replaced by two new individuals, called the offspring; the parameters of each of the offspring are randomly chosen from those of its parents.

❖ *Mutation.* The offspring undergo a mutation. In other words, some of their parameters, randomly chosen, are changed by addition of random values. The mutation probability is such that the average number of mutated parameters per individual is 0.5.

❖ *Elitism.* These mutated descendents replace the less fit individuals in the original population. The 10 % best performing individuals are not replaced. The resulting population is used for a new iteration, beginning with the evaluation of the norms, as explained above.

❖ *Algorithm termination.* Iterations stop after a certain maximum number of iterations (500 in this case) or after a certain maximum number of iteration without improvement in the results (in this case 50 for the H_2 norm and 30 for the H_∞ norm; this last value was smaller because calculations were slower).

The Genetic Algorithm Toolbox for Matlab[105] was used to implement this algorithm.

6.4.3. Results and conclusions

The following H_2 controller was found, with a norm of the matrix in (5.82) equal to 1.7905:

$$K(s)=\frac{5.704}{s^{5/4}+10s+10s^{3/4}+9.999^{1/2}+9.422s^{1/4}-5.399} \tag{6.117}$$

The following H_∞ controller was found, resulting in a norm of the matrix in (5.82) equal to 6.7115:

$$K(s)=\frac{7.448}{s^{5/4}+9.383s+8.642s^{3/4}-2.316s^{1/2}+9.227s^{1/4}-4.736} \tag{6.118}$$

The relevant Bode and singular value plots for both cases are found in Figure 6.22, Figure 6.23 and Figure 6.24.

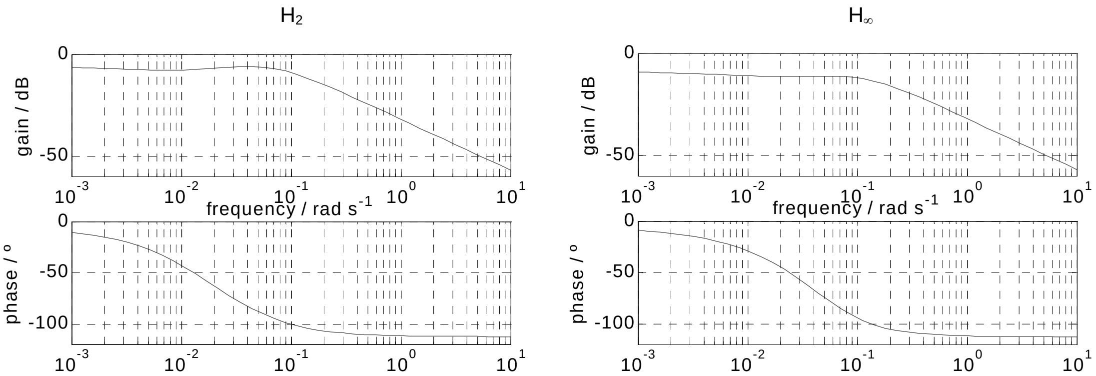


Figure 6.22 — Bode diagrams of SG_1

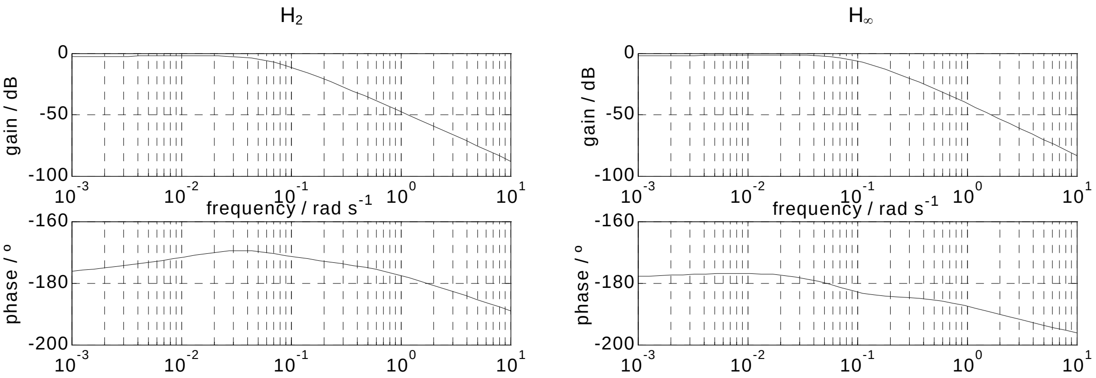


Figure 6.23 — Bode diagrams of SG_2K

[105] Chipperfield *et al.* (1994).

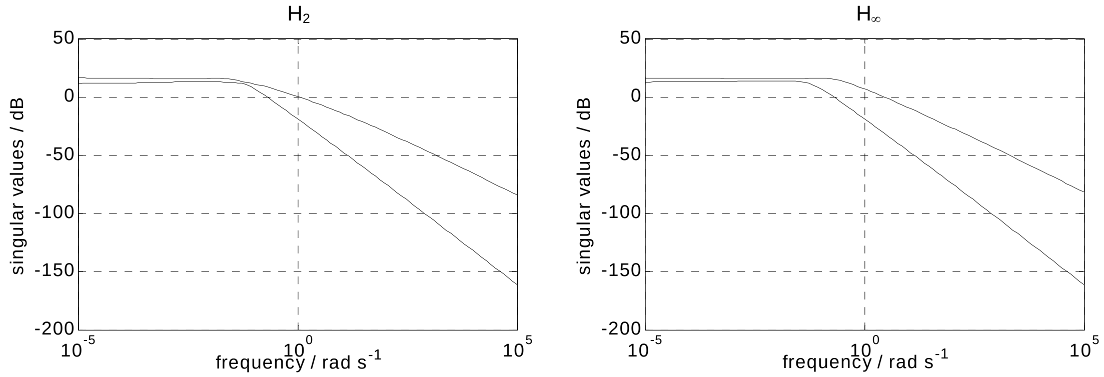


Figure 6.24 — Singular values of loop transfer function in (5.82)

For both controllers, the magnitude of SG_1 is always below –6 dB (actually its maximum is –6.05 dB for the H_2 controller and –7.13 dB for the H_∞ controller). SG_2K decays with –45 dB/decade with both controllers. This means the objectives were attained in both cases. This is also shown by simulations of the resulting control loops. Changes in plant parameters are also handled by the resulting controllers.

Other values for weights W_1 to W_4 allow obtaining different results and shaping the loop in other ways. Thus it would be possible to cope with different performance specifications.

Two important questions arise from the use of a numerical minimisation method. The first is the time needed to reach a solution. For the H_2 case of the problem above, this means about 1 minute and 38 seconds of computation per iteration are needed (in a Pentium IV @ 2.53 GHz); for the H_∞ case, this may mean up to about 2 minutes and 50 seconds per iteration (in the same machine), but depends on how close is the mesh of frequencies used to estimate the norm. These are bearable values, though faster results would, of course, be desirable.

Still concerning this point, it is worth noticing that an increase in the dimension of the matrix in (5.82) reflects very heavily on the computational effort needed. Dimensions above 4 often prevent the numerical algorithm from reaching a solution.

The second question concerns how far the optimisation went when it stops. The best validation possible is to use the algorithm for integer plants, for which there is an analytical solution available, and then compare both results. This was tried for several cases, and the numerical method always got very close to the analytical result. Assuming the same to happen with the fractional case, it seems that further possible reductions in norms are not very relevant.

7. Conclusions

> »Die Wasser fragen dich«, verkündete Fuchur, »ob du alle Geschichten, die du in Phantásien begonnen hast, auch zu Ende geführt hast.«
> »Nein«, sagte Bastian, »eigentlich keine.«
> Fuchur horchte eine Weile. Sein Gesicht nahm einen bestürzten Ausdruck an.
> »Sie sagen, dann wird dich die weiße Schlange nicht durchlassen. Du musst zurück nach Phantásien und alles zu Ende bringen.«
> »Alle Geschichten?«, stammelte Bastian. »Dann kann ich nie mehr zurück. Dann war alles umsonst.«
> Fuchur lauschte gespannt.
> »Was sagen sie?«, wollte Bastian wissen.
> »Still!«, sagte Fuchur.
> Nach einer Weile seufzte er und erklärte:
> »Sie sagen, es ist nicht zu ändern, es sei denn, es fände sich jemand, der diese Aufgabe für dich übernimmt.«
> »Aber es sind unzählige Geschichten«, rief Bastian, »und aus jeder kommen immer neue. So eine Aufgabe kann niemand übernehmen.«
>
> Michael Ende, *Die unendliche Geschichte*, XXVI

As claimed in section 1.1, the work reported in this thesis validates, enlarges and applies knowledge on fractional order control. Two main conclusions may be therefrom drawn.

The first is not original, though it is expected that it be further supported by all the above: fractional calculus is a useful tool for control. Many plants can be accurately described by fractional models, and, while in those cases integer models could usually be used as well, their complexity would have to be significant, or their performance poor. Fractional controllers achieve good performances, both for integer and fractional plants. They often achieve a significant degree of robustness.

This is of course not to say that fractional controllers are *always* better. First of all, what one classifies as *better* depends on the case. There may be circumstances when a fractional controller will achieve the best performance, but will also be too complex; some simpler integer controller having a poorer but acceptable performance will then certainly be preferred.

Even not taking that into account, things are also not clear in what performance is concerned. Sometimes fractional controllers are appropriate and perform adequately, sometimes not. In all previous chapters, it was tried to assess the cases when different methods may be applied. Whenever they are, it is necessary to check if fractional controllers really perform adequately. If they do, suitable methods of finding integer controllers should be also experimented, and the results compared, so as to choose among all possibilities one well-performing enough and simple.

The second conclusion is that there is still much work to do in this area. Again, this is no new conclusion, though it is expected that the above improves on what was known (as stated in section 1.4).

In what concerns the theory of fractional calculus, a consensual geometrical interpretation of what fractional derivatives stand for is still missing.

In what concerns the implementation of fractional plants, it does not seem that additional formulas be wanted; actually those existing are already very numerous. It is not impossible that new approximations will do better, but that is unlikely; a reasonable future work, however, may be to shed further light on relative merits and demerits of the various approximations, so as to make conditions when each is to be preferred even more clear-cut.

In what concerns identification methods for finding fractional models, future work should include adapting integer identification methods known to be less sensitive to noise, and of stochastic nature, such as, for instance, those proposed by Schoukens *et al.* (1988) or those proposed by Helmicki *et al.* (1991). Just as there are several identification methods for integer orders, so several for fractional orders may be expected to coexist, being necessary to know when each may be applied and to use a few to retain the best result obtained.

Synthesis of fractional controllers is certainly the field where more innovations are possible and welcome. This is especially so for MIMO plants. Of all methods presented, fractional PIDs and H_2 and H_∞ controllers are not confined to SISO plants, but their tuning is, to a great extent, still based upon the brute force of numerical optimisation methods. The development of more refined analytical methods for both these types of controllers would be a significant advance. In what concerns H_2 and H_∞ controllers, this probably involves the development of methods for reckoning these norms more similar to those in use for integer plants—based upon state-space representations. In what concerns fractional PIDs, another possible improvement would be the development of additional tuning rules. Rules similar to the second Ziegler-Nichols rule (making use of a closed-loop response of the plant) are certainly possible. Rules specific for non-minimum phase plants may also be of interest.

In what concerns applications, those shown certainly admit improvement. The most relevant is that of the flexible robot, both because the plant exists in laboratory and because of the importance flexible robots are expected to assume soon. Future work in that plant includes improving the model of the robot to obtain an increased accuracy of simulations, applying the method to more complicated robots with more links, and refining the parameter tuning. Other applications of fractional control are of course also possible; for instance, active noise control is expected to become one of them.

Appendix A. Mathematical issues

> He said that new systems of nature were but new fashions, which would vary in every age; and even those who pretend to demonstrate them from mathematical principles, would flourish but a short period of time, and be out of vogue when that was determined.
>
> Jonathan Swift, *Gulliver's Travels* III, 8

Some non-trivial mathematical details needed for several issues addressed by the thesis are summed up in this appendix so as not to overload the main text.

A.1. Complex calculus

This section generalises usual formulas of real calculus for complex numbers.

A.1.1. Trigonometric functions

Since

$$\cos x = \sum_{n=0}^{+\infty} \frac{(-1)^n x^{2n}}{(2n)!} \tag{A.1}$$

we will have, for an imaginary argument,

$$\cos jx = \sum_{n=0}^{+\infty} \frac{(-1)^n j^{2n} x^{2n}}{(2n)!} = \sum_{n=0}^{+\infty} \frac{(-1)^n (-1)^n x^{2n}}{(2n)!} = \sum_{n=0}^{+\infty} \frac{x^{2n}}{(2n)!} \tag{A.2}$$

This last series is the definition of the hyperbolic cosine; thus

$$\boxed{\cos jx = \cosh x} \tag{A.3}$$

Similarly, since

$$\sin x = \sum_{n=0}^{+\infty} \frac{(-1)^n x^{2n+1}}{(2n+1)!} \tag{A.4}$$

we will have, for an imaginary argument,

$$\sin jx = \sum_{n=0}^{+\infty} \frac{(-1)^n j^{2n+1} x^{2n+1}}{(2n+1)!} = \sum_{n=0}^{+\infty} \frac{(-1)^n j (-1)^n x^{2n+1}}{(2n+1)!} = j\sum_{n=0}^{+\infty} \frac{x^{2n+1}}{(2n+1)!} \tag{A.5}$$

This last series is the definition of the hyperbolic sine; thus

$$\sin jx = j \sinh x \qquad \text{(A.6)}$$

A.1.2. Exponential, logarithmic and power functions

Since

$$e^x = \sum_{n=0}^{+\infty} \frac{x^n}{n!} \qquad \text{(A.7)}$$

we will have, for an imaginary argument,

$$e^{jx} = \sum_{n=0}^{+\infty} \frac{j^n x^n}{n!} \qquad \text{(A.8)}$$

It is now expedient to separate this summation into its odd and even terms:

$$e^{jx} = \sum_{n=0}^{+\infty} \frac{j^{2n} x^{2n}}{(2n)!} + \sum_{n=0}^{+\infty} \frac{j^{2n+1} x^{2n+1}}{(2n+1)!} = \sum_{n=0}^{+\infty} \frac{(-1)^n x^{2n}}{(2n)!} + \sum_{n=0}^{+\infty} \frac{j(-1)^n x^{2n+1}}{(2n+1)!} \qquad \text{(A.9)}$$

These series are those of (A.1) and (A.4). Thus

$$e^{jx} = \cos x + j \sin x \qquad \text{(A.10)}$$

Let us reckon the complex number $a+jb$ such that

$$\log j = a + jb \qquad \text{(A.11)}$$

Taking the exponential of this equality, we will have

$$j = e^{a+jb} = e^a e^{jb} = e^a (\cos b + j \sin b) = e^a \cos b + j e^a \sin b \qquad \text{(A.12)}$$

For the equality to hold, the real part of this last complex must verify

$$e^a \cos b = 0 \Rightarrow \cos b = 0 \Leftrightarrow b = \frac{\pi}{2} + k\pi, \quad k \in \mathbb{Z} \qquad \text{(A.13)}$$

and hence the imaginary part must verify

$$\begin{aligned} & e^a \sin b = 1 \Leftrightarrow e^a \sin\left(\frac{\pi}{2} + k\pi\right) = 1 \Leftrightarrow \\ & \Leftrightarrow e^a \sin\left(\frac{\pi}{2} + 2k\pi\right) = 1 \vee e^a \sin\left(-\frac{\pi}{2} + 2k\pi\right) = 1, \quad k \in \mathbb{Z} \end{aligned} \qquad \text{(A.14)}$$

The last condition is impossible since e^a is always positive while the sinus will be

–1, and so we are left with

$$e^a \sin\left(\frac{\pi}{2} + 2k\pi\right) = 1 \Leftrightarrow e^a = 1 \Leftrightarrow a = 0, \quad k \in \mathbb{Z} \tag{A.15}$$

Putting this together gives

$$\log j = j\frac{\pi}{2} + j2k\pi, \quad k \in \mathbb{Z} \tag{A.16}$$

The branch of this multi-valued function that results in $\log j = j\frac{\pi}{2}$ is the one more often used.

