UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA ELÉCTRICA

FRACTIONALIZED ADAPTIVE METHODS IN AUTOMATIC ENGINEERING PROBLEMS

TESIS PARA OPTAR AL GRADO DE
DOCTOR EN CIENCIAS DE LA INGENIERÍA, MENCIÓN ELÉCTRICA

JAVIER ANDRÉS GALLEGOS VÉLIZ

PROFESOR GUÍA:
MANUEL DUARTE MERMOUD

MIEMBROS DE LA COMISIÓN:
NÉSTOR BECERRA YOMA
CARLOS CONCA ROSSENDE
RICARDO PÉREZ CORREA

SANTIAGO DE CHILE
2020

RESUMEN DE LA MEMORIA PARA OPTAR
AL TÍTULO DE
POR: JAVIER ANDRÉS GALLEGOS VÉLIZ
FECHA: 27 DE NOVIEMBRE, 2020
PROF. GUÍA: MANUEL DUARTE MERMOUD

FRACTIONALIZED ADAPTIVE METHODS IN AUTOMATIC ENGINEERING PROBLEMS

In this work, we focus on viewing the order of derivation and/or integration in the differential equations of parametric adaptation as a parameter itself that can be also adjusted. This entails the use of fractional-order derivatives/integrals in adaptive schemes and the tuning of the derivation/integration order. It is shown that the estimation performance is objectively enhanced with a carefully designed tuning of the derivation parameter. Specifically, we establish that the fractionalization of the gradient method brings transient and robustness improvements to the estimation process. Moreover, when this estimation is used as a part of a controller or observer scheme for a given process, these advantages are inherited to the control or observer aim.

En este trabajo, nos enfocamos en ver el orden de derivación y/o integración en las ecuaciones diferenciales de adaptación paramétrica como un parámetro en sí mismo que también se puede ajustar. Esto implica el uso de derivadas/integrales de orden fraccionario en esquemas adaptativos y el ajuste del orden de derivación/integración. Se muestra que el rendimiento de la estimación se mejora objetivamente con un ajuste cuidadosamente diseñado del parámetro de derivación. En concreto, establecemos que la fraccionalización del método de gradiente aporta mejoras transitorias y de robustez al proceso de estimación. Además, cuando esta estimación se utiliza como parte de un esquema de controlador o observador para un proceso dado, estas ventajas se heredan al objetivo del control o del observador.

Acknowledgments

This work has been partly funded by ANID under the grants PCHA No. 21181187 and 11170154, and partly by a grant from the University of Chile.

Contents

Chapter 1

Introduction

1.1 Context

Adaptive schemes are dynamical systems with elements being adjusted to satisfy some criteria for the overall system in which they are embedded. In automation engineering, they are mainly employed to deal with systems where uncertainties, which always occur in models of real processes, cannot be ignored for security (air-plane navigation), precision (robotic manipulations) or stabilizing concerns (nonlinear models) (e.g., see the examples in [AKO07] or in [NA05, Chapter 11]). Correspondingly, the main adaptation criteria are lowering the energy consumption, smoothing the transient behaviour, lowering the stabilization time, and/or robustifying the response.

Some other elements in adaptive schemes remain unchanged despite their effects on the overall performance. It is heuristically known the connection between the adaptation gain –usually assumed a constant parameter– and the rate of the convergence. However, there are elements that also can provide benefits when carefully chosen, even though they do not seem modifiable at first sight. For instance, in the problem of asymptotically tracking a reference, the reference signal can be arbitrarily chosen in the transient stage; in the state observer, the dimension of the model can be enlarged or decreased to cope with unmodelled dynamics or transient improvements; in the case of the parameters tuning, the differential equation describing the tuning can be chosen of fractional type.

An intelligent adaptive scheme should choose the best way of adapting its elements in view of performance, which amounts to adapting those other elements too. In fact, the current trend in adaptive theory has shifted the emphasis from the Lyapunov stabilization to the performance features (e.g, transient response, robustness, settling time, relaxing of assumptions). For instance, to enlarge the class of uncertain systems that can be controlled, instead of the asymptotic tracking, bounds for the transient and the permanent error together with practical ways to reduce them are established in [TV18]. Weaker notions of stability (e.g. $\lambda-$stability [MD91]) show that the Lyapunov stability is not a cornerstone when dealing with robustness or solvability problems. On the other hand, Lyapunov stability does not cover the transient performance, which is one of the main practical concern. Antecedents of this approach in the discrete-time case can be found in [BK05] where adaptive optimization of the least-means-square algorithm is studied and in [Bay92] where the optimization is performed off-line.

In this work, we focus on viewing the order of derivation in the differential equations of the adaptation as a parameter to be adjusted, which implies the introduction of fractional-order derivatives or integrals in adaptive schemes. These operators play here a similar role of the complex numbers in linear system theory: they help us to solve engineering problems, but the question of their *physical interpretation* is not relevant. For our purpose, it is enough that they can be computationally implementable. The main motivation for their introduction comes from the fact that fractional differential equations allow including non-local effects, which enriches the complexity of the adaptation scheme and their capability to cope with uncertainties. In addition, being the convolution of the derivative, the fractional derivative has an averaging effect which removes short-term oscillations from the measured data.

The formal introduction of the fractional calculus in adaptive problems appeared in [VIPC02] where a simulation study was performed for the control of a first-order plant. Experimental validation of this control strategy for a coaxial rotor has been performed in [KAAY16]. However, little theoretical progress has been made. On the one hand, authors have focused on adaptive problems involving fractional-order systems with the same derivation order for the adjustment of parameters, which is not the subject of this work. On the other hand, the convergence for fractional adaptations has remained an open problem since no similar statement as the Barbalat's Lemma, which plays a key role in the integer-order case, has been found for the fractional case and only weak statements are known [Gall15a]. Essentially, it is due to the convolutional character of the fractional integral, which yields a *losing of excitation*. Even the stability property becomes problematic because there are no general methods for systems composed of integer and fractional subsystems (see for instance an integral approach in our work [AGD19]), in which we are interested. In addition, quantitative or even qualitative formal statements of the *hereditary* or averaging properties of fractional derivatives described above are not available.

1.2 Contributions

The major results of this thesis are as follows:

(1) We establish adaptive schemes defined with non-integer operators that solve classic engineering problems. We address the problems of designing estimators, controllers, and observers for uncertain real processes that can be modelled using classic physical laws. Since these laws are formulated with integer operators, the resulting models are nonlinear differential equations of integer order. We solve these problems by introducing fractional operators and proving stability properties under similar assumptions than classic solutions defined with integer operators. As a result, we are showing that fractional operators are admissible for classic problems and providing tools to analyse the overall behaviour of systems composed of subsystems with different derivation orders.

(2) We prove that the proposed fractional schemes have objective advantages as compared

with classical integer-order solutions. This means that the obtained advantages cannot be overcome by varying other parameters in the adaptive scheme. Specifically, we show in a precise mathematical sense that the transient performance is improved and the robustness is enlarged by using non-integer schemes. We also establish objective disadvantages of fractional schemes related to the convergence and the ways to fix them without losing the advantages.

(3) We develop adaptive schemes in which the derivation order is tuned like any other adaptive parameters in view of taking advantage of both fractional and integer-order features while preserving the stability. It is shown that the resulting scheme achieves better performance than fixed-order schemes in transient and robustness aspects. Moreover, the design of the scheme is flexible in the sense that other performance criteria can be incorporated too. This corresponds to the striving for building intelligent adaptive schemes in which the minimum of a priori knowledge of the system is needed to get optimal performance.

1.2.1 Hypothesis & Objectives

The above contributions entail the following hypothesis and objective.

Hypothesis: Fractional systems have properties that enhance the performance of adaptive schemes when applied to solve classic automatic problems formulated in integer calculus.

Objective: Design general enough adaptive schemes with fractional operators such that when applied to solve such automatic problems exhibit objective advantages.

1.2.2 Summary

These contributions are organized as follows. In Chapter 2, we review some features of fractional calculus. In Chapter 3, the estimation designs using fractional operators are established. These designs are applied to adaptive control and observer problems in Chapter 4. Chapter 5 explores quantitatively the qualitative advantages already established. A discussion of the results is provided in Chapter 6. An Appendix with the proofs of the main results wraps up this work.

Caveat: some mathematical contributions developed by the author to analyse fractional-order systems will be only referenced in this work to focus on engineering applications.

1.3 Notation & Preliminaries

In this section, we pile up several notation and definitions that will be used throughout the paper. The reader can skip this section and return to it when needed.

1.3.1 Algebra

$\mathbb{R}$, $\mathbb{C}$ denote the set of real and complex numbers, respectively. For $x \in \mathbb{R}^n$, $x^\top$ or x' denotes the transpose of x. For a symmetric matrix $P \in \mathbb{R}^{n\times n}$, $\underline{\lambda}_P$ and $\bar{\lambda}_p$ denote its lower and upper eigenvalue, respectively. For a symmetric matrix $A \in \mathbb{R}^{n\times n}$, $A > 0$ or $A \geq 0$ means that $\underline{\lambda}_A > 0$ or $\underline{\lambda}_A \geq 0$, respectively.

$\|\cdot\|$ is the Euclidean norm on $\mathbb{R}^n$. $\mathrm{d}(A,B)$ denotes the Euclidean distance between the sets A, B. For $A \in \mathbb{R}^{n\times n}$, $\|x\|_A := x'Ax, \forall x \in \mathbb{R}^n$ is a metric when $A > 0$.

1.3.2 Fractional Calculus

For a measurable function $x : [a,b] \to \mathbb{R}$ with $\int_a^b |x(s)|\mathrm{d}s < \infty$, its Riemann-Liouville fractional integral of order $\alpha > 0$ is defined by

$$I_{0^+}^{\alpha} x(t) = \int_0^t k_\alpha(t-s)x(s)\mathrm{d}s, \quad t \in [a,b], \tag{1.1}$$

where $k_\alpha(t) := \frac{1}{\Gamma(\alpha)} t^{\alpha-1}$ with $\Gamma(\alpha) := \int_0^\infty \tau^{\alpha-1}\exp(-\tau)\mathrm{d}\tau$ is the gamma function [Die10].

There is no unique way to define a fractional derivative; each definition having defects and virtues. The most suited for our purpose is the Caputo derivative, which is defined by

$${}^{C}D_{0^+}^{\alpha} x(t) = \int_0^t k_{1-\alpha}(t-\tau)\dot{x}(\tau)\mathrm{d}\tau \tag{1.2}$$

when $0 < \alpha \leq 1$ and $x \in AC([a,b],\mathbb{R})$, the space of absolutely continuous functions (see [Die10] for the definition with $\alpha > 1$). Often, we omit the indexes $0^+, C$. Less often, we also write $x^{(\alpha)} := D^\alpha x(t)$.

The Mittag–Leffler function $E_\alpha : \mathbb{C} \to \mathbb{C}$ is defined by

$$E_\alpha(z) := \sum_{k=0}^{\infty} \frac{z^k}{\Gamma(1+\alpha k)}.$$

For any $\alpha \in (0,1]$, the restriction $E_\alpha : \mathbb{R} \to \mathbb{R}_{>0}$ is strictly monotonically increasing. Notice that $E_1(z) = \exp(z)$ and that $E_\alpha(A)$ can be defined similarly as $\exp(A)$ for any $A \in \mathbb{R}^{n\times n}$. The function $E_\alpha(At^\alpha)x_0$ is the solution to

$$\begin{aligned} D_{0^+}^{\alpha} x(t) &= Ax(t) \\ x(0) &= x_0. \end{aligned}$$

Most of the properties of fractional-order linear systems stem from the properties of the Mittag–Leffler function [Die10] and have a parallel with the properties of integer-order systems.

Fractionalize a system is to allow that its derivation order takes real values. So, if we have the system

$$\dot{x}_{\mathrm{i}} = f_{\mathrm{i}}(x,t), \quad \mathrm{i} = 1,\ldots,n,$$

then its fractionalization is given by

$$D^{\alpha_{\mathrm{i}}} x_{\mathrm{i}} = f_{\mathrm{i}}(x, t), \quad \mathrm{i} = 1, \ldots, n.$$

1.3.3 Function Sets

For a function $f : [0, \infty) \to \mathbb{R}^n$, its $L_p(0, T)-$norm with $p \in \{1, 2, \ldots\} \cup \infty$ is defined by $\|f\|_T^p := \int_0^T \|f(t)\|^p \mathrm{d}t$. L_p is the space of functions with finite L_p-norm for $T = \infty$. Similarly, $\|f\|_{\beta,T}^p := \left(I^\beta \|f\|^p\right)(T)$ and $L_{p,\beta}$ is the space of functions f such that $\limsup_{T\to\infty} \|f\|_{\beta,T}^p < \infty$ for any $T > 0$.

$\mathcal{O}(g)$ is the set of functions f such that $\|f\| \leq C\|g\|$ for some constant number C.

The function $f : \mathbb{R}_{\geq 0} \to \mathbb{R}^n$ holds the PE condition if there exist $\varepsilon, T > 0$ such that for any $t_0 > 0$

$$\int_{t_0}^{t_0+T} f f' \mathrm{d}\tau \geq \varepsilon I_n. \tag{1.3}$$

If f holds the PE condition, then f is called persistently exciting function.

For a function $x : [0, \infty) \to \mathbb{R}$, its RMS (root mean square) value is defined by $x_{RMS}(t)^2 := \frac{1}{t} \int_0^t x(s)^2 \mathrm{d}s$.

1.3.4 Stability

Let $M \subset \mathbb{R}^n$ and consider a set of trajectories denoted by $x(\cdot; x_0) : [0, \infty) \times \mathbb{R}^n \to \mathbb{R}^n$ with $x(0; x_0) = x_0$.

M is an *stable set at* $t = 0$ if for any $\varepsilon > 0$ there exists δ such that $\mathrm{d}\left(x_0, M\right) < \delta$ implies $\mathrm{d}\left(x\left(t; x_0\right), M\right) < \varepsilon$ for all $t > 0$. M is *attractive* if there exists δ_0 such that $\mathrm{d}\left(x_0, M\right) < \delta_0$ implies $\lim_{t\to\infty} \mathrm{d}\left(x\left(t; x_0\right), M\right) = 0$. M is *asymptotically stable* if it is attractive and stable. Sometimes, we speak of stability when meaning to asymptotic stability.

Chapter 2

Fractional Systems

The aim of this section is to establish some properties of fractional operators that can have applicability in engineering problems. It is clear that the fractionalization of a given control or observer system provides an additional degree of freedom to the designer. We will show in what sense this degree can become relevant.

2.1 Memory

It is a common feature of fractional derivatives the use of nonlocal past values of a function in the computation of the derivative at any given point (e.g., see expression (1.2)). This yields the well-known non-local, long-memory, or hereditary effects in a system defined by a fractional derivative. It is less highlighted, however, that the fractional integral has a shorter memory than the integer integral, providing a similar effect than the forgetting factor or fading memory in system theory.

2.1.1 Forgetting factor

Forgetting factor, as a concept, is the attenuation of the past data to give relevance to the current data. It is used in robust and adaptive control designs to counteract noisy and time-varying effects (e.g., see [SL91, Chapter 8]). It is also used to define a performance index with fading memory. A way to exploit the forgetting factor intrinsic to the definition of the fractional integral for $\alpha < 1$ (see expression (1.1)) is to fractionalize the integrals appearing in classic designs.

For instance, consider a system of output y with an integral control u (of arbitrary dimension determined by η) given by

$$\dot{\eta} = f(\eta, y),$$
$$u = h(\eta, y) = h(\int f(\eta, y)\mathrm{d}\tau, y).$$

Then, its fractionalization is given by

$$D^\alpha \eta = f(x, y),$$
$$u = h(\eta, y) = h(I^\alpha f(\eta, y), y).$$

Thus, when the measurement of y is noisy, the fractional control has the advantage of attenuating past data due to the forgetting factor effect. A particular instance is the PID control whose fractionalization, the $PI^\alpha D$ control, was studied in [GAD20b]. In that paper, we prove the stability, robustness, and convergence properties when controlling nonlinear systems and apply the control on a boost converter. Another example, relying on a backstepping scheme, was obtained in [GD18b] to control feedback linearizable systems by changing $\dot{u} = f(\eta, y)$ by $D^\alpha u = f(\eta, y)$. In both cases, the robustness was enhanced in comparison to the integer-order case when noise is present.

The forgetting factor entails a loosing of information or dissipative property for a fractional system. The paradigmatic example is the fractional capacitor, which can be discharged even though no charge is connected to it [Gall20e, Proposition 1]. It can be proved that if a system is dissipative with respect to a convex differentiable energy function, then its fractionalization is also dissipative (see [GD18c, Proposition 1.iii]). Recalling that the dissipativeness is additive for feedback or neutral connections, one can easily generalize the passivity-based control strategy by including fractional elements. This is what ensures the stability for the $PI^\lambda D$ controllers [GAD20b].

Finally, a measure of the forgetting effect is that the set of functions with bounded fractional integral strictly increases as the order of integration decreases. This is relevant because, roughly speaking, additive external disturbances with bounded square integral does not alter the stability of the system.

2.1.2 Remembering property

A concrete manifestation of the long-memory is that the initial condition can be recovered in the following sense. Consider the scalar equation

$$D^\alpha_{0+} x(t) = -a(t)x(t),$$

where a is a non-negative continuous function and $\alpha \in (0, 1)$. In [GD16a], it was proved that if $I^\alpha_{0+} a(t) \to 0$, then $x(t) \to x(0)$ as $t \to \infty$. In particular, if $a(t)$ is compactly supported, then $\lim_{t\to\infty} x = x(0)$ for any $\alpha < 1$. However, when $\alpha = 1$ and since a is non-negative, x is non-decreasing and does not return to $x(0)$.

Seeing $a(t)$ as a multiplicative perturbation for the system, this property tell us that local perturbations (i.e. compactly supported) 'affect' more when $\alpha = 1$ than one with $0 < \alpha < 1$ since the system does not recover its initial configuration when a vanishes.

2.2 Stability

It was soon recognized that the set of matrices A that render stable the origin of the system

$$D^{\alpha}x = Ax, \quad 0 < \alpha \leq 1, \tag{2.1}$$

is strictly enlarged (in the inclusion sense) as α decreases (see a formal exposition in [CDT16]). Therefore, the following conjectures have motivated much research on this field:

(a1) The fractionalization of any stable system is stable,

(a2) The fractionalization of some unstable systems is stable,

(b1) The fractionalization of a stabilizing controller (or observer) yields a closed-loop stable system.

(b2) The fractionalization of some unstable controller (or observer) yields a closed-loop stable system.

Most of the literature is concerned with proving instances of (a1). Conjecture (b), which involves the stability problem of a multi-order system (the fractional controller/observer system and the integer-order system to be controlled/observed) and that contains the topic of this work, has been poorly developed besides contributions of the author [GD18d, GAD20a, GAD20b]. Conjecture (b2) entails the searching of novel controller/observer designs.

