XXXVII SUMMER SCHOOL
ON MATHEMATICAL PHYSICS
RAVELLO, ITALY, SEPTEMBER 17–29, 2012

Gruppo Nazionale per la Fisica Matematica
(GNFM-INdAM)

Fractional Calculus: Methods for Applications

(Lecture notes)

Igor Podlubny

September 30, 2012

Contents

Preface

These lecture notes have been prepared for the participants of the XXXVII-th Summer School on Mathematical Physics, Ravello, Italy, September 17–29, 2012. They are foucsed on methods that simplify applications of the fractional-order modelling in various fields of science and engineering, and correspond to the second part of the two week course "Fractional Calculus: Theory and Applications"; the first part was taught by Professor Francesco Mainardi from the University of Bologna. An exciting addition to this course was the slecture given by Professor Michele Caputo.

These lecture notes are primarily based on the following works:

1 Podlubny, I.: Geometric and physical interpretation of fractional integration and fractional differentiation. *Fractional Calculus and Applied Analysis*, vol. 5, no. 4, 2002, pp. 367–386. [Preprint: arXiv:math/0110241)

- Heymans, N., and Podlubny, I.: Physical interpretation of initial conditions for fractional differential equations with Riemann-Liouville fractional derivatives. *Rheologica Acta*, vol. 45, no. 5, June 2006, pp. 765772. [Preprint: arXiv:math-ph/0512028]
- Podlubny, I.: Matrix approach to discrete fractional calculus, *Fractional Calculus and Applied Analysis*, vol. 3, no. 4, 2000, pp. 359386.
- Podlubny I., Chechkin A.V., Skovranek T., Chen YQ, Vinagre B.: Matrix approach to discrete fractional calculus II: partial fractional differential equations. *Journal of Computational Physics*, vol. 228, no. 8, 1 May 2009, pp. 3137-3153. [Preprint: arXiv:0811.1355]
- Jiao Zh., Chen YQ, Podlubny I.: *Distributed-Order Dynamic Systems: Stability, Simulation, Applications and Perspectives*. Springer, London, 2012, ISBN 9781447128519.
- Podlubny I., Skovranek T., Vinagre B., Petras I., Verbitsky V., Chen Y.Q.: Matrix approach to discrete fractional calculus III: non- equidistant grids, variable step length, and distributed orders. *Philosophical Transactions of the Royal Society A*(accepted).

- Sierociuk D., Podlubny I., Petras I.: Experimental evidence of variable-order behavior of ladders and nested ladders. *IEEE Transactions on Control Systems Technology*, 15 February 2012 (Early Access), doi:10.1109/TCST.2012.2185932. [Preprint: arXiv:1107.2575]
- Petras I., Sierociuk D., Podlubny, I.: Identification of parameters of a half-order system. *IEEE Transactions on Signal Processing*, vol. 60, no. 10, 2012, pp. 5561-5566.
- Petras, I., Podlubny, I.: State space description of national economies: the V4 countries. *Computational Statistics & Data Analysis*, vol 52, no. 2, 2007, pp. 1223-1233. [Preprint: arXiv:math/0609789]
- Petras I., Podlubny I.: Least squares or least circles? *CHANCE*, vol. 23, no. 2, pp. 3842.
- Skovranek T., Podlubny I., Petras I.: Modeling of the national economies in state-space: a fractional calculus approach. *Economic Modelling*, vol. 29, no. 4, 2012, pp. 13221327.

I am thankful to Professor Tommaso Ruggeri and Professor Salvatore Rionero for organizing this school, to Professor Mainardi for suggesting the idea of our course, to Professors Franco Cardin, Luciano Teresi, and Laurent Desvillettes, who delivered the parallel courses at this summer school that gave me additional motivation and ideas, and especially to the wonderful participants from different countries, who actively collaborated during our meetings. I am also obliged to all my co-authors of the works used in these lecture notes: Ivo Petras, YangQuan Chen, Blas Vinagre, Nicole Heymans, Aleksei Chechkin, Tomas Skovranek, Viktor Verbitsky, and Dominik Sierociuk.