Complex powers are handled thus:

$$k^{a+jb} = e^{\log\left(k^{a+jb}\right)} = e^{(a+jb)\log k} = e^{a\log k + jb\log k} = e^{\log k^a} e^{jb\log k} \tag{A.17}$$

Using (A.10) this becomes

$$k^{a+jb} = k^a\left[\cos\left(b\log k\right) + j\sin\left(b\log k\right)\right] \tag{A.18}$$

A.1.3. Important derivatives

Let x and b be complex-valued column vectors of dimension n. Then

$$\begin{aligned} b^T x &= \sum_{k=1}^{n} b_k x_k = \\ &= \sum_{k=1}^{n} \left(\mathrm{Re}[b_k]\mathrm{Re}[x_k] - \mathrm{Im}[b_k]\mathrm{Im}[x_k]\right) + j\left(\mathrm{Re}[b_k]\mathrm{Im}[x_k] + \mathrm{Im}[b_k]\mathrm{Re}[x_k]\right) \end{aligned} \tag{A.19}$$

Our purpose is to reckon

$$\frac{\partial b^T x}{\partial x} = \begin{bmatrix} \frac{\partial b^T x}{\partial x_1} \\ \frac{\partial b^T x}{\partial x_2} \\ \vdots \\ \frac{\partial b^T x}{\partial x_n} \end{bmatrix} \tag{A.20}$$

So as to use the Cauchy-Riemann conditions[106] it is necessary to find

[106] These say that for a complex-valued function $f(z)$ to have a derivative it is necessary that

$$\frac{\partial \operatorname{Re}\left[b^T x\right]}{\partial \operatorname{Re}[x]} = \begin{bmatrix} \operatorname{Re}[b_1] \\ \operatorname{Re}[b_2] \\ \vdots \\ \operatorname{Re}[b_n] \end{bmatrix} = \operatorname{Re}[b] \tag{A.21}$$

$$\frac{\partial \operatorname{Re}\left[b^T x\right]}{\partial \operatorname{Im}[x]} = \begin{bmatrix} -\operatorname{Im}[b_1] \\ -\operatorname{Im}[b_2] \\ \vdots \\ -\operatorname{Im}[b_n] \end{bmatrix} = -\operatorname{Im}[b] \tag{A.22}$$

$$\frac{\partial \operatorname{Im}\left[b^T x\right]}{\partial \operatorname{Re}[x]} = \begin{bmatrix} \operatorname{Im}[b_1] \\ \operatorname{Im}[b_2] \\ \vdots \\ \operatorname{Im}[b_n] \end{bmatrix} = \operatorname{Im}[b] \tag{A.23}$$

$$\frac{\partial \operatorname{Im}\left[b^T x\right]}{\partial \operatorname{Im}[x]} = \begin{bmatrix} \operatorname{Re}[b_1] \\ \operatorname{Re}[b_2] \\ \vdots \\ \operatorname{Re}[b_n] \end{bmatrix} = \operatorname{Re}[b] \tag{A.24}$$

The Riemann-Cauchy conditions being verified, it is possible to write

$$\boxed{\frac{\partial b^T x}{\partial x} = \operatorname{Re}[b] + j \operatorname{Im}[b] = b} \tag{A.25}$$

And since

$$x^T b = \sum_{k=1}^{n} b_k x_k = b^T x \tag{A.26}$$

$$\begin{cases} \dfrac{\partial \operatorname{Re}[f(z)]}{\partial \operatorname{Re}[z]} = \dfrac{\partial \operatorname{Im}[f(z)]}{\partial \operatorname{Im}[z]} \\ -\dfrac{\partial \operatorname{Re}[f(z)]}{\partial \operatorname{Im}[z]} = \dfrac{\partial \operatorname{Im}[f(z)]}{\partial \operatorname{Re}[z]} \end{cases}$$

If the above partial derivatives are continuous in a neighbourhood of z, the derivative is given by

$$\begin{bmatrix} \dfrac{\partial \operatorname{Re}[f(z)]}{\partial \operatorname{Re}[z]} \\ \dfrac{\partial \operatorname{Im}[f(z)]}{\partial \operatorname{Re}[z]} \end{bmatrix} \equiv \begin{bmatrix} \dfrac{\partial \operatorname{Im}[f(z)]}{\partial \operatorname{Im}[z]} \\ -\dfrac{\partial \operatorname{Re}[f(z)]}{\partial \operatorname{Im}[z]} \end{bmatrix}$$

See Apostol (1974, p. 118-120).

it will be also

$$\frac{\partial x^T b}{\partial x} = b \tag{A.27}$$

Let us now assume that x is real and reckon

$$\begin{aligned}
&\frac{\partial}{\partial x} x^T b\overline{b^T} x = \frac{\partial}{\partial x}\left(\sum_{k=1}^{n} b_k x_k\right)\overline{b^T} x = \\
&= \frac{\partial}{\partial x}\begin{bmatrix} \sum_{k=1}^{n} x_k b_k \overline{b_1} & \sum_{k=1}^{n} x_k b_k \overline{b_2} & \cdots & \sum_{k=1}^{n} x_k b_k \overline{b_n} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \frac{\partial}{\partial x}\sum_{p=1}^{n}\sum_{k=1}^{n} x_k b_k \overline{b_p} x_p = \\
&= \begin{bmatrix} 2x_1 b_1 \overline{b_1} + \sum_{k=2}^{n} b_1 \overline{b_k} x_k + \sum_{p=2}^{n} x_p b_p \overline{b_1} \\ 2x_2 b_2 \overline{b_2} + \sum_{\substack{k=1 \\ k\neq 2}}^{n} b_2 \overline{b_k} x_k + \sum_{\substack{p=1 \\ p\neq 2}}^{n} x_p b_p \overline{b_2} \\ \vdots \\ 2x_n b_n \overline{b_n} + \sum_{k=1}^{n-1} b_n \overline{b_k} x_k + \sum_{p=1}^{n-1} x_p b_p \overline{b_n} \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^{n} b_1 \overline{b_k} x_k \\ \sum_{k=1}^{n} b_2 \overline{b_k} x_k \\ \vdots \\ \sum_{k=1}^{n} b_n \overline{b_k} x_k \end{bmatrix} + \begin{bmatrix} \sum_{p=1}^{n} x_p b_p \overline{b_1} \\ \sum_{p=1}^{n} x_p b_p \overline{b_2} \\ \vdots \\ \sum_{p=1}^{n} x_p b_p \overline{b_n} \end{bmatrix} = \\
&= b\overline{b^T} x + \left(x^T b\overline{b^T}\right)^T = b\overline{b^T} x + \overline{b} b^T x
\end{aligned} \tag{A.28}$$

A.2. Transcendental functions

This section deals with some transcendental functions needed for fractional calculus.

A.2.1. Function Γ

For a positive x, function Γ (gamma function[107]) is defined as

$$\Gamma(x) \overset{\text{def}}{=} \int_0^{+\infty} e^{-y} y^{x-1} dy \tag{A.29}$$

Thus

$$\Gamma(1) = \int_0^{+\infty} e^{-y} dy = \left[-e^{-y}\right]_0^{+\infty} = 0 - (-1) = 1 \tag{A.30}$$

[107] Hildebrand (1976, p. 77-78); Wider (1947, p. 303-306).

$$\Gamma(x+1)=\int_0^{+\infty}e^{-y}y^x dy=\left[-e^{-y}y^x\right]_{y=0}^{y=+\infty}-\int_0^{+\infty}-e^{-y}xy^{x-1}dy=$$
$$=-0+0+x\int_0^{+\infty}e^{-y}y^{x-1}dy=x\Gamma(x) \tag{A.31}$$

Those two properties show that

$$\Gamma(n+1)=n!,\ n\in\mathbb{N} \tag{A.32}$$

The second property above, (A.31), may be used recursively:

$$\Gamma(x)=\frac{\Gamma(x+n)}{x(x+1)\dots(x+n-1)}=\frac{\Gamma(x+n)}{\prod_{i=0}^{n-1}(x+i)},\quad n\in\mathbb{N} \tag{A.33}$$

This allows expanding the definition (A.29) of function Γ for $x\in\mathbb{R}^-\setminus\mathbb{Z}$ as follows:

$$\Gamma(x)\overset{\text{def}}{=}\begin{cases}\int_0^{+\infty}e^{-y}y^{x-1}dy, \text{ if } x\in\mathbb{R}^+\\ \dfrac{\Gamma(x+n)}{\prod_{i=0}^{n-1}(x+i)},\ n=-\mathscr{C}(x), \text{ if } x\in\mathbb{R}^-\setminus\mathbb{Z}\end{cases} \tag{A.34}$$

The expression above shows that function Γ cannot be continuously generalised for $x\in\mathbb{Z}_0^-$ since the ratio in the second branch of (A.34) diverges to infinity in such points.

Theorem A. 1: For any $n\in\mathbb{Z}$,

$$\Gamma(x)\Gamma(-x+1)=(-1)^n\Gamma(-x-n+1)\Gamma(x+n) \tag{A.35}$$

Proof: This result is obtained for positive and negative integers separately, both times by mathematical induction. If $n=0$, the equality is obvious.

Using property (A.31) twice, it can be established that (A.35) holds for $n=1$:

$$\Gamma(x)\Gamma(-x+1)=\frac{\Gamma(x+1)}{x}\left[-x\Gamma(-x)\right]=-\Gamma(-x)\Gamma(x+1) \tag{A.36}$$

The inductive step is proved applying (A.36) to (A.35):

$$\Gamma(x)\Gamma(-x+1)=(-1)^{-n}\Gamma(-x-n+1)\Gamma(x+n)=$$
$$=(-1)^n\left[-\Gamma(-x-n)\Gamma(x+n+1)\right]=(-1)^{n+1}\Gamma(-x-(n+1)+1)\Gamma(x+(n+1)) \tag{A.37}$$

Similarly, it can be established that (A.35) holds for $n=-1$:

$$\Gamma(x)\Gamma(-x+1)=(x-1)\Gamma(x-1)\frac{\Gamma(-x+2)}{-x+1}=-\Gamma(-x+2)\Gamma(x-1) \tag{A.38}$$

The inductive step is proved applying (A.38) to (A.35):

$$\begin{aligned}&\Gamma(x)\Gamma(-x+1)=(-1)^{-n}\Gamma(-x-n+1)\Gamma(x+n)=\\&=(-1)^{n}\left[-\Gamma(-x-n+2)\Gamma(x+n-1)\right]=\\&=(-1)^{n-1}\Gamma\big(-x-(n-1)+1\big)\Gamma\big(x+(n-1)\big)\end{aligned} \tag{A.39}$$

Q.E.D.

Corollary: From (A.33) and (A.35) we obtain

$$\prod_{i=0}^{n-1}(x+i)=\frac{\Gamma(x+n)}{\Gamma(x)}=\frac{\Gamma(-x+1)}{\Gamma(-x-n+1)}(-1)^{n} \tag{A.40}$$

Remark: Thanks to (A.18), definition (A.29) may still be used for complex arguments:

$$\begin{aligned}&\Gamma(a+jb)=\int_0^{+\infty}e^{-y}y^{a+jb-1}dy=\int_0^{+\infty}e^{-y}y^{a-1}\left[\cos(b\log y)+j\sin(b\log y)\right]dy=\\&=\int_0^{+\infty}e^{-y}y^{a-1}\cos(b\log y)\,dy+j\int_0^{+\infty}e^{-y}y^{a-1}\sin(b\log y)\,dy\end{aligned} \tag{A.41}$$

A.2.2. Combinations

Function Γ allows defining combinations as

$$\binom{a}{b}\overset{\text{def}}{=}\frac{a!}{b!(a-b)!}=\frac{\Gamma(a+1)}{\Gamma(b+1)\Gamma(a-b+1)} \tag{A.42}$$

This formula holds for all $a,b\in\mathbb{R}\setminus\mathbb{Z}^{-}$. It cannot be used when negative integers appear since Γ would have to be reckoned for values where it is not defined. But we can use (A.35) in (A.42) and make

$$\binom{a}{b}\overset{\text{def}}{=}\frac{(-1)^{b}\Gamma(b-a)}{\Gamma(b+1)\Gamma(-a)} \tag{A.43}$$

which still makes sense if $a\in\mathbb{Z}^{-}\wedge b\in\mathbb{N}\wedge b-a>0$.

Putting (A.42) and (A.43) together,

$$\binom{a}{b} \overset{\text{def}}{=} \begin{cases} \dfrac{\Gamma(a+1)}{\Gamma(b+1)\Gamma(a-b+1)}, \text{ if } a,b \in \mathbb{R}\setminus\mathbb{Z}^- \\ \dfrac{(-1)^b \Gamma(b-a)}{\Gamma(b+1)\Gamma(-a)}, \text{ if } a \in \mathbb{Z}^- \wedge b \in \mathbb{N} \wedge b-a>0 \end{cases} \tag{A.44}$$

A.2.3. Functions $_pF_q$

Hypergeometric functions F are defined[108] by means of series that depend not only on variable x but also on p coefficients $a_1, a_2, \ldots, a_p$ and on q coefficients $b_1, b_2, \ldots, b_q$:

$$_pF_q\left(a_1,a_2,\ldots,a_p;b_1,b_2,\ldots,b_q;x\right) \overset{\text{def}}{=} 1+\sum_{k=1}^{+\infty}\left[\frac{x^k}{k!}\prod_{n=0}^{k-1}\frac{(a_1+n)(a_2+n)\ldots(a_p+n)}{(b_1+n)(b_2+n)\ldots(b_q+n)}\right] \tag{A.45}$$

Applying (A.40) we can express $_pF_q$ alternatively as

$$_pF_q\left(a_1,a_2,\ldots,a_p;b_1,b_2,\ldots,b_q;x\right) = 1+\sum_{k=1}^{+\infty}\left[\frac{x^k}{k!}\frac{\frac{\Gamma(a_1+k)}{\Gamma(a_1)}\frac{\Gamma(a_2+k)}{\Gamma(a_2)}\cdots\frac{\Gamma(a_p+k)}{\Gamma(a_p)}}{\frac{\Gamma(b_1+k)}{\Gamma(b_1)}\frac{\Gamma(b_2+k)}{\Gamma(b_2)}\cdots\frac{\Gamma(b_q+k)}{\Gamma(b_q)}}\right]=$$

$$=\sum_{k=0}^{+\infty}\left[\frac{x^k}{k!}\frac{\Gamma(a_1+k)\Gamma(a_2+k)\ldots\Gamma(a_p+k)\Gamma(b_1)\Gamma(b_2)\ldots\Gamma(b_q)}{\Gamma(a_1)\Gamma(a_2)\ldots\Gamma(a_p)\Gamma(b_1+k)\Gamma(b_2+k)\ldots\Gamma(b_q+k)}\right]=$$

$$=\frac{\Gamma(b_1)\Gamma(b_2)\ldots\Gamma(b_q)}{\Gamma(a_1)\Gamma(a_2)\ldots\Gamma(a_p)}\sum_{k=0}^{+\infty}\frac{x^k}{k!}\frac{\Gamma(a_1+k)\Gamma(a_2+k)\ldots\Gamma(a_p+k)}{\Gamma(b_1+k)\Gamma(b_2+k)\ldots\Gamma(b_q+k)} \tag{A.46}$$

Indexes p and q may be omitted when it is clear what values they have. In what follows they will be omitted for the particular case

$$_2F_1(a,b;c;x) \equiv F(a,b;c;x) \overset{\text{def}}{=}$$

$$=1+\frac{ab}{c}x+\frac{a(a+1)b(b+1)}{c(c+1)}\frac{x^2}{2!}+\frac{a(a+1)(a+2)b(b+1)(b+2)}{c(c+1)(c+2)}\frac{x^3}{3!}+\ldots=$$

$$=1+\sum_{n=1}^{+\infty}\frac{\prod_{i=0}^{n-1}(a+i)\prod_{j=0}^{n-1}(b+j)}{\prod_{k=0}^{n-1}(c+k)}\frac{x^n}{n!}=\frac{\Gamma(c)}{\Gamma(a)\Gamma(b)}\sum_{n=0}^{+\infty}\frac{\Gamma(a+n)\Gamma(b+n)}{\Gamma(c+n)}\frac{x^n}{n!} \tag{A.47}$$

which, without further qualification, is *the* hypergeometric function.

108 Wall (1948, p. 335); Miller *et al.* (1993, p. 303-304).

Another particular case of interest for what follows is

$$ {}_1F_1(a;c;x) = \lim_{b\to+\infty} F(a,b;c;x) \overset{\text{def}}{=} $$

$$ \overset{\text{def}}{=} \frac{\Gamma(c)}{\Gamma(a)} \sum_{n=0}^{+\infty} \frac{\Gamma(a+n)}{n!\Gamma(c+n)} x^n = 1 + \sum_{n=1}^{+\infty} \frac{\prod_{i=0}^{n-1}(a+i)}{\prod_{k=0}^{n-1}(c+k)} \frac{x^n}{n!} \tag{A.48} $$

A.2.4. Functions γ, P, Q, Γ and γ^*

There are five functions called incomplete gamma function[109]:

$$ \gamma(x,z) = \int_0^z e^{-y} y^{x-1} dy, \quad x > 0 \tag{A.49} $$

$$ P(x,z) = \frac{1}{\Gamma(x)} \int_0^z e^{-y} y^{x-1} dy, \quad x > 0 \tag{A.50} $$

$$ Q(x,z) = 1 - \frac{1}{\Gamma(x)} \int_0^z e^{-y} y^{x-1} dy, \quad x > 0 \tag{A.51} $$

$$ \Gamma(x,z) = \int_z^{+\infty} e^{-y} y^{x-1} dy \tag{A.52} $$

$$ \gamma^*(x,z) = \frac{z^{-x}}{\Gamma(x)} \int_0^z e^{-y} y^{x-1} dy \tag{A.53} $$

Obviously,

$$ \Gamma(x) = \lim_{z\to+\infty} \gamma(x,z) \tag{A.54} $$

$$ \gamma(x,z) = \Gamma(x) P(x,z), \quad x > 0 \tag{A.55} $$

$$ \gamma(x,z) = \Gamma(x) - \Gamma(x,z), \quad x > 0 \tag{A.56} $$

$$ \gamma(x,z) = \gamma^*(x,z) \Gamma(x) z^x, \quad x > 0 \tag{A.57} $$

$$ \gamma(x,z) = \Gamma(x) \left[1 - Q(x,z)\right], \quad x > 0 \tag{A.58} $$

The following result is quoted without proof[110]:

$$ \gamma^*(x,z) = e^{-z} \sum_{n=0}^{+\infty} \frac{z^n}{\Gamma(x+n+1)} \tag{A.59} $$

---

[109] This makes the subject somewhat confusing, since it is necessary to check which of the five is meant when «incomplete gamma function» is mentioned. See Hildebrand (p. 675, 1976) and Abramowitz *et al.* (p. 260, 1964) about $\gamma(x,z)$; see Press *et al.* (p. 216, 1992) and Abramowitz *et al.* (p. 260, 1964) about $P(x,z)$; see Press *et al.* (p. 216, 1992) about $Q(x,z)$; see Abramowitz *et al.* (p. 260, 1964) about $\Gamma(x,z)$; and see Abramowitz *et al.* (p. 260, 1964) and Miller *et al.* (p. 300, 1993) about $\gamma^*(x,z)$.

[110] Abramowitz *et al.* (1964, p. 262); Miller *et al.* (1993, p. 309); Press *et al.* (1992, p. 217).

Theorem A. 2: The incomplete gamma function and the hypergeometric function $_1F_1$ are related by[111]

$$\gamma^*(x,z)=\frac{e^{-z}}{\Gamma(x+1)}\,{}_1F_1(1;x+1;z) \tag{A.60}$$

Proof: The right-hand member of (A.60) is equal to

$$\frac{e^{-z}}{\Gamma(x+1)}\Gamma(x+1)\sum_{n=0}^{+\infty}\frac{\Gamma(1+n)}{\Gamma(1+n)\Gamma(x+1+n)}x^n=e^{-z}\sum_{n=0}^{+\infty}\frac{x^n}{\Gamma(x+1+n)} \tag{A.61}$$

and this is the right-hand side of (A.59). **Q.E.D.**

A.2.5. Function E_t

Function E_t (Mittag-Leffler function) is defined as[112]

$$E_t(\nu,a)=t^\nu e^{at}\gamma^*(\nu,at) \tag{A.62}$$

Theorem A. 3: When one of the parameters is zero E_t reduces to

$$E_t(0,a)=e^{at} \tag{A.63}$$

$$E_t(\nu,0)=\frac{t^\nu}{\Gamma(\nu+1)} \tag{A.64}$$

Proof: Using (A.60) and (A.48), we get

$$E_t(0,a)=e^{at}\gamma^*(0,at)=e^{at}\frac{e^{-at}}{\Gamma(0)}\,{}_1F_1(1;1;at)=\sum_{n=0}^{+\infty}\frac{\Gamma(n+1)}{n!\Gamma(n+1)}(at)^n \tag{A.65}$$

which, by (A.7), is the right-hand side of (A.63). We also get

$$E_t(\nu,0)=t^\nu\gamma^*(\nu,0)=t^\nu\frac{1}{\Gamma(\nu+1)}\,{}_1F_1(1;\nu+1;0)=\frac{t^\nu}{\Gamma(\nu+1)}\left(1+\sum_{n=1}^{+\infty}0\right) \tag{A.66}$$

Q.E.D.

Theorem A. 4: The (integer) derivatives of E_t are given by

---

[111] Miller *et al.* (1993, p. 309).