Conjecture (a) is of curse true for linear systems like (2.1). Using a linearization argument (see e.g. [CDT16, GD17b]), (a1) and (a2) hold locally for a large class of nonlinear systems. More generally, it was proved in [GD19a] that whenever the stability of a (time-varying) system is exponential then its fractionalization is asymptotically stable. Unfortunately, statement (a1) and hence statement (b1) are not true in general; for the time-varying version of (2.1), i.e. $A = A(t)$, a counterexample for the asymptotic stability was shown in [Gall15a]. This is related to an incomplete version of the LaSalle and Lyapunov Theorems as compared to the statement for the integer-order case [GD16b]. A candidate to hold conjecture (b2) is the fractionalization of a PID controller in [FC14], even if the stability of the corresponding PID was not treated.

2.3 Growth and Decayment

In general, there exists a combination of an extremely fast start with a subsequent sub-exponential (i.e. slower than exponential) decay for systems with derivation order $\alpha < 1$ (including the case with conformable fractional derivative). The sub-exponential growth/decay can be related to the long-memory component of the derivative that acts like an 'inertial' effect and yields a similar behaviour than delayed systems, which are known to have slower responses. This long-memory effect only appears as time grows so that in the transient time dominates the fact that, as seen from definition (1.2), the derivative of the solution must be singular so that the integral at zero does not vanish. These two effects should yield an enhanced transient response, in the sense of small overshoot, for the fractionalization of an integer-order controller/observer. Moreover, according to the heuristic trade-off between convergence's speed and noise-sensitivity, the latter effect should also yield enhanced robustness.

We will detail these effects for the scalar equation

$$x^{(\alpha)} = f(x, t), \tag{2.2}$$

where f is a $\mathcal{C}^1$ function.

2.3.1 Transient

For $\alpha < 1$, it has been proved in [GAD19b] that the solutions of (2.2) hold $|\frac{\mathrm{d}}{\mathrm{d}t}x(t)| = \mathcal{O}(t^{\alpha-1})$ as $t \to 0^+$, whenever $f(x(0), 0) \neq 0$. On the other hand, the sign of $\frac{\mathrm{d}}{\mathrm{d}t}x(t)$ is the same that the sign of $f(x, t)$ for a small enough interval $[0, \varepsilon)$, as can be seen from Definition 1.2. Let $x(0) > 0$. Then, using the mean value theorem and the continuity (differentiability) of the solution, we have that for a small enough interval $[0, \varepsilon)$, $x_\alpha(t) < x_1(t)$ for any $t \in [0, \varepsilon)$ whenever $f(x(0), 0) < 0$, where x_α, x_1 denote the solution of (2.2) for derivation order $\alpha, 1$, respectively. Similarly, $x_\alpha(t) > x_1(t)$ whenever $f(x(0), 0) > 0$. The same arguments can be repeated for $x(0) < 0$. Therefore, the fractional solutions are faster (i.e. faster decay or faster grows, respectively) than the integer one in a small enough transient period.

2.3.2 Order of growth/decay

It has been proved in [GD19a] that the solutions of (2.2) can decay at most in sub-exponential order to zero for system with derivation order $0 < \alpha < 1$. On the other hand, when $|f(x, t)| \leq \lambda|x|$ and assuming that x grows (so that w.l.o.g. $x \geq 0$), we have

$$x^{(\alpha)} \leq \lambda x.$$

A comparison argument shows that $x(t) \leq E_\alpha(\lambda t^\alpha)$. This proves that for such systems there is no finite time escape.

2.3.3 Rate of decay

Although the order of convergence is sub-exponential in fractional-order systems, we still have some control on its speed in the following sense. Consider for simplicity the Mittag-Leffler decayment $\|x(t)\| \leq E_\alpha(-\lambda t^\alpha)$ for any $t > 0$, where $\lambda > 0$. Since the function $t \to E_\alpha(-\lambda t^\alpha)$ is monotonically decreasing [Die10], it follows that the function $\lambda \to E_\alpha(-\lambda t^\alpha)$ is also monotonically decreasing. Therefore, for any $\varepsilon > 0$ and any fixed time t_0, there exist a large enough λ such that $\|x(t)\| < \varepsilon$ for any $t > t_0$, providing a way to manage the speed performance via the parameter λ.

2.4 Modelling

Due to the long-memory property and sub-exponential (power-law) behaviour, fractional-order systems have become a relatively popular alternative to get accurate models for different kinds of phenomena (see [P08, Chapter 10] and references therein). Among other features, fractional calculus has provided finite-dimensional models of linear systems with irrational transfer functions (hence, of infinite dimension) depending on non-integers exponents of the Laplace variable s (as in the heat equation [VMLO15]), because s^q is D^q in the time domain. For the same reason, it has enlarged the class of possible control designs based on frequency analysis for linear systems (see e.g. [Ous91], [LC12]). In a great extent, the fact that fractional linear systems yield input-output relationships of power-law type in the frequency domain has explained the interest for fractional derivative as many biological or social processes exhibit non-exponential behaviour.

From the physics point of view, these models should be considered phenomenological and of reduced-order kind. For, on the one hand, they may have further properties which might not be observed in the real processes; e.g., the conservation of mass or energy is violated in some fractional transport equations, the discharge of a fractional capacitor depends on how it was charged, the singularity that occurs at the initial time of the fractional derivative. On the other hand, the reduced-order nature can be deduced from the following asymptotic result [Ous91]

$$s^\alpha = \lim_{N\to\infty} \prod_{n=1}^{N} \frac{1 + s/\omega_n}{1 + s/\nu_n},$$

where ω_n, ν_n depend on α. Thus, a finite-dimensional fractional-order system is equivalent to an infinite-dimensional integer-order one (for null initial condition). Since the discrete version of the above expression entails infinite delays, there would be also an order-reduction when modelling infinite-dimensional delayed phenomena. It must be noted that actual processes are infinite-dimensional.

The above expression is also notable in providing a way to approximate (and to simulate) fractional systems, and hence, to estimate the computational complexity in using fractional instead of integer operators with respect to the degree of approximation required.

Thus, from an engineering point of view, the fractionalization of a finite-dimensional model of a real process could help to match the observed data when this process is, in fact, infinite-dimensional. For instance, consider the problem of controlling a process. A fairly standard technique is to perform a finite-dimensional linearization around an operation point and to employ classic control designs for linear systems. This solution works at the same extent that the linear finite-dimensional model approximates the real process. Therefore, the use of a fractional model should yield better performance when dealing with large-dimensional processes. This idea has been explored theoretical and numerically by the author in two differing settings: in an observer model [DGAC18b] and in a predictive model [GMM20d].

2.5 Robustness

Due to practical considerations, any proposed controller or observer must be robust because any model used in its design is only an approximation of the real underlying process. We will revise some arguments showing that the fractionalization of a controller or observer can enhance its robustness.

a) It is a well-known problem that the computation of the derivative of a measured signal is inexact, introduces delays and/or amplifies the noise. In fact, this is the reason why state observers are needed and why the control design becomes a complex problem. However, some controllers perform derivation on variables of the system, being the PID scheme the most popular example. Since the fractional derivative is the integral of the integer derivative and integrals attenuate the effects of disturbances, a smoothing or averaging effect is to be expected in the fractionalization of the derivative of a given signal. The fractionalization of the PID control, the so-called $PI^{\lambda}D^{\mu}$, is one of the earliest applications of fractional operators in control problems [P08] and has been studied to control linear systems relying on frequency domain techniques. Another example occurs in boundary control of partial differential equations where the integer derivative is replaced by a fractional one [MM95].

b) Due to the long-memory effect in the actualization of the variables for a system defined with fractional derivatives, a sudden increment of an external disturbance has an attenuated effect in comparison with integer-order ones where the effect is directly transferred to the next instant due to the fundamental theorem of the calculus. Dually, expressing the equation of the system in its integral form, this attenuation can be explained by the forgetting factor.

c) Since the solutions of fractionalized systems become slower as time goes, under-reaction can be expected to external disturbances. Put in other words, this means a larger attenuation of disturbances.

Chapter 3

Fractional Adaptation

Adaptive schemes are designed to achieve control or estimation aims by adjusting some of their parameters or their structure. Mainly, the parameter adjustment consists of assigning values to the parameter derivatives according to the following criteria: the cancellation of cross-terms in Lyapunov functions [KK95], the domination of the overall dynamic [WC02], or the optimization of a performance index guided by the gradient descent or the Least-Square technique [SB94]. However, these schemes present general speaking a poor performance regarding robust and transient aspects. In the cancellation and optimization cases, it seems obvious that any perturbation blurs their designing principle. The domination case admits more robust designs, but, for the same reason, conservatism in the transient performance should be expected. Since adaptive schemes are justified in its capability to deal with uncertain settings, the design of high-performance robust schemes becomes one of the main focus for adaptive researchers. This section is devoted to studying the transient and robust effects of the assignation of values to the fractional rather than the integer derivative of the parameters.

3.1 Regression Model

To establish a general framework with the abstraction of any specific problematic, consider the regression model

$$y = m'\theta, \tag{3.1}$$

where $y : [0, \infty) \to \mathbb{R}$ and $m : [0, \infty) \to \mathbb{R}^n$ are measurable functions and $\theta \in \mathbb{R}^n$ is a vector of unknown constant parameters. To the best of our knowledge, it can be shown that for all previously studied adaptive problems with linearly parametrized uncertainty one can define a regression equation of the type (3.1) (see Section 6.1).

Let $\hat{\theta} = \hat{\theta}(t)$ be an estimation of θ at time t and define

$$\hat{y} := m'\hat{\theta}, \tag{3.2}$$

$$\mathrm{e} := \hat{y} - y = m'(\hat{\theta} - \theta) =: m'\tilde{\theta}. \tag{3.3}$$

In this setting, the adaptive goal is to minimize $|\mathrm{e}|$ by adapting $\hat{\theta}$, and the estimation goal is to achieve $\lim_{t\to\infty}\hat{\theta}(t) = \theta$.

3.2 Fractionalized Gradient Descent

One of the most effective methods to solve optimization problems is the gradient descent. Basically, for the minimization of a performance index J, it suggests the dynamic

$$\dot{\hat{\theta}}(t) = -\lambda \nabla_{\hat{\theta}} J(t), \tag{3.4}$$

where $\nabla_{\hat{\theta}}$ is the gradient with respect to $\hat{\theta}$, and λ is a designer-chosen nonnegative function that we will assume scalar for simplicity.

Intuitively, (3.4) yields the fastest convergence of $\hat{\theta}$ to the optimal argument θ^* for the optimal value J^* because $\hat{\theta}$ follows the steepest descent of J. However, the gradient-descent has rather poor robust and transient properties. As a matter of fact, modifications to gradient designs, such as projection, switching, dead zone, σ, e_1 or θ-modification, must be considered in real applications where adaptive schemes with a bad transient response or with poor robustness are utterly impractical [OT89, p.661]. Although J could include specific terms to improve the performance in these aspects, the convergence study of the resulting scheme becomes more difficult as witnessed in the fact researchers in this field have preferred the above modifications.

Based on the arguments in Section 2.3 and 2.5, we can expect transient and robustness improvements with a fractionalized gradient estimation. On the other hand, as a modification of the gradient-descent, we should also expect that the optimization will be impoverished in other aspects such as the speed or even the convergence itself.

We start by writing (3.4) as

$$\hat{\theta}(t) = -\lambda \mathcal{H}(s)[\nabla_{\theta} J](t),$$

where s denotes the Laplace domain variable or the differential operator $\mathrm{d}/\mathrm{d}t$. Thus, $\mathcal{H}(s) = 1/s$ leads to (3.4), $\mathcal{H}(s) = 1/(s+\sigma)$ to the gradient descent with *leakage*, which is used to enhance robustness as it entails a filtering of the measured signals [OT89, Section 2.4.4], $\mathcal{H}(s) = k_p + k_{\mathrm{i}} s^{-1}$ leads to a *proportional-integral* (PI) gradient descent, and $\mathcal{H}(s) = 1/s^{\alpha}$ to the fractionalized gradient (e.g. [VIPC02]). Although combinations among them can also be studied, our aim is to understand what the simplest fractionalization brings to the estimation problem. The key seems to be the expression for s^{α} in Section 2.4, which yields a connection with the leakage gradient. In fact, the leakage also introduces memory effects in the adaptation [HC89, p.409].

Choosing the index $J = \mathrm{e}^2$, we must thus consider

$$D^{\alpha}_{0^+}\hat{\theta}(t) = -\lambda(t)\,\mathrm{e}(t)\,m(t), \tag{3.5}$$

where $\hat{\theta}(0) \in \mathbb{R}^n$ and $\hat{\theta}(0) \neq \theta$ as otherwise the solution is trivial. Recalling the definition of the fractional derivative, notice that (3.5) does not specify the rate of the parameter change

($\dot{\theta}$) but the way in that the past rates are to be weighted (quantified by α) at the current time. In this sense, we say that (3.5) is a *a nonlocal* estimation. Conditions for existence, uniqueness, and continuity can be deduced from [GAD19b]. For our purpose, it is enough requiring continuity or locally integrability on m, λ. The choice $\alpha = 1$ and $\lambda(t) = t^{\alpha-1}\lambda'(t)$ yields the conformable version of the gradient descent. That is, this case can be written as $T^{\alpha}_{0^+}\hat{\theta}(t) = -\lambda'(t)\,\mathrm{e}(t)\,m(t)$, where $T^{\alpha}_{0^+}$ is the conformable derivative – a kind of local fractional derivative introduced in [Abd15].

Formally, the problem is to determine the role of the parameter α in the minimization of e^2. The following result, whose proof is given in Section 6.2.1, establishes its role in the transient and robust performance.

Theorem 1 For the estimator (3.5), the following statements hold:

i. $\hat{\theta} \in L_\infty$ and $\|\tilde{\theta}(t)\| \leq \|\tilde{\theta}(0)\|$ for all $t > 0$.
ii. If $\lambda > 0$, then $\mathrm{e}^2 \in L_{1,\alpha}$ and $\lim_{t\to\infty} \mathrm{e}_{RMS} = 0$
iii. Suppose that m is a bounded and uniformly continuous function. Then, $\mathrm{e} \to 0$ as $t \to \infty$ and $\lambda \to \infty$.
iv. There exists a bounded uniformly continuous signal m such that $\mathrm{e} \to 0$ as $t \to \infty$ when $\alpha = 1$, but e does not converge to zero when $\alpha < 1$ for bounded λ.
v. If $m \in PE$ is bounded, then $\hat{\theta} \to \theta$ and $\mathrm{e} \to 0$ in sub-exponential order for $\alpha < 1$ and in exponential order for $\alpha = 1$ as $t \to \infty$.
vi. There exists a transient period where $\|\tilde{\theta}(t)\|$ is lesser when using $\alpha < 1$ than when using $\alpha = 1$, even if one allows λ to be arbitrarily large but bounded.
vii. For any additive perturbation ν in the measurement of y with $\nu \in L_{2,\alpha}$, $\hat{\theta}$ remains bounded. Moreover, the space of non-destabilizing perturbations is enlarged as α decreases.

Remark 1 The following comments must be underlined.

- Item (i) provides a bound for the transient behaviour, which is suited to apply multiple-models for further improvement of the transient, i.e. to employ several estimators of the form (3.5) with different initial conditions.
- Item (ii) means that arbitrary prescribed performance in the error magnitude can be achieved by increasing λ.
- Items (iv) and (v) represent objective disadvantages in the choice $\alpha < 1$ and are a consequence of the modification to the gradient descent. Item (iv) and the convergence slower than exponential are true even for more general adaptive laws of type $D^\alpha\hat{\theta} = f(\mathrm{e}, \hat{\theta}, m)$. For the former, this occurs essentially because of the failure of conjecture (a2) in Section 2.2, see more details[Gall15b]; for the latter, see Section 2.3.
- A milder condition than (v) for the convergence of e is by requiring the PE condition on some rather than the full components of m (see Section 6.2.2).
- Items (vi) and (vii) provide expressions for the transient and robustness enhancement due to the fractionalization and represent objective advantages in comparison to the case $\alpha = 1$ in the sense that they cannot be overcome by choosing bounded λ. Item (vi) can be intuitively explained from the fact that the singularity $\|\dot{\theta}(t)\| \sim t^{\alpha-1}$ as $t \to 0^+$

yields an effective adaptive gain larger than any finite one, and hence, the fractional estimator converges initially faster than any other with $\alpha = 1$. Item (vii) can be explained by the forgetting effect of the fractional integral, which entails an attenuation of old data, together with the fact that $\hat{\theta}$ is the fractional integral of $\lambda(t)\,\mathrm{e}(t)\,m(t)$. In fact, in the limit case $\alpha = 0$, in which the forgetting is instantaneous since $I^0 x = x$, $\hat{\theta}$ remains bounded for any bounded perturbation when m is bounded since

$$\begin{aligned}\hat{\theta}(t) &= -mm'(t)(\hat{\theta}(t) - \theta) + \nu(t)\omega(t)\\ &= (I + mm'(t))^{-1}[mm'\theta + \nu(t)m(t)].\end{aligned}$$

3.3 Adaptation with Memory

It must be recalled that convergence is a prerequisite in many adaptive applications (e.g. in synchronization for secure communications). The shortcoming of fractional adaptation in this regard (see Theorem 1(iii)) can be critical as the obtained way to ensure the error convergence is given by a strong condition on m (see Theorem 1(iv)). We devote this subsection to relax this requirement retaining the already obtained transient and robustness features.

The trick to weakening the PE condition is storing excitation with a memory element given by an integral. To preserve the advantages of the fractionalization established in Theorem 1, the adaptation is obtained by adding such a memory term to the gradient. Being the automatic adjustment to changes in the environment, the adaptation is benefited from short-memory designs. Therefore, to enhance the performance of the adaptation, the memory element must be used with a forgetting factor, which will be accomplished by letting the integral to be fractional. This yields the fractionalization of the integral adaptation (e.g., see [PKD18] and the pioneering work [Kre77]).

Specifically, $\hat{\theta}$ is determined by the following equations (we have replaced m by u).

$$D^{\alpha}_{0+}\hat{\theta} = -\Gamma\big(\gamma_1 \mathrm{e} u + \gamma_2\kappa(\Psi\hat{\theta} - \Upsilon)\big), \tag{3.6a}$$

$$\Psi(t) := \int_0^{t_a(t)} k(t_a(t), s)u(s)u'(s)\mathrm{d}s, \tag{3.6b}$$

$$\Upsilon(t) := \int_0^{t_a(t)} k(t_a(t), s)u(s)y(s)\mathrm{d}s = \Psi(t)\theta, \tag{3.6c}$$

where $\gamma_{1,2}$ are normalizing functions, which we choose as $\gamma_1 := \frac{1}{1+\|u\|^2}$ and $\gamma_2 := \frac{1}{1+\bar{\Psi}}$, with $\bar{\Psi}(t) = \int_0^{t_a(t)} k(t_a(t), s)\|u(s)\|\mathrm{d}s + \int_0^{t_a(t)} k(t_a(t), s)\|u(s)\|^2\mathrm{d}s$. The numbers $0 < \alpha \leq 1$ and $\kappa > 0$, the matrix $\Gamma \in \mathbb{R}^{n\times n}, \Gamma > 0$, the non-decreasing nonnegative function $t_a = t_a(t)$ and the locally integrable function k are designer-chosen.