Igor Podlubny
Ravello, September 30, 2012.

Lecturers of the second week and participants.

Chapter 1

Geometric and Physical Interpretation of Fractional Integration and Fractional Differentiation

1.1 Introduction

It is generally known that integer-order derivatives and integrals have clear physical and geometric interpretations, which significantly simplify their use for solving applied problems in various fields of science.

However, in case of fractional-order integration and differentiation, which represent a rapidly growing field both in theory and in applications to real-world problems, it is not so. Since the appearance of the idea of differentiation and integration of arbitrary (not necessary integer) order there was not any acceptable geometric and physical interpretation of these operations for more than 300 years. The lack of these interpretations has been acknowledged at the first international conference on the fractional calculus in New Haven (USA) in 1974 by including it in the list of open problems [21]. The question was unanswered, and therefore repeated at the subsequent conferences at the University of Strathclyde (UK) in 1984 [15] and at the Nihon University (Tokyo, Japan) in 1989 [19]. The round-table discussion [13, 10, 14] at the conference on transform methods and special functions in Varna (1996) showed that the problem was still unsolved, and since that time the situation, in fact, still did not change.

Fractional integration and fractional differentiation are generalisations of notions of integer-order integration and differentiation, and include n-th derivatives and n-folded integrals (n denotes an integer number) as particular cases. Because of this, it would

be ideal to have such physical and geometric interpretations of fractional-order operators, which will provide also a link to known classical interpretations of integer-order differentiation and integration.

Since the need for the aforementioned geometric and physical interpretations is generally recognised, several authors attempted to provide them. Probably due mostly to linguistical reasons, much effort have been devoted to trying to relate *fractional* integrals and derivatives, on one side, and *fractal geometry*, on the other [18, 27, 9, 16, and others]. However, it has been clearly shown by R. Rutman [22, 23] that this approach is inconsistent.

Besides those "fractal-oriented" attempts, some considerations regarding interpretation of fractional integration and fractional differentiation were presented in [16]. However, those considerations are, in fact, only a small collection of selected examples of applications of fractional calculus, in which hereditary effects and self-similarity are typical for the objects modelled with the help of fractional calculus. Although each particular problem, to which fractional derivatives or/and fractional integrals have been applied, can be considered as a certain illustration of their meaning, the paper [16] cannot be considered as a definite answer to the posed question.

A different approach to geometric interpretation of fractional integration and fractional differentiation, based on the idea of the contact of α-th order, has been suggested by F. Ben Adda [1, 2]. However, it is difficult to speak about an acceptable *geometric interpretation* if one cannot see any picture there.

Obviously, there is still a lack of geometric and physical interpretation of fractional integration and differentiation, which is comparable with the simple interpretations of their integer-order counterparts.

In this chapter we present a new approach to solution of this challenging old problem.

We start with introducing a simple and really geometric interpretation of several types of fractional-order integration: the left-sided and the right-sided Riemann–Liouville fractional integration, the Riesz potential, and the Feller potential.

Based on this, a physical interpretation of the Riemann–Liouville fractional integration is proposed in terms of inhomogeneous and changing (non-static, dynamic) time scale. Moreover, on this way we give a new physical interpretation of the Stieltjes integral. We also try to persuade the readers that the suggested physical interpretation of fractional integration is in line with the current views on space–time in physics. We also suggest physical interpretation for the Riemann-Liouville fractional differentiation and for the Caputo fractional differentiation. Finally, we show that the suggested approach to geometric interpretation of fractional integration can be used for providing a new geometric and physical interpretation for convolution integrals of the Volterra type.

1.2 Geometric interpretation of fractional integration: Shadows on the walls

In this section we first give a geometric interpretation of left-sided and right-sided Riemann–Liouville fractional integrals, and then consider the Riesz potential.