[112] Miller *et al.* (1993, p. 48). Actually this is the *two*-parameter Mittag-Leffler function. There is also a *one*-parameter Mittag-Leffler function, which is not needed or addressed here.

$$\frac{d^n}{dt^n}E_t(\nu,a)=E_t(\nu-n,a) \tag{A.67}$$

Proof: This is yet another proof by mathematical induction. First we prove that (A.67) holds for the first derivative by using (A.59):

$$E_t(\nu,a)=t^\nu e^{at}e^{-at}\sum_{k=0}^{+\infty}\frac{(at)^k}{\Gamma(\nu+k+1)}=\sum_{k=0}^{+\infty}\frac{a^k t^{k+\nu}}{\Gamma(\nu+k+1)} \tag{A.68}$$

Thus

$$\frac{d}{dt}E_t(\nu,a)=\sum_{k=0}^{+\infty}\frac{(k+\nu)a^k t^{k+\nu-1}}{\Gamma(\nu+k+1)}=\sum_{k=0}^{+\infty}\frac{a^k t^{k+\nu-1}}{\Gamma(\nu+k)}=E_t(\nu-1,a) \tag{A.69}$$

To prove the inductive step, we make

$$\frac{d^{n+1}}{dt^{n+1}}E_t(\nu,a)=\frac{d}{dt}\frac{d^n}{dt^n}E_t(\nu,a)=\frac{d}{dt}E_t(\nu-n,a)=E_t(\nu-n-1,a) \tag{A.70}$$

Q.E.D.

A.3. Power series expansions

In this section some power series expansions needed somewhere else are given[113].

Theorem A. 5: The MacLaurin series expansion of $(x+a)^\nu$ is given by

$$(x+a)^\nu=\sum_{i=0}^{+\infty}a^{\nu-i}\frac{\Gamma(\nu+1)}{\Gamma(i+1)\Gamma(\nu-i+1)}x^i \tag{A.71}$$

Proof: The derivatives of $(x+a)^\nu$ are given by

$$\frac{d^k}{dx^k}(x+a)^\nu=(x+a)^{\nu-k}\prod_{i=0}^{k-1}(\nu-i) \tag{A.72}$$

This is proved by mathematical induction as follows. (A.72) holds for the first derivative, since

$$\frac{d}{dx}(x+a)^\nu=\nu(x+a)^{\nu-1} \tag{A.73}$$

The inductive step is proved thus:

[113] All power series in this section have an infinite number of terms and thus the question of whether they converge or diverge to infinity arises. This issue, however, will not be addressed.

$$\frac{d}{dx}\left[\left(x+a\right)^{\nu-k}\prod_{i=0}^{k-1}\left(\nu-i\right)\right]=\left(x+a\right)^{\nu-k-1}\left(\nu-k\right)\prod_{i=0}^{k-1}\left(\nu-i\right)=$$
$$=\left(x+a\right)^{\nu-(k+1)}\prod_{i=0}^{k}\left(\nu-i\right) \qquad (A.74)$$

Replacing (A.72) into the expression of a MacLaurin series expansion of a function *f*(*x*)

$$f\left(x\right)=\sum_{n=0}^{+\infty}\left.\frac{d^n f\left(x\right)}{dx^n}\right|_{x=0}\frac{x^n}{n!} \qquad (A.75)$$

gives (A.71). **Q.E.D.**

Corollary 1: **The** The MacLaurin series expansion of $\left(-x+a\right)^{\nu}$ is given by

$$\left(-x+a\right)^{\nu}=\sum_{i=0}^{+\infty}a^{\nu-i}\left(-1\right)^{i}\frac{\Gamma\left(\nu+1\right)}{\Gamma\left(i+1\right)\Gamma\left(\nu-i+1\right)}x^{i} \qquad (A.76)$$

Corollary 2: The MacLaurin series expansion of $\left(1-z^{-1}\right)^{\nu}$ is given by

$$\left(1-z^{-1}\right)^{\nu}=\sum_{i=0}^{+\infty}\left(-1\right)^{i}\frac{\Gamma\left(\nu+1\right)}{\Gamma\left(i+1\right)\Gamma\left(\nu-i+1\right)}z^{-i} \qquad (A.77)$$

Corollary 3: The MacLaurin series expansion of $\left(\frac{1-z^{-1}}{1+z^{-1}}\right)^{\nu}$ is given by

$$\left(\frac{1-z^{-1}}{1+z^{-1}}\right)^{\nu}=\Gamma\left(\nu+1\right)\Gamma\left(-\nu+1\right)\times$$
$$\times\sum_{k=0}^{+\infty}z^{-k}\left[\sum_{j=0}^{k}\frac{\left(-1\right)^{j}}{\Gamma\left(\nu-j+1\right)\Gamma\left(j+1\right)\Gamma\left(k-j+1\right)\Gamma\left(-\nu+j-k+1\right)}\right] \qquad (A.78)$$

Proof: From (A.71) we get

$$\left(1+z^{-1}\right)^{-\nu}=\sum_{i=0}^{+\infty}\frac{\Gamma\left(-\nu+1\right)}{\Gamma\left(i+1\right)\Gamma\left(-\nu-i+1\right)}z^{-i} \qquad (A.79)$$

Putting (A.77) and (A.79) together,

$$\left(\frac{1-z^{-1}}{1+z^{-1}}\right)^{\nu}=\left[\sum_{j=0}^{+\infty}\left(-1\right)^{j}\frac{\Gamma\left(\nu+1\right)}{\Gamma\left(j+1\right)\Gamma\left(\nu-j+1\right)}z^{-j}\right]\left[\sum_{i=0}^{+\infty}\frac{\Gamma\left(-\nu+1\right)}{\Gamma\left(i+1\right)\Gamma\left(-\nu-i+1\right)}z^{-i}\right]=$$

$$= \sum_{j=0}^{+\infty} \sum_{i=0}^{+\infty} (-1)^j \frac{\Gamma(\nu+1)}{\Gamma(j+1)\Gamma(\nu-j+1)} \frac{\Gamma(-\nu+1)}{\Gamma(i+1)\Gamma(-\nu-i+1)} z^{-i-j} \tag{A.80}$$

By letting $k = i + j \Leftrightarrow i = k - j$ this becomes

$$\left(\frac{1-z^{-1}}{1+z^{-1}}\right)^{\nu} = \sum_{j=0}^{+\infty} \sum_{k=j}^{+\infty} \frac{(-1)^j \Gamma(\nu+1)}{\Gamma(\nu-j+1)\Gamma(j+1)} \frac{\Gamma(-\nu+1)}{\Gamma(k-j+1)\Gamma(-\nu+j-k+1)} z^{-k} \tag{A.81}$$

and changing the order of the summations the right-hand member of (A.78) is obtained. **Q.E.D.**

Corollary 4: The MacLaurin series expansion of $\left(\dfrac{1-z^{-2}}{1+4z^{-1}+z^{-2}}\right)^{\nu}$ is given by

$$\left(\frac{1-z^{-2}}{1+4z^{-1}+z^{-2}}\right)^{\nu} = \Gamma(\nu+1)\left[\Gamma(-\nu+1)\right]^2 \sum_{q=0}^{+\infty} z^{-q} \left[\sum_{n=0}^{\mathscr{C}\left(\frac{q}{2}\right)} \sum_{p=2n}^{q} \frac{(-1)^n}{\Gamma(n+1)\Gamma(\nu-n+1)} \right.$$

$$\left. \frac{\left(2-\sqrt{3}\right)^{-\nu-p+2n} \left(2+\sqrt{3}\right)^{-\nu-q+p}}{\Gamma(p-2n+1)\Gamma(-\nu-p+2n+1)\Gamma(q-p+1)\Gamma(-\nu-q+p+1)} \right] \tag{A.82}$$

Proof: Since

$$\left(\frac{1-z^{-2}}{1+4z^{-1}+z^{-2}}\right)^{\nu} = \left(1-z^{-2}\right)^{\nu} \left(z^{-1}+2-\sqrt{3}\right)^{-\nu} \left(z^{-1}+2+\sqrt{3}\right)^{-\nu} \tag{A.83}$$

and

$$\left(1-z^{-2}\right)^{\nu} = \sum_{n=0}^{+\infty} (-1)^n \frac{\Gamma(\nu+1)}{\Gamma(n+1)\Gamma(\nu-n+1)} z^{-2n} \tag{A.84}$$

$$\left(z^{-1}+2-\sqrt{3}\right)^{-\nu} = \sum_{i=0}^{+\infty} \left(2-\sqrt{3}\right)^{-\nu-i} \frac{\Gamma(-\nu+1)}{\Gamma(i+1)\Gamma(-\nu-i+1)} z^{-i} \tag{A.85}$$

$$\left(z^{-1}+2+\sqrt{3}\right)^{-\nu} = \sum_{j=0}^{+\infty} \left(2+\sqrt{3}\right)^{-\nu-j} \frac{\Gamma(-\nu+1)}{\Gamma(j+1)\Gamma(-\nu-j+1)} z^{-j} \tag{A.86}$$

we get

$$\left(\frac{1-z^{-2}}{1+4z^{-1}+z^{-2}}\right)^{\nu}=\sum_{n=0}^{+\infty}\sum_{i=0}^{+\infty}\sum_{j=0}^{+\infty}\left[\frac{(-1)^n\Gamma(\nu+1)}{\Gamma(n+1)\Gamma(\nu-n+1)}\right.$$
$$\left.\frac{\left(2-\sqrt{3}\right)^{-\nu-i}\Gamma(-\nu+1)}{\Gamma(i+1)\Gamma(-\nu-i+1)}\frac{\left(2+\sqrt{3}\right)^{-\nu-j}\Gamma(-\nu+1)}{\Gamma(j+1)\Gamma(-\nu-j+1)}z^{-2n-j-i}\right] \tag{A.87}$$

By letting $p=i+2n\Leftrightarrow i=p-2n$ this becomes

$$\left(\frac{1-z^{-2}}{1+4z^{-1}+z^{-2}}\right)^{\nu}=\Gamma(\nu+1)\left[\Gamma(-\nu+1)\right]^2\sum_{n=0}^{+\infty}\sum_{p=2n}^{+\infty}\sum_{j=0}^{+\infty}\left[\frac{(-1)^n}{\Gamma(n+1)\Gamma(\nu-n+1)}\right.$$
$$\left.\frac{\left(2-\sqrt{3}\right)^{-\nu-p+2n}}{\Gamma(p-2n+1)\Gamma(-\nu-p+2n+1)}\frac{\left(2+\sqrt{3}\right)^{-\nu-j}}{\Gamma(j+1)\Gamma(-\nu-j+1)}z^{-p-j}\right] \tag{A.88}$$

and this, in turn, by letting $q=j+p\Leftrightarrow j=q-p$, becomes

$$\left(\frac{1-z^{-2}}{1+4z^{-1}+z^{-2}}\right)^{\nu}=\Gamma(\nu+1)\left[\Gamma(-\nu+1)\right]^2\sum_{n=0}^{+\infty}\sum_{p=2n}^{+\infty}\sum_{q=p}^{+\infty}\left[\frac{(-1)^n}{\Gamma(n+1)\Gamma(\nu-n+1)}\right.$$
$$\left.\frac{\left(2-\sqrt{3}\right)^{-\nu-p+2n}}{\Gamma(p-2n+1)\Gamma(-\nu-p+2n+1)}\frac{\left(2+\sqrt{3}\right)^{-\nu-q+p}}{\Gamma(q-p+1)\Gamma(-\nu-q+p+1)}z^{-q}\right] \tag{A.89}$$

Changing the order of the summations the right-hand member of (A.82) is obtained. **Q.E.D.**

Corollary 5: The MacLaurin series expansion of $\left(3-4z^{-1}+z^{-2}\right)^{\nu}$ is given by

$$\left(3-4z^{-1}+z^{-2}\right)^{\nu}=\left[\Gamma(\nu+1)\right]^2\times$$
$$\times\sum_{k=0}^{+\infty}z^{-k}\sum_{j=0}^{k}\frac{3^{\nu-j}(-1)^k}{\Gamma(j+1)\Gamma(\nu-j+1)\Gamma(k-j+1)\Gamma(\nu-k+j+1)} \tag{A.90}$$

Proof: Since

$$\left(3-4z^{-1}+z^{-2}\right)^{\nu}=\left(1-z^{-1}\right)^{\nu}\left(3-z^{-1}\right)^{\nu} \tag{A.91}$$

and

$$\left(3-z^{-1}\right)^{\nu}=\sum_{j=0}^{+\infty}3^{\nu-j}\frac{\Gamma(\nu+1)}{\Gamma(j+1)\Gamma(\nu-j+1)}(-1)^j z^{-j} \tag{A.92}$$

we get (using also (A.77))

$$\left(3-4z^{-1}+z^{-2}\right)^{\nu}=\sum_{j=0}^{+\infty}\sum_{i=0}^{+\infty}\frac{3^{\nu-j}\Gamma(\nu+1)}{\Gamma(j+1)\Gamma(\nu-j+1)}\frac{\Gamma(\nu+1)(-1)^{i+j}z^{-i-j}}{\Gamma(i+1)\Gamma(\nu-i+1)} \tag{A.93}$$

By letting $k=i+j\Leftrightarrow i=k-j$ this becomes

$$\left(3-4z^{-1}+z^{-2}\right)^{\nu}=\sum_{j=0}^{+\infty}\sum_{k=j}^{+\infty}\frac{3^{\nu-j}\left[\Gamma(\nu+1)\right]^{2}}{\Gamma(j+1)\Gamma(\nu-j+1)}\frac{(-1)^{k}\left(z^{-1}\right)^{k}}{\Gamma(k-j+1)\Gamma(\nu-k+j+1)} \tag{A.94}$$

Changing the order of the summations the right-hand member of (A.90) is obtained. **Q.E.D.**

Corollary 6: The MacLaurin series expansion of $\left(11-18z^{-1}+9z^{-2}-2z^{-3}\right)^{\nu}$ is given by

$$\left(11-18z^{-1}+9z^{-2}-2z^{-3}\right)^{\nu}=2^{\nu}\left[\Gamma(\nu+1)\right]^{3}\sum_{k=0}^{+\infty}z^{-k}\left[\sum_{t=0}^{k}\sum_{p=0}^{t}\frac{(-1)^{p}}{\Gamma(p+1)\Gamma(\nu-p+1)}\right.$$
$$\left.\frac{\left(-\frac{7}{4}-\frac{\sqrt{39}}{4}j\right)^{\nu-t+p}\left(-\frac{7}{4}+\frac{\sqrt{39}}{4}j\right)^{\nu-k+t}}{\Gamma(t-p+1)\Gamma(\nu-t+p+1)\Gamma(k-t+1)\Gamma(\nu-k+t+1)}\right] \tag{A.95}$$

Proof: Since

$$\left(11-18z^{-1}+9z^{-2}-2z^{-3}\right)^{\nu}=$$
$$=\left(-2z^{-1}+2\right)^{\nu}\left(z^{-1}-\frac{7}{4}-\frac{\sqrt{39}}{4}j\right)^{\nu}\left(z^{-1}-\frac{7}{4}+\frac{\sqrt{39}}{4}j\right)^{\nu} \tag{A.96}$$

and

$$\left(-2z^{-1}+2\right)^{\nu}=\sum_{p=0}^{+\infty}2^{\nu-p}\frac{\Gamma(\nu+1)}{\Gamma(p+1)\Gamma(\nu-p+1)}\left(-2z^{-1}\right)^{p} \tag{A.97}$$

$$\left(z^{-1}-\frac{7}{4}-\frac{\sqrt{39}}{4}j\right)^{\nu}=\sum_{q=0}^{+\infty}\left(-\frac{7}{4}-\frac{\sqrt{39}}{4}j\right)^{\nu-q}\frac{\Gamma(\nu+1)}{\Gamma(q+1)\Gamma(\nu-q+1)}z^{-q} \tag{A.98}$$

$$\left(z^{-1}-\frac{7}{4}+\frac{\sqrt{39}}{4}j\right)^{\nu}=\sum_{r=0}^{+\infty}\left(-\frac{7}{4}+\frac{\sqrt{39}}{4}j\right)^{\nu-r}\frac{\Gamma(\nu+1)}{\Gamma(r+1)\Gamma(\nu-r+1)}z^{-r} \tag{A.99}$$

we get

$$\left(11-18z^{-1}+9z^{-2}-2z^{-3}\right)^{\nu} = \sum_{p=0}^{+\infty}\sum_{q=0}^{+\infty}\sum_{r=0}^{+\infty}\left[\frac{2^{\nu}(-1)^{p}\left[\Gamma(\nu+1)\right]^{3}}{\Gamma(p+1)\Gamma(\nu-p+1)} \frac{\left(-\frac{7}{4}-\frac{\sqrt{39}}{4}j\right)^{\nu-q}}{\Gamma(q+1)\Gamma(\nu-q+1)} \frac{\left(-\frac{7}{4}+\frac{\sqrt{39}}{4}j\right)^{\nu-r}}{\Gamma(r+1)\Gamma(\nu-r+1)} z^{-p-q-r}\right] \tag{A.100}$$

By letting $t = p+q \Leftrightarrow q = t-p$ this becomes

$$\left(11-18z^{-1}+9z^{-2}-2z^{-3}\right)^{\nu} = \sum_{p=0}^{+\infty}\sum_{t=p}^{+\infty}\sum_{r=0}^{+\infty}\left[\frac{2^{\nu}(-1)^{p}\left[\Gamma(\nu+1)\right]^{3}}{\Gamma(p+1)\Gamma(\nu-p+1)} \frac{\left(-\frac{7}{4}-\frac{\sqrt{39}}{4}j\right)^{\nu-t+p}}{\Gamma(t-p+1)\Gamma(\nu-t+p+1)} \frac{\left(-\frac{7}{4}+\frac{\sqrt{39}}{4}j\right)^{\nu-r}}{\Gamma(r+1)\Gamma(\nu-r+1)} z^{-t-r}\right] \tag{A.101}$$

and this, in turn, by letting $k = t+r \Leftrightarrow r = k-t$, becomes

$$\left(11-18z^{-1}+9z^{-2}-2z^{-3}\right)^{\nu} = \sum_{p=0}^{+\infty}\sum_{t=p}^{+\infty}\sum_{k=t}^{+\infty}\left[\frac{2^{\nu}(-1)^{p}\left[\Gamma(\nu+1)\right]^{3}}{\Gamma(p+1)\Gamma(\nu-p+1)} \frac{\left(-\frac{7}{4}-\frac{\sqrt{39}}{4}j\right)^{\nu-t+p}}{\Gamma(t-p+1)\Gamma(\nu-t+p+1)} \frac{\left(-\frac{7}{4}+\frac{\sqrt{39}}{4}j\right)^{\nu-k+t}}{\Gamma(k-t+1)\Gamma(\nu-k+t+1)} z^{-k}\right] \tag{A.102}$$

Changing the order of the summations the right-hand member of (A.95) is obtained. **Q.E.D.**

Remark: Even though the imaginary unit shows up in (A.95), the coefficients are all real, since the imaginary parts cancel each other for every value of *k*.