The rationale of the adjustment (3.6) comes from the following facts. Let

$$J(\hat{\theta}(t)) := \mathrm{e}^2(t) + \int_0^{t_a} k(t, s)\big(m'(s)\hat{\theta}(t) - y(s)\big)^2\mathrm{d}s$$

be a performance index measuring the current output error and the weighted past error when it is used the current estimated $\hat{\theta}(t)$. Since $m'(s)\hat{\theta}(t) - y(s) = m'\tilde{\theta}$, the second term provides a measure of the parametric error. As the right-hand side of (3.6a) is proportional to $-\nabla_{\hat{\theta}(t)} J(\hat{\theta}(t))$, $\hat{\theta}(t)$ follows the steepest descent to minimize e^2 and (a measure of) $\tilde{\theta}'\tilde{\theta}$ when $\alpha = 1$.

The choice of k aims to cope with noisy measurements and parametric variations by acting like a forgetting factor in the sense that it attenuates the past data in the convolution. Choosing $k_\beta(t,s) = (t-s)^{-\beta+1}$, we obtain $\Upsilon = I^\beta uy$ and $\Psi = I^\beta uu'$, which can be implemented with the following fractional differential equations

$$\begin{aligned} D^\beta_{0+}\Upsilon(t) &= m(t)y(t), \quad \Upsilon(0) = 0, \\ D^\beta_{0+}\Psi(t) &= m(t)m'(t), \quad \Psi(0) = 0, \end{aligned} \tag{3.7}$$

for any $\beta > 0$. The implementation through filters for the case $k(\tau, s) = \exp(-q(t-s))$ for any $q > 0$ was studied in [Kre77]. Both $\exp(-q(t-s))$ and $(t-s)^{-\beta+1}$ allow introducing forgetting factors in the estimation with the difference that the latter has a longer memory and does not need a filtering step when $\beta < 1$ representing an objective computational advantage.

The choice of t_a aims to keep the excitation. In this paper, we consider the choice

$$t_a(t) = \max\{\arg\max_{\tau\in[0,t]} \gamma_2(\tau)\, \lambda_{min}\big(\int_0^\tau k(\tau,s)u(s)u'(s)\mathrm{d}s\big).\} \tag{3.8}$$

To ensure convergence, we will make an excitation assumption similar to that in [PKD18], which requires the following definition.

Definition 1 The function $u : [0,\infty) \to \mathbb{R}^n$ has finite excitation (FE) if there exist $\varepsilon, \gamma > 0$ such that

$$\int_0^\varepsilon u(s)u'(s)\mathrm{d}s \geq \gamma I_n \tag{3.9}$$

It is clear that FE is strictly weaker than PE since the latter requires the uniform fulfilment of (3.9) on the real line. Roughly speaking, the memory element moves the FE to each interval so that a kind of effective PE is obtained. The following result formalizes this idea (see the proof in Section 6.3.1).

Theorem 2 Consider the adaptive law (3.6). Then, the following statements hold:

i. $\hat{\theta}, I^\alpha_{0+}\big(\gamma_1 e^2\big) \in L_\infty$, and $\|\tilde{\theta}(t)\|_{\Gamma^{-1}} \leq \|\tilde{\theta}(0)\|_{\Gamma^{-1}}$ for all $t \geq 0$.

ii. (Transient) If $0 < \alpha_1 < \alpha_2 \leq 1$, then there exists $\varepsilon > 0$ such that $\|\tilde{\theta}_{\alpha_1}(t)\|_{\Gamma^{-1}} < \|\tilde{\theta}_{\alpha_2}(t)\|_{\Gamma^{-1}}$, $\forall t \in [0,\varepsilon]$, where $\tilde{\theta}_\gamma := \hat{\theta}_\gamma - \theta$ and $\hat{\theta}_\gamma$ is the solution of (3.6) with $\alpha = \gamma$.

iii. (Intrinsic robustness) For any additive perturbation $v \in L_{2,\alpha}$ in the measurement of y, $\hat{\theta}$ remains bounded. Moreover, the space of non-destabilizing perturbations is enlarged as α decreases.

iv. (Convergence) Suppose that u has finite excitation for ε, γ, and t_a is given by (3.8). Then, $\tilde{\theta}(t) \to 0$ as $t \to \infty$ at Mittag-Leffler order of a rate of convergence regulated through κ, α and Γ.

v. (Robustness due to excitation) Suppose that u has finite excitation for ε, γ, and t_a is given by (3.8), If instead of (3.1) the model is given by

$$y = u'\theta + \nu, \tag{3.10}$$

then $\hat{\theta}$ and e remains bounded when ν is bounded, and they converge to zero when ν converges to zero.

Remark 2 Some important points are highlighted below.

a. Since α, β play a role to optimize the estimation performance, it is convenient to let them taking free values in each coordinate. Section 6.3.2 is devoted to this problem. Notice that the stability analysis for the case when α is a vector cannot be reduced to a scalar equation through Lyapunov functions, and therefore, (3.6) is a candidate to satisfy conjecture (b2) in Section 2.2.

b. The occurrence of signals as in Theorem 1(iii) that avoid the convergence of the error is precluded since the adaptive law (3.6) is not algebraic in e, m (see the third item in Remark 1).

3.4 Finite-Time Adaptation

For the same reason that the convergence is required in some adaptive designs, the order and speed of the convergence can become a decisive criterion to evaluate their performances. For instance, an error function of power-law order of convergence as the obtained in Theorem 2 can have unbounded L_2-norm, which would mean unbounded energy in practical applications. The slower than exponential convergence introduced by the already studied fractional adaptive laws can be overcome if the estimator ensures finite-time convergence. To counteract the speed limitations established in Section 2.3, a discontinuous or switching procedure must be considered. This section is devoted to design a finite-time estimator using a discontinuous mechanism.

We will use the DREM procedure [OAPAB19] that reduces the estimation problem to a scalar one. Consider arbitrary linear operators $H_\mathrm{i} : \mathcal{L}_\infty \to \mathcal{L}_\infty$, for $\mathrm{i} = 1, \ldots, n$. Applying them to (3.1), we get $y_{H_\mathrm{i}} = m'_{H_\mathrm{i}}\theta$, where $(\cdot)_{H_\mathrm{i}} := H_\mathrm{i}(\cdot)$ is component-wise applied. Let $Y_\mathrm{e} : [0, \infty) \to \mathbb{R}^n$ be the vector function with components $y_{H_1}, \ldots, y_{H_q}$ and $M : [0, \infty) \to \mathbb{R}^{n\times n}$ the matrix function of rows $m'_{H_1}, \ldots, m'_{H_n}$. Then, $Y_\mathrm{e} = M\theta$ and the following holds

$$\begin{aligned} Y := adj(M)Y_\mathrm{e} &= adj(M)M\theta, \\ &= \det(M)I_n\theta, \\ &= \det(M)\theta =: \Delta\theta, \end{aligned}$$

where the identity $adj(A)A = \det(A)I_n$ for any square matrix $A \in \mathbb{R}^{n\times n}$ was used. Thus, the estimation problem is equivalent to the following set of scalar regression equations

$$Y_\mathrm{i} = \Delta\theta_\mathrm{i}, \quad \mathrm{i} = 1, \ldots, n. \tag{3.11}$$

To take advantage of previously established features in Theorem 1, we consider the fractionalization of a gradient-like estimator for (3.11), i.e.

$$D_{0+}^{\alpha_{\rm i}}\hat{\theta}_{\rm i}(t) = \lambda_{\rm i}(t)\Delta(t)\big(Y_{\rm i}(t) - \Delta(t)\hat{\theta}_{\rm i}(t)\big), \tag{3.12}$$

where $\alpha_{\rm i} > 0$ and $\lambda_{\rm i} \geq 0$ are designer-chosen variables. Since they are n independent scalar equations, we can drop the index i.

Theorem 3 For some designer-chosen parameter $0 < \mu < 1$, suppose that there exists $\delta > 0$ such that

$$\big(I_{0+}^{\alpha}\lambda\Delta^2\big)(t = \delta) \geq \frac{\mu}{1-\mu}. \tag{3.13}$$

Let w be the solution to $D_{0+}^{\alpha}w = -\lambda\Delta^2 w$ satisfying $w(0) = 1$. Define the estimator $\hat{\theta}^{FT}$ as $\hat{\theta}$ on $[0, t_f]$ where t_f is such that $w(t_f) = 1 - \mu$, and

$$\hat{\theta}^{FT}(t) = \frac{1}{1 - w(t)}\big(\hat{\theta}(t) - w(t)\hat{\theta}(0)\big), \tag{3.14}$$

for $t > t_f$. Then, the following statements hold:

i. $\hat{\theta}^{FT}$ is a well-defined piecewise continuous function.
ii. (Transient) If $0 < \alpha_1 < \alpha_2 \leq 1$, then there exists $\varepsilon > 0$ such that $|\tilde{\theta}_{\alpha_1}(t)| < |\tilde{\theta}_{\alpha_2}(t)|$, $\forall t \in [0, \varepsilon]$, where $\tilde{\theta}_{\gamma} := \theta_{\gamma} - \theta$ and θ_{γ} is the solution of (3.12) with $\alpha = \gamma$.
iii. (Adaptation gain) If $0 \leq \lambda_1 \leq \lambda_2$, then $|\tilde{\theta}_1| \geq |\tilde{\theta}_2|$ where $\tilde{\theta}_{\rm i} := \theta_{\rm i} - \theta$ and $\theta_{\rm i}$ is the solution of (3.12) with $\lambda = \lambda_{\rm i}$.
iv. (Finite-Time Convergence) There exists $t_f < \delta < \infty$ such that $\hat{\theta}^{FT}(t) = \theta$ and $\mathrm{e}(t) = 0$ for any $t \geq t_f$.

The proof of Theorem 3 and that of the following properties are in Section 6.4. In the next two propositions, dealing with sufficiency for condition (3.13), notice the independence of the derivation order.

Proposition 1 Let m be a continuous function. A necessary and sufficient condition to have a finite-time estimator of the type (3.14) is that Δ not be the zero function. If the components of m are linearly independent functions on an arbitrarily small interval, then Δ is a non-zero function.

The fact that λ appears in the convergence condition (3.13) on an equal footing than Δ can be exploited to arbitrary reduce the convergence time t_f. More precisely,

Proposition 2 Suppose that Δ is continuous and not the zero function. Then, t_f can be arbitrarily reduced by increasing λ.

Consider that model (3.1) is affected by an unknown additive disturbance ν, i.e.

$$y = m'\theta + \nu. \tag{3.15}$$

Due to the linearity of the filters, the DREM procedure for (3.15) yields

$$Y = \Delta\theta + \Delta V, \tag{3.16}$$

where V is the vector of components $\nu_{H_{\rm i}} := H_{\rm i}(\nu)$ for $\rm i = 1, \dots, q$. Thus, the estimation problem is again essentially scalar.

Since ν is assumed unknown, the estimator is required to be robust meaning that it is not destabilized when ν is bounded and preserves the already established features when $\nu \equiv 0$. To satisfy this, we must modify the estimator (3.14) because, in the fractional case, w is not monotonous and can decrease to 1 when Δ goes to zero (see the proof of Theorem 3(i)), so the denominator of (3.14), which is not compensated with the numerator term as in the case $\nu \equiv 0$, can become arbitrarily small.

The proposed modification will be guided by the heuristic "do not fit to bad data" [OT89, p.662] avoiding adjusting to the noise when the excitation is not strong enough, which in our case means Δ near to zero. Specifically, we actualize the estimator to (3.14) only when condition (3.13) is satisfied and hold the previous value otherwise after the first time that this condition was achieved. Formally,

$$\hat{\theta}^{RFT}(t) := \begin{cases} \hat{\theta}(t) & \text{if } t < \delta, \\ \hat{\theta}^{FT}(t) & \text{if } t : \left(I^{\alpha}_{0+}\lambda\Delta^2\right)(t) \geq \frac{\mu}{1-\mu}, \\ \hat{\theta}^{RFT}(t^-) & \text{otherwise,} \end{cases} \tag{3.17}$$

where $\hat{\theta}^{FT}(t)$ is given by the algebraic expression (3.14), δ is defined as in Theorem 3, and $\hat{\theta}^{RFT}(t^-)$ means the left-hand limit of $\hat{\theta}^{RFT}$ evaluated at t. Were $\hat{\theta}$ obtained from (3.12), we should state a persistent excitation (PE) condition to ensure robustness [SB94]; however, PE conditions are qualitatively different to the finite-time condition (3.13) whose aim is precisely to relax them (see the Introduction). Instead, we will use the fractionalized projective gradient as it is the slightest robust modification to the fractionalized gradient (3.12). In the scalar case, it is given by

$$D^{\alpha}_{0+}\hat{\theta}(t) = \mathcal{P}_{[a,b]}[\lambda(t)\Delta(t)\big(Y(t) - \Delta(t)\hat{\theta}(t)\big)], \tag{3.18}$$

where $0 < \alpha \leq 1$, and $\mathcal{P}_{[a,b]}[x] = x$ when $\hat{\theta} \in (a, b)$, $\mathcal{P}_{[a,b]}[\cdot] = \varepsilon$ when $\hat{\theta} = a$, and $\mathcal{P}_{[a,b]}[\cdot] = -\varepsilon$ when $\hat{\theta} = b$ for $\varepsilon > 0$ a designer chosen parameter (see [Tao03] for the case $\alpha = 1$; there are also continuous versions). This operator is aimed to trap $\hat{\theta}$ in the interval $[a, b]$ provided that θ is known to lie in its interior.

Proposition 3 Consider $0 < \alpha \leq 1$. Then

a) For any additive perturbation $v \in \mathcal{L}_{2,\alpha}$ in the measurement of y, $\hat{\theta}^{FT}$ remains bounded. Moreover, the space of non-destabilizing perturbations is enlarged as the derivation order decreases.

b) Suppose that there are known numbers a, b such that $\theta \in (a, b)$ and choose $\hat{\theta}(0) \in (a, b)$. Then, the estimator (3.17) for the model (3.15) is robust in the sense that for any disturbance ν, $\hat{\theta}$ remains bounded, and if $\nu \equiv 0$ and (3.13) holds, then the estimator converges in finite time to θ.

c) The alertness, i.e. the capability to track parametric variations, is improved as α decreases.

It follows that the use of λ entails the trade-off 'convergence against robustness'. In some cases, it could not be problematic since the speed of convergence can be arbitrarily lowered down after the transient by decreasing λ. For the general case, the trade-off can be relaxed recalling the freedom to choose the derivation order α, its role to improve the transient performance (Theorem 3(ii)) and its role to enhance the robustness.

3.5 Switched Adaptation

As we said, the trick to avoid the slow convergence speed of fractional systems is by introducing a discontinuous or switching procedure. In the finite-time estimator, we used the discontinuity; here, we will explore a switching mechanism that is able to keep the continuity. Moreover, we will show that the proposed design can address the problem of estimating an optimal derivation order whatsoever the optimality criteria are.

The key feature of the precedent adaptive laws is that their stability analysis can be carried out using Lyapunov functions that are independent of the derivation order. This implies, in the Switching Theory language, that they are common Lyapunov functions for arbitrary switching in the derivation order.

General speaking, we study a switching system with the following form

$$ {}_{a_{\tau(t)}}D^{\alpha_{\varpi(t)}}x(t) = f_{\sigma(t)}\left(x,t\right), \quad x(t_0) \in \mathbb{R}^n \tag{3.19}$$

where $x : [0,\infty) \to \mathbb{R}^n$; the functions $\sigma : [0,\infty) \to I_f$, $\varpi : [0,\infty) \to I_\alpha$ and $\tau : [0,\infty) \to I_a$ are right-continuous and piecewise constant taking values on the index sets I_f, I_α, I_a, respectively, at each switching instant $t_{\rm i}$ belonging to the set $E := \{t_{\rm i}\}_{{\rm i}\in\mathbb{N}} \subset [0,\infty)$. The sets I_f, I_α, I_a index the sets $Q_f \subset \{f | f : \mathbb{R}^n \times [0,\infty) \to \mathbb{R}^n\}$, $Q_\alpha \subset (0,1]$, and $Q_a \subset [0,\infty)$, respectively. Q_f is called the set of driving fields. An impulsive behaviour will be specified by an impulsive rule function $\delta : [0,\infty) \to Q_\delta \subset \mathbb{R}^n$ sharing the same properties than σ, ϖ, τ. The switching nature of (3.19) implies that the sequence $\{t_{\rm i}\}_{{\rm i}\in\mathbb{N}}$ and the functions $\sigma, \varpi, \tau, \delta$ are unknown and can depend on the initial condition and/or on the trajectory.

However, switched fractional systems presents several problems of ill-definition –both theoretical and computationally– due to the nonlocality of the initial condition. Fortunately, the following instance of them can be made well-defined, and it will be enough for our purpose.

Definition 2 System (3.19) is in the *resetting mode* if for each realization of $\{t_{\rm i}\}_{{\rm i}\in\mathbb{N}}$, the initial times of the derivative are given by $t_{\rm i}$ i.e. $a_{\tau(t_{\rm i})} = t_{\rm i}$ and there exists an arbitrarily small number T_0 such that $t_{{\rm i}+1} - t_{\rm i} > T_0 > 0$ for any $t_{\rm i} \in E$. In particular, the solution to (3.19) can be expressed in terms of a sequence of solutions to non-switching fractional systems given by

$$\begin{aligned} x(t) &= \xi_{\rm i}(t), \quad \forall t \in [t_{\rm i}, t_{{\rm i}+1}), t_{\rm i} \in E, \\ {}_{t_{\rm i}}D^{\alpha_{\varpi(t_{\rm i})}}\xi_{\rm i}(t) &= f_{\sigma(t_{\rm i})}(\xi_{\rm i}, t), \ \ \textit{with} \ \ \xi_{\rm i}(t_{\rm i}) = x(t_{\rm i}^-) + \delta(t_{\rm i}) \in \mathbb{R}^n, \end{aligned}$$

where $x(t_i^-)$ is the left-hand limit of $x(t)$ at t_i. Each ξ_i-system will be called a subsystem of (3.19).

In Definition 2, we can distinguish the following two cases:

Definition 3 System (3.19) in the *resetting mode* is impulsive if δ is not the zero function, and is non-impulsive otherwise.

System (3.19) in the resetting mode with Caputo derivative is mathematically well-defined because the existence, uniqueness, and continuity (among switching times in the impulsive case) of its solutions are assured with the existence, uniqueness, and continuity of solutions for each non-switching system associated according to Definition 2, which in turn can be ensured with a Lipschitz continuity requirement in the first argument and continuity in the second argument on each $f_{i\in S}$ (see e.g. [Die10]). Moreover, system (3.19) is computationally well-defined using standard software as shown in Section 6.5.1.

With this definition, we are able to design a variable order estimator for (3.1). From precedent theorems, we know that the transient is improved with a derivation order $\alpha < 1$, while the speed and convergence conditions are improved with $\alpha = 1$. Therefore, enhanced performance is achieved by taking initially a lower-than-one derivation order and, after some small enough time, force the derivation order to be 1 (see the simulations in the papers made by this author in [AG19] and [AG20] for the adaptive control of a linear system).