1.2.1 Left-sided Riemann–Liouville fractional integral

Let us consider the left-sided Riemann–Liouville fractional integral [20, 24] of order α,

$$ {}_0I_t^\alpha f(t) = \frac{1}{\Gamma(\alpha)} \int_0^t f(\tau)(t-\tau)^{\alpha-1} d\tau, \tag{1.1} $$

and write it in the form

$$ {}_0I_t^\alpha f(t) = \int_0^t f(\tau) dg_t(\tau), \tag{1.2} $$

$$ g_t(\tau) = \frac{1}{\Gamma(\alpha+1)} \Big\{ t^\alpha - (t-\tau)^\alpha \Big\}. \tag{1.3} $$

The function $g_t(\tau)$ has an interesting scaling property. Indeed, if we take $t_1 = kt$ and $\tau_1 = k\tau$, then

$$ g_{t_1}(\tau_1) = g_{kt}(k\tau) = k^\alpha g_t(\tau). \tag{1.4} $$

Now let us consider the integral (1.2) for a fixed t. Then it becomes simply a Stieltjes integral, and we can utilize G. L. Bullock's idea [3].

Let us take the axes τ, g, and f. In the plane (τ, g) we plot the function $g_t(\tau)$ for $0 \le \tau \le t$. Along the obtained curve we "build a fence" of the varying height $f(\tau)$, so the top edge of the "fence" is a three-dimensional line $(\tau, g_t(\tau), f(\tau))$, $0 \le \tau \le t$.

This "fence" can be projected onto two surfaces (see Fig. 1.1):

- the area of the projection of this "fence" onto the plane (τ, f) corresponds to the value of the integral
$$ {}_0I_t^1(t) = \int_0^t f(\tau)d\tau; \tag{1.5}$$
- the area of the projection of the same "fence" onto the plane (g, f) corresponds to the value of the integral (1.2), or, what is the same, to the value of the fractional integral (1.1).

In other words, our "fence" throws two shadows on two walls. The first of them, that on the wall (τ, f), is the well-known "area under the curve $f(\tau)$", which is a standard geometric interpretation of the integral (1.5). The "shadow" on the wall (g, f) is a geometric interpretation of the fractional integral (1.1) for a fixed t.

Obviously, for $g_t(\tau) = \tau$ both "shadows" are equal. This shows that classical definite integration is a particular case of the left-sided Riemann–Liouville fractional integration even from the geometric point of view.

What happens when t is changing (namely growing)? As t changes, the "fence" changes simultaneously. Its length and, in a certain sense, its shape changes. For illustration, see Fig. 1.2. If we follow the change of the "shadow" on the wall (g, f), which is changing simultaneously with the "fence" (see Fig.1.3), then we have a dynamical geometric interpretation of the fractional integral (1.1) as a function of t.

1.2.2 Right-sided Riemann–Liouville fractional integral

Let us consider the right-sided Riemann–Liouville fractional integral [20, 24],

$$ {}_tI_0^\alpha f(t) = \frac{1}{\Gamma(\alpha)} \int_t^b f(\tau)(\tau - t)^{\alpha-1} d\tau, \tag{1.6}$$

and write it in the form

$$ {}_tI_0^\alpha f(t) = \int_t^b f(\tau) dh_t(\tau), \tag{1.7}$$

$$ h_t(\tau) = \frac{1}{\Gamma(\alpha + 1)} \Big\{ t^\alpha + (\tau - t)^\alpha \Big\}. \tag{1.8}$$

Then we can provide a geometric interpretation similar to the geometric interpretation of the left-sided Riemann–Liouville fractional integral. However, in this case there is no any fixed point in the "fence" base – the end, corresponding to $\tau = b$, moves along the line $\tau = b$ in the plane (τ, g) when the "fence" changes its shape. This movement can be