A.4. Continued fractions

This section deals with continued fractions, entities that have the form

$$a_0+\cfrac{b_1}{a_1+\cfrac{b_2}{a_2+\cfrac{b_3}{a_3+\dots}}} \tag{A.103}$$

Coefficients *a* and *b* may be either constants of functions. Furthermore, there may be a finite number of such; or they may go on indefinitely. In both cases, the continued fraction may be truncated after a certain number of terms, providing an approximation to its value[114].

The notation above may become excessively nested, and one of the two following is usually preferred[115]:

$$a_0+\frac{b_1}{a_1+}\,\frac{b_2}{a_2+}\,\frac{b_3}{a_3+}\,\frac{b_4}{a_4+}\dots \tag{A.104}$$

$$\left[a_0;\frac{b_1}{a_1},\frac{b_2}{a_2},\frac{b_3}{a_3},\frac{b_4}{a_4},\dots\right]=\left[a_0;\frac{b_i}{a_i}\right]_{i=1}^{+\infty} \tag{A.105}$$

(A.104) is known as the Abramowitz notation and (A.105) as the Pringsheim notation. Of course, in this last case, the upper limit can be either infinity or some suitable natural number.

Both real numbers and real valued functions can be expanded into continued fractions.

The subsections that follow are on the following subjects:

- continued fraction expansions of real numbers;
- continued fraction expansions of real valued functions;
- how to numerically evaluate a continued fraction;
- derivation of important continued fraction expansions.

A.4.1. Continued fraction expansions of real numbers

Every real non-integer number x_0 may transformed as follows:

$$x_0=\mathscr{C}(x_0)+\frac{1}{x_1},\quad x_1=\frac{1}{x_0-\mathscr{C}(x_0)} \tag{A.106}$$

This same transformation may now be applied to x_1 and so on. The general

[114] If there is an infinite number of coefficients the problem arises of knowing whether the continued fraction converges or not. This problem is extensively dealt with in Wall (1948) but shall not be addressed here; all continued fractions are assumed to be convergent, and those that will appear indeed are.

[115] Radok (1999); Wall (1948, p.17). This last reference also presents the notation

$$a_0+\mathop{\mathrm{K}}_{i=1}^{+\infty}\frac{b_i}{a_i}$$

The *K* stands here for *Kettenbruch*, the German word for continued fraction.

expressions are[116]

$$x_k = \frac{1}{x_{k-1} - \mathscr{C}\left(x_{k-1}\right)}, \quad k \in \mathbb{N} \tag{A.107}$$

$$x_0 = \mathscr{C}\left(x_0\right) + \frac{1}{\mathscr{C}\left(x_1\right)+}\frac{1}{\mathscr{C}\left(x_2\right)+}\frac{1}{\mathscr{C}\left(x_3\right)+}\frac{1}{\mathscr{C}\left(x_4\right)+}\ldots = \left[\mathscr{C}\left(x_0\right); \frac{1}{\mathscr{C}\left(x_i\right)}\right]_{i=1}^{+\infty} \tag{A.108}$$

Numbers *x* are called remainders. If x_0 is a fraction, there will be a time when the process stops and (A.107) cannot be applied further because $x_{k-1} \in \mathbb{N}$. If, on the contrary, $x_0 \in \mathbb{R} \setminus \mathbb{Q}$, then the continued fraction will have an infinite number of coefficients *a* and *b*.

Whatever the case, truncating the expansion after some terms provides a good approximation[117] of x_0.

A.4.2. Continued fraction expansions of real valued functions

Let $f_0\left(x\right)$ be a real valued function given by the ratio of two polynomials[118]:

$$f_0\left(x\right) = \frac{c_{10} + c_{11}x + c_{12}x^2 + c_{13}x^3 + \ldots}{c_{00} + c_{01}x + c_{02}x^2 + c_{03}x^3 + \ldots} \tag{A.109}$$

This can be rewritten as

$$f_0\left(x\right) = \frac{1}{\frac{c_{00}}{c_{10}} + \frac{c_{00} + c_{01}x + c_{02}x^2 + c_{03}x^3 + \ldots}{c_{10} + c_{11}x + c_{12}x^2 + c_{13}x^3 + \ldots} - \frac{c_{00}}{c_{10}}} =$$

$$= \frac{c_{10}}{c_{00} + c_{10}\frac{c_{00} + c_{01}x + c_{02}x^2 + c_{03}x^3 + \ldots}{c_{10} + c_{11}x + c_{12}x^2 + c_{13}x^3 + \ldots} - c_{00}} =$$

$$= \frac{c_{10}}{c_{00} + \frac{c_{10}c_{00} - c_{00}c_{10} + c_{10}c_{01}x - c_{00}c_{11}x + c_{10}c_{02}x^2 - c_{00}c_{12}x^2 + c_{10}c_{03}x^3 - c_{00}c_{13}x^3 + \ldots}{c_{10} + c_{11}x + c_{12}x^2 + c_{13}x^3 + \ldots}} =$$

$$= \frac{c_{10}}{c_{00} + x\frac{\left(c_{10}c_{01} - c_{00}c_{11}\right) + \left(c_{10}c_{02} - c_{00}c_{12}\right)x + \left(c_{10}c_{03} - c_{00}c_{13}\right)x^2 + \ldots}{c_{10} + c_{11}x + c_{12}x^2 + c_{13}x^3 + \ldots}} \tag{A.110}$$

[116] Radok (1999).

[117] This approximation is the best possible for the order of magnitude of the denominator. In other words, no fractional approximation with a denominator of the same order of magnitude can be found closer to the desired value (Beskin, p. 85-95, 2001).

[118] Radok (1999).

So, if we now let

$$c_{2,i} = c_{10}c_{0,i+1} - c_{00}c_{1,i+1} \tag{A.111}$$

$$f_1(x) = \frac{c_{20} + c_{21}x + c_{22}x^2 + c_{23}x^3 + \ldots}{c_{10} + c_{11}x + c_{12}x^2 + c_{13}x^3 + \ldots} \tag{A.112}$$

we can write

$$f_0(x) = \frac{c_{10}}{c_{00} + xf_1(x)} \tag{A.113}$$

Function $f_1(x)$ can be transformed in the very same manner:

$$f_1(x) = \frac{c_{20}}{c_{10} + xf_2(x)}$$

$$f_2(x) = \frac{c_{30} + c_{31}x + c_{32}x^2 + c_{33}x^3 + \ldots}{c_{20} + c_{21}x + c_{22}x^2 + c_{23}x^3 + \ldots} \tag{A.114}$$

$$c_{3,i} = c_{20}c_{1,i+1} - c_{10}c_{2,i+1}$$

The general expressions will be

$$f_0(x) = 0 + \frac{c_{10}}{c_{00}+}\frac{c_{20}x}{c_{10}+}\frac{c_{30}x}{c_{20}+}\frac{c_{40}x}{c_{30}+}\ldots = \left[0; \frac{c_{10}}{c_{00}}, \left\{\frac{c_{j+1,0}x}{c_{j,0}}\right\}\right]_{j=1}^{+\infty} \tag{A.115}$$

$$c_{j,i} = c_{j-1,0}c_{j-2,i+1} - c_{j-2,0}c_{j-1,i+1} = -\begin{vmatrix} c_{j-2,0} & c_{j-2,i+1} \\ c_{j-1,0} & c_{j-1,i+1} \end{vmatrix} \tag{A.116}$$

(Only the last fraction in (A.115) is to be repeated: this is what is meant by the curly brackets.)

If $f_0(x)$ is the ratio of two polynomials composed of a finite number of monomials, then its continued fraction expansion will have a finite number of terms.

If we want to apply this method to a function $f_0(x)$ that is not the ratio of two polynomials, a solution will be to expand it into a Taylor series, which is the ratio of two polynomials (the polynomial in the denominator being the monomial 1).

However, in some instances, it may be possible to find by analytical means explicit formulas for the continued fraction expansions of important functions. Two such expansions are presented below in section A.4.4.

Just as with real numbers, continued fraction expansions may be truncated after some terms providing a good approximation for $f_0(x)$.

A.4.3. Evaluation of a continued fraction

Two ways of evaluating a continued fraction will be described here[119]. The first (and most obvious one) is evaluating it from its innermost denominator out, that is to say, from right to left. If the fraction has infinite coefficients, this requires truncating it somewhere, so as to obtain a reasonable approximation: but where that truncation should be done is very difficult to ascertain beforehand.

The second (and more efficient) approach is to reckon the so-called convergents according to the following recursive expressions:

$$P_{-1} = 1 \tag{A.117}$$
$$P_0 = a_0 \tag{A.118}$$
$$P_k = a_k P_{k-1} + b_k P_{k-2}, \quad k \in \mathbb{N} \tag{A.119}$$
$$Q_{-1} = 0 \tag{A.120}$$
$$Q_0 = 1 \tag{A.121}$$
$$Q_k = a_k Q_{k-1} + b_k Q_{k-2}, \quad k \in \mathbb{N} \tag{A.122}$$
$$R_k = \frac{P_k}{Q_k} \tag{A.123}$$

Theorem A. 6: The sequence R_k is the sequence of approximations of the continued fraction that would be obtained truncating it after *k* terms.

Proof: This is proved by induction as follows. For $k = 1$ we have

$$R_1 = \frac{P_1}{Q_1} = \frac{a_1 a_0 + b_1}{a_1 + 0} = a_0 + \frac{b_1}{a_1} \tag{A.124}$$

as desired, so the formula is valid in this case. To prove the inductive step, notice that taking into account a further pair of coefficients means that the last denominator considered, which is a_k, will be added a new fraction, which is $\frac{b_{k+1}}{a_{k+1}}$; so where we had a_k we will have $a_k + \frac{b_{k+1}}{a_{k+1}}$. So, if the formula is valid for a given integer *k*,

$$R_k = \frac{P_k}{Q_k} = \frac{a_k P_{k-1} + b_k P_{k-2}}{a_k Q_{k-1} + b_k Q_{k-2}} \tag{A.125}$$

we shall have, replacing a_k by $a_k + \frac{b_{k+1}}{a_{k+1}}$,

[119] There are other algorithms available. See Radok (1999); Wall (1948, p. 15); Press *et al.*, (1992, p. 170). This last reference includes other more complicated (but also more numerically reliable) algorithms.

$$R_{k+1}=\frac{\left(a_k+\dfrac{b_{k+1}}{a_{k+1}}\right)P_{k-1}+b_kP_{k-2}}{\left(a_k+\dfrac{b_{k+1}}{a_{k+1}}\right)Q_{k-1}+b_kQ_{k-2}}=\frac{a_ka_{k+1}P_{k-1}+b_{k+1}P_{k-1}+a_{k+1}b_kP_{k-2}}{a_ka_{k+1}Q_{k-1}+b_{k+1}Q_{k-1}+a_{k+1}b_kQ_{k-2}}=$$
$$=\frac{a_{k+1}\left(a_kP_{k-1}+b_kP_{k-2}\right)+b_{k+1}P_{k-1}}{a_{k+1}\left(a_kQ_{k-1}+b_kQ_{k-2}\right)+b_{k+1}Q_{k-1}}=\frac{a_{k+1}P_k+b_{k+1}P_{k-1}}{a_{k+1}Q_k+b_{k+1}Q_{k-1}}=\frac{P_{k+1}}{Q_{k+1}} \tag{A.126}$$

Q.E.D.

A.4.4. Important continued fraction expansions

From the definition (A.47) of F it is possible to find several important continued fraction expansions[120]. First we show that

$$F\left(a,b+1;c+1;x\right)-\frac{a\left(c-b\right)}{c\left(c+1\right)}xF\left(a+1,b+1;c+2;x\right)=$$
$$=1+\sum_{n=1}^{+\infty}\frac{\prod_{i=0}^{n-1}\left(a+i\right)\prod_{j=1}^{n}\left(b+j\right)}{n!\prod_{k=1}^{n}\left(c+k\right)}x^n-\frac{a\left(c-b\right)}{c\left(c+1\right)}x\left[1+\sum_{n=1}^{+\infty}\frac{\prod_{i=1}^{n}\left(a+i\right)\prod_{j=1}^{n}\left(b+j\right)}{n!\prod_{k=2}^{n+1}\left(c+k\right)}x^n\right]=$$
$$=1+\sum_{n=1}^{+\infty}\frac{\prod_{i=0}^{n-1}\left(a+i\right)\prod_{j=1}^{n}\left(b+j\right)}{n!\prod_{k=1}^{n}\left(c+k\right)}x^n-\frac{a\left(c-b\right)}{c\left(c+1\right)}x-\sum_{n=1}^{+\infty}\frac{a\left(c-b\right)}{c\left(c+1\right)}\frac{\prod_{i=1}^{n}\left(a+i\right)\prod_{j=1}^{n}\left(b+j\right)}{n!\prod_{k=2}^{n+1}\left(c+k\right)}x^{n+1}=$$
$$=1-\frac{a\left(c-b\right)}{c\left(c+1\right)}x+\sum_{n=1}^{+\infty}\frac{\prod_{i=0}^{n-1}\left(a+i\right)\prod_{j=1}^{n}\left(b+j\right)}{n!\prod_{k=1}^{n}\left(c+k\right)}x^n-\sum_{n=1}^{+\infty}\frac{\left(c-b\right)\prod_{i=0}^{n}\left(a+i\right)\prod_{j=1}^{n}\left(b+j\right)}{n!\prod_{k=0}^{n+1}\left(c+k\right)}x^{n+1}=$$
$$=1-\frac{a\left(c-b\right)}{c\left(c+1\right)}x+\frac{a\left(b+1\right)}{c+1}x+\ldots$$
$$\ldots+\sum_{n=2}^{+\infty}\frac{\prod_{i=0}^{n-1}\left(a+i\right)\prod_{j=1}^{n}\left(b+j\right)}{n!\prod_{k=1}^{n}\left(c+k\right)}x^n-\sum_{n=2}^{+\infty}\frac{\left(c-b\right)\prod_{i=0}^{n-1}\left(a+i\right)\prod_{j=1}^{n-1}\left(b+j\right)}{\left(n-1\right)!\prod_{k=0}^{n}\left(c+k\right)}x^n=$$

[120] Wall (1948, p. 335-343). Even though these expansions may be used when the variable is complex, here we are concerned with the real case only.

$$= 1 + \frac{abc + ac - ac + ab}{c(c+1)} x + \sum_{n=2}^{+\infty} \frac{\prod_{i=0}^{n-1}(a+i)\prod_{j=1}^{n-1}(b+j)}{(n-1)!\prod_{k=1}^{n}(c+k)} \left(\frac{b+n}{n} - \frac{c-b}{c} \right) =$$

$$= 1 + \frac{abc + ab}{c(c+1)} x + \sum_{n=2}^{+\infty} \frac{\prod_{i=0}^{n-1}(a+i)\prod_{j=1}^{n-1}(b+j)}{(n-1)!\prod_{k=1}^{n}(c+k)} \frac{bc + nc - cn + bn}{nc} =$$

$$= 1 + \frac{ab(c+1)}{c(c+1)} x + \sum_{n=2}^{+\infty} \frac{\prod_{i=0}^{n-1}(a+i)\prod_{j=1}^{n-1}(b+j)}{n!\prod_{k=0}^{n}(c+k)} b(c+n) =$$

$$= 1 + \frac{ab}{c} x + \sum_{n=2}^{+\infty} \frac{\prod_{i=0}^{n-1}(a+i)\prod_{j=0}^{n-1}(b+j)}{n!\prod_{k=0}^{n-1}(c+k)} =$$

$$= 1 + \sum_{n=1}^{+\infty} \frac{\prod_{i=0}^{n-1}(a+i)\prod_{j=0}^{n-1}(b+j)}{n!\prod_{k=0}^{n-1}(c+k)} = F(a,b;c;x) \qquad \text{(A.127)}$$

This identity can be rewritten as

$$F(a,b;c;x) = F(a,b+1;c+1;x) - \frac{a(c-b)}{c(c+1)} xF(a+1,b+1;c+2;x) \Leftrightarrow$$

$$\Leftrightarrow \frac{F(a,b;c;x)}{F(a,b+1;c+1;x)} = 1 - \frac{a(c-b)}{c(c+1)} x \frac{F(a+1,b+1;c+2;x)}{F(a,b+1;c+1;x)} \Leftrightarrow$$

$$\Leftrightarrow \frac{F(a,b+1;c+1;x)}{F(a,b;c;x)} = \frac{1}{1 - \frac{a(c-b)}{c(c+1)} x \frac{F(a+1,b+1;c+2;x)}{F(a,b+1;c+1;x)}} \qquad \text{(A.128)}$$

From the definition (A.47) of F it may be seen that interchanging the first two coefficients does not alter the value of the function. So the last equality yields

$$\frac{F(a+1,b;c+1;x)}{F(a,b;c;x)} = \frac{1}{1 - \frac{b(c-a)}{c(c+1)} x \frac{F(a+1,b+1;c+2;x)}{F(a+1,b;c+1;x)}} \qquad \text{(A.129)}$$

This, evaluated with b+1 instead of b and c+1 instead of c, gives

$$\frac{F(a+1,b+1;c+2;x)}{F(a,b+1;c+1;x)}=\frac{1}{1-\frac{(b+1)(c-a+1)}{(c+1)(c+2)}x\frac{F(a+1,b+2;c+3;x)}{F(a+1,b+1;c+2;x)}} \tag{A.130}$$

which may be replaced into the last equality of (A.128). But that equality, evaluated with a+1 instead of a, b+1 instead of b and c+2 instead of c, becomes

$$\frac{F(a+1,b+2;c+3;x)}{F(a+1,b+1;c+2;x)}=\frac{1}{1-\frac{(a+1)(c-b+1)}{(c+2)(c+3)}x\frac{F(a+2,b+2;c+4;x)}{F(a+1,b+2;c+3;x)}} \tag{A.131}$$

which may be replaced into (A.130). By increasing the coefficients in (A.130) to a+1, b+1 and c+2 again, a new equality may be found that may be replaced into (A.131): and this process may go on indefinitely, leading to a continued fraction known as the continued fraction of Gauß:

$$\frac{F(a,b+1;c+1;x)}{F(a,b;c;x)}=0+\frac{1}{1-}\frac{\frac{a(c-b)}{c(c+1)}x}{1-}\frac{\frac{(b+1)(c-a+1)}{(c+1)(c+2)}x}{1-}\cdots$$
$$\cdots\frac{\frac{(a+1)(c-b+1)}{(c+2)(c+3)}x}{1-}\frac{\frac{(b+2)(c-a+2)}{(c+3)(c+4)}x}{1-}\frac{\frac{(a+2)(c-b+2)}{(c+4)(c+5)}x}{1-}\cdots \tag{A.132}$$