To take advantage of the robustness properties of lower than one derivation orders, notice that the perceptible way in which the noise or parasitic signals affect the behaviour is by preventing the error function to the decay to zero. Thus, we can force a progressive decreasing of the derivation order, say 0.1 each time, on the sequence $\{t_k\}_{k\in\mathbb{N}}$ defined by

$$t_{k+1} = inf_{t\geq t_k+T_0}\{|\mathrm{e}(t)| \geq c(t-t_k, \alpha(t_k))\},$$

where $T_0 > 0$ prevents Zeno phenomenon and can be chosen so small as needed, $t_0 = 0$ is not included in the sequence, $c(\cdot, \alpha)$ is a function that converges to zero for $\alpha = 1$, and $c(t, \cdot)$ is a decreasing function. For instance, one can choose $c(t, 1) = A\exp(-\mu t)$ and $c(t, \alpha) = B1/\alpha$. The following features are thus obtained.

i. Suppose that after a transient period, which can be included in $[0, T_0]$, the derivation order was fixed at $\alpha = 1$. If there is no disturbance and the convergence conditions are satisfied, the choice of $c(\cdot, \alpha)$ ensures that the error will converge to zero and no switching will occur.

ii. If there is a significant disturbance, then the error will not converge to zero and switching will occur. Due to the choice of $c(t, \cdot)$, the derivation order decreases as $|\mathrm{e}|$ increases, which is desirable since, due to Theorem 1(vii), the estimator becomes more robust as the derivation order decreases. If $|\mathrm{e}|$ becomes eventually bounded, the switching sequence $\{t_k\}_{k\in\mathbb{N}}$ is finite.

iii. If the disturbance disappears when (say) $\alpha = \alpha(t_k)$, then no more switching will occur. If the convergence condition for $\alpha = \alpha(t_k)$ holds, then the rate of convergence will be determined by this order. Due to this, one should increase again the derivation order to eventually get $\alpha = 1$. For instance, consider the rising sequence $\{t'_k\}_{k\in\mathbb{N}}$

$$t'_{k+1} = t_c + T_1 \quad where\ t_c > t_k \ \&\ |\mathrm{e}(t)| \leq c'(t - t_c, \alpha(t_k)), \forall t \in [t_c, t_c + T_1],$$

where $c'(t - t'_k, \alpha(t_k))$ decays polynomially at order $\alpha(t_k)$ and $T_1 > 0$ prevents Zeno phenomena. Due to the choice of c, $t'_{k+1} \neq t_{k+1}$, and thus the rising and the lowering sequences are well-defined.

The following result is proven in Appendix 6.5.2.

Theorem 4 Consider the estimator with non-impulsive switching in the resetting mode

$$D^{\alpha(t)}_{a(t)^+}\hat{\theta}\,(t) = -F(\mathrm{e}, m)\,(t)\,, \tag{3.20}$$

where F is given by (3.6) or (3.14), and $\alpha(\cdot)$ is defined by the lowering and rising order sequences $\{t_k\}_{k\in\mathbb{N}}$ and $\{t'_k\}_{k\in\mathbb{N}}$. Suppose that m satisfies the excitation condition of Theorem 2 or Theorem 3, respectively. Then (3.20) ensures boundedness of $\hat{\theta}$, which has the same convergence properties of $\alpha = 1$, and the same transient and robustness properties of any $\alpha < 1$.

Remark 3 Notice that the resulting scheme (3.20) provides a midway to avoid long-memory effects, which seem to some degree unsuited for adaptation, without losing completely the non-locality effect that should yield an averaging effect useful to attenuate noise and without losing the transient effect.

3.6 Optimal Second-Level Adaptation

In section 3.5, we have seen a knowledge-guided way to find a (sub-)optimal time-varying derivation order for the estimator equation. The procedure relies on the obtained properties for the derivation order; however, there could be additional properties that have not been established. Moreover, the adaptation gain was not optimized despite its role as described e.g., in Theorem 3(iii) and Proposition 3. In the following, we will propose a way to fix these issues in order to find an optimal estimator.

From the proposed adaptive designs, we draw two conclusions: (a) the adaptive gain λ and the differentiation order α affect the performance in such a way that the choice $\alpha < 1$ or $\lambda \neq 1$ can be optimal, and (b) the use of time-varying λ and α do not affect the stability. It follows that a second-level adaptation to enhance the performance is possible without destabilizing the adaptive process. As a second-level adaptation, the adjustment of λ, α should be concerned with the optimization of a performance index J. Unlike the estimation in the first-level, there is no a priori λ_0, α_0 for which λ, α should converge.

Suppose that we have the explicit dependence $J = J(\lambda(t), \alpha(t))$ for $J \geq 0$ a differentiable function. If we want to minimize J, then we set

$$\dot{\alpha}(t) = -\mathcal{P}_{[a,b]}\{\frac{\partial J}{\partial \alpha(t)}\},$$
$$\dot{\lambda}(t) = -\mathcal{P}_{[0,\infty)}\{\frac{\partial J}{\partial \lambda(t)}\}, \tag{3.21}$$

where $\mathcal{P}_{[a,b]}$ is the projective operator retaining the variable in the set $[a,b]$. This operator is the identity in the interior of the domain $\mathcal{D} := [a,b] \times [0,\infty)$ and only acts on the boundary (see [Tao03]). Independently of the convexity or even the existence of extreme points of J, we have the following result (see Appendix 6.6).

Proposition 4 J is nonincreasing along (3.21), strictly decreasing when $\frac{\partial J}{\partial \alpha} \neq 0$ or $\frac{\partial J}{\partial \lambda} \neq 0$ in the interior of $\mathcal{D}$, and λ, α remain on $\mathcal{D}$.

However, due to the fact that λ, α appear in the differential equation of $\hat{\theta}$, such an explicit dependence is in general not manifest after the integration. For instance, according to (3.5), we would have

$$\hat{\theta}(t) = -\int_0^t k_{\alpha(s)}(t-s)\lambda(s)\mathrm{e}(s)m(s)\mathrm{d}s,$$

where we have dropped the initial condition as it is a constant. Hence, $\hat{\theta}(t)$ and $\mathrm{e} = m'(\hat{\theta} - \theta)$ do not depend explicitly on $\lambda(t), \alpha(t)$. This can be fixed by considering them piecewise constants functions with their values actualized at the end of small enough intervals. Indeed, in the interior of such intervals, we can always write

$$\hat{\theta}(t) = -\int_0^t k_{\alpha(t)}(t-s)\lambda(t)\mathrm{e}(s)m(s)\mathrm{d}s,$$

where $\alpha(t), \lambda(t)$ are constant. Consider $J = \mathrm{e}^2$, then

$$\frac{\partial J}{\partial \alpha(t)} = -\mathrm{e}(t)m'(t)\int_0^t \frac{\partial}{\partial \alpha(t)}\big(k_{\alpha(t)}(t-s)\mathrm{e}(s)\big)\lambda(t)m(s)\mathrm{d}s,$$
$$\frac{\partial J}{\partial \lambda(t)} = -\mathrm{e}(t)m'(t)\int_0^t \frac{\partial}{\partial \lambda(t)}\big(\lambda(t)\mathrm{e}(s)\big)k_{\alpha(t)}(t-s)m(s)\mathrm{d}s,$$

To analyse these laws, we will ignore the explicit dependence of $\mathrm{e}(s)$ on values at $t > s$. We obtain

$$\frac{\partial J}{\partial \alpha(t)} = -\mathrm{e}(t)m'(t)\int_0^t K_{\alpha(t)}(t-s)\mathrm{e}(s)\lambda(t)m(s)\mathrm{d}s,$$
$$\frac{\partial J}{\partial \lambda(t)} = -\mathrm{e}(t)m'(t)\int_0^t k_{\alpha(t)}(t-s)\mathrm{e}(s)m(s)\mathrm{d}s,$$

where $K_\alpha(t) := \frac{\partial}{\partial \alpha}k_\alpha(t) = -\frac{\Gamma'(\alpha)}{\Gamma^2(\alpha)}t^{\alpha-1} - \frac{1}{\Gamma(\alpha)}\log(t)t^{\alpha-1}$. We see that K_α flips its sign from $+$ to $-$ as t grows. Therefore, α is forced to decrease and then to increase, meanwhile λ is forced only to increase, where we use $sign(\mathrm{e}(t)m(t)) = sign(\mathrm{e}(s)m(s))$ component-wise for s near of t and provided that they are not zero. This confirms that the transient is improved by starting with a lower derivation order, that the gradient ensures faster convergence and that there exists a relation of λ with the convergence's speed. Also, this might reveal a role of $\alpha < 1$ in reducing the overshoot since the latter occurs when e changes its signs, a property which is otherwise hard to obtain formally.

Chapter 4

Applications

In this section, we present some cases exploiting particular features of fractionalization, and applications of the designs in Section 3 in control and observer problems for nonlinear systems with parametric uncertainty. Notice that we have already shown an application in the parameter estimation for regression models; in fact, in the applications of this section, the problem is finding a suited regression equation in each case. We consider nonlinear problems as they include the linear case.

4.1 Case 1: Modelling

Engineering models stemming from classical physical laws, which are no other things than expressions of causal (local) relationships, are formulated in integer derivatives. However, parameters appearing in these models seem to obey other laws as they must be measured or estimated in practice. In this sense, there is room to introduce models for the parameters based on fractional derivatives. For it is common to attribute a slower dynamic (e.g., by using a different time-scale) to the parameters' evolution that could be better described by polynomial rather than exponential convergence.

Consider the regression model (3.1) with an evolution of its parameters described by

$$D^{\alpha_0}\theta(t) = f_1(m, y, t)\theta(t) + f_2(m, y, t), \tag{4.1}$$

where f_1, f_2 are known functions and $\alpha_0 \in (0, 1)$ is also known (thus, the unknown is $\theta(0)$). Due to the nonlocality and unlike the case $\alpha_0 = 1$, it is not enough to known $\theta(t_0)$ to predict future values when $t_0 \neq 0$.

Since θ is completely unknown, we cannot estimate θ using an observer design when f_1 does not ensure asymptotic stability. Moreover, if we want to track a variable having a given order of decreased/increased, then we cannot employ an estimator with a faster or slower order because such a difference hampers the parametric error convergence. Due to the speed hierarchy explained in Section 2.5, we are compelled to use α_0 in the following modified observer

$$D^{\alpha_0}\hat{\theta} = f_1(m,t)\hat{\theta} + f_2(m,t) - P^{-1}F(\mathrm{e},m), \tag{4.2}$$

where F is any of the designs of Section 3 and $P \in \mathbb{R}^{n\times n}$ is specified in the following result (see the proof in 6.7.1).

Proposition 5 Suppose that the parameters are ruled by (4.1). If there exists $P > 0$ such that $Pf_1 + f_1'P \leq 0$, then the same statements of Theorem 1, 2 or 3 hold when using the estimator (4.2).

Remark 4 The hypothesis of Proposition 5 is just a stability condition on f_1. Obviously, if f_1 ensures asymptotic stability, then we obtain $\lim_{t\to\infty}(\hat{\theta}(t) - \theta(t) = 0$.

4.2 Case 2: Long-memory

The remembering property of fractional systems, described in Section 2.1.2 and that established that the solution returns to the initial condition when its fractional derivative is made zero, can be used to simplify the projective gradient method since it will be not necessary computing the normal to the trapping surface, which avoids getting stuck on the boundary.

We will employ the designs of Section 3 with the choice of the adaptation gain as follows

$$\lambda(t) := \lambda(t, \hat{\theta}(t)) := \mathbf{1}_\Omega(\hat{\theta}(t))\lambda'(t), \tag{4.3}$$

where $\mathbf{1}_\Omega(\hat{\theta})$ is the indicator function which is zero whenever $\hat{\theta} \notin \Omega \subset \mathbb{R}^n$ and $\lambda' > 0$ is an arbitrary scalar function. Without loss of generality, we assume that Ω is a closed hypercube. The following result is proved in Section 6.7.2.

Proposition 6 Suppose that $\hat{\theta}(0) \in \Omega \subset \mathbb{R}^n$. Then, the use of (4.3) in the designs of Section 3 ensures $\hat{\theta}(t) \in \Omega$ when $\alpha \in (0,1]$ and $\hat{\theta} \in int(\Omega)$ a.e. when $\alpha \in (0,1)$ for any $t > 0$ and any kind of disturbance altering those adaptive schemes. Moreover, the properties in Section 3 are kept.

4.3 Application 1: Control Problem

Consider the class of nonlinear systems that can be written in the following form

$$\begin{aligned} \dot{x}_\mathrm{i} &= x_{\mathrm{i}+1}, \quad 1 \leq \mathrm{i} \leq n-1, \\ \dot{x}_n &= \beta_0(x)u + \theta'\varphi(x,t) + \nu, \\ y &= x_1, \end{aligned} \tag{4.4}$$

where $x = (x_1, \ldots, x_n)'$ is measurable, $\theta \in \mathbb{R}^p$ is an unknown vector, and φ, β_0 are known continuously differentiable functions of suited dimensions with $\beta_0(x) \neq 0$ for any $x \in \mathbb{R}^n$

(i.e. (4.4) is controllable). ν is a time-varying disturbance that together with $\varphi(x,t)'\theta$ can describe a general nonlinear function according to the universal approximation property (e.g., see [Cot90]). For linear systems, (4.4) is just their canonical controllable form. (4.4) is the state representation of input-output systems of the type

$$y^{(n)} + f(y, \dot{y}, \ldots, y^{(n-1)}) + g(y, \dot{y}, \ldots, y^{(n-1)})u = 0,$$

where the unknown function f is approximated by $\varphi'\theta + \nu$. Although g could be approximated similarly, this would lead to a more complicated solution (e.g., by using projection) to preclude the loss of controllability (no invertibility of g or its approximation in every instant). Besides linear systems (e.g., basic electric circuits) or commonly used linearization of nonlinear systems (aircraft longitudinal dynamics [SL03]), an important class of systems included in the above representation are mechanical systems in which the input-output relation is obtained by the Euler-Lagrange equations (e.g., see the robotic system in [Tao03, p.477]).

The control problem is to design an input function u such that the output y tracks a given reference y_m, i.e., such that the tracking error $y - y_m$ goes to zero asymptotically, and all the closed-loop signals remain bounded for arbitrary initial conditions. The reference signal y_m is generated by the following known reference model

$$\begin{aligned} \dot{x}_{m,\mathrm{i}} &= x_{m,\mathrm{i}+1}, \quad 1 \leq \mathrm{i} \leq n-1 \\ \dot{x}_{m,n} &= -m'x_m + k_m r, \\ y_m &= x_{m,1}, \end{aligned}$$

where $m \in \mathbb{R}^n$ is such that $s^n + m_{n-1}s^{n-1} + \ldots + m_1 s + m_0$ is Hurwitz, $k_m > 0$ and r is a bounded piecewise continuous function. By choosing $r = 0$ and assuming observability, this problem includes the regularization target (i.e. $x \to 0$ as $t \to \infty$).

We solve this problem by using an estimation-based approach separating the estimation from the control design. For the latter, we employ a robust controller ensuring boundedness when the estimation of θ is bounded. It can be done using the design in [KK95, Eqns. (3.1) & (3.9)], which takes the following form for system (4.4)

$$\begin{aligned} z_\mathrm{i} &= x_\mathrm{i} - x_{m,\mathrm{i}} - v_{\mathrm{i}-1}, \\ v_0 &= 0, \\ v_\mathrm{i} &= -z_{\mathrm{i}-1} - c_\mathrm{i} z_\mathrm{i} - \hat{\theta}' w_\mathrm{i} + \sum_{k=1}^{\mathrm{i}-1} \left(\frac{\partial v_{\mathrm{i}-1}}{\partial x_k} x_{k+1} + \frac{\partial v_{\mathrm{i}-1}}{\partial y_r^{(k-1)}} y_r^{(k)} \right) - s_\mathrm{i} z_\mathrm{i} \\ s_\mathrm{i} &= \kappa_\mathrm{i} |w_\mathrm{i}| + g_\mathrm{i} \left| \frac{\partial v_{\mathrm{i}-1}}{\partial \hat{\theta}} \right|^2, \quad \kappa_\mathrm{i}, g_\mathrm{i} > 0 \\ w_j &= 0, \quad j = 1, \ldots, n-1, \\ w_n &= \varphi, \\ u &= \frac{1}{\beta_0(x)} [v_n - m_0 x_{m,1} - \ldots - m_{n-1} x_{m,n} + k_m r], \end{aligned}$$

for $\mathrm{i} = 1, \ldots, n$. This controller yields the equation [KK95, Eqn. (3.2)] (notice that in our case $D = 0$),

$$\dot{z} = A(z, \hat{\theta}, t) z + W\tilde{\theta}, \tag{4.5}$$

where $W = [0 \ \dots \ 0 \ w_n]$, and A ensures input-to-state stability (ISS) with respect to the input $\hat{\theta}$ [KK95, Lemma 3.2] due to the damping terms $s_{\rm i}$. For the estimator, the following regression equation can be obtained from (4.4) (via the swapping technique [KK95, Section V.B] or Section 6.1),

$$\varepsilon = \Omega' \tilde{\theta} + \tilde{\varepsilon} + \bar{\nu}, \tag{4.6}$$

where ε is a measurable signal, $\bar{\nu}$ is a bounded function vanishing when $\nu \equiv 0$, and the functions $\Omega \in \mathbb{R}^{p \times n}$ and $\tilde{\varepsilon} \in \mathbb{R}^n$ satisfy $\dot{\Omega}' = \bar{A}\Omega' + \phi'$ and $\dot{\tilde{\varepsilon}} = \bar{A}\tilde{\varepsilon}$, respectively, for $\bar{A} \in \mathbb{R}^{n \times n}$ an asymptotically stable designer-chosen matrix and ϕ a matrix with zero entries but the last column which is φ. Since $\tilde{\varepsilon}$ converges exponentially to zero, we can pick the last component of ε to obtain a regression equation and the following result (see the proof in Section 6.7.3).

Proposition 7 Suppose that the last column of Ω has FE. Then, the control and the identification goals are achieved asymptotically using the estimator (3.6), i.e. y converges to y_m and $\hat{\theta}$ converges to θ as $t \to \infty$, and all the closed-loop signals remain bounded for arbitrary initial conditions. In addition, the robust and transient properties of the estimator are inherited in the closed-loop performance.

The hypothesis of Proposition 7 is unsatisfactory in the sense that it depends on closed-loop signals. However, in contrast to hypotheses involving a persistent excitation condition, this hypothesis can be satisfied with the help of an external excitation without compromising the control aim. Roughly speaking, we can use the reference $r_0 = r + r_{\rm e}$ where $r_{\rm e}$ is a 'rich enough' signal guaranteeing the fulfilment of the hypothesis and vanishes once that it is achieved (say) at $t = T$. Since $r_0 = r$ for any $t > T$, y still converges to y_m. The following result formalizes this idea (see the proof in Section 6.7.4) and, when particularized to the linear case, the condition is weaker than the one obtained with the fractionalized gradient (see [GD18a]).

Proposition 8 The hypothesis of Proposition 7 is verifiable *a priori* using external excitation. In particular, in the linear case, it is enough for $r_{\rm e}$ to have p different frequencies.

4.4 Application 2: Observer Problem

In the control example above, we made the assumption of full knowledge of the state. It is common, however, that only a function of the state is available and that its derivatives cannot be computed without introducing noise. This motivates the problem of estimating the state (i.e. the derivatives of the output) for systems with parametric uncertainty.