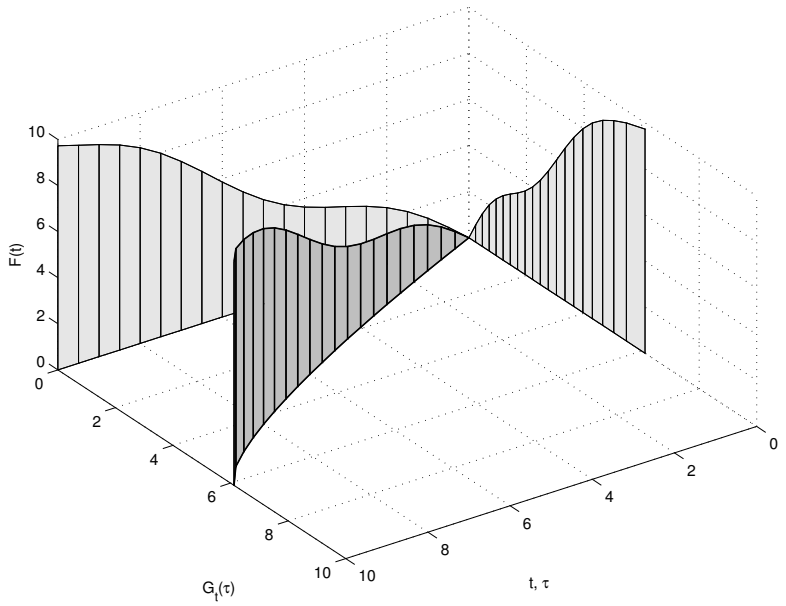


Figure 1.1: The "fence" and its shadows: ${}_0I_t^1 f(t)$ and ${}_0I_t^\alpha f(t)$, for $\alpha = 0.75$, $f(t) = t + 0.5\sin(t)$

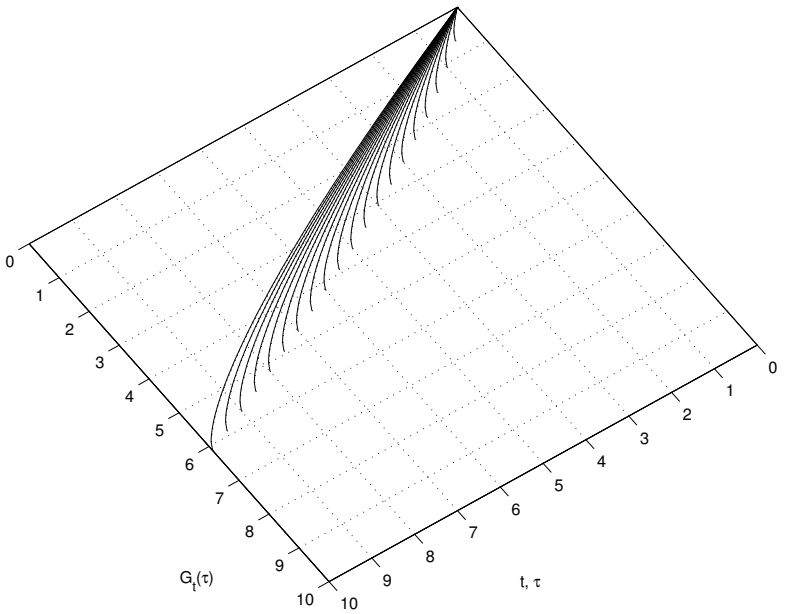


Figure 1.2: The process of change of the fence basis shape for ${}_0I_t^\alpha f(t)$, $\alpha = 0.75$.