By letting $b=0$ and replacing c with c–1, this becomes

$$\frac{F(a,1;c;x)}{F(a,0;c-1;x)}=\frac{F(a,1;c;x)}{1+\sum_{n=1}^{+\infty}0}=F(a,1;c;x)=$$
$$=0+\frac{1}{1-}\frac{\frac{a}{c}x}{1-}\frac{\frac{c-a}{c(c+1)}x}{1-}\frac{\frac{(a+1)c}{(c+1)(c+2)}x}{1-}\frac{\frac{2(c-a+1)}{(c+2)(c+3)}x}{1-}\frac{\frac{(a+2)(c+1)}{(c+3)(c+4)}x}{1-}\cdots \tag{A.133}$$

Now from (A.71), (A.40) and definition (A.47) it can be seen that

$$\left(1-z^{-1}\right)^{\nu}=\sum_{k=0}^{+\infty}(-1)^k\frac{\Gamma(\nu+1)}{\Gamma(\nu-k+1)}\frac{z^{-k}}{\Gamma(k+1)}=1+\sum_{k=1}^{+\infty}(-1)^k\frac{\Gamma(\nu+1)}{\Gamma(\nu-k+1)}\frac{z^{-k}}{\Gamma(k+1)}=$$
$$=1+\sum_{k=1}^{+\infty}\left[\prod_{i=0}^{k-1}(-\nu+i)\right]\frac{\left(z^{-1}\right)^k}{k!}=F\left(-\nu,1,1;z^{-1}\right) \tag{A.134}$$

By replacing z^{-1} with $-z^{-1}$, $-z$ or z in this equality, we also see that

$$\left(1+z^{-1}\right)^{\nu}=F\left(-\nu,1;1;-z^{-1}\right) \tag{A.135}$$

$$\left(1-z\right)^{\nu}=F\left(-\nu,1;1;z\right) \tag{A.136}$$

$$\left(1+z\right)^{\nu}=F\left(-\nu,1;1;-z\right) \tag{A.137}$$

This allows equation (A.133) to yield[121]

$$\left(1-z^{-1}\right)^{\nu}=0+\frac{1}{1-}\frac{-\nu z^{-1}}{1-}\frac{\frac{1+\nu}{1\times 2}z^{-1}}{1-}\frac{\frac{1-\nu}{2\times 3}z^{-1}}{1-}\frac{\frac{2(2+\nu)}{3\times 4}z^{-1}}{1-}\ldots$$
$$\ldots\frac{\frac{2(2-\nu)}{4\times 5}z^{-1}}{1-}\frac{\frac{3(3+\nu)}{5\times 6}z^{-1}}{1-}\frac{\frac{3(3-\nu)}{6\times 7}z^{-1}}{1-}\ldots=$$
$$=\left[0;\frac{1}{1},\frac{\nu z^{-1}}{1},\left\{\frac{-\frac{i(i+\nu)}{(2i-1)2i}z^{-1}}{1},\frac{-\frac{i(i-\nu)}{2i(2i+1)}z^{-1}}{1}\right\}\right]_{i=1}^{+\infty} \tag{A.138}$$

(In this last expression, curly brackets are meant to show that each value of dumb variable i adds two terms to the continued fraction.) Expansions similar to (A.138) may be easily found for the functions (A.135), (A.136) and (A.137).

So as to obtain the next desired result it is expedient to prove first what follows.

Lemma A. 1: From the continued fraction of Gauß (A.132) we obtain, by letting $a=\frac{1-k}{2}$, $b=-\frac{k}{2}$, $c=\frac{1}{2}$ and $x=z^2$,

$$\frac{F\left(\frac{1-k}{2},\frac{2-k}{2};\frac{3}{2};z^2\right)}{F\left(\frac{1-k}{2},-\frac{k}{2};\frac{1}{2};z^2\right)}=0+\frac{1}{1-}\frac{\frac{\frac{1-k}{2}\cdot\frac{1+k}{2}}{\frac{1}{2}\cdot\frac{3}{2}}z^2}{1-}\frac{\frac{\frac{2-k}{2}\cdot\frac{2+k}{2}}{\frac{3}{2}\cdot\frac{5}{2}}z^2}{1-}\ldots$$
$$\ldots\frac{\frac{\frac{3-k}{2}\cdot\frac{3+k}{2}}{\frac{5}{2}\cdot\frac{7}{2}}z^2}{1-}\frac{\frac{\frac{4-k}{2}\cdot\frac{4+k}{2}}{\frac{7}{2}\cdot\frac{9}{2}}z^2}{1-}\frac{\frac{\frac{5-k}{2}\cdot\frac{5+k}{2}}{\frac{9}{2}\cdot\frac{11}{2}}z^2}{1-}\ldots=$$

121 For complete mathematical rigour, it is necessary to impose that, in (A.138), $z^{-1}\notin\left]1;+\infty\right[$ (Wall, p. 343, 1948); otherwise, the continued fraction will not converge.

$$=\cfrac{1}{1+\cfrac{\frac{k^2-1^2}{1\times 3}z^2}{1+\cfrac{\frac{k^2-2^2}{3\times 5}z^2}{1+\cfrac{\frac{k^2-3^2}{5\times 7}z^2}{1+\cfrac{\frac{k^2-4^2}{7\times 9}z^2}{1+\cfrac{\frac{k^2-5^2}{9\times 11}z^2}{1+\dots}}}}}}=\cfrac{1}{1+\cfrac{\left(k^2-1^2\right)z^2}{3+\cfrac{\left(k^2-2^2\right)z^2}{5+\cfrac{\left(k^2-3^2\right)z^2}{7+\cfrac{\left(k^2-4^2\right)z^2}{9+\cfrac{\left(k^2-5^2\right)z^2}{11+\dots}}}}}} \tag{A.139}$$

Lemma A. 2: From the definition of the hypergeometric series (A.47), we obtain, by letting $a=\frac{1-k}{2}$, $b=\frac{2-k}{2}$, $c=\frac{3}{2}$ and $x=z^2$,

$$F\left(\frac{1-k}{2},\frac{2-k}{2};\frac{3}{2};z^2\right)=1+\sum_{n=1}^{+\infty}\frac{\prod_{i=0}^{n-1}\left(\frac{1-k}{2}+i\right)\prod_{j=0}^{n-1}\left(\frac{2-k}{2}+j\right)}{\prod_{m=0}^{n-1}(1+m)\prod_{p=0}^{n-1}\left(\frac{3}{2}+p\right)}z^{2n}=$$

$$=1+\sum_{n=1}^{+\infty}z^{2n}\frac{\prod_{i=0}^{n-1}(2i+1-k)(2i+2-k)}{\prod_{m=0}^{n-1}(2m+2)(2m+3)}=1+\sum_{n=1}^{+\infty}z^{2n}\frac{\prod_{i=0}^{2n-1}(i-k+1)}{\prod_{m=0}^{2n-1}(m+2)} \tag{A.140}$$

Lemma A. 3: From the definition of the hypergeometric series (A.47), we obtain, by letting $a=\frac{1-k}{2}$, $b=-\frac{k}{2}$, $c=\frac{1}{2}$ and $x=z^2$,

$$F\left(\frac{1-k}{2},-\frac{k}{2};\frac{1}{2};z^2\right)=1+\sum_{n=1}^{+\infty}\frac{\prod_{i=0}^{n-1}\left(\frac{1-k}{2}+i\right)\prod_{j=0}^{n-1}\left(-\frac{k}{2}+j\right)}{\prod_{m=0}^{n-1}(1+m)\prod_{p=0}^{n-1}\left(\frac{1}{2}+p\right)}z^{2n}=$$

$$=1+\sum_{n=1}^{+\infty}z^{2n}\frac{\prod_{i=0}^{n-1}(2i+1-k)(2i-k)}{\prod_{m=0}^{n-1}(2m+2)(2m+1)}=1+\sum_{n=1}^{+\infty}z^{2n}\frac{\prod_{i=0}^{2n-1}(i-k)}{\prod_{m=0}^{2n-1}(m+1)} \tag{A.141}$$

We are now in position to prove the following:

Theorem A. 7: The following continued fraction expansion holds:

$$\frac{(1+x)^k-(1-x)^k}{(1+x)^k+(1-x)^k}=$$
$$=\frac{kz}{1+}\frac{(k^2-1^2)z^2}{3+}\frac{(k^2-2^2)z^2}{5+}\frac{(k^2-3^2)z^2}{7+}\frac{(k^2-4^2)z^2}{9+}\frac{(k^2-5^2)z^2}{11+}\dots \quad \text{(A.142)}$$

Proof: From (A.71) and (A.76) follows that the left-hand side of (A.142) may be expanded as

$$\frac{(1+x)^k-(1-x)^k}{(1+x)^k+(1-x)^k}=\frac{\sum_{i=0}^{+\infty}\frac{\Gamma(k+1)}{\Gamma(i+1)\Gamma(k-i+1)}x^i-(-1)^i\frac{\Gamma(k+1)}{\Gamma(i+1)\Gamma(k-i+1)}x^i}{\sum_{i=0}^{+\infty}\frac{\Gamma(k+1)}{\Gamma(i+1)\Gamma(k-i+1)}x^i+(-1)^i\frac{\Gamma(k+1)}{\Gamma(i+1)\Gamma(k-i+1)}x^i}=$$
$$=\frac{\sum_{i=0}^{+\infty}\frac{\Gamma(k+1)}{\Gamma(i+1)\Gamma(k-i+1)}x^i\left[1-(-1)^i\right]}{\sum_{i=0}^{+\infty}\frac{\Gamma(k+1)}{\Gamma(i+1)\Gamma(k-i+1)}x^i\left[1+(-1)^i\right]}=$$
$$=\frac{2\sum_{j=0}^{+\infty}\frac{\Gamma(k+1)}{\Gamma(2j+1+1)\Gamma\left[k-(2j+1)+1\right]}x^{2j+1}}{2\sum_{n=0}^{+\infty}\frac{\Gamma(k+1)}{\Gamma(2n+1)\Gamma(k-2n+1)}x^{2n}}=\frac{\sum_{j=0}^{+\infty}\frac{\Gamma(k+1)}{\Gamma(2j+2)\Gamma(k-2j)}x^{2j+1}}{\sum_{n=0}^{+\infty}\frac{\Gamma(k+1)}{\Gamma(2n+1)\Gamma(k-2n+1)}x^{2n}}=$$
$$=\frac{\sum_{j=0}^{+\infty}\frac{1}{\Gamma(2j+2)\Gamma(k-2j)}x^{2j+1}}{\sum_{n=0}^{+\infty}\frac{1}{\Gamma(2n+1)\Gamma(k-2n+1)}x^{2n}}=\frac{\frac{x}{\Gamma(2)\Gamma(k)}+\sum_{j=1}^{+\infty}\frac{1}{\Gamma(2j+2)\Gamma(k-2j)}x^{2j+1}}{\frac{\Gamma(1)}{\Gamma(1)\Gamma(k+1)}+\sum_{n=1}^{+\infty}\frac{1}{\Gamma(2n+1)\Gamma(k-2n+1)}x^{2n}}=$$
$$=\frac{kx+\sum_{j=1}^{+\infty}\frac{\Gamma(k+1)}{\Gamma(2j+2)\Gamma(k-2j)}x^{2j+1}}{1+\sum_{n=1}^{+\infty}\frac{\Gamma(k+1)}{\Gamma(2n+1)\Gamma(k-2n+1)}x^{2n}}=kx\frac{1+\sum_{j=1}^{+\infty}x^{2j}\frac{\Gamma(k)}{\Gamma(2j+2)\Gamma(k-2j)}}{1+\sum_{n=1}^{+\infty}x^{2n}\frac{\Gamma(k+1)}{\Gamma(2n+1)\Gamma(k-2n+1)}} \quad \text{(A.143)}$$

Now (A.40), together with (A.140) and (A.141), lets us write

$$kz\frac{F\left(\frac{1-k}{2},\frac{2-k}{2};\frac{3}{2};z^2\right)}{F\left(\frac{1-k}{2},-\frac{k}{2};\frac{1}{2};z^2\right)}=kz\frac{1+\sum_{n=1}^{+\infty}z^{2n}\frac{\prod_{i=0}^{2n-1}(i-k+1)}{\prod_{m=0}^{2n-1}(m+2)}}{1+\sum_{n=1}^{+\infty}z^{2n}\frac{\prod_{i=0}^{2n-1}(i-k)}{\prod_{m=0}^{2n-1}(m+1)}}=$$

$$=kz\frac{1+\sum_{n=1}^{+\infty}z^{2n}\frac{\frac{\Gamma(k-1+1)}{\Gamma(k-1+1-2n)}(-1)^{2n}}{\frac{\Gamma(2+2n)}{\Gamma(2)}}}{1+\sum_{n=1}^{+\infty}z^{2n}\frac{\frac{\Gamma(k+1)}{\Gamma(k-2n+1)}(-1)^{2n}}{\frac{\Gamma(1+2n)}{\Gamma(1)}}}=kz\frac{1+\sum_{n=1}^{+\infty}z^{2n}\frac{\Gamma(k)}{\Gamma(2+2n)\Gamma(k-2n)}}{1+\sum_{n=1}^{+\infty}z^{2n}\frac{\Gamma(k+1)}{\Gamma(k-2n+1)\Gamma(1+2n)}} \quad (A.144)$$

which is the same as (A.143). And from (A.139) follows that the left-hand side of (A.144) is equal to the right-hand side of (A.142). **Q.E.D.**

Theorem A. 8: The following continued fraction expansion holds:

$$\frac{(x+1)^k}{(x-1)^k}-1=\frac{2k}{x-k-}\,\frac{1^2-k^2}{3x-}\,\frac{2^2-k^2}{5x-}\,\frac{3^2-k^2}{7x-}\,\frac{4^2-k^2}{9x-}\,\frac{5^2-k^2}{11x-}\cdots \quad (A.145)$$

Proof: (A.142) is equivalent to

$$\frac{(1+z)^k-(1-z)^k}{(1+z)^k+(1-z)^k}=\cfrac{k}{\cfrac{1}{z}-\cfrac{1^2-k^2}{\cfrac{3}{z}-\cfrac{2^2-k^2}{\cfrac{5}{z}-\cfrac{3^2-k^2}{\cfrac{7}{z}-\cfrac{4^2-k^2}{\cfrac{9}{z}-\cfrac{5^2-k^2}{\cfrac{11}{z}-\dots}}}}}} \quad (A.146)$$

and by letting $x=\frac{1}{z}$ we get

$$\cfrac{k}{x-\cfrac{1^2-k^2}{3x-\cfrac{2^2-k^2}{5x-\cfrac{3^2-k^2}{7x-\cfrac{4^2-k^2}{9x-\cfrac{5^2-k^2}{11x-\dots}}}}}}=\frac{\left(1+\frac{1}{x}\right)^k-\left(1-\frac{1}{x}\right)^k}{\left(1+\frac{1}{x}\right)^k+\left(1-\frac{1}{x}\right)^k}=\frac{(x+1)^k-(x-1)^k}{(x+1)^k+(x-1)^k} \quad (A.147)$$

So as to simplify notation, let

$$N=\cfrac{1^2-k^2}{3x-\cfrac{2^2-k^2}{5x-\cfrac{3^2-k^2}{7x-\cfrac{4^2-k^2}{9x-\cfrac{5^2-k^2}{11x-\dots}}}}} \tag{A.148}$$

Hence (A.147) results in

$$\frac{k}{x-N}=\frac{(x+1)^k-(x-1)^k}{(x+1)^k+(x-1)^k} \Leftrightarrow N=x-\frac{(x+1)^k+(x-1)^k}{(x+1)^k-(x-1)^k}k \tag{A.149}$$

The right-hand side of (A.145) is

$$\frac{2k}{x-k-N}=\frac{2k}{x-k-x+\frac{(x+1)^k+(x-1)^k}{(x+1)^k-(x-1)^k}k}=\frac{2}{-1+\frac{(x+1)^k+(x-1)^k}{(x+1)^k-(x-1)^k}}=$$

$$=\frac{2}{\frac{-(x+1)^k+(x-1)^k+(x+1)^k+(x-1)^k}{(x+1)^k-(x-1)^k}}=\frac{(x+1)^k-(x-1)^k}{(x-1)^k}=\frac{(x+1)^k}{(x-1)^k}-1 \tag{A.150}$$

which is the left-hand side of (A.145). **Q.E.D.**

Corollary: By letting $z^{-1}=-\frac{1}{x}$ in (A.145) we get[122]

$$\left(\frac{1-z^{-1}}{1+z^{-1}}\right)^{\nu}=1+\frac{2\nu}{-\frac{1}{z^{-1}}-\nu-}\frac{1-\nu^2}{-\frac{3}{z^{-1}}-}\frac{4-\nu^2}{-\frac{5}{z^{-1}}-}\frac{9-\nu^2}{-\frac{7}{z^{-1}}-}\dots=$$

$$=\left[1;\frac{2\nu}{-\frac{1}{z^{-1}}-\nu},\left\{\frac{\nu^2-i^2}{-\frac{2i+1}{z^{-1}}}\right\}\right]_{i=1}^{+\infty} \tag{A.151}$$

(Once more, the curly brackets intend to show that only the last term is to be repeated with each iteration of i.)