Consider the nonlinear system

$$\begin{aligned} \dot{x} &= f(x) + \big(g_0(x) + \sum_{\rm i=1}^{m} g_{\rm i}(x)\theta_{\rm i}\big)u + \sum_{\rm i=1}^{m} q_{\rm i}(x)\theta_{\rm i} + \nu, \\ y &= h(x), \end{aligned} \tag{4.7}$$

where $x : [0, \infty) \to \mathbb{R}^n$ is not measurable, ν is a disturbance, and $y, u : [0, \infty) \to \mathbb{R}$ are measurable. Let $\theta \in \mathbb{R}^m$ be the vector of the unknown parameters θ_{i}. Notice that system (4.7) is more general than (4.4) because no controllability issue is involved *a priori*.

Suppose that $f, g_0, g_{\mathrm{i}}, q_{\mathrm{i}} : \mathbb{R}^n \to \mathbb{R}^n$ and $h : \mathbb{R}^n \to \mathbb{R}^p$ are known smooth functions satisfying the geometric (i.e. coordinate-free) conditions in [MT93, Th. 2.1, p.19] for $1 \leq \mathrm{i} \leq m$. These conditions trivially holds in the linear case when the system is observable. Then, there exist a global diffeomrphic transformation $\xi = T(x)$, which is independent of θ, and a function M such that (4.7) is input-to-output equivalent to (see e.g. [MST01])

$$\begin{aligned} \dot{z} &= Az + \Psi_0(y, u) + \mathrm{d}\beta'(y, u)\theta + \Phi(y, u)w, \\ y &= Cz, \end{aligned} \tag{4.8}$$

where $z = \xi + M\theta$, $\beta' = C(AM + \Psi)$, Ψ_0, Ψ, Φ are known functions of suited dimensions, A, C are in canonical observable form with zeros in the last row of A, d is known, the triple (A, d, C) is strictly positive real, w is the disturbance term coming from ν, and $M : [0, \infty) \to \mathbb{R}^{n \times p}$ is the solution to

$$\begin{aligned} \dot{M} &= (I_n - \mathrm{d}C)AM + (I_n - \mathrm{d}C)\Psi, \\ M(0) &= [0, N_0']', N_0 \in \mathbb{R}^{n-1 \times p}. \end{aligned}$$

Since (4.8) can be seen as a linear system with nonlinear inputs, the following Luenberger observer is a natural choice for the state estimator:

$$\begin{aligned} \dot{\hat{z}} &= A\hat{z} + \Psi_0(y, u) + \mathrm{d}\beta'(y, u)\hat{\theta} - K(\hat{y} - y), \\ \hat{y} &= C\hat{z}, \end{aligned} \tag{4.9}$$

where K is such that $A_o := A - KC$ is Hurwitz, and $\hat{\theta}$ is an estimation of θ.

To obtain a regression equation, notice that (4.8) can be written in Laplace domain as $y = W_1(s)\big(\beta'(y, u)\theta\big) + W_2(s)\big(\Psi_0(y, u) + \Phi(y, u)w\big)$, where $W_1 = C(sI - A)^{-1}\mathrm{d}$ and $W_2 = C(sI - A)^{-1}$ are known filters. By defining $\mathsf{y} := y - W_2(s)\big(\Psi_0(y, u)\big)$, $\bar{\nu} := \Phi(y, u)w$ and $\mathsf{m} := W_1(s)\big(\beta(y, u)'\big)$, we obtain

$$\mathsf{y} := \mathsf{m}'\theta + \bar{\nu}. \tag{4.10}$$

Due to the motivation coming from the control problem, we use the finite-time estimator (3.14) for the regression equation (4.10) because it ensures a fast convergence to the true state as a consequence of the finite-time of parametric convergence. Thus, we obtain the following result (see Section 6.7.5 for the proof).

Proposition 9 Suppose that m satisfies the condition of Theorem 3. Then, the function $\hat{x} := T^{-1}(\hat{z} - M\hat{\theta})$ robustly estimate x, meaning that $\hat{x}(t) - x(t) \to 0$ as $t \to \infty$ at exponential order when $\nu \equiv 0$, and $\hat{x} - x$ remains bounded or converges to zero when ν is bounded or ν converges to zero, respectively, for y, u bounded. In addition, $\hat{\theta}$ robustly estimate θ and the performance properties of Theorem 3 are inherited by the observer.

Notice that the hypothesis of Proposition 9 depends ultimately on the signals u, y. Thus, using Proposition 1 and the arguments of the proof of Proposition 8, the a priori verification of the hypothesis amounts to know the spectral relation between y, u. In the linear case, where the frequencies in the input are transferred to the output, this hypothesis can be easily recast in terms of the spectral content of u. Although non-a-priori hypotheses are also acceptable in adaptive literature (cf. [MST01, Proposition 1]), we remark that the condition of Theorem 3 can be verified online in finite-time since one could add frequencies through u until the condition is satisfied, which cannot be done with persistent excitation conditions.

Turning back to the control problem of system (4.4), it follows that the condition $\varphi = \varphi(u, y)$ ensures the existence of the observer (4.8). Moreover, the resulting observer has the same feedback form that (4.4), which implies that the a priori verifiability of Proposition 9 follows from Proposition 8 when using the same control structure (with the external reference) with x replaced by $\hat{x}$. From Proposition, 9 we have that $\hat{x}(t) \to x(t)$ as $t \to \infty$ and, due to the smoothness of the functions v_{i} and $\beta > 0$, $u(\hat{x}) \to u(x)$ as $t \to \infty$. Due to the robustness property of the controller established in Proposition 7, we obtain $y \to y_m$ as $t \to \infty$ and all the closed-loop signals remain bounded.

Chapter 5

Simulation Study

The aim of this section is to present qualitative and quantitatively the role of the free parameters in the adaptive scheme, coming from the fractional operators, in a concrete application of the theoretical results already obtained. To this aim, we consider the tracking problem for the system

$$\begin{aligned}\dot{x}_1 &= x_2,\\ \dot{x}_2 &= u + \theta x_1^2,\\ y &= x_1,\end{aligned}$$

where θ is unknown and the reference to track is $y_r \equiv 1$. Despite its simplicity, this system exhibits finite-escape time when a gradient-like estimation together with a certainty equivalent control is used due to the non-Lipschitz term (see e.g. [SM92, Example 1]). This is handled with the damping term s_{i} providing the robust part of the controller Section 4.3.

We will simulate the fractional operators using a commonly accepted algorithm that relies on the approximating expression in 2.4. The simulations will confirm our theoretical results, but they are just approximations whose precision degree we will not discuss. If every simulation is an approximation, then those of fractional operators are approximations of approximations. This is one of the reasons to provide theoretical arguments for the advantages/disadvantages that have been observed in simulations in the revised literature.

5.1 State feedback adaptive control

We start by assuming that $x = (x_1, x_2)'$ is available. It turns out that the hypothesis of Proposition 7 in the scalar case is verified if x_1 is not the zero function, which is ensured because the reference is not the zero function. Therefore, we can solve the problem using Proposition 7. Recall that the estimator with memory (3.6) has introduced two degrees of freedom whose effect on the performance is the subject of the simulations below.

5.1.1 Role of β

As β decreases, $k_\beta(t,s) = (t-s)^{-\beta+1}$ becomes larger when $(t-s) \to 0^+$ and smaller when $(t-s) \to \infty$. This means that k_β acts as a forgetting factor in the convolution integrals of (3.6) and β can be seen as a memory parameter quantifying the amount of past data to be considered. In particular, $\beta = 1$ indicates that all past is considered uniformly and $\beta = 0$ that the relevance is assigned to more recent data; in other words, the forgetting is faster as β decreases.

Recalling the heuristic "forgetting fast improves robustness" [HC89], β should be a relevant parameter to enhance the robustness. More precisely, the choice of β has importance when older data should be ignored. The latter occurs in two practical situations: when the parameters are time-varying and when the measurements contain noise. To test the first case, we simulate a change in θ from $2 \to -1$ at $t = 3$. Figure 5.1 depicts how the lowering of β improves the alertness of the estimator to track this parametric variation. To test the second case, we introduce an additive white noise ν in the dynamic equation of x_2. Table 5.1 shows that the lowering of the forgetting number β improves the robustness as measured in the $L_2(0,T)-$norm of $(y - y_r, x_2)$. In both cases, we have considered a fixed value $\alpha = 1$.

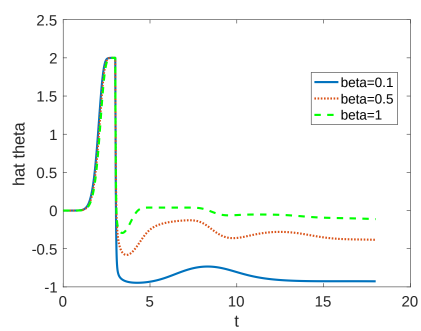

Figure 5.1: $\beta-$dependence of the estimator when θ changes at $t = 3$ from 2 to -1

$\beta \setminus$ NP	0.001	0.01	0.1
1	4.279	4.506	6.162
0.5	3.966	4.26	5.828
0.1	**3.647**	**3.951**	**5.461**

Table 5.1: The dependence of the $L_2(0,T)-$norm of $(y - y_r, x_2)$ regarding the parameter β for white noise disturbances of noise power NP. Simulation time was $T = 20$.

5.1.2 Role of α

Now, we set $\beta = 0.1$ and study the role of α. We will illustrate that the claim of Theorem 2 on the transient behaviour of the estimator holds also for the transient performance of the controlled output as stated in Proposition 7. Figure 5.2 shows the transient improvement (i.e. $\hat{\theta}$ becomes closer to θ) in the estimation process as α decreases. Figures 5.3 and 5.4

depict the transient improvement and the overshoot reduction as α decreases in the closed-loop performance. Notice that Figures 5.3 and 5.4 also show that the solution with a greater α becomes closer to the tracking aim than the solutions with lower α as time goes. This corresponds to the fact that the convergence speed's order increases as α increases, which stems from the relation $E_\alpha(-\lambda t^\alpha) \sim t^{-\alpha}$, the proof of Theorem 2, and the fact that the convergence is exponential for $\alpha = 1$.

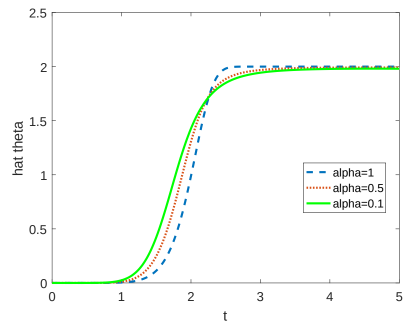

Figure 5.2: $\alpha-$dependence of the estimator at the transient period.

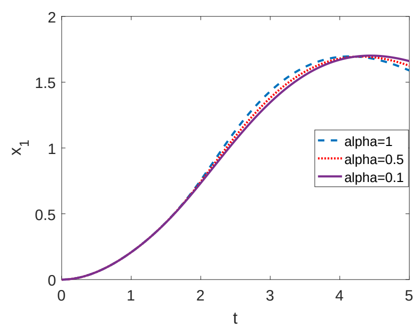

Figure 5.3: Transient behaviour of x_1 depending on the choice of α.

The fact that a slower convergence is obtained when using a smaller α can be exploited to enhance the robustness, for it must be recalled the heuristic "slow adaptation generally improves robustness" [Ous91, HC89]. In addition, since the robustness is enlarged according to Theorem 2 in the sense that some signals making to diverge the scheme when α is large ensure boundedness when α is small, it should be expected attenuation of the disturbances as α decreases. Indeed, Table 5.2 confirms that a smaller than one value of α yields the smallest value of the $L_2(0,T)-$norm of $(y - y_r, x_2)$ when noise is introduced in the dynamic equation of x_2. However, the noise attenuation is counteracted in the computation of the L_2-norm with the slower convergence; that is why the best value of α obtained from Table 5.2 is not the smallest one. This compensation is also shown when the noise becomes more relevant than the convergence (last column in Table 5.2) because $\alpha = 0.1$ yields a lesser norm value than $\alpha = 1$ despite the exponential convergence.

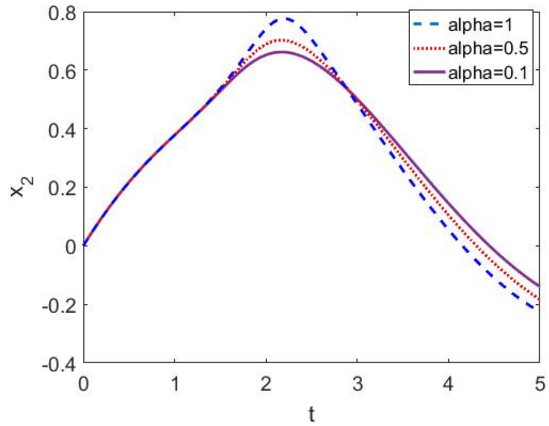

Figure 5.4: Transient behaviour of x_2 depending on the choice of α.

$\alpha \setminus$ NP	0.001	0.01	0.1
1	3.649	3.953	5.461
0.5	**3.571**	**3.851**	**5.297**
0.1	3.746	3.99	5.347

Table 5.2: The dependence of the $L_2(0,T)-$norm of $(y-y_r, x_2)$ regarding the parameter α under white noise disturbances of noise power NP. Simulation time was $T=20$.

5.2 Output feedback adaptive control

We now assume that x is not available and design the controller with an estimation $\hat{x}$ of x obtained only from input-output measurements. Due to the scalar nature of the uncertainty, we can use Proposition 9 because its hypothesis is verified if x_1 is not the zero function. Using the same controller than above with $\hat{x}$ instead of x, this hypothesis holds because the reference is not the zero function.

α	$\|\tilde{x}\|_2^2$
1	2.926
0.5	2.925
0.1	**2.923**

Table 5.3: The dependence of the $L_2(0,T)-$norm of $\tilde{x}=(\hat{x}_1-x_1, \hat{x}_2-x_2)$ regarding the parameter α . Simulation time was $T=2$.

Figure 5.5 depicts the finite-time estimation of the parameter as a function of the derivation order α using (3.14). It is verified the statement of Theorem 3 on the transient period since the performance is improved as α decreases. Notice that the switched time occurs at a common instant $t_f=0.3$; if we had use a common μ in the notation of Theorem 3, then the convergence time would be decreased as α decreased. We took this choice because of the easiness of computation. Figure 5.6 and 5.7 show the convergence of the state estimation error. Due to the short time of convergence to the unknown parameter, the transient dependence on the derivation order of the estimators is better appreciated in Table 5.3 by using the $L_2(0,T)-$norm of the estimation error $\tilde{x}$ for $T = 2$. This confirms the performance heritability claimed in Proposition 9.

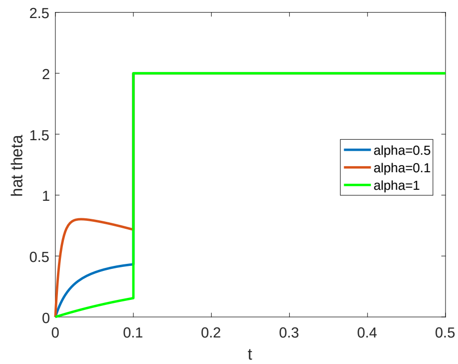

Figure 5.5: $\alpha-$dependence of the estimator at the transient period.

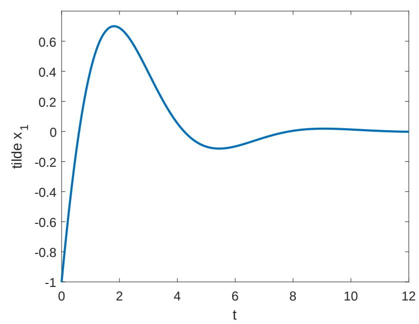

Figure 5.6: Convergence of the state estimation error $\tilde{x}$ for $\alpha = 0.1$.

Figure 5.8 shows the controlled output when the controllers associated with each adaptive observers obtained for different values of α are applied. Since the transient performance of the state and parameter estimation is improved as α decreases, the transient in the controlled output is also improved as shown in the decreasing of the overshoot in Figure 5.9 and in the first row of Table 5.4 in which the $L_2(0,T)-$norm of e $:= (x_1 - 1, x_2)$ was computed for $T = 2$.

$\alpha \setminus$ NP	0	0.001	0.01	0.1
1	7.744	16.3	16.45	17.1
0.5	7.72	16.25	16.31	17.04
0.1	**7.689**	**16.19**	**16.26**	**16.96**

Table 5.4: The dependence of the $L_2(0,T)-$norm of e $:= (x_1 - 1, x_2)$ regarding the parameter α under white noise disturbances of noise power NP. Simulation time was $T = 2$ for the first column and $T = 10$ for the rest.

So far, transient improvements have been verified. We now study the robust behaviour of the closed-loop when white noise is introduced in the equation of x_2. Since the robustness properties are inherited from that of the finite-time estimator, we will look only on the

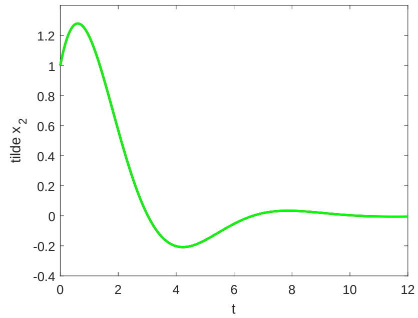

Figure 5.7: Convergence of the state estimation error $\tilde{x}$ for $\alpha = 0.1$.

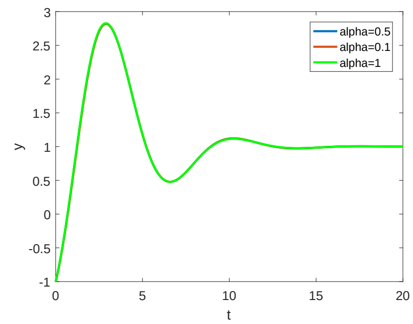

Figure 5.8: Controlled output as a function of α.

aggregated effect on the closed-loop system. Table 5.4 shows that the noise attenuation in the $L_2(0,T)-$norm of e is increased as the derivation order of the estimator is decreased. In contrast to Table 5.2, the noise attenuation as α decreases is not (significantly) counteracted by the speed of convergence of the estimation since we are using the finite-time estimator. As a consequence, the lowest α achieves the best attenuation.

Moreover, since the controller is of feedback type, this noise attenuation should yield an attenuation in the $L_2(0,T)-$norm of the control, which is effectively observed in Table 5.4. We see a larger noise attenuation than the observed in Table 5.4 or even in Table 5.3 when α is decreased. This could be explained as the control mainly acts at the transient time when the fractional order plays a crucial role (see the first column of Table 5.3 and how the gap is kept after noise addition). This effect is relevant in practice as the $L_2(0,T)-$norm is a measure of the energy consumption. Therefore, we have obtained as an indirect consequence the minimization of the control energy by lowering the derivation order. This, together with the minimization of the $L_2(0,T)-$norm of e^2, entails a kind of LQ control optimized in the derivation order of the observer.

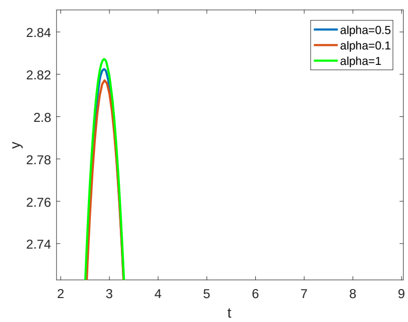

Figure 5.9: Zoom-in to the transient behaviour of the output as a function of α.