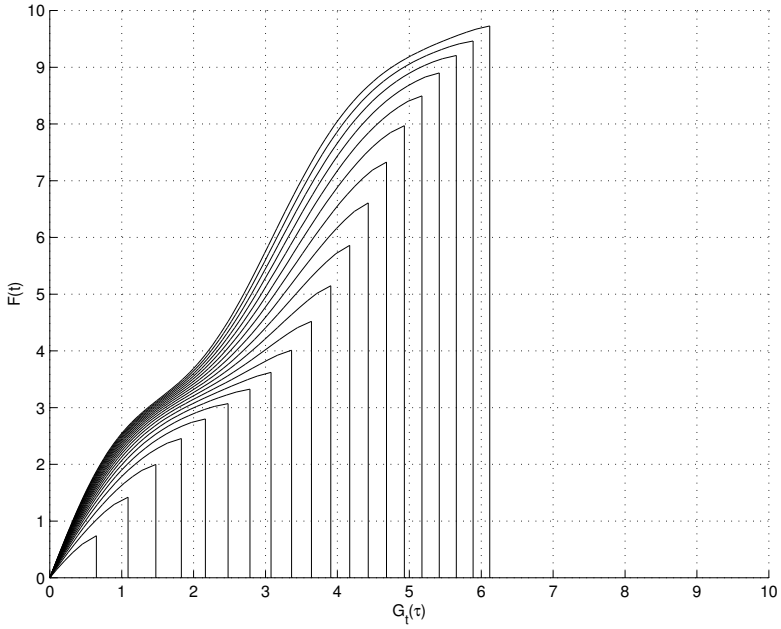


Figure 1.3: Snapshots of the changing "shadow" of changing "fence" for ${}_0I_t^\alpha f(t)$, $\alpha = 0.75$, $f(t) = t + 0.5\sin(t)$, with the time interval $\Delta t = 0.5$ between the snapshops.

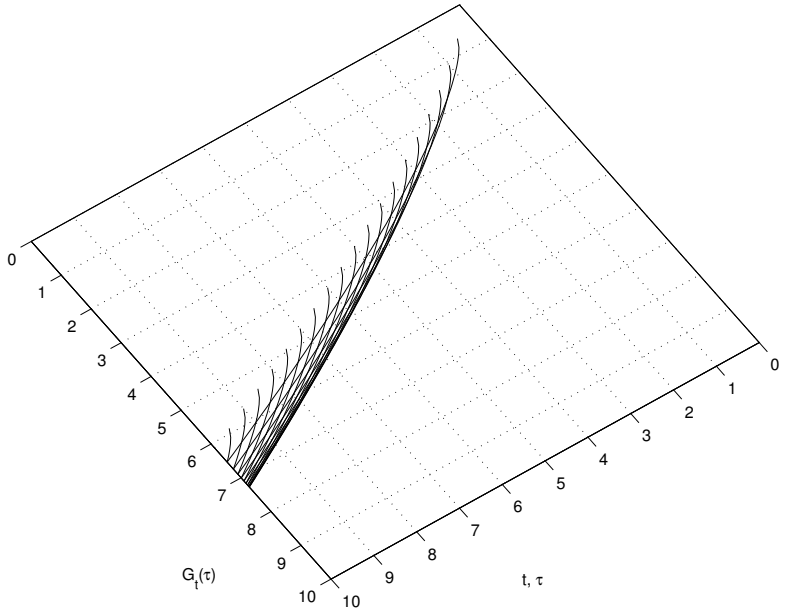


Figure 1.4: The process of change of the fence basis shape for ${}_tI^{\alpha}_{10}f(t)$, $\alpha = 0.75$.

observed in Fig. 1.4. (In the case of the left-sided integral, the left end, corresponding to $\tau = 0$, is fixed and does not move.)

All other parts of the geometric interpretation remain the same: the "fence" changes its shape as t changes from 0 to b, and the changing shadows of this "fence" on the walls (g, f) and (τ, f) represent correspondingly the right-sided Riemann–Liouville fractional integral (1.6) and the classical integral with the moving lower limit:

$$ {}_tI^1_b(t) = \int_t^b f(\tau)d\tau; \tag{1.9}$$

Obviously, for $g_t(\tau) = \tau$ both "shadows" are equal. Therefore, we see that not only the left-sided, but also the right-sided Riemann-Liouville fractional integration includes the classical definite integration as a particular case even from the geometrical point of view.