A.5. An important trigonometric relation

Lemma A. 4: The tangent of a sum is given by

[122] Again, for full mathematical rigour, we have to impose $z^{-1}\in[-1;+1]$ (Wall, p. 346-347, 1948); otherwise, the continued fraction will not converge.

$$\operatorname{tg}\left(\operatorname{arctg} x_1 + \operatorname{arctg} x_2\right) = \frac{x_1 + x_2}{1 - x_1 x_2} \tag{A.152}$$

Theorem A. 9: The tangent of a summation is given by

$$\operatorname{tg}\left(\sum_{i=1}^{a} \operatorname{arctg} x_i\right) = \frac{\sum_{k \in I_a} (-1)^{\mathscr{C}(k/2)} {}_a\mathscr{S}_k}{1 + \sum_{k \in P_a} (-1)^{\mathscr{C}(k/2)} {}_a\mathscr{S}_k} \tag{A.153}$$

where I_a and P_a are the sets of odd and even naturals smaller than or equal to a

$$I_a = \left\{n \in \mathbb{N} : n \le a \wedge \exists_{p \in \mathbb{N}} n = 2p - 1\right\} \tag{A.154}$$

$$P_a = \left\{n \in \mathbb{N} : n \le a \wedge \exists_{p \in \mathbb{N}} n = 2p\right\} \tag{A.155}$$

and

$$ {}_a\mathscr{S}_1 = x_1 + x_2 + \ldots + x_a \tag{A.156}$$

$$\begin{aligned} {}_a\mathscr{S}_2 = {} & x_1 x_2 + x_1 x_3 + \ldots + x_1 x_a + \\ & + x_2 x_3 + \ldots + x_2 x_a + \\ & + \ldots + \\ & + x_{a-1} x_a \end{aligned} \tag{A.157}$$

$$\begin{aligned} {}_a\mathscr{S}_3 = {} & x_1 x_2 x_3 + x_1 x_2 x_4 + \ldots + x_1 x_2 x_a + \\ & + x_1 x_3 x_4 + \ldots + x_1 x_3 x_a + \\ & + \ldots + \\ & + x_1 x_{a-1} x_a + \\ & + x_2 x_3 x_4 + \ldots + x_2 x_3 x_a + \\ & + \ldots + \\ & + x_2 x_{a-1} x_a + \\ & \vdots \\ & + x_{a-2} x_{a-1} x_a \end{aligned} \tag{A.158}$$

and so on; or, generally,

$$ {}_a\mathscr{S}_k = \sum \prod_{i \in \mathscr{P}_{a,k}} x_i \tag{A.159}$$

$$\begin{cases} \#\left(\mathscr{P}_{a,k}\right) = k \\ \mathscr{P}_{a,k} \subset \left\{n \in \mathbb{N} : n \le a\right\} \end{cases} \tag{A.160}$$

Proof: This is proved by mathematical induction. For $a = 2$,

$$\operatorname{tg}\left(\sum_{i=1}^{2}\operatorname{arctg} x_i\right)=\frac{\sum_{k\in\{1\}}(-1)^{\mathscr{E}(k/2)}\,{}_2\mathscr{S}_k}{1+\sum_{k\in\{2\}}(-1)^{\mathscr{E}(k/2)}\,{}_2\mathscr{S}_k}=\frac{(-1)^0\,{}_2\mathscr{S}}{1+(-1)^1\,{}_2\mathscr{S}_2}=\frac{x_1+x_2}{1-x_1x_2} \tag{A.161}$$

The inductive step is proved applying (A.152) and (A.153) as follows:

$$\operatorname{tg}\left(\operatorname{arctg} x_{a+1}+\sum_{i=1}^{a}\operatorname{arctg} x_i\right)=\frac{x_{a+1}+\dfrac{\sum_{k\in I_a}(-1)^{\mathscr{E}(k/2)}\,{}_a\mathscr{S}_k}{1+\sum_{k\in P_a}(-1)^{\mathscr{E}(k/2)}\,{}_a\mathscr{S}_k}}{1-x_{a+1}\dfrac{\sum_{k\in I_a}(-1)^{\mathscr{E}(k/2)}\,{}_a\mathscr{S}_k}{1+\sum_{k\in P_a}(-1)^{\mathscr{E}(k/2)}\,{}_a\mathscr{S}_k}}=$$

$$=\frac{x_{a+1}+\sum_{k\in P_a}(-1)^{\mathscr{E}(k/2)}x_{a+1}\,{}_a\mathscr{S}_k+\sum_{k\in I_a}(-1)^{\mathscr{E}(k/2)}\,{}_a\mathscr{S}_k}{1+\sum_{k\in P_a}(-1)^{\mathscr{E}(k/2)}\,{}_a\mathscr{S}_k-\sum_{k\in I_a}(-1)^{\mathscr{E}(k/2)}x_{a+1}\,{}_a\mathscr{S}_k} \tag{A.162}$$

The three terms in the numerator of (A.162) sum up to $\sum_{k\in I_{a+1}}(-1)^{\mathscr{E}(k/2)}\,{}_{a+1}\mathscr{S}_k$ and the last two terms in the denominator sum up to $\sum_{k\in P_{a+1}}(-1)^{\mathscr{E}(k/2)}\,{}_{a+1}\mathscr{S}_k$; hence

$$\operatorname{tg}\left(\operatorname{arctg} x_{a+1}+\sum_{i=1}^{a}\operatorname{arctg} x_i\right)=\frac{\sum_{k\in I_{a+1}}(-1)^{\mathscr{E}(k/2)}\,{}_{a+1}\mathscr{S}_k}{1+\sum_{k\in P_{a+1}}(-1)^{\mathscr{E}(k/2)}\,{}_{a+1}\mathscr{S}_k} \tag{A.163}$$

Q.E.D.

A.6. Minimisation algorithms

In this section two minimisation algorithms are described: the Nelder-Mead simplex algorithm and genetic algorithms. They both share the important characteristics of being iterative and of not requiring (unlike other methods) the knowledge of the derivative of the function to minimise.

A.6.1. The Nelder-Mead simplex algorithm

A *simplex* is a geometrical figure in $\mathbb{R}^n$ defined by n + 1 vertices. (This corresponds to a *triangle* in $\mathbb{R}^2$ and to a *tetrahedron* in $\mathbb{R}^3$.) A non-degenerate simplex encloses a non-null n-dimensional volume. (A 2-dimensional volume is an area; a 3-dimensional volume is a volume in the ordinary sense of the word.)

The Nelder-Mead simplex algorithm uses a simplex to sweep the solution space looking for a minimum. (Other types of optimisation problems may be dealt with by reconverting them into minimisation problems.) A short description follows[123]. J is the function to minimise, or objective function[124].

❖ *Initialisation.* The algorithm begins with a non-degenerate simplex, the $n + 1$ vertices of which are possible solutions of the minimisation problem. If some good candidates to solution are known it is reasonable to include them. Other vertices may be distributed around such candidates or may be randomly found within the space of possible solutions.

❖ Values for four coefficients must be established beforehand. These are the *reflection coefficient* ρ, the *expansion coefficient* χ, the *contraction coefficient* γ and the *shrinkage coefficient* σ. They must verify the following relations:

$$\rho > 0 \tag{A.164}$$
$$\chi > 1 \tag{A.165}$$
$$\chi > \rho \tag{A.166}$$
$$0 < \gamma < 1 \tag{A.167}$$

It is usual to choose the following values:

$$\rho = 1 \tag{A.168}$$
$$\chi = 2 \tag{A.169}$$
$$\gamma = \frac{1}{2} \tag{A.170}$$
$$\sigma = \frac{1}{2} \tag{A.171}$$

❖ *Algorithm termination.* Several iterations will then be performed. Each consists in finding a new simplex, different from the former in at least one vertex. Iterations may stop when an individual is found with a value of J below some threshold, when several consecutive iterations fail to significantly improve results, or after some fixed number of iterations.

❖ *Iterations.* The first thing to do in each iteration is to order the several vertexes $\mathbf{x}$ according to their value of the objective function, so that

$$J(\mathbf{x}_1) \le J(\mathbf{x}_2) \le \ldots \le J(\mathbf{x}_n) \le J(\mathbf{x}_{n+1}) \tag{A.172}$$

Of course $\mathbf{x}_1$ will be the best point and $\mathbf{x}_{n+1}$ the worst.

❖ The *centroid* of the n best points is reckoned:

$$\overline{\mathbf{x}} = \sum_{i=1}^{n} \mathbf{x}_i \tag{A.173}$$

---

[123] Lagarias *et al.* (1998); Press *et al.* (1992, p. 408-412); Jang (1997, p. 189-193). Different authors describe this method with slight variations.

[124] When the algorithm is used with the purpose of identifying a fractional model from a time response, as mentioned in section 4.3, the objective function will decrease when the appropriate time response of the individual gets closer to the data we want to fit.

❖ The *reflection point* is reckoned:

$$\mathbf{x}_r = \bar{\mathbf{x}} + \rho\left(\bar{\mathbf{x}} - \mathbf{x}_{n+1}\right) \tag{A.174}$$

There are now four different possibilities depending on the value of $J\left(\mathbf{x}_r\right)$.

❖ *First*. If $J\left(\mathbf{x}_1\right) \le J\left(\mathbf{x}_r\right) < J\left(\mathbf{x}_n\right)$, vertex $\mathbf{x}_{n+1}$ is replaced by $\mathbf{x}_r$ and we proceed to the next iteration.

❖ *Second*. If $J\left(\mathbf{x}_r\right) < J\left(\mathbf{x}_1\right)$, the *expansion point* is reckoned:

$$\mathbf{x}_e = \bar{\mathbf{x}} + \chi\left(\mathbf{x}_r - \bar{\mathbf{x}}\right) \tag{A.175}$$

Then, if $J\left(\mathbf{x}_e\right) < J\left(\mathbf{x}_r\right)$, vertex $\mathbf{x}_{n+1}$ is replaced by $\mathbf{x}_e$; if $J\left(\mathbf{x}_r\right) \le J\left(\mathbf{x}_e\right)$, vertex $\mathbf{x}_{n+1}$ is replaced by $\mathbf{x}_r$. In both cases we proceed to the next iteration.

❖ *Third*. If $J\left(\mathbf{x}_n\right) \le J\left(\mathbf{x}_r\right) < J\left(\mathbf{x}_{n+1}\right)$, the *outside contraction point* is reckoned:

$$\mathbf{x}_{oc} = \bar{\mathbf{x}} + \gamma\left(\mathbf{x}_r - \bar{\mathbf{x}}\right) \tag{A.176}$$

If $J\left(\mathbf{x}_{oc}\right) \le J\left(\mathbf{x}_r\right)$, vertex $\mathbf{x}_{n+1}$ is replaced by $\mathbf{x}_{oc}$ and we proceed to the next iteration. But if $J\left(\mathbf{x}_r\right) < J\left(\mathbf{x}_{oc}\right)$ a shrink will be performed (see below).

❖ *Fourth*. If $J\left(\mathbf{x}_r\right) \ge J\left(\mathbf{x}_{n+1}\right)$, the *inside contraction point* is reckoned:

$$\mathbf{x}_{ic} = \bar{\mathbf{x}} - \gamma\left(\bar{\mathbf{x}} - \mathbf{x}_{n+1}\right) \tag{A.177}$$

If $J\left(\mathbf{x}_{ic}\right) < J\left(\mathbf{x}_{n+1}\right)$, vertex $\mathbf{x}_{n+1}$ is replaced by $\mathbf{x}_{ic}$ and we proceed to the next iteration. But if $J\left(\mathbf{x}_{n+1}\right) \le J\left(\mathbf{x}_{ic}\right)$ a shrink will be performed (see below).

❖ *Shrink*. If one of the two conditions mentioned in points three and four is verified, vertices $\mathbf{x}_2$ to $\mathbf{x}_{n+1}$ are replaced by new vertexes given by

$$\tilde{\mathbf{x}}_i = \mathbf{x}_1 + \sigma\left(\mathbf{x}_i - \mathbf{x}_1\right), \quad i = 2, \ldots, n+1 \tag{A.178}$$

Only vertex $\mathbf{x}_1$ is kept. Then we proceed to the next iteration.

This algorithm can (and should) be completed by some criterion to take into account what the order of vertices will be if there are several equal values of *J* in the original simplex or after a shrink.

A.6.2. Genetic algorithms

In few words, a genetic algorithm is an iterative optimisation method, consisting in a refined trial and error search, and inspired in the evolutionary principle of survival of the fittest.

A general description of a genetic algorithm follows[125]. It is suited for a minimisation problem; other types of optimisation are handled similarly or reconverted to a minimisation problem. J will again be the function to minimise, or objective function.

❖ *Initialisation.* First of all a population is created. A population is a set of individuals. Each individual is a possible solution of the minimisation problem. When the purpose is to identify a fractional model, each individual will consist in a set of parameters of the model that are to be minimise.

❖ The initial population is usually randomly generated. If some good candidates to solution are known it is reasonable to include them. Other individuals may be distributed around such candidates or may be uniformly spread over the space of possible solutions. This algorithm easily copes with constraints in values variables may assume: impossible solutions should never exist in the population; if, for instance, there is a parameter which is known to have to vary within some range, no individual in the population is to be allowed to have an out-of-range parameter.

❖ There are basically two ways of storing individuals. The original versions of genetic algorithms stored numbers in binary format. With such a codification, each individual consisted of several binary numbers, stored in sequence, just as each living individual (in the real, biological world) has a unique binary number codified in the nuclei of its cells in DNA molecules[126]. The sequence of binary numbers is called the genetic code of the individual, or its chromosome[127]. Recently decimal codification is being increasingly used. That alternative will be described below.

❖ *Algorithm termination.* Several iterations will then be performed. Each consists in finding a new population, from the population of the previous iteration. It is expected that with each passing iteration there will be more and more individuals in the population corresponding to low values of the objective function. Iterations may stop when an individual is found with a value of the objective function below some threshold, when several consecutive iterations fail to significantly improve results, or after some fixed number of iterations. To keep the biological analogy, iterations are also called generations.

❖ *Iterations.* The population of a given generation (after the first one) is obtained from that of the previous generation by four means: elitism, crossover, mutation and spontaneous generation.

❖ *Elitism.* This consists in allowing the best-performing individuals of one generation to make it into the next one. (The original genetic algorithms did not allow any individual to survive from one generation to the other. But this meant that some very good individuals—potential solutions, probably close to acceptable values—could be discarded.) There should be not too many individuals qualifying for this, otherwise solutions will hardly improve.

❖ *Crossover.* This consists in picking two individuals from the previous iteration (the parents) and combining them into one or two new individuals. The simplest way to do this is to cut the two genetic codes after a certain number of bits and swapping the

---

[125] Jang (1997, p. 174-180).

[126] In DNA (deoxyribonucleic acid) macromolecules, pairs of molecules of adenine and thymine and pairs of molecules of guanine and cytosine alternate. If we attribute value 1 to one of these pairs and value 0 to the other, each DNA molecule will correspond to a binary number—a very long one with milliards of digits. (Asimov, p. 578-586, 1987)

[127] This in spite of most living beings having more than one chromosome; humans, for instance, have forty-six. (Asimov, p. 559, 1987)

two first halves. The two resulting recombinations are called the offspring, the children or the descendents. (One of them may be eliminated if only one descendent is sought.) It is reasonable to cut the genetic code somewhere that makes sense; for instance, after one of the parameters (and not in the middle of some number). This is called one-point crossover, because there is one cut-point for each parent, and so parents are split into two parts. It is possible to conceive a two-point crossover, splitting the genetic codes of the parents into three parts; or a three-point crossover, and so forth. It is even possible to conceive a crossover involving three or even more parents, each descendent having some part of the genetic code of each of the several parents, but this is seldom done because it is complicated and unnecessary. Parents should be chosen within the population according to their value of the objective function. An idea is to have the probability of an individual qualifying for parent increase as the objective function decreases. Another is to specify some threshold and letting all individuals with values of the objective function below that threshold be parents.

❖ *Mutation.* This consists in picking some of the descendents obtained by crossover and then changing them randomly, in the hope that the change will improve performance. The idea is to allow each bit to mutate (pass from 0 to 1 or from 1 to) with a certain probability. Bits that mutate may change one (or more) of the parameters marginally or severely depending on the position they have. Mutation probability should be low; otherwise the algorithm will fall short of nothing more than a random search (with a consequent loss of efficiency of the algorithm). But as an alternative, mutations may be applied to individuals from the previous generation. In that case, mutation probability has to be higher (otherwise individuals would pass from one iteration to the other without change), and individuals that mutate are to be selected among the best performing of the previous generation. The probability of an individual being chosen to mutate may increase as its objective function gets lower, or a threshold may be chosen and all individuals better than the threshold be selected to mutate.

❖ *Spontaneous generation.* This consists in creating some new individuals out of nothing, as was done for the first generation. Spontaneous generation allows a more thorough exploration of the space of possible solutions, but should not be abused under penalty of the algorithm becoming little more than a random search[128].

❖ The number of individuals in each generation may be fixed, those discarded being always replaced, or allowed to vary. It is not a good idea to let it *increase*, for that would increase the computational burden, but as performances get better it may be unnecessary to keep as many as before, and the population may be allowed to get less numerous.

When decimal codification is used, the algorithm is changed as follows:

❖ Numbers are stored in decimal format, and each individual consists in a set of decimal numbers[129].

❖ Mutation consists in changing one of the parameters (or eventually more) randomly. A random number with a zero mean distribution is generated and added to some randomly chosen parameter; or some randomly chosen parameter is multiplied by

---

[128] Valério *et al.* (2003c); Valério *et al.* (2004d).

[129] Of course computers store *everything* as binary numbers—*internally*. The question here is how the *programmer* will have access to the data. With binary codification access to each bit of information is assumed—and necessary. With decimal codification numbers may well be internally represented as binary numbers, but the programmer will not try to access bits individually and will treat decimal numbers for what they are.

a random number with a distribution centred on one.

❖ Crossover consists essentially in the same, but now it is really impossible to split genetic codes in unsuitable places.

One of the best ways to understand genetic algorithms is to analyse one. Some that may be taken as examples are given above in subsections 6.1.3 (using binary codification), 6.2.3 (using decimal codification) and 6.4.2 (also using decimal codification but differing significantly from the former on the order of steps).

It is also to be noticed that tuning the parameters of the algorithm (such as mutation probabilities, number of elements in each generation, and so on) is—as is always the case with methods such as this one—an important issue for which some trial and error is necessary. Selecting more general features of the algorithm (presence or absence of elitism, number of points for crossover…) is also a matter for which more than one solution is admissible.

Appendix B. The Ninteger toolbox

This appendix briefly describes Ninteger, a toolbox for MatLab intended to help developing fractional order controllers and assess their performance.

The toolbox was developed alongside with the work reported in the rest of this thesis, for validating results, testing alternative methods, developing controllers, plotting figures, and so forth. It was thought good to make it available to other researchers in the area. The address of the support site in the Internet, where the code is available together with a manual[130], is

```
http://www.gcar.dem.ist.utl.pt/pessoal/dvalerio/ninteger/ninteger.htm
```

The toolbox is free to encourage academic use of fractional control, and because in this wise it is easier to profit from suggestions and improvements from other users. Its code may be freely distributed and altered (with mild restraints essentially related to acknowledging the source).

The version of the toolbox currently available is version 2.3. The toolbox requires MatLab version 6.5 or above (actually most functions run under version 5, but the graphical interface does not), together with the Control and Optimisation toolboxes. It is expected that it may be of use to several researchers in the area, and that the feedback they may give will help improving it[131].

B.1. Approximations of fractional derivatives

The Ninteger toolbox implements all approximations given in section 3.3. An attempt was made to ask the user for the same information irrespective of the approximation chosen, so that the results may be easily compared. A function to compute numerical approximations of a function's fractional derivatives using Theorem 2.8 is also provided.

B.2. Frequency responses of fractional plants

The Ninteger toolbox implements functions for finding the frequency response of fractional transfer functions and plotting it in Bode, Nichols and Nyquist diagrams. These functions are the fractional counterparts of already-existing Matlab functions for integer plants.

B.3. Identification of fractional models

The Ninteger toolbox implements all identification methods given in chapter 4. Numerical optimisation methods mentioned in subsections 4.3 and A.6 are of course not

[130] Valério (2005).

[131] It is to be noticed that there are other software packages for fractional control available. The Crone toolbox for Matlab (Cois *et al.*, 2002; Oustaloup *et al.*, 2002), developed by the Crone team at the University of Bordeaux, is probably the best known one. That toolbox is of much superior quality, but is not free.

implemented; there are good, easily available implementations thereof already.