$\alpha \setminus$ NP	0	0.001	0.01	0.1
1	1.762	5.207	6.345	16.63
0.5	1.675	5.122	6.214	16.53
0.1	**1.544**	**5.009**	**6.107**	**16.41**

Table 5.5: The dependence of the $L_2(0,T)-$norm of u regarding the parameter α under white noise disturbances of noise power NP. Simulation time was $T = 2$ for the first column and $T = 10$ for the rest.

Chapter 6

Proofs

6.1 Regression equations

The problem is to associate a regression form, built from measurable signals, to a given dynamic equations. The first case represents a general nonlinear system with a linear-in-parameter term. That is, consider

$$\dot{x}(t) = f(x, u, t) + \phi(x, u, t)'\theta,$$

where $x : \mathbb{R}_{\geq 0} \to \mathbb{R}^n$, the control function u and $\phi : \mathbb{R}_{\geq 0} \to \mathbb{R}^{n\times m}$ are measurable signals. By integration, we have for any $t \geq t_0$

$$x(t) - x(t_0) = \int_{t_0}^{t} f\mathrm{d}\tau + (\int_{t_0}^{t} \phi\mathrm{d}\tau)'\theta.$$

Define $y(t) := x(t) - x(t_0) - \int_{t_0}^{t} f\mathrm{d}\tau$ and $m(t) := \int_{t_0}^{t} \phi\mathrm{d}\tau$. Then,

$$y(t) = m(t)'\theta.$$

The second case represents a error equation of the type

$$\dot{\mathrm{e}} = A\mathrm{e} + W_1\tilde{\theta} + W_2\dot{\hat{\theta}}, \tag{6.1}$$

where $\mathrm{e} : \mathbb{R}_{\geq 0} \to \mathbb{R}^n$, $W_{1,2} : \mathbb{R}_{\geq 0} \to \mathbb{R}^{n\times m}$ are measurable signals and $A = A(\mathrm{e}, \hat{\theta}, t)$ is a designer-chosen matrix such that (6.1) is input-to-state stable (ISS), with $(\tilde{\theta}, \dot{\hat{\theta}})$ seen as the input. This model appears in adaptive backstepping designs. Common dynamic errors model such as

$$\mathrm{e}^{(m)} + a_1\mathrm{e}^{(m-1)} + \ldots + a_m\mathrm{e} = \tilde{\theta}'w$$

where $\mathrm{e} : \mathbb{R}_{\geq 0} \to \mathbb{R}$, can be written as (6.1), with $W_2 \equiv 0$ and A a constant stable matrix, by redefining the error function as a vector of components $\mathrm{e}^{(\mathrm{i})}$.

The regression formulation via swapping is obtained by defining $\dot{\chi}_0 = A\chi_0 + W_1\hat{\theta} - W_2\dot{\hat{\theta}}$, $\dot{\chi} = A\chi + W_1$, $\mathsf{y} := \mathrm{e} + \chi_0$ and $\varepsilon = \mathrm{e} + \chi_0 - \chi\theta$. Then $\mathsf{y} = \varepsilon + \chi'\theta$ and $\dot{\varepsilon} = A\varepsilon$. Choosing $\chi_0(0) = -z(0)$ and $\chi(0) = 0$, we arrive to

$$\mathsf{y} = \chi\theta$$

6.2 Proofs for the fractionalized gradient

6.2.1 Proof of Theorem 1

Proof. The following holds

$$\begin{aligned}
D^{\alpha}_{0+}\tilde{\theta} &= -\lambda m m'\tilde{\theta} \\
\tilde{\theta}' D^{\alpha}_{0+}\tilde{\theta} &= -\lambda \mathrm{e}^2 \\
D^{\alpha}_{0+}\tilde{\theta}'\tilde{\theta} &\leq -2\lambda \mathrm{e}^2 \leq 0 \\
&\implies \tilde{\theta}'\tilde{\theta}(t) \leq \tilde{\theta}'\tilde{\theta}(0), \quad \forall t > 0 \\
&\implies I^{\alpha}_{0+}\mathrm{e}^2(t) \leq \frac{1}{\lambda(t)}\tilde{\theta}'\tilde{\theta}(0), \quad \forall t > 0,
\end{aligned}$$

where the first inequality is due to [TT18]. The first implication above, obtained by fractional integration, yields statement (i) and tell us that the transient behaviour is uniformly bounded by $\tilde{\theta}'\tilde{\theta}(0)$. The second implication, obtained also by integration and the fact that $\tilde{\theta}'\tilde{\theta} \geq 0$, and the result in [GD17a] yield statement (ii). It follows that $I^{\alpha}_{0+}\mathrm{e}^2$ can be arbitrarily reduced by enlarging λ. Since $I^{\alpha}_{0+}\mathrm{e}^2(t) \to 0$ implies $\mathrm{e} \to 0$ under uniform continuity [Gall15a], the asymptotic behaviour can be completely controlled through λ (the conditions on m and the boundedness ensure that e is uniformly continuous). This proves item (iii).

Item (iv) can be found in [Gall15a]. Item (v) follows from the second example in [GD19a]. The first part of item (vii) follows from [GD18a, Theorem 1]. For the second part, consider the particular case $m = -\mathrm{e}$ and $n = 1$. Then, $\mathrm{e} = \nu/(1 + \tilde{\theta})$ and

$$D^{\alpha}\tilde{\theta} = \mathrm{e}^2 = \frac{\nu^2}{(1 + \tilde{\theta})^2}.$$

Picking $\nu = \nu_0(1 + \tilde{\theta})$, we conclude that $\hat{\theta}$ remains bounded when $\nu_0 \in L_{2,\alpha}$ and $\hat{\theta} \to \infty$ as $t \to \infty$ when $\nu_0 \in L_{2,\beta} \setminus L_{2,\alpha}$ for any $\beta < \alpha$. The claim follows by noting that $L_{2,\alpha} \subsetneq L_{2,\beta}$ [GD18a, Remark 1], and hence, $L_{2,\beta} \setminus L_{2,\alpha}$ is not empty. (For well definition, we can take any particular $\hat{\theta}(0)$ such that $\tilde{\theta}(0) > -1$ since $D^{\alpha}\tilde{\theta} > 0$ ensures $\tilde{\theta}(t) > -1$).

Finally, we prove (vi). Let $V = \tilde{\theta}'\tilde{\theta}$. Without loss of generality, we assume $D^{\alpha}_{0+}\tilde{\theta}(0) \neq 0$; otherwise, one can redefine the initial time. Then, $|\frac{\mathrm{d}}{\mathrm{d}t}V(t)| = \mathcal{O}(t^{\alpha-1})$ as $t \to 0^+$ (see e.g., [GAD19b]), Moreover, the sign of $\frac{\mathrm{d}}{\mathrm{d}t}V(t)$ must be nonpositive for at least a small enough interval $[0, \varepsilon)$ as it can be seen from the fact that $D^{\alpha}V = I^{1-\alpha}\dot{V} \leq 0$. Using that $V(0) > 0$, the differentiability of $\hat{\theta}$ for $t > 0$ [GAD19b] and the mean value theorem (MVT), we can find a small enough number $\varepsilon < \varepsilon$ such that $V_\alpha(t) < V_\beta(t)$ for any $t \in [0, \varepsilon)$, where V_α and V_β denote the function V evaluated on the solutions of (3.5) for derivation order α, β, respectively, and fixed initial conditions. This is because the MVT ensures that $V_\gamma(t) - V_\gamma(0^+) = V_\gamma(t) - V_0 = \dot{V}_\gamma(\xi)t$ for $\xi \in (0, t)$, any $\gamma \in (0, 1]$; this and the fact that $\dot{V}_\gamma(\xi)$ grows as γ decreases for ξ near of 0^+ yield the claim. Moreover, when m does not depend on θ (e.g. in identification problems), it follows that $\mathrm{e}_\alpha(t) < \mathrm{e}_1(t)$ for a sufficient small interval $[0, \varepsilon]$. □

6.2.2 Proof for the Remark 1

Suppose there exist $T, \varepsilon(t) > \varepsilon_0, r < n$ and an orthonormal matrix P such that m satisfies

$$\int_t^{t+T} mm' \mathrm{d}\tau = P \begin{bmatrix} \varepsilon(t) I_r & 0 \\ 0 & 0 \end{bmatrix} P', \quad \forall t > 0.$$

Then, we obtain $\mathrm{e} = \tilde{\theta}' m = \tilde{\theta}' P P' m =: \tilde{\psi}' \bar{m}$. Also, $\int_t^{t+T} \bar{m}\bar{m}' \mathrm{d}\tau = \begin{bmatrix} \varepsilon I_r & 0 \\ 0 & 0 \end{bmatrix}$. With obvious definitions, we have that $\bar{m}_1 \in PE(r)$, $\bar{m}_2 = 0$ and $\mathrm{e} = \tilde{\psi}_1' m_1$. The application of Theorem 1(v) for $\mathrm{e} = \tilde{\psi}_1' m_1$ yields the convergence.

6.3 Proofs for the estimator with memory

6.3.1 Proof of Theorem 2

(i). Define $u_n = \gamma_1 u$, $\Psi_n = \gamma_2 \Psi$, and $\Upsilon_n = \gamma_2 \Upsilon$. Using (3.6), we have

$$\begin{aligned} D_{0^+}^\alpha \hat{\theta} &= -\Gamma\big(\mathrm{e} u_n + \kappa(\Psi_n \hat{\theta} - \Psi_n \theta)\big), \\ D_{0^+}^\alpha \tilde{\theta} = D_{0^+}^\alpha \hat{\theta} &= -\Gamma\big(u_n u' \tilde{\theta} - \kappa \Psi_n \tilde{\theta}\big), \end{aligned}$$

where we use that θ is constant so that $D^\alpha \theta = 0$ when using Caputo derivative. Then

$$\begin{aligned} \tilde{\theta}' \Gamma^{-1} D_{0^+}^\alpha \tilde{\theta} &= -\tilde{\theta}' [u_n u' \tilde{\theta} + \kappa \Psi_n \tilde{\theta}], \\ &\leq -\gamma_1 \mathrm{e}^2 - \kappa \lambda_{min}(\Psi_n) \tilde{\theta}' \tilde{\theta}, \end{aligned}$$

where we have used that $\Psi_n \geq 0$ since $uu' \geq 0$. Since y, u were assumed continuous, we obtain $D^\alpha \tilde{\theta}' \Gamma^{-1} \tilde{\theta} \leq 2 \tilde{\theta}' \Gamma^{-1} D^\alpha \tilde{\theta}$. Calling $V(t) = \tilde{\theta}'(t) \Gamma^{-1} \tilde{\theta}(t)$, we get

$$D_{0^+}^\alpha V \leq -2\gamma_1 \mathrm{e}^2 - \frac{2\kappa \lambda_{min}(\Psi_n)}{\lambda_{max}(\Gamma^{-1})} V. \tag{6.2}$$

In particular, $D_{0^+}^\alpha V \leq -2\gamma_1 \mathrm{e}^2$. By applying $I_{0^+}^\alpha$ in both sides, we get $V + 2 I^\alpha \gamma_1 \mathrm{e}^2 \leq V(0)$. Then $I^\alpha \gamma_1 \mathrm{e}^2 < \infty$ and $V(t) \leq V(0)$ for any $t > 0$. Since $V(t) = \|\tilde{\theta}(t)\|^2_{\Gamma^{-1}}$, we obtain $\|\tilde{\theta}(t)\|_{\Gamma^{-1}} \leq \|\tilde{\theta}(0)\|_{\Gamma^{-1}}$ and, in particular, $\hat{\theta} \in \mathcal{L}_\infty$.

(ii) Similar to the proof of Theorem 1(vi).

(iii) The first part follows from [GD18a, Theorem 1]. For the second part, consider the particular case $u = -\mathrm{e}$, $\kappa = 0$ and $n = 1$. Then, $\mathrm{e} = v/(1+\tilde{\theta})$ and the rest is similar to the proof of Theorem 1(vii)

(iv) Since u has finite excitation, there exist $\varepsilon, \gamma > 0$ s.t. $\int_0^\varepsilon u(s)u'(s)\mathrm{d}s \geq \gamma I_n$. Using that $k_\beta(t,\cdot)$ is increasing and $u(s)u'(s) \geq 0$, we have for $k_{min} := k(\varepsilon, 0)$

$$\int_0^\varepsilon k_\beta(\varepsilon, s)u(s)u'(s)\mathrm{d}s \geq k_{min}\int_0^\varepsilon u(s)u'(s)\mathrm{d}s \geq k_{min}\gamma I_n =: \gamma_0 I_n. \tag{6.3}$$

Then, the following inequality holds,

$$\begin{aligned}\Psi_n(\varepsilon) &\geq \frac{1}{1+max_{\tau\in[0,\varepsilon]}\bar{\Psi}(\tau)}\int_0^\varepsilon k(t_a,t)uu'\mathrm{d}s,\\ &\geq \frac{1}{1+max_{\tau\in[0,\varepsilon]}\bar{\Psi}(\tau)}\gamma_0 I_n,\\ &=: \bar{\gamma}I_n,\end{aligned}$$

Using the definition of t_a, it follows that $\Psi_n(t) \geq \bar{\gamma}I_n$ for any $t \geq \varepsilon$. Replacing in (6.2), we have

$$\begin{aligned}D^\alpha_{0+}V(t) &\leq -\frac{2\kappa\bar{\gamma}}{\lambda_{max}(\Gamma^{-1})}V + f_0(t),\\ &=: -cV + f_0(t),\end{aligned}$$

where $c = \frac{2\kappa\bar{\gamma}}{\lambda_{max}(\Gamma^{-1})} > 0$ and $f_0(t) = -2\gamma_1\mathrm{e}^2(t) + \frac{2\kappa\bar{\gamma}}{\lambda_{max}(\Gamma^{-1})}V(t) - \frac{2\kappa\lambda_{min}(\Psi_n)}{\lambda_{max}(\Gamma^{-1})}V(t)$ for $t < \varepsilon$ and $f(t) = 0$ for $t \geq \varepsilon$. Using that $V \geq 0$ and a comparison argument, to know the asymptotic behaviour is enough studying the solution to $D^\alpha_{0+}V(t) = -cV + f_0(t)$, which is given by [Die10]

$$V(t) = V(0)E_\alpha(-ct^\alpha) + \int_0^t (t-\tau)^{\alpha-1}E_{\alpha,\alpha}(-c(t-\tau)^\alpha)f_0(t)\mathrm{d}\tau.$$

Since f_0 converges to zero and $t^{\alpha-1}E_{\alpha,\alpha}(-ct^\alpha) \in \mathcal{L}^1$ [BP00], their convolution converges to zero and one can prove that V converges to zero. As we are also interested in its convergence order, we further develop to get

$$\begin{aligned}V(t) &= V(0)E_\alpha(-ct^\alpha) + \int_0^t (t-\tau)^{\alpha-1}E_{\alpha,\alpha}(-c(t-\tau)^\alpha)f_0(t)\mathrm{d}\tau,\\ &\leq V(0)E_\alpha(-ct^\alpha) + C\int_{t-\varepsilon}^t \tau^{\alpha-1}E_{\alpha,\alpha}(-c\tau^\alpha)\mathrm{d}\tau,\end{aligned}$$

where $C = \max_{t\in[0,t_s]}|f_0(t)|$ and a change of the integration variable in the convolution was made. Using [CDT17, Lemma 2](ii), we obtain for t large enough

$$\begin{aligned}V(t) &\leq V(0)E_\alpha(-ct^\alpha) + Cm(\alpha,c)\int_{t-t_s}^t \tau^{-(1+\alpha)}\mathrm{d}\tau,\\ &\leq V(0)E_\alpha(-ct^\alpha) + Cm(\alpha,c)(t-t_s)^{-(1+\alpha)}t_s,\end{aligned}$$

where $m(\alpha, c)$ is independent of time. Thus, the convergence order of V, corresponding to the slower term, is $E_\alpha(-ct^\alpha) \sim t^{-\alpha}$ [CDT17, Lemma 2](i). It is clear that if u is bounded, then $\mathrm{e} = \tilde{\theta}' u \to 0$ at the same rate. Since $E_\alpha(-ct^\alpha)$ is monotonically decreasing [P08], the increasing of c through Γ, κ, yields a faster decays.

(v) When an additive disturbance ν appears in the measurement of y, the following modification to the previous developments is verified

$$\begin{aligned} D^\alpha V &\leq -cV + |f_0| - 2\tilde{\theta}' u_n \nu + 2\|\tilde{\theta}\|\|\nu\|_\infty \\ &\leq -c\lambda_{min}(\Gamma^{-1})\|\tilde{\theta}\|^2 + 2\|\tilde{\theta}\|(\|u_n\|\|\nu\| + \|\nu\|_\infty) + |f_0|. \end{aligned}$$

where the term $\|\nu\|_\infty = \sup_{t\geq 0} |\nu(t)|$ appears thanks to the normalizing term γ_2 as $\|\gamma_2 \int ku\nu\| \leq \|\nu\|_\infty \gamma_2 \int k\|u\| \leq \|\nu\|_\infty$. We now prove that $\tilde{\theta}$ is bounded. Using that for any $x, y \in \mathbb{R}$ and any $\varepsilon > 0$ it holds that $2xy \leq \varepsilon x^2 + \varepsilon^{-1} y^2$, we have

$$D^\alpha V \leq -c\lambda_{min}(\Gamma^{-1})\|\tilde{\theta}\|^2 + \varepsilon\|\tilde{\theta}\|^2 + \varepsilon^{-1}(\|u_n\|\|\nu\| + \|\nu\|_\infty)^2 + |f_0|.$$

Since u_n, ν, f_0 are bounded, there exists a constant number C such that

$$D^\alpha V \leq -\lambda_{max}(\Gamma^{-1})(c\lambda_{min}(\Gamma^{-1}) - \varepsilon)V + C.$$

Taking $\varepsilon < c\lambda_{min}(\Gamma^{-1})$, it follows that $t^{\alpha-1}E_{\alpha,\alpha}(-(c-\varepsilon)t^\alpha) \in \mathcal{L}^1$. Using this, a comparison argument and the solution as above, we conclude that $V, \tilde{\theta}$ are bounded since the convolution of $t^{\alpha-1}E_{\alpha,\alpha}(-(c-\varepsilon)t^\alpha)$ is now with a constant. Moreover, if ν converges to zero, then we can sharp the bound C with a decaying to zero function $C(t)$, and hence, the convolution term will converges to zero yielding the convergence of $\tilde{\theta}$ to zero.

6.3.2 Multi-order

First, notice that from (3.1) we have for any $\mathrm{i} = 1, \ldots, n$ and $t > 0$

$$\begin{aligned} y(t)u_\mathrm{i}(t) &= u_\mathrm{i}(t)u'(t)\theta, \\ \int_0^t k_\mathrm{i}(t,\tau)y(\tau)u_\mathrm{i}(\tau)\mathrm{d}\tau &= \int_0^t k_\mathrm{i}(t,\tau)u_\mathrm{i}(\tau)u'(\tau)\theta\mathrm{d}\tau. \end{aligned}$$

Redefining $\Upsilon(t)$ as the vector function of components $\int_0^t k_\mathrm{i}(t,\tau)yu_\mathrm{i}(\tau)\mathrm{d}\tau$ and $\Psi(t)$ as the matrix function of rows $\int_0^t k_\mathrm{i}(t,\tau)u_\mathrm{i}(\tau)u'(\tau)\mathrm{d}\tau$ for $\mathrm{i} = 1, \ldots, n$, the same relation $\Upsilon(t) = \Psi(t)\theta$ is obtained. When particularizing to the fractional kernel, it follows that the implementation (3.7) has the same form with now β a vector and $D^\beta x$ meaning the vector of components $D^{\beta_\mathrm{i}} x_\mathrm{i}$. Hence, the generalization of Theorem 2 is straightforward.