1.2.3 Riesz potential

The Riesz potential [20, 24]

$$ {}_0R_b^\alpha f(t) = \frac{1}{\Gamma(\alpha)} \int_0^b f(\tau)|\tau - t|^{\alpha-1} d\tau \tag{1.10} $$

is the sum of the left-sided and the right-sided Riemann–Liouville fractional integrals:

$$ {}_0R_b^\alpha f(t) = \frac{1}{\Gamma(\alpha)} \int_0^t f(\tau)(t-\tau)^{\alpha-1} d\tau + \frac{1}{\Gamma(\alpha)} \int_t^b f(\tau)(\tau - t)^{\alpha-1} d\tau. \tag{1.11} $$

The Riesz potential (1.10) can be written in the form

$$ {}_0R_b^\alpha f(t) = \int_0^b f(\tau) dr_t(\tau), \tag{1.12} $$

$$ r_t(\tau) = \frac{1}{\Gamma(\alpha+1)} \Big\{ t^\alpha + \operatorname{sign}(\tau - t)\, |\tau - t|^\alpha \Big\}. \tag{1.13} $$

The shape of the "fence", corresponding to the Riesz potential, is described by the function $r_t(\tau)$. In this case the "fence" consists of the two parts: one of them (for $0 < \tau < t$) is the same as in the case of the left-sided Riemann–Liouville fractional integral, and the second (for $t < \tau < b$) is the same as for the right-sided Riemann–Liouville integral, as shown in Fig. 1.5. Both parts are joined smothly at the inflection point $\tau = t$.

The shape of the "fence", corresponding to the Riesz potential, is shown in some of its intermediate position by the bold line in Fig. 1.5. Obviously, Fig. 1.5 can be obtained by laying Fig. 1.4 over Fig. 1.2, which is a geometric interpretation of the relationship (1.11).

The shadow of this "fence" on the wall (g, f) represents the Riesz potential (1.10), while the shadow on the wall (τ, f) corresponds to the classical integral

$$ I(t) = \int_0^b f(\tau) d\tau. \tag{1.14} $$

For $\alpha = 1$ both "shadows" are equal. This shows that the classical definite integral (1.14) is a particular case of the Riesz fractional potential (1.10) even from the geometric point of view. We have already seen this inclusion in the case of the left-sided and the right-sided Riemann–Liouville fractional integration. This demonstrates the strength of the suggested geometric interpretation of these three types of generalization of the notion of integration.

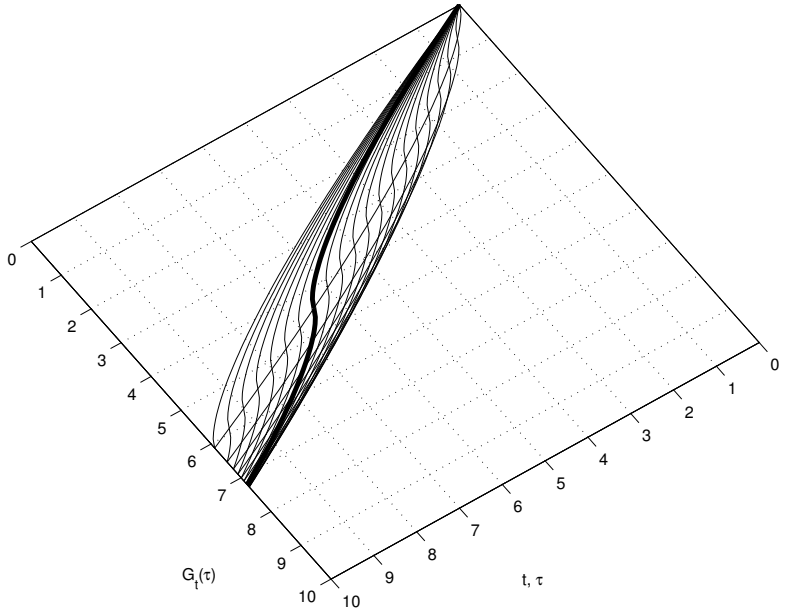


Figure 1.5: The process of change of the fence basis shape for the Riesz potential ${}_0R^{\alpha}_{10}f(t)$, $\alpha = 0.75$.