B.4. Synthesis of fractional controllers

The Ninteger toolbox implements Crone controllers of all three generations as described in sections 5.1 to 5.3. Approximations of fractional PIDs based upon the implementations of approximations of fractional derivatives described above in section B.1 are also provided. And functions for finding the H_2 and H_∞ norms of fractional plants as documented in section 5.5 are also given.

B.5. Graphical interface

The kernel of the Ninteger toolbox consists of functions that may be called from the command prompt. This is a sound choice that allows implemented algorithms to be easily reused in other programs and functions. On the other hand, a graphical interface may assist the user in choosing parameters, making reasonable suggestions, and immediately showing the results achieved, allowing an interactive design of controllers. That is why the toolbox includes a graphical interface built upon the kernel of command prompt functions, for assisting the user and rending the fine-tuning of parameters easier.

Some screenshots of this graphical interface may be seen in the figures below.

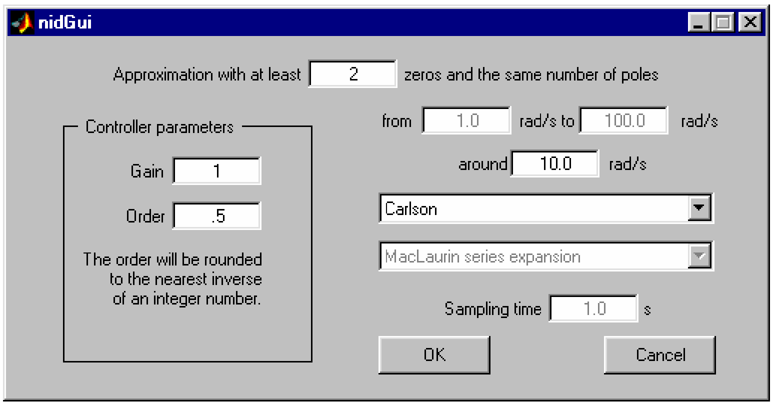


Figure B. 1 — Dialogue for building an approximation of (3.18)

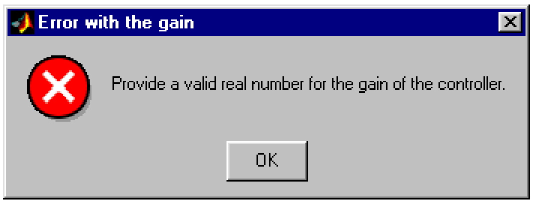


Figure B. 2 — Example of an error message

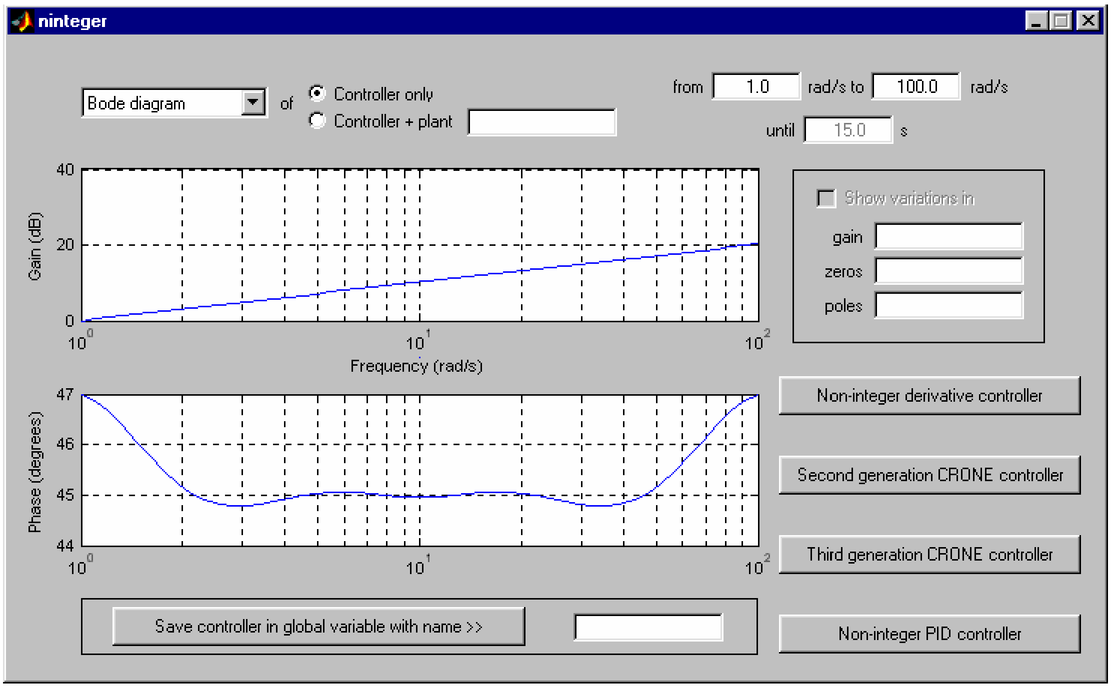


Figure B. 3 — The main dialogue displaying a Bode diagram of an approximation of (3.18)

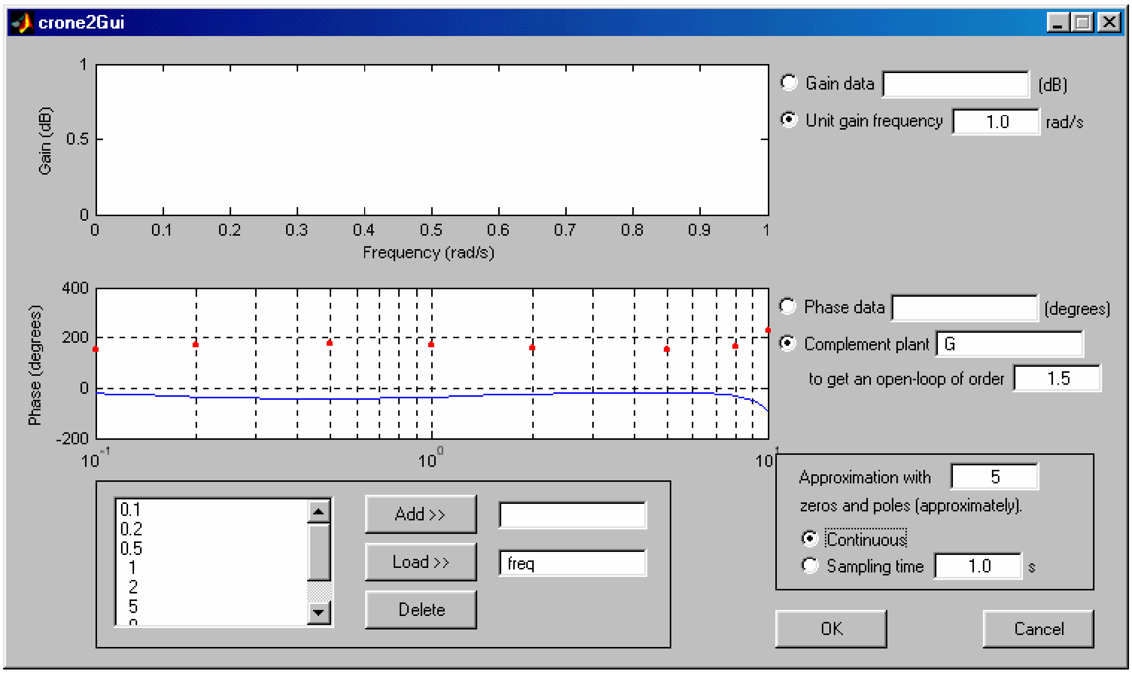


Figure B. 4 — Dialogue for building a second generation Crone controller

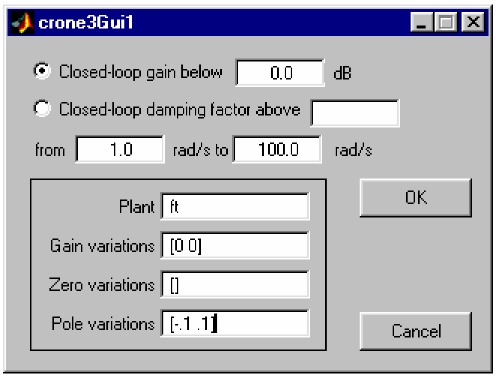


Figure B. 5 — A dialogue for building a third generation Crone controller

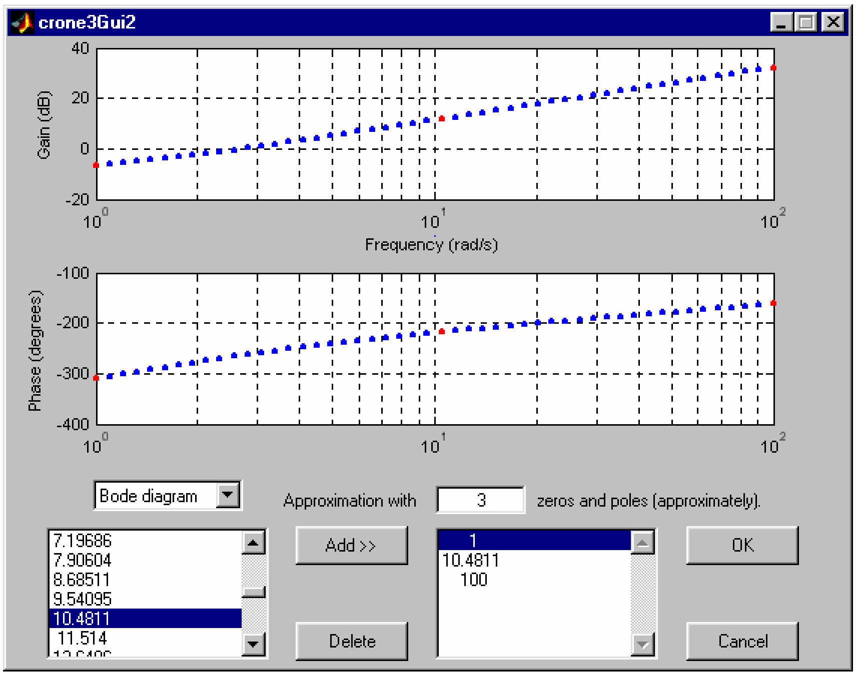


Figure B. 6 — Another dialogue for building a third generation Crone controller

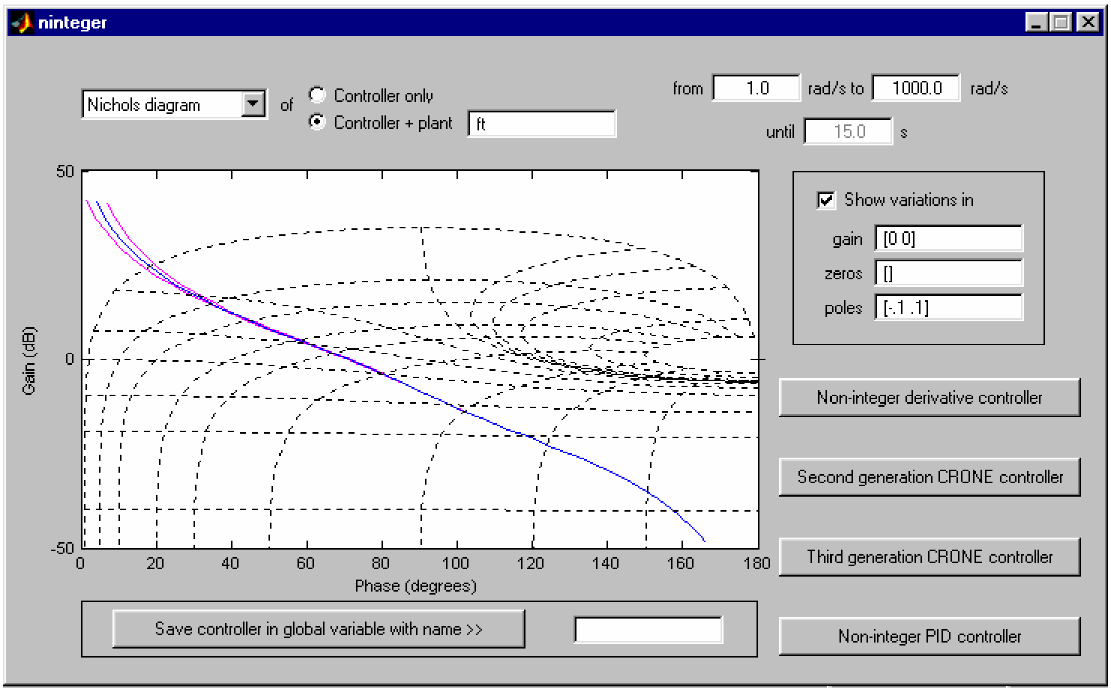


Figure B. 7 — The main dialogue displaying a Bode diagram of an open-loop consisting of a third generation Crone controller and a plant with a pole known with uncertainty

B.6. Simulink library

The Ninteger toolbox includes a library with Simulink blocks for implementing a general fractional transfer function and a fractional PID controller. Such blocks make use of the command prompt functions described above.

B.7. Continued fractions

Since several of the functions described above make use of continued fractions, the Ninteger toolbox includes functions for expanding real numbers and rational functions into continued fractions and for evaluating continued fractions.

Bibliography

More than these, my son, require not. Of making many books there is no end: and much study is an affliction of the flesh.

Ecclesiastes xii, 12 (Douay-Rheims)

ABRAMOWITZ, Milton; STEGUN, Irene A. — Handbook of mathematical functions with formulas, graphs, and mathematical tables. [S.l.]: Dover Books, 1970.

ALMEIDA, Guilherme de — Sistema Internacional de unidades (SI): grandezas e unidades físicas, terminologia, símbolos e recomendações. Lisboa: Plátano, 2002.

APOSTOL, T. M. — Mathematical analysis. — London: Addison-Wesley, 1974.

ASIMOV, Isaac — Asimov's new guide to science. — Harmondsworth: Penguin Books, 1987.

BAGLEY, Ronald L.; TORVIK, Peter J. — Fractional calculus: a different approach to the analysis of viscoelastic damped structures. AIAA journal. 21:5 (1983) 741-748.

BARBOSA, Ramiro S.; MACHADO, J. A. Tenreiro; FERREIRA, Isabel M. — A fractional calculus perspective of PID tuning. In Proceedings of ASME 2003 design engineering technical conferences and Computers and information in engineering conference. Chicago: ASME, 2003.

BARBOSA, Ramiro S.; MACHADO, J. A. Tenreiro; FERREIRA, Isabel M. — Tuning of PID controllers based on Bode's ideal transfer function. Nonlinear dynamics. 38 (2004a) 305-321.

BARBOSA, Ramiro S.; MACHADO, J. A. Tenreiro; FERREIRA, Isabel M. — PID controller tuning using fractional calculus concepts. Fractional calculus & applied analysis. 7:2 (2004b) 119-134.

BENNET, Stuart — A brief history of automatic control. IEEE control systems. 16:3 (1996) 17-25.

BENOSMAN, M.; BOYER, F.; LE VEY, G. — Flexible links manipulators: from modeling to control. Journal of intelligent and robotic systems. 34 (2002) 381-414.

BENSON, David; TADJERAN, Charles; MEERSCHAERT, Mark M.; FARNHAM, Irene; POHLL, Greg — Radial fractional-order dispersion through fractured rock. Water resources research. 40 (2004) 1-9. http://www.maths.otago.ac.nz/~mcubed/frade_wrr.pdf

BESKIN, N. M. — Fracções contínuas. Lisboa: Ulmeiro, 2001. Translation.

CALDERÓN, A. J.; VINAGRE, B. M.; FELIÚ, V. — Fractional sliding mode control of a DC-DC buck converter with application to DC motor drives. In Proceedings of the

11th international conference on advanced robotics. Coimbra: IEEE, 2003. p. 252-257.

CANDAN, Çağatay; OZAKTAS, Haldun M. — Sampling and series expansion theorems for fractional Fourier and other transforms. Signal processing. 83 (2003) 2455-2457.

CAPONETTO, R.; FORTUNA, L.; PORTO, D. — Parameter tuning of a non integer order PID controller. In Electronic proceedings of the fifteenth international symposium on mathematical theory of networks and systems. 2002. http://www.nd.edu/~mtns/papers/7434.pdf

CAPONETTO, Riccardo; FORTUNA, Luigi; PORTO, Domenico — A new tuning strategy for a non integer order PID controller. In First IFAC workshop on fractional differentiation and its applications. Bordeaux: IFAC, 2004.

CARLSON, G. E.; HALIJAK, C. A. — Approximation of fractional capacitors $(1/s)^{1/n}$ by a regular Newton process. IEEE transactions on circuit theory. 7 (1964) 210-213.

CHEN, W. — A new definition of the fractional Laplacian. ArXiv Computer Science e-prints. 2002. http://arxiv.org/ftp/cs/papers/0209/0209020.pdf

CHEN, W.; HOLM, S. — Physical interpretation of fractional diffusion-wave equation via lossy media obeying frequency power law. ArXiv Mathematical Physics e-prints. 2003. http://arxiv.org/ftp/math-ph/papers/0303/0303040.pdf

CHEN, YangQuan; VINAGRE, Blas M. — A new IIR-type digital fractional order differentiator. Signal processing. 83 (2003) 2359-2365.

CHEN, YangQuan; MOORE, Kevin L.; VINAGRE, Blas M.; PODLUBNY, Igor — Robust PID controller autotuning with a phase shaper. In First IFAC workshop on fractional differentiation and its applications. Bordeaux: IFAC, 2004.

CHIPPERFIELD, Andrew; FLEMING, Peter; POHLHEIM, Hartmut; FONSECA, Carlos — Genetic algorithm toolbox for use with Matlab; version 1.2 user's guide. Sheffield: University of Sheffield, 1994.

COIS, O.; LANUSSE, P.; MELCHIOR, P.; DANCLA, F.; OUSTALOUP, A. — Fractional systems toolbox for Matlab: applications in system identification and Crone CSD. In 41st IEEE conference on decision and control. Las Vegas: IEEE, 2002.

COOPER, Gordon; COWAN, Duncan — The application of fractional calculus to potential field data. Exploration geophysics. 34 (2003) 51-56.

DAUPHIN, Gabriel — Application des représentations diffusives au temps discret. Paris: École Nationale Supérieure des Télécomunications, 2001. Doctorate thesis.

DEL-CASTILLO-NEGRETE, D.; CARRERAS, B. A.; LYNCH, V. E. — Fractional diffusion in plasma turbulence. Physics of plasmas. 11:8 (2004) 3584.

DIAZ, J. B.; OSLER, T. J. — Differences of fractional order. Mathematics of computation. 28:125 (1974) 185-202.

FELLAH, Z. E. A.; BERGER, S.; LAURIKS, W.; DEPOLLIER, C. — Verification of Kramers-Kronig relationship in porous materials having a rigid frame. Journal of sound and vibration. 270 (2004) 865-885.