Now consider the case $\alpha \in \mathbb{R}^n$ with components in $(0, 1]$. For illustration's sake, we consider $n = 2$. The first step of the proof of Theorem 2 can be repeated to obtain

$$D^\alpha \tilde{\theta} = -uu'\tilde{\theta} - \kappa\Psi\tilde{\theta}, \tag{6.4}$$

where $\Gamma = I_n$ for simplicity. By assuming that u has FE, we can write $\Psi(t) = \Psi_0 + f(t)$ where Ψ_0 is a diagonal matrix with positive entries γ_i, which are proportional to κ, and f a matrix that vanishes after some finite time T when choosing $t_a(t) = \varepsilon$ for any $t > 0$. Then

$$\begin{bmatrix} D^{\alpha_1}\tilde{\theta}_1 \\ D^{\alpha_2}\tilde{\theta}_2 \end{bmatrix} = -\begin{bmatrix} u_1^2 + \gamma_1 & u_1 u_2 \\ u_1 u_2 & u_2^2 + \gamma_2 \end{bmatrix} \begin{bmatrix} \tilde{\theta}_1 \\ \tilde{\theta}_2 \end{bmatrix} - \kappa f(t) \begin{bmatrix} \tilde{\theta}_1 \\ \tilde{\theta}_2 \end{bmatrix}.$$

Let $V_1 = \frac{1}{2}\tilde{\theta}_1^2$, $V_2 = \frac{1}{2}\tilde{\theta}_2^2$ and $F(t) := \|f(t)\|\|(\tilde{\theta}_1(t), \tilde{\theta}_2(t))|^2$. Using normalization if necessary, there exists a constant $C > 0$ such that $|u_1 u_2 \tilde{\theta}_1 \tilde{\theta}_2| \leq C(V_1 + V_2)$, and hence,

$$\begin{bmatrix} D^{\alpha_1} V_1 \\ D^{\alpha_2} V_2 \end{bmatrix} \preceq \begin{bmatrix} -u_1^2 - \gamma_1 + C & C \\ C & -u_2^2 - \gamma_2 + C \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} + F(t) =: \Lambda \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} + F(t),$$

where $\preceq$ means the relation $\leq$ component-wise. Clearly, Λ is Metzler (i.e. nonegative off-diagonal entries), and using κ large enough (which yields γ_i large enough), it is also Hurwitz. Since $F(t)$ vanishes for $t > T$, it follows from [GAD20a, Theorem 2] that V_1, V_2 converge to zero, and hence, $\tilde{\theta}$ converges to zero as $t \to \infty$. The same arguments can be carried out for the $n > 2$ case.

6.4 Proofs for the Finite-Time Estimator

Proof of Theorem 3

i. According to the continuity assumption on m and λ, the solution to (3.12) exists, is unique and continuous (see e.g. [GAD19b]). For the same reason, w exists and is continuous. Moreover, it is given by $w(t) = 1 - (I^\alpha \lambda \Delta^2)(t)$, as $I^\alpha_{0+} D^\alpha_{0+} w(t) = w(t) - w(0)$ (see e.g. [Die10]). Therefore, $\hat{\theta}^{FT}$ is piecewise continuous. Since condition (3.13) holds, it follows that $\lambda\Delta^2$ is not the zero function; hence, $w(t) < 1$ for any $t > \delta$ because $\lambda \geq 0$ and even if $\lambda\Delta^2 = 0$ after $t > \delta$, the solution would be strictly increasing (because and the fractional integral would be strictly decreasing for $t > \delta$) but does not attain 1 in finite time because if $w(t_1) = 1$ then w cannot be strictly increasing. Therefore, $\hat{\theta}^{FT}(t) \in \mathbb{R}$ for any $t \geq 0$.

ii. Using equations (3.11), (3.12) and the fact that $D^\alpha_{0+}\theta = 0$, we have

$$D^\alpha_{0+}\tilde{\theta}^F = -\gamma\Delta^2\tilde{\theta}. \tag{6.5}$$

That is, w and $\tilde{\theta}$ share the same equation. In particular, due to the linearity of the operator $D^\alpha_{t_0^+}$, we have $\tilde{\theta}(t) = w(t)\tilde{\theta}(0)$, i.e.

$$\theta = \frac{1}{1 - w(t)}\big(\hat{\theta}(t) - w(t)\hat{\theta}(0)\big). \tag{6.6}$$

Suppose that $w(t) > 1 - \mu$ for any $t > 0$. Then

$$\begin{aligned} w(\delta) &= 1 - \big(I^\alpha_{0+}\lambda\Delta^2 w\big)(\delta) \\ &< 1 - (1-\mu)\big(I^\alpha_{0+}\lambda\Delta^2\big)(\delta) \\ &\leq 1 - \mu, \end{aligned}$$

where in the last inequality we use (3.13) and in the previous one the fact that all the integrands are nonnegative. This contradicts the assumption that $w(t) > 1-\mu$ for any $t > 0$. Therefore, there exists $t_f < \delta$ such that $w(t_f) = 1-\mu$ where we use that the integrands are continuous (bounded in $[0, \delta]$, in particular), and hence its convolution is continuous. This means that $\hat{\theta}^{FT}$ is equal to the right-hand of (3.14) for any $t > t_f$. Comparing (3.14) with (6.6), we conclude that $\hat{\theta}^{FT}(t) = \theta$ for any $t > t_f$.

iii. Since $\tilde{\theta}^F$ satisfies $D^{\alpha}_{0+}\tilde{\theta}^F = -\lambda\Delta^2\tilde{\theta} \leq 0$, it follows that $\tilde{\theta}$ does not change its sign (see e.g. [Die10, Chapter 7]). Hence, $\lambda_1\Delta^2\tilde{\theta}_{\mathrm{i}} \leq \lambda_2\Delta^2\tilde{\theta}$ when $\tilde{\theta}(0) \geq 0$. A comparison argument (see e.g. [Die10, Chapter 6]) and reversing the inequalities when $\tilde{\theta}(0) \leq 0$, yield the result.

iv. Consider the transient period, i.e. $t \in [0, t_f]$, where $w(t) > 1-\mu$ and hence $\hat{\theta}^{FT}(t) = \hat{\theta}$. The fact $D^{\alpha}_{t_0^+}w(t) \leq 0$ implies, by applying I^{α}_{0+} in both sides, that $w \leq 1$. Since $\tilde{\theta}(t) = w(t)\tilde{\theta}(0)$, it follows that $|\hat{\theta}(t) - \theta| \leq |\hat{\theta}(0) - \theta|$ and hence, $|\hat{\theta}^{FT}(t) - \theta| \leq |\hat{\theta}^{FT}(0) - \theta|$ for all $t \geq 0$. The rest is similar to the proof of Theorem 1(vi).

Proof of Proposition 1

(Sufficiency) Notice that by sending $\mu \to 0^+$ –recall that μ is a designer chosen parameter– condition (3.13) can be restated as $\left(I^{\alpha}_{0+}\lambda\Delta^2\right)(t = \delta) > 0$. When u is continuous, this is equivalent to require that Δ is not the zero function provided that λ is chosen not to be the zero function. (Necessity) If $\Delta \equiv 0$, then $Y \equiv 0$ according to (3.11) and no information of θ can be extracted.

Let $I = (a, b)$ be the interval where the components of m are linearly independent and t_0 a point in the interior of I. Since the filters H_{i} are arbitrarily-chosen operators, we can choose the shift operators $H_{\mathrm{i}}(x)(t) = x(t - \delta_{\mathrm{i}})$ for some $\delta_{\mathrm{i}} > 0$ such that $t_0 - \delta_{\mathrm{i}} \in I$. It is easy to see that these shift operators are linear, map continuous functions to continuous functions, and that the matrix M is given by

$$M(t) = \begin{pmatrix} m_1(t_1) & \cdots & m_q(t_1) \\ \vdots & \ddots & \vdots \\ m_1(t_q) & \cdots & m_q(t_q) \end{pmatrix},$$

where $t_{\mathrm{i}} = t - \delta_{\mathrm{i}}$ for $\mathrm{i} = 1, \ldots, q$. According to the choice of δ_{i}, we have $t_1, \ldots, t_q \in I$ when M is computed at $t = t_0$. Since the components of m are linearly independent in I, there exists a choice of δ_{i} for $\mathrm{i} = 1, \ldots, q$ such that $M(t_0)$ is invertible. To see this, note that the span of the range of the function $g : t \in (a, t_0) \to (m_1(t), \ldots, m_q(t))$ is $\mathbb{R}^q$ because its orthogonal is $\{0\}$ as the functions m_{i} are linearly independents in (a, t_0). Hence, there exists $t_1, \ldots, t_q$ such that $g(t_{\mathrm{i}})$ are linearly independents vectors for $\mathrm{i} = 1, \ldots, q$, which entails the existence of δ_{i} for $\mathrm{i} = 1, \ldots, q$. Therefore, $\Delta(t_0) \neq 0$ and the claim follows from the continuity assumption.

Proof of Proposition 2

Since Δ is continuous and not the zero function, it satisfies (3.13) for some μ whenever λ is not chosen zero. Note then that the increasing of λ reduces the time δ when μ is fixed in expression (3.13) due to the continuous assumption on Δ. Since $t_f < \delta$, the claim follows.

Proof of Proposition 3

a) It follows from the use of the gradient estimator, Theorem 1(vi) and the algebraic relation to obtain $\hat{\theta}^{FT}$.

b) We first prove by contradiction that $\hat{\theta}(t) \in [a,b]$ for any $t > 0$. Since $\hat{\theta}(0) \in (a,b)$ and $\mathcal{P}_{[a,b]}$ is the identity when $\hat{\theta} \in (a,b)$, if there exists t_1 such that $\hat{\theta}(t_1) < a$, then there must exist $t' < t_1$ such that $\hat{\theta}(t') = a$ and $\hat{\theta}(t) > a$ for any $t < t'$. From expression (1.2) and the fact that $0 < \alpha \leq 1$, the assumption $\dot{\hat{\theta}}(t') < 0$, needed to prove the existence of t_1, is contradictory because, on the one hand, $D^{\alpha}_{0+}\hat{\theta}(t') = \varepsilon$ from the definition of $\mathcal{P}_{[a,b]}$ and, on the other, there must exist a small enough $\varepsilon > 0$ such that $\dot{\hat{\theta}}(t) \leq 0$ for $t \in (t' - \varepsilon, t')$ from the fact that $\hat{\theta}(t) > a$ for any $t < t'$, which would yield $D^{\alpha}_{0+}\hat{\theta}(t') < 0$ if $\dot{\hat{\theta}}(t') < 0$. Then, the case $\hat{\theta}(t_1) < a$ for $t_1 > 0$ cannot occur. A similar reasoning shows that the case $\hat{\theta}(t_1) > b$ for $t_1 > 0$ cannot occur either. Therefore, $\hat{\theta}(t) \in [a,b]$ for any $t > 0$ i.e. $\hat{\theta}$ is bounded.

Recall that w is bounded according to the proof of Theorem 3. Then, $\hat{\theta}^{RFT}$ is bounded since it is obtained from algebraic operations of bounded functions and (3.14) is used only when $w < 1 - \mu$ so that its denominator is bounded away from zero.

Let $\nu \equiv 0$. Since $\mathcal{P}_{[a,b]}$ is the identity when $\hat{\theta} \in (a,b)$ and $\hat{\theta}(0) \in (a,b)$, we can use Theorem 3(ii) to show that $|\hat{\theta} - \theta|$ cannot increase i.e. $\hat{\theta}$ remains in (a,b). Therefore, the same arguments for the fractionalized gradient (3.12) show that $\hat{\theta}^{RFT}$ exhibits finite-time convergence when (3.13) is satisfied. Moreover, all the above propositions hold for (3.17).

c) The direct relation between the fast decaying of w and the impoverishing to track parametric variations of DREM based estimators has been stressed before in [OAPAB19]. We have that w is nonincreasing when $\alpha = 1$; but, when $\alpha < 1$, w can increase and even when decreasing, the convergence is strictly slower than exponential as α decreases. The case for $\alpha = 1$ follows from the facts that $\dot{w} = -\lambda\Delta^2 w \leq 0$, $w(0) = 1$ and that w does not change sign. The fact that $D^{\alpha}w = -\lambda\Delta^2 w \leq 0$ does not imply monotony was established in [GD16b, Proposition 3]. In particular, if $D^{\alpha}_{t_0^+}w(t) = 0$ for $t > t_f$ then w begins to increase and converges to 1. The claim on the convergence slower than exponential can be found in [GD19a].

6.5 Proofs for the Switched Adaptation

6.5.1 Implementability

We want to simulate

$$ {}_{t_0}D^{\alpha}x = f(t,x) \tag{6.7} $$

for $x(t_0) = x_0$. To do this, we use Ninteger toolbox for Matlab, which allows to using the NID block to simulate the system

$$ {}_{0}D^{\alpha}x = f(t,x) \tag{6.8} $$

for $x(0) = x_0$. We claim that (6.7) can be simulated with

$$ {}_{0}I^{\alpha}f(t,x) \tag{6.9} $$

with $f(t,x) = 0$ for $t < t_0$ and adding $x(t_0) = x_0$ for $t > t_0$. Indeed, in that case (6.9) becomes

$$ \int_{t_0}^{t} (t-\tau)^{\alpha-1} f(\tau,x)\mathrm{d}\tau \tag{6.10} $$

and hence,

$$ \begin{aligned} x(t) &= x(t_0) + {}_{0}I^{\alpha}f(t,x) &(6.11)\\ &= x(t_0) + \int_{t_0}^{t} (t-\tau)^{\alpha-1} f(\tau,x)\mathrm{d}\tau &(6.12) \end{aligned} $$

which is the solution to (6.7).

6.5.2 Proof of Theorem 4

The fact that m satisfies the excitation condition in Theorem 2 allows us to choose c as the exponential and c' as the Mittag-Leffler functions with specific rates of decay since, according to Theorem 2, e converges at those orders, respectively, under no disturbances. The same goes for the case when it satisfies Theorem 3 and one uses (3.14). The boundedness of $\hat{\theta}$ follows from noting that $V = \tilde{\theta}'\tilde{\theta}$ is a common Lyapunov function for any derivation order in $(0,1]$ and using [GAD20c, Corollary 1] (moreover, $\tilde{\theta} = 0$ is stable). The rest of the proof follows from the construction of the rising and lowering sequences.

6.6 Proof of Proposition 4

In the interior of $\mathcal{D}$, we have

$$ \begin{aligned} \dot{J} &= \frac{\partial J'}{\partial \alpha}\dot{\alpha} + \frac{\partial J'}{\partial \lambda}\dot{\lambda}, \\ &= -\frac{\partial J'}{\partial \alpha}\frac{\partial J}{\partial \alpha} - \frac{\partial J'}{\partial \lambda}\frac{\partial J}{\partial \lambda} \leq 0. \end{aligned} $$

The effect of the projection on the boundary is such that $\dot{J} \leq 0$ since, in the worst case where J decreases outside the domain and increases inside, the parameters can be forced to stay on the boundary, by adding a conditional whereby $\dot{J} = 0$. Finally, by construction, λ, α remain bounded.

6.7 Proofs of Section 4

6.7.1 Proof of Proposition 5

Proof. We made the proof for the fractionalized gradient. Notice that

$$D^\alpha \tilde{\theta} = f_1(m,t)\tilde{\theta} + P^{-1}\lambda e m.$$

Defining $V = \tilde{\theta}' P \tilde{\theta}$, we get

$$\begin{aligned} D^\alpha V &\leq \tilde{\theta}'(P f_1(m,t) + f_1' P)\tilde{\theta}\lambda \mathrm{e}^2, \\ &\leq -\lambda \mathrm{e}^2. \end{aligned}$$

The rest is similar to the proof of Theorem 1.

6.7.2 Proof of Proposition 6

Proof. Since Ω is a hypercube, $\hat{\theta} \notin \Omega \subset \mathbb{R}^n$ occurs when some component (say) $\hat{\theta}_{\mathrm{i}_0}$ of $\hat{\theta}$ crosses the boundary of the hypercube. Without loss of generality, we assume $\hat{\theta}_{\mathrm{i}_0}(T) = \theta_+$ where $\theta_- \leq \hat{\theta}_{\mathrm{i}_0} \leq \theta_+$ is the hypercube condition for i_0. Suppose that $\hat{\theta} \in \Omega^c$ for $t > T$. We have

$$\begin{aligned} \hat{\theta}_{\mathrm{i}_0}(t) &= \hat{\theta}_{\mathrm{i}_0}(0) + \left(I^\alpha_{0+} D^\alpha \hat{\theta}_{\mathrm{i}_0}\right)(t) \\ &= \hat{\theta}_{\mathrm{i}_0}(0) + \int_0^t k_\alpha(t-\tau) D^\alpha \hat{\theta}_{\mathrm{i}_0}(\tau)\mathrm{d}\tau \\ &= \hat{\theta}_{\mathrm{i}_0}(0) + \int_0^T k_\alpha(t-\tau) D^\alpha \hat{\theta}_{\mathrm{i}_0}(\tau)\mathrm{d}\tau, \end{aligned}$$

where the last inequality is due to the null value of the indicator function in Ω^c. Since $\hat{\theta}_{\mathrm{i}_0}(0) < \theta_+$ by assumption, we have that $D^\alpha \hat{\theta}_{\mathrm{i}_0}(t) > 0$ for any $t \in [T-\varepsilon, T)$ with $\varepsilon > 0$ small enough due to the continuity for $t \leq T$. However, these values are underweight by the kernel function as t grows since $\frac{\mathrm{d}}{\mathrm{d}t} k_\alpha(t-\tau) = -\frac{\alpha-1}{\Gamma(\alpha)}(t-\tau)^{\alpha-2}$. This means that $\hat{\theta}_{\mathrm{i}_0}(t) < \hat{\theta}_{\mathrm{i}_0}(T) < \theta_+$, which contradicts the assumption $\hat{\theta} \in \Omega^c$ for $t > T$.

Using that $\|\tilde{\theta}\| \leq \|\tilde{\theta}(0)\|$ and $\theta, \hat{\theta} \in \Omega$, it follows that $\hat{\theta}$ remains in Ω when no disturbance occurs. □

6.7.3 Proof of Proposition 7

Proof. Since the last column of Ω has FE, the last component of ε in (4.6) has the regression form that allows to estimate θ with $\hat{\theta}$ satisfying (3.6). By Theorem 2, the identification aim is achieved i.e. $\hat{\theta} \to \theta$ as $t \to \infty$. In particular, $\hat{\theta}$ is bounded. Using the ISS property [KK95, Lemma 3.2], z is also bounded. This implies that $v_{\rm i}, x, W$ are also bounded for $\rm i = 1, \ldots, n$. Using the ISS property and the fact that $\tilde{\theta}$ converges to zero, we conclude that z converges to zero. In particular, $z_1 = y - y_m$ converges to zero.

When $\nu \neq 0$, bounded additive terms appears in (4.5) and (4.6). Using the ISS property and Theorem 2, we conclude the boundedness of z, x, y when ν is bounded, and the convergence $y \to y_m$ when ν converges to zero. The last claim follows from the additive form in which $\tilde{\theta}$ appears in (4.5). □

6.7.4 Proof of Proposition 8

Proof. Let ω be the last column of Ω. We show first that if the components of ω are linearly independent continuous functions in some interval I, then ω has FE. Due to the ISS property and the boundedness of the estimator, we can restrict the verification of the positive definiteness condition (3.9) to an arbitrarily large compact space $C \subset \mathbb{R}^p$. Since the components of ω are linearly independent functions, $c'\omega$ is not the zero function on I for any $c \in C - \{0\}$. Using the continuity, there exists a small enough number $\gamma(c)$ such that $\int_I (c'\omega)^2 {\rm d}t \geq \gamma(c)c'c > 0$. From the compactness of C, there exists a constant $\gamma > 0$ such that $\int_I (c'\omega)^2 {\rm d}t \geq \gamma c'c > 0$ for any $c \in C - \{0\}$. Then, $\int_I \omega\omega' {\rm d}t \geq \gamma I_p$ on any compact subset, which is enough to establish Theorem 2. Since ω is a linear filtering of φ, the independence of ω follows from that of φ (we can always take null initial condition for the filter, so that $\omega = h * \varphi$).

Thus, we must analyse the linear independence of φ on some interval. If φ contains p different frequencies, i.e. $c'\varphi(t)$ can be written as $\sum_{\rm i=1}^{p} {\rm d_i}(c) sin(w_{\rm i}t + \phi_{\rm i})$ for any c, then this independence is verified. Since φ is a function of x, a verifiable condition on φ is that if x contains r different frequencies then φ contains p different frequencies. For instance, in the linear case, $\phi(x)' = (x_1, \ldots, x_n)$, it is enough to take $r = n$. In general, nonlinearities enlarge the number of frequencies so that $r < p$ (e.g. $\varphi = x^3$).

Now, we show the a priori verifiability. Notice that r appears additively in the equation of x_n, and the other components of x are integrals of x_n. Therefore, the introduction of an external signal $r_{\rm e}$ with r different frequencies makes x a vector with r different frequencies. □

6.7.5 Proof of Proposition 9

Proof. Let $\rm e = \hat{z} - z$. Then

$$\dot{\rm e} = A_o {\rm e} + {\rm d}\beta'(y, u)\tilde{\theta} + \Phi(y, u)w.$$

Suppose that $\nu \equiv 0$ (i.e. $w \equiv 0$). According to Theorem 3, $\tilde{\theta}(t) = 0$ for any $t > t_f$. Since β is continuous, $\mathrm{d}\beta'(y(t), u(t))\tilde{\theta}(t) = 0$ for any $t > t_f$. Since A_o is Hurwitz, $\mathrm{e} \to 0$ as $t \to \infty$ at exponential order. Therefore, $\hat{\xi} := \hat{z} - M\hat{\theta} \to z - M\theta = \xi = T(x)$. Noting that T is a diffeomorphism and applying T^{-1}, we have $\hat{x}(t) \to x(t)$ as $t \to \infty$. When $\nu \neq 0$, we use Proposition 3(ii), the boundedness of y, u, the smoothness of β, Φ and the BIBO stability stemming from the Hurwitz property, to conclude the boundedness or convergence of $\hat{x} - x$. □

Chapter 7

Conclusions

In this work, we have provided affirmative answers to the following nested questions: Can fractional operators be used to analytically solve classic engineering problems in which the uncertainties in the model cannot be ignored? Are there objective advantages when solving such a problem with fractional operators? Are there automatic methods, which do not depend on additional prior knowledge, to choose which fractional operators are suited for a specific problem ?. In answering these questions, we have taken several choices that will now discuss.

a) The objective advantages were obtained by resorting to the singularity of the kernel function in the defining expression of the fractional derivative (yielding the transient improvement) and to the forgetting factor effect of the fractional integration (yielding the robustness improvement). In particular cases, the long memory property of fractional systems could provide further advantages: in periodic settings such as the tracking or rejecting of periodic exogenous signals; in the repetitive task problems treated in Adaptive Learning; in Network Theory to enhance the complexity by the introduction of long-memory elements, which could improve the approximations of functional relations or reduce the order of linear parametrizations that have poorer performance than nonlinear ones as the order grows.

b) The objective disadvantages (slow speed of convergence, the loosing of the gradient method entailing convergence hindering) demanded alternative designs to cope with them. These designs were based on well-established modifications of the gradient. Although an absolutely new method could present the same problems (the slow speed of convergence and the convergence issues seem to be unavoidable, as argued in this document), other advantages may appear that remain unnoticed.

c) The estimation proposed in this work is part of an adaptive control strategy. Hence, there are specific features that one looks for in the estimator, which preclude the use of statistic-like estimators: On-line (adaptive) estimation by which current information is actively used in the place of a priori information which is unreliable or inaccurate when controlling physical systems. Stability, robustness and transient behaviour rather than statistical properties become paramount in control specifications and the estimator, being part of the control system, requires them.

d) There are alternative ways to answer those questions. On the one hand, many adaptive schemes include a filtering step (e.g. see [CH08]) whose filters could be chosen of a fractional type without affecting the stability. Also, the desired performance could be defined as the output of a fractional system or the performance index could be written using fractional integrals. These solutions (explored by the author in the MRAC context [GD18a]) should be not considered suited because they do not address the adaptation problem, which consists of tuning the parameters, and could be applied in non-adaptive settings. On the other hand, the local fractional derivative can be used instead of the Caputo since the resulting adaptive scheme shares the main features: noise rejection, transient improvement, and sub-exponential convergence. We have excluded this result because such a scheme must be implemented with a standard gradient adaptation with a time-varying adaptive gain having a polynomial singularity at $t = 0$, which makes this scheme of scarce interest from a practical viewpoint.

e) We focus on engineering problems formulated in integer-order derivatives to avoid issues of modelling when using fractional derivatives and to stress the applicability of the proposed methods. As a matter of fact, our results can be easily extended to include fractional systems, which yields a more general adaptive framework (see the linear case in our works [GD18a, DGAC18b]). However, a realistic fractional application requires additional but standard steps since the empirically soundest models occur in diffusion processes where fractional (stochastic) partial equations, both in spacial and time, are employed (e.g., see [DPT20]).

f) Uncertainties can be alternatively handled with robust designs relying on worst-case analysis. In general, robust schemes should show conservative performance in comparison to adaptive ones when dealing with structured uncertainty and/or performance, while adaptive schemes should show more conservative responses to unstructured uncertainty. Schemes combining robust structure and adaptive fine-tuning were used in the control application of this work, but the fractionalization was only used in the adaptive part. A fractionalized structure for the observer was proposed in our work [DGAC18b] which exploits the filtering property of the fractional operator; a fractionalized reference model in [GD18a] exploits the enlargement of dynamics references.

g) We have relied on the gradient descent to propose the fractionalized adaptive scheme. An alternative optimization technique is Least Square whose main feature is the homogenization of the rate of convergence with respect to each component. Preliminary results, which involve the development of stability tools to deal with time-varying Lyapunov functions, show that no decisive additional advantage is obtained with its fractionalization. Moreover, the homogenization is already achieved with the DREM scheme in Theorem 3.

Chapter 8

Bibliography

[Abd15] Abdeljawad, T. (2015). On conformable fractional calculus. Journal of Computational and Applied Mathematics, 279, 57–69.

[AGD19] N. Aguila-Camacho, J. A. Gallegos, M.A. Duarte-Mermoud, Analysis of fractional order error models in adaptive systems: Mixed order cases. Fractional Calculus and Applied Analysis 22 (2019), 1113–1132.

[AG19] Aguila-Camacho N., Gallegos, J.A. Switched Fractional Order Model Reference Adaptive Control for first order plants. In Chilecon 2019, Valparaíso, Chile (2019).

[AG20] Aguila-Camacho N., Gallegos, J.A. Switched Fractional Order Model Reference Adaptive Control for unknown linear time invariant systems. Accepted for presentation at IFAC 2020 World Congress, Berlin, Germany (2020), Jul. 1217.

[AKO07] Astolfi, A., Karagiannis, D., Ortega, R. Nonlinear and Adaptive Control with Applications. Communications and Control Engineering. Springer (2007)

[Bay92] D. S. Bayard, "Transient analysis of an adaptive system for optimization of the design parameters," IEEE Transactions on Automatic Control, vol. 37, pp. 842–848, 1992

[BK05] R. Buche, H. J. & Kushner, (2005). Adaptive optimization of least-squares tracking algorithms: with applications to adaptive antenna arrays for randomly time-varying mobile communications systems. IEEE Transactions on Automatic Control, 50(11), 1749–1760. doi:10.1109/tac.2005.858682

[BP00] C Bonnet, JR. Partington, Coprime factorizations and stability of fractional differential systems. *System and Control Letters* 2000; 41: 167–174.

[CDT16] N.D. Cong, T.S. Doan, H.T. Tuan. Asymptotic stability of linear fractional systems with constant coefficients and small time dependent perturbations. arXiv:1601.06538v1 [math.DS], 2016.

[CDT17] N Cong, T Doan, HT Tuan . Perron-type theorem for fractional differential systems.

Electronic Journal of Differential Equations 2017; 2017(142): 1–13.

[CH08] C. Cao and N. Hovakimyan, "L1 Adaptive Output Feedback Controller for Systems of Unknown Dimension," in IEEE Transactions on Automatic Control, vol. 53, no. 3, pp. 815-821, April 2008

[CM09] R. Curtain and K. Morris, Transfer functions of distributed parameter systems: A tutorial, Automatica, vol. 45, no. 5, pp. 1101-1116, 2009.

[Cot90] N. E. Cotter, (1990). The Stone-Weierstrass theorem and its application to neural networks. IEEE Transactions on Neural Networks, 1(4), 290–295. doi:10.1109/72.80265

[DGAC18b] M Duarte-Mermoud, J Gallegos, N Aguila-Camacho, R Castro-Linares, Mixed Order Fractional Observers for Minimal Realizations of Linear Time-Invariant Systems, Algorithms 11 (9), 136, 18 pp, 2018. https://doi.org/10.3390/a11090136. Published first on line 9 September 2018.

[Die10] K. Diethelm, The analysis of fractional differential equations. An application-oriented exposition using differential operators of Caputo type, Lecture Notes in Mathematics 2004, Springer-Verlag, Berlin (2010)

[DPT20] H. Du, P. Perre, I. Turner, Modelling fungal growth with fractional transport models, Communications in Nonlinear Science and Numerical Simulation, Volume 84, May 2020, 105157

[FC14] Feliu-Batlle, V., Castillo-García, F. J. (2014). On the robust control of stable minimum phase plants with large uncertainty in a time constant. A fractional-order control approach. Automatica, 50(1), 218–224. doi:10.1016/j.automatica.2013.10.002

[GAD19b] Gallegos J, Aguila-Camacho N, Duarte-Mermoud MA. Smooth solutions to mixed-order fractional differential systems with applications to stability analysis. J. Integral Equations Appl 2019; 31: 59–84.

[GAD20a] Gallegos J, Aguila-Camacho N, Duarte-Mermoud MA. Vector Lyapunov-like functions for multi-order fractional systems with multiple time-varying delays, Communications in Nonlinear Science and Numerical Simulation, Volume 83, 2020

[GAD20b] Gallegos J, Aguila-Camacho N, Duarte-Mermoud MA. Robust adaptive passivity-based $PI^{\lambda}D$ control. International Journal of Adaptive Control and Signal Processing. Accepted, 2020.

[GAD20c] Gallegos J, Aguila-Camacho N, Duarte-Mermoud MA. Switched systems with changing derivation order: stability and applications, preprint

[GD16a] J.A. Gallegos, M.A. Duarte-Mermoud, Boundedness and Convergence on Fractional Order Systems. *Journal of Computational and Applied Mathematics*, **296**, Issue C (2016), 815-826.

[GD16b] J. A. Gallegos, M. Duarte-Mermoud. On Lyapunov theory for fractional system, *Applied Mathematics and Computation*, **287** (2016), 161-170.

[GD17a] Gallegos, J.A. Duarte-Mermoud, M.A. Convergence of fractional adaptive systems using gradient approach. *ISA Trans.* **2017**, *69*, pp. 31-42, DOI 10.1016/j.isatra.2017.04.021

[GD17b] J.A. Gallegos, M.A. Duarte-Mermoud, Robustness and convergence of fractional systems and their applications to adaptive systems. Fractional Calculus and Applied Analysis 20 (2017), 895-913.

[GD18a] Gallegos J, Duarte-Mermoud M. Mixed order robust adaptive control for general linear time invariant systems. *Journal of the Franklin Institute* 2018; 355(8): 3399–3422.

[GD18b] J. A. Gallegos, M. A. Duarte-Mermoud, Robust mixed order backstepping control of non-linear systems, IET Control Theory &, Applications 12 (9), 1276-1285, 2018.

[GD18c] J. A. Gallegos, M. A. Duarte-Mermoud, A dissipative approach to the stability of multi-order fractional systems, IMA Journal of Mathematical Control and Information, 2018. Published first on line November 01, 2018. DOI: https://doi.org/10.1093/imamc/dny043.

[GD18d] J. A. Gallegos, M. A. Duarte-Mermoud, Attractiveness and stability for Riemann-Liouville fractional systems, Electronic journal of qualitative theory of differential equations, 2018 (73) 1-16, 2018. DOI: 10.14232/ejqtde.2018.1.73. Published First Online August 29, 2018.

[GD19a] J. Gallegos, Manuel Duarte-Mermoud, Converse theorems in Lyapunov's second method and applications for fractional order systems, Turk J Math, 43, (2019), 1626-1639

[Gall15a] J.A. Gallegos et al., On fractional extensions of Barbalat Lemma, Syst. Control Lett., 84 (2015), 7–12

[Gall15b] J. A. Gallegos, Convergencia en esquemas adaptativos usando operadores fraccionarios, Master's tesis, 2015

[Gall20e] Gallegos J, Stability for Grunwald-Letnikov systems, preprint

[GMM20d] Gallegos J, Munoz-Carpintero D, Martinez B, Predictive control of nonlinear systems using nonlocal linearization, preprint

[HC89] Hsu, L. and R. R. Costa, "Variable structure model reference adaptive control using only input and output measurement: Part I," International Journal of Control, vol. 49, no. 2, pp.399–416, 1989.

[KAAY16] G. Kavuran, B. B. Alagoz, A. Ates, C. Yeroglu, Implementation of Model Reference Adaptive Controller with Fractional Order Adjustment Rules for Coaxial Rotor Control Test System, Balkan Journal of Electrical & Computer Engineering, vol. 4, no. 2, 2016.

[Kre77] Kreisselmeier, G. Adaptive observers with exponential rate of convergence. *IEEE Trans. Automat. Control* **1977**, *22*, pp. 2-8, DOI 10.1109/TAC.1977.1101401

[KK95] M. Krstic and P. Kokotovic “Adaptive nonlinear design with controller-identifier separation and swapping," *IEEE Trans. Autom. Control*, vol. 40, no. 3, 426–440, 1995.

[LC12] Y. Luo and Y.Q. Chen, Stabilizing and robust fractional order PI controller synthesis for first order plus time delay systems, Automatica, vol. 48, no. 9, pp. 2159-2167, Sep. 2012.

[MD91] D. E. Miller, E. J. and Davison, (1991) An adaptive controller which provides an arbitrarily good transient and steady-state response. IEEE Trans. Autom. Control AC-36,68-81.

[MM95] Mbodje, Montsey, Boundary fractional derivative control of the wave equation, IEEE Transactions on Automatic Control (Volume: 40 , Issue: 2 , Feb 1995), 10.1109/9.341815

[MST01] R. Marino, G. L. Santosuosso, and P. Tomei, Robust adaptive observers for nonlinear systems with bounded disturbances, IEEE Trans. on Automatic Control, 46 (2001) 967-972.

[MT93] R. Marino, P. Tomei(1993). Global adaptive output-feedback control of nonlinear systems. I. Linear parameterization. IEEE Trans Autom Control 38(1):17–32.

[NA05] K.S. Narendra and A.M. Annaswamy, Stable Adaptive Systems, Dover Publications, 2005.

[OAPAB19] R. Ortega, S. Aranovskiy, A. A. Pyrkin, A. Astolfi and A. A. Bobtsov, New Results on Parameter Estimation via Dynamic Regressor Extension and Mixing: Continuous and Discrete-time Cases, 2019, 1908.05125, arXiv, eess.SY

[OT89] R. Ortega, & Tang, Y. (1989). Robustness of adaptive controllers—A survey. Automatica, 25(5), 651–677.

[Ous91] Oustaloup A. La commande CRONE: commande robuste d’ordre non entier. Hermes; 1991.

[P08] Podlubny I., Fractional-order systems and $PI^{\lambda}D^{\mu}$ controllers, IEEE Transations on Automatic Control 44 (1) (1999) 208–214. doi:10.1109/9.739144.

[PKD18] A. Parikh, R. Kamalapurkar, Warren E. Dixon, Integral concurrent learning: Adaptive control with parameter convergence using finite excitation, Int J Adapt Control Signal Process. 2018; 1-13.

[SB94] S. Sastry, M. Bodson, Adaptive Control: Stability, Convergence and Robustness, Prentice Hall, 1994.

[SL91] J.-J. E. Slotine and W. Li, Applied Nonlineur Control. Englewood Cliffs, NJ: Prentice Hall. 1991.

[SL03] B. L. Stevens and F. L. Lewis, Aircraft Control and Simulation, 2nd ed. New York: Wiley, 2003

[SM92] Schwartz, C. A., & Mareels, I. M. Y. (1992). Comments on “Adaptive control of linearizable systems" by S.S. Sastry and A. Isidori. IEEE Transactions on Automatic Control, 37(5), 698–701.

[Tao03] G. Tao, Adaptive Control Design and Analysis, Wiley-Interscience, Hoboken, NJ, USA, 2003

[TV18] P. Tomei, CM Verrelli. Advances on adaptive learning control: the case of non-minimum phase linear systems. Syst Control Lett 2018; 115: 55–62

[TT18] H. Tuan, H. Trinh, Stability of fractional-order nonlinear systems by Lyapunov direct method, IET Control Theory A. 12 (17) (2018) 2417 – 2422.

[VIPC02] B. M. Vinagre, I. Petra, I. Podlubny, Y. Q. Chen, Using Fractional Order Adjustment Rules and Fractional Order Reference Models in Model-Reference Adaptive Control, Nonlinear Dynamics, vol. 29, no. 1, pp. 269–279, 2002.

[VMLO15] S. Victor, P. Melchior, J. Lévine, A. Oustaloup, (2015). Flatness for linear fractional systems with application to a thermal system. Automatica, 57, 213–221

[WC02] L Wei, Q Chunjiang, Adaptive control of nonlinearly parameterized systems: the smooth feedback case. *IEEE Transactions on Automatic Control* 2002; 47(8): 1249–1266.