FERREIRA, N. M. Fonseca; MACHADO, J. A. Tenreiro — Fractional-order hybrid control of robotic manipulators. In Proceedings of the 11th international conference on advanced robotics. Coimbra: IEEE, 2003. p. 393-398.

GARCÍA-FINÃNA, Marta; CRUZ-ORIVE, Luis M. — Fractional trend of the variance in Cavalieri sampling. Image analysis and stereology. 19:2 (2000) 71-79. http://www.general.monash.edu.au/ss/pdf/vol19no2/Garcia_Finana_001.pdf

GOODWIN, Graham Clifford; GRAEBE, Stefan F.; SALGADO, Mario E. — Control system design. Upper Saddle River: Prentice Hall, 2001.

GORENFLO, Rudolf; MAINARDI, Francesco; RABERTO, Marco; SCALAS, Enrico — Fractional diffusion in finance: basic theory. In Modelli dinamici in economia e finanza. Urbino: MDF, 2000. http://www.econ.uniurb.it/bischi/MDEF2000/MainardiMDEF.pdf

GORENFLO, Rudolf; VIVOLI, Alessandro — Fully discrete random walks for space-time fractional diffusion equations. Signal processing. 83 (2003) 2411-2420.

GRADSHTEYN, I. S.; RYZHIK, I. M.; JEFFREY, Allan — Table of integrals, series and products. Boston: Academic Press, 1980.

HÄGGLUND, Tore; ÅSTRÖM, Karl J. — Automatic tuning of PID controllers. In LEVINE, William S. — The control handbook. Boca Raton: CRC Press, 1996, p. 817-826.

HARTLEY, Tom T.; LORENZO, Carl F. — Fractional-order system identification based on continuous order-distributions. Signal processing. 83 (2003) 2287-2300.

HELMICKI, Arthur J.; JACOBSON, Clas A.; NETT, Carl N. — Control oriented system identification: a worst-case / deterministic approach in H_∞. IEEE transactions on automatic control. 36:10 (1991) 1163-1176.

HERRMANN, Felix — Quantitative tools for seismic stratigraphy and lithology characterization. In Annual report of the industry consortia of Earth Resources Laboratory. Cambridge: Massachussets Institute of Technology, 2001. http://www.eos.ubc.ca/~felix/Preprint/herrmann.pdf

HILDEBRAND, Francis Begnaud — Advanced calculus for applications. Englewood Cliffs: Prentice-Hall, 1976.

JANG, J.-S. R. — Derivative-free optimization. In JANG, J.-S. R.; SUN, C.-T.; MIZUTANI E. — Neuro-fuzzy and soft computing. Upper Saddle River: Prentice Hall, 1997, p. 173-196.

KLEINZ, Marcia; OSLER, Thomas J. — A child's garden of fractional derivatives. The college mathematics journal. 31:2 (2000) 82-88.

KHATIB, Ousssama — A unified approach for motion and force control of robot manipulators : the operations space formulation. IEEE journal of Robotics and Automation. RA-3 :1 (1987) 43-53.

LAGARIAS, J.C. ; REEDS, J. A.; WRIGHT, M. H. ; WRIGHT, P. E. — Convergence properties of the Nelder-Mead simplex method in low dimensions. SIAM journal of optimization. 9:1 (1998) 112-147.

LANDAU, I. D.; REY, D.; KARIMI, A.; VODA, A.; FRANCO, A. — A flexible transmission system as a benchmark for robust digital control. European journal of control. 1 (1995) 77-96.

LANUSSE, P.; POINOT, T.; COIS, O.; OUSTALOUP, A.; TRIGEASSOU, J. C. — Tuning of an active suspension system using a fractional controller and a closed-loop tuning. In Proceedings of the 11th international conference on advanced robotics. Coimbra: IEEE, 2003. 258-263.

LAVOIE, J. L.; OSLER, T. J.; TREMBLAY, R. — Fractional derivatives and special functions. SIAM review. 18:2 (1976) 240-268.

LAWRENCE, P. J.; ROGERS, G. J. — Sequential transfer-function synthesis from measured data. Proceedings of the IEE. 126:1 (1979) 104-106.

LEVY, E. — Complex curve fitting. IRE transactions on automatic control. 4 (1959) 37-43.

LORENZO, Carl F.; HARTLEY, Tom T. — Initialization in fractional order systems. In Proceedings of the European control conference. Porto: EUCA, 2001. p. 1471-1476.

LOVERRO, Adam — Fractional calculus: history, definitions and applications for the engineer. 2004. http://www.nd.edu/~msen/Teaching/UnderRes/FracCalc.pdf

LU, Silong; MOLZ, Fred J.; FIX, George J. — Possible problems of scale dependency in applications of the three-dimensional fractional advection-dispersion equation to natural porous media. Water resources research. 38:9 (2002) 1-9. http://www.geotransinc.com/publications/Possible_problems_of_scale.pdf

LUBLIN, L.; GROCOTT, S.; ATHANS, M. — H_2 (LQG) and H_∞ control. In LEVINE, William S. — The control handbook. Boca Raton: CRC Press, 1996, p. 651-661.

MACHADO, J. A. Tenreiro — Analysis and design of fractional-order digital control systems. Journal of systems analysis, modeling and simulation. 27 (1997) 107-122.

MACHADO, J. A. Tenreiro; AZENHA, Abílio — Fractional-order hybrid control of robot manipulators. In 1998 IEEE international conference on systems, man and cybernetics: intelligent systems for humans in a cyberworld. [S.l.]: IEEE, 1998. p. 788-

793.

MACHADO, J. A. Tenreiro — On the implementation of fractional-order control systems through discrete-time algorithms. In Bánki Donát Polytechnic műszaki főiskola: jubilee international conference proceedings. Bánki Donát: Bánki Donát Polytechnic, 1999. p. 39-44.

MACHADO, J. A. Tenreiro — Discrete-time fractional-order controllers. Fractional calculus & applied analysis. 4:1 (2001) 47-66.

MACHADO, J. A. Tenreiro — A probabilistic interpretation of the fractional-order differentiation. Fractional calculus & applied analysis. 6:1 (2003) 73-80.

MALTI, Rachid; AOUN, Mohamed; COIS, Olivier; OUSTALOUP, Alain; LEVRON, François — H_2 norm of fractional differential systems. In Proceedings of ASME 2003 design engineering technical conferences and Computers and information in engineering conference. Chicago: ASME, 2003.

MARTINS, Jorge Manuel Mateus — Modeling and identification of flexible manipulators towards robust control. Lisboa: Universidade Técnica de Lisboa, 2000. Masters thesis.

MATHIEU, B.; MELCHIOR, P.; OUSTALOUP, A.; CEYRAL, Ch. — Fractional differentiation for edge detection. Signal processing. 83 (2003) 2421-2432.

MATSUDA, K.; FUJII, H. — H_∞ optimized wave-absorbing control: analytical and experimental results. Journal of guidance, control and dynamics. 16:6 (1993) 1146-1153.

MILLER, Kenneth S.; ROSS, Bertram — An introduction to the fractional calculus and fractional differential equations. New York: John Wiley and Sons, 1993.

MONJE, C. A.; VINAGRE, B. M.; CHEN, Y. Q.; FELIU, V.; LANUSSE, P.; SABATIER, J. — Proposals for fractional $PI^\lambda D^\mu$ tuning. In First IFAC workshop on fractional differentiation and its applications. Bordeaux: IFAC, 2004.

MORARI, Manfred; ZAFIRIOU, Evanghelos — Robust process control. London: Prentice Hall, 1999.

NABAIS, João Miguel Lemos Chasqueira — Controlo por modo de deslizamento de um robô flexível. Lisboa: Universidade Técnica de Lisboa, 2002. Masters thesis.

NIGMATULLIN, Raoul R.; LE MEHAUTE, Alain — The geometrical and physical meaning of the fractional integral with complex exponent. International journal of science georesources. 1:8 (2004) 2-9.

OGATA, Katsuhiko — Modern control engineering. 3ª ed. London: Prentice Hall International, 1997.

ORTIGUEIRA, Manuel Duarte — Introduction to fractional linear systems II: discrete-time case. IEE proceedings — vision, image and signal processing. 147:1 (2000) 71-78. http://www2.uninova.pt/~mdortigueira/publ_files/fls_iee_d.pdf

ORTIGUEIRA, Manuel D.; MATOS, Carlos J. C. — New results on fractional linear prediction. Initialization in fractional order systems. In Proceedings of the European control conference. Porto: EUCA, 2001.

ORTIGUEIRA, Manuel Duarte — On the initial conditions in continuous-time fractional linear systems. Signal processing. 83 (2003) 2301-2309.

ORTIGUEIRA, Manuel D.; MACHADO, J. A. Tenreiro; SÁ DA COSTA, José — Which differintegration? IEE proceedings vision, Image & signal processing. 2005. Accepted for publication.

OUSTALOUP, Alain — La commande CRONE: commande robuste d'ordre non entier. Paris: Hermès, 1991.

OUSTALOUP, Alain; MATHIEU, B.; LANUSSE, P. — The CRONE control of resonant plants: application to a flexible transmission. European journal of control. 1 (1995) 113-121.

OUSTALOUP, Alain; LEVRON, François; MATHIEU, Benoît; NANOT, Florence M.. — Frequency-band complex noninteger differentiator: characterization and synthesis. IEEE transactions on circuits and systems I: Fundamental theory and applications. 47:1 (2000) 25-39.

OUSTALOUP, A.; MELCHIOR, P.; LANUSSE, P.; COIS, O.; DANCLA, F. — The CRONE toolbox for Matlab. In 41st IEEE conference on Decision and Control. Las Vegas: IEEE, 2002.

OUSTALOUP, A.; ORSONI, B.; MELCHIOR, P.; LINARÈS, H. — Path planning by fractional differentiation. Robotica. 21 (2003) 59-69.

PETRAS, I.; HYPIUSOVA, M. — Design of fractional-order controllers via H_∞ norm minimisation. In MIKLES, J.; VESELY, V. — Selected topics in modelling and control. Bratislava: Slovak University of Technology Press, 2002a. vol. 3, p. 50-54.

PETRÁŠ, Ivo; VINAGRE, Blas M. — Practical application of digital fractional-order controller to temperature control. Acta montanistica slovaca. 7:2 (2002b) 131-137 http://actamont.tuke.sk/pdf/2002/n2/11petras.pdf

PETRÁŠ, Ivo; VINAGRE, Blas M.; DORČÁK, L'ubomir; FELIU, Vincent — Fractional digital control of a heat solid: experimental results. In International carpathian control conference. Malenovice: ICCC, 2002. 365-370. http://www.icc-conf.cz/Conference/ICCC2002/Proceedings/papers/365.pdf

PINA, Heitor — Métodos numéricos. Lisboa: McGraw-Hill, 1995.

PODLUBNY, Igor — Fractional differential equations: an introduction to fractional derivatives, fractional differential equations, to methods of their solution and some of their applications. San Diego: Academic Press, 1999.

POMMIER, Valérie; MUSSET, Roland; LANUSSE, Patrick; OUSTALOUP, Alain — Study of two robust controls for an [sic] hydraulic actuator. In Proceedings of the European Control Conference. Cambridge: EUCA, 2003.

PRESS, William H.; TEUKOLSKY, Saul A.; VETTERLING, William T.; FLANNERY, Brian P. — Numerical recipes in C — the art of scientific computing. 2nd edition. Cambridge: Cambridge University Press, 1992. http://www.library.cornell.edu/nr/cbookcpdf.html

RADOK, Rainer — Continued fractions. 199[9?]. http://mpec.sc.mahidol.ac.th/numer/step1aa.htm

SABATIER, Jocelyn; MELCHIOR, Pierre; OUSTALOUP, Alain — Réalisation d'un banc d'essais thermique pour l'enseignement des systèmes non entiers. In Colloque sur l'enseignement des technologies et des sciences de l'information et des systèmes. Toulouse: Université Paul Sabatier, 2003. p. 361-364.

SAMKO, Stefan G.; KILBAS, Anatoly A.; MARICHEV, Oleg I. — Fractional integrals and derivatives: theory and applications. Yverdon: Gordon and Breach Science Publishers, 1993. Reviewed and improved translation of: Интегралы и производые дробного порядка и некоторые их прилжения.

SANATHANAN, C. K.; KOERNER, J. — Transfer function synthesis as a ratio of two complex polynomials. IEEE transactions on automatic control. 8 (1963) 56-58.

SCHOUKENS, J.; PINTELON, Rik; RENNEBOOG, J. — A maximum likelihood estimator for linear and nonlinear systems—a practical application of estimation techniques in measurement problems. IEEE transactions on instrumentation and measurement. 37:1 (1988) 10-17.

SILVA, C. A.; SOUSA, J. M.; RUNKLER, T.; SÁ DA COSTA, J. — A logistic process scheduling problem: genetic algorithms or ant colony optimization? in IFAC Congress. 2005. Accepted for presentation.

STANISLAVSKY, A. A. — Probability interpretation of the integral of fractional order. Theoretical and mathematical Physics. 138:3 (2004) 418-431.

SUÁREZ, J. I.; VINAGRE, B. M.; CHEN, Y. Q. — Spatial path tracking of na autonomous industrial vehicle using fractional order controllers. In Proceedings of the 11th international conference on advanced robotics. Coimbra: IEEE, 2003. p. 405-410.

TAJAHUERCE, Enrique; SAAVEDRA, Genaro; FURLAN, Walter D.; SICRE, Enrique E.; ANDRÉS, Pedro — White-light optical implementation of the fractional Fourier transform with adjustable order control. Applied optics. 39:2 (2000) 238-245.

TSENG, Chien-Cheng — Design of fractional order digital FIR differentiators. IEEE signal processing letters. 8:3 (2001) 77-79.

UNSER, Michael; BLU, Thierry — Fractional splines and wavelets. SIAM review. 42:1 (2000) 43-67.

VALÉRIO, Duarte — Fractional order robust control: an application. In Student Forum. Porto: EUCA, 2001a. p. 25-28.

VALÉRIO, Duarte Pedro Mata de Oliveira — Controlo robusto de ordem não inteira: síntese em frequência. Lisboa: Instituto Superior Técnico da Universidade Técnica de Lisboa, 2001b. Master's thesis.

VALÉRIO, Duarte; SÁ DA COSTA, José — Time domain implementations of fractional order controllers. In Controlo 2002. Aveiro: 5th Portuguese Conference on Automatic Control, 2002. p. 353-358.

VALÉRIO, Duarte; SÁ DA COSTA, José — Digital implementation of fractional control and its application to a two-link robotic arm. In Proceedings of the European control conference. Cambridge: EUCA, 2003a.

VALÉRIO, Duarte; SÁ DA COSTA, José — Optimisation of fractional order control parameters for a robotic arm. In Proceedings of the 11th international conference on advanced robotics. Coimbra: IEEE, 2003b. p. 393-398.

VALÉRIO, Duarte; SÁ DA COSTA, José — A method for identifying digital models and its application to non-integer order control. In Controlo 2004. Faro: 6th Portuguese Conference on Automatic Control, 2004a.

VALÉRIO, Duarte; SÁ DA COSTA, José — Ninteger: a non-integer control toolbox for Matlab. In First IFAC workshop on fractional differentiation and its applications. Bordeaux: IFAC, 2004b.

VALÉRIO, Duarte; SÁ DA COSTA, José — Non-integer control of a flexible robot. In First IFAC workshop on fractional differentiation and its applications. Bordeaux: IFAC, 2004c.

VALÉRIO, Duarte Pedro Mata de Oliveira — Ninteger v. 2.3: fractional control toolbox for MatLab. 2005. http://www.gcar.dem.ist.utl.pt/pessoal/dvalerio/ninteger/manual.pdf

VALÉRIO, Duarte; SÁ DA COSTA, José — Time-domain implementation of fractional order controllers. IEE proceedings: control theory & applications. 2005a. Accepted for publication.

VALÉRIO, Duarte; SÁ DA COSTA, José — Levy's identification method extended to commensurate fractional order transfer functions. In Fifth EUROMECH Nonlinear Dynamics conference. Eindhoven: EUROMECH, 2005b.

VALÉRIO, Duarte; SÁ DA COSTA, José — Ziegler-Nichols type tuning rules for fractional PID controllers. In Proceedings of ASME 2005 design engineering technical conferences and Computers and information in engineering conference. Long Beach: ASME, 2005c.

VALÉRIO, Duarte; SÁ DA COSTA, José — Fractional control of a flexible robot. In LE MEHAUTE, Alain; MACHADO, J. A. Tenreiro; TRIGEASSOU, Jean Claude; SABATIER, Jocelyn — Fractional differentiation and its applications. Bordeaux: LAPS, 2005d. Accepted for publication.

VALÉRIO, Duarte; SÁ DA COSTA, José — Tuning of fractional controllers minimising H_2 and H_∞ norms. Acta polytechnica hungarica. 2005e. Submitted for publication.

VALÉRIO, Duarte; SÁ DA COSTA, José — Tuning of fractional PID controllers with Ziegler-Nichols type rules. Signal processing. 2005f. Submitted for publication.

VINAGRE, B. M.; PODLUBNY, I.; HERNÁNDEZ, A.; FELIU, V. — Some approximations of fractional order operators used in control theory and applications. Fractional calculus & applied analysis. 3 (2000) 231-248.

VINAGRE JARA, Blas Manuel — Modelado y control de sistemas dinámicos caracterizados por ecuaciones íntegro-diferenciales de orden fraccional. Madrid: Universidad Nacional de Educación a Distancia, 2001. Doctorate thesis.

VINAGRE, B. M.; PETRAS, I.; MERCHAN, P.; DORCAK, L. — Two digital realizations of fractional controllers: application to temperature control of a solid. In Proceedings of the European control conference. Porto: EUCA, 2001. p. 1764-1767.

WALL, H. S. — Analytic theory of continued fractions. Princeton: D. Van Nostrand Company, 1948.

WANG, J.; BRAS, R. L. — Ground heat flux estimated from surface soil temperature. Journal of hydrology. 216 (1999) 214-226.

WHEELER, Nicholas — Construction and physical application of the fractional calculus. Reeds College Physics Seminar, March 5, 1997. Unpublished.

WIDDER, David V. — Advanced calculus. NewYork: Prentice-Hall, 1947.

ZIEGLER, J. G.; NICHOLS, N. B. — Optimum settings for automatic controllers. Transactions of the ASME. 64 (1942) 759-768.

> Achar se ham neste liuro algũs erros, assi de faltas de letras, como tãbem algũas mudadas: porem sam tã conhecidos os erros, que facilmente poderaa o discreto lector suprilos. E portanto se nam faz aqui errata delles porq̃ parece q̃ yr buscar o erro ao fim do liuro he cousa muy prolixa.
>
> *Copilaçam de todalas obras de Gil Vicente*