1.2.4 Feller potential

The Feller potential operator $\Phi^{\alpha}f(t)$ is, similarly to the Riesz potential, also a linear combination of the left- and right-sided Riemann–Liouville fractional integrals, but with general constant coefficients c, d [24, Chap. 3]:

$$\Phi^{\alpha}f(t) = c\,{}_aI^{\alpha}_t f(t) + d\,{}_tI^{\alpha}_b f(t). \tag{1.15}$$

The geometric interpretation of the Feller potential can be easily obtained by properly scaling and then superimposing Fig. 1.4 and Fig. 1.2. The "fence" obtained in this way is, in general, discontinuous at $\tau = t$. Its shadow on the wall (τ, f) is equal to the classical definite integral (1.14). The shadow on the wall (g, f) consists, in general, of the two areas, which may overlap depending on the values of the coefficients c and d.

1.3 Two kinds of time – I

The geometric interpretation of fractional integration, given in the previous sections, is substantially based on adding the third dimension (for $g_t(\tau)$) to the classical pair τ, $f(\tau)$. If we consider τ as time, then $g(\tau)$ can be interpreted just as a "deformed" time scale. What could be the meaning of having – and using – two time axes? To answer

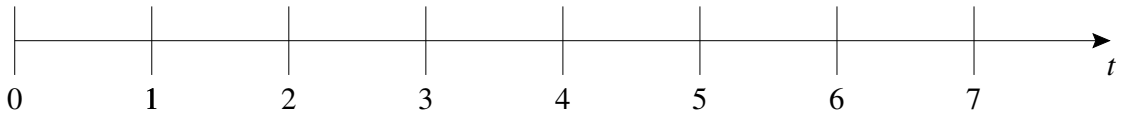


Figure 1.6: Homogeneous time axes.

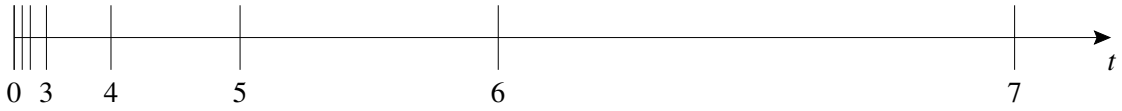


Figure 1.7: Time slowing down.

this question, let us recall some facts of the history of the development of the notion of time.

That were contributions of Barrows and Newton to the development of mathematics and physics in the XVII century which led to the appearance of the "mathematical time", which is postulated to "flow equably" and which is usually depicted as a semi-infinite straight line [26].

Newton himself postulated [17]:

> "Absolute, true and mathematical time of itself, and from its own nature, flows equably without relation to anything external."

Such a postulate was absolutely necessary for developing Newton's differential calculus and applying it to problems of mechanics [26]:

> "The outstanding mathematical achievement associated with the geometrization of time was, of course, the invention of the calculus of fluxions by Newton."
>
> "Mathematically, Newton seems to have found support for his belief in absolute time by the need, in principle, for an ideal rate-measurer."

The invention of differential and integral calculus and today's use of them is the strongest reason for continuing using homogeneous equably flowing time.

Time is often depicted using the time axis, and the geometrically equal intervals of the time axis are considered as corresponding to equal time intervals (Fig. 1.6).

This assumption, however, cannot be neither proved nor rejected by experiment. Two lengths of geometric intervals can be measured and compared, since they are available for measurement simultaneously, at the same time (or, more precisely, at the same time and at the same place). Two time intervals can never be compared, because they are available to us for measurement (or for observation) only sequentially.

Indeed, how do we measure time intervals? Only by observing some processes, which we consider as regularly repeated. G. Clemence wrote [7]:

> "The measurement of time is essentially a process of counting. Any recurring phenomenon whatever, the occurences of which can be counted, is in fact a measure of time."

Clocks, including atomic clocks, repeat their "ticks", and we simply count those ticks, calling them hours, minutes, seconds, milliseconds, etc. But we are not able to verify if the *absolute* time which elapsed between, say, the fifth and the sixth tick (the sixth "second") is exactly the same as the time, which elapsed between the sixth and the seventh tick (the seventh "second"). This possible inhomogeneity of the time scale is illustrated in Fig. 1.7.

The fact that time measurement as a process of counting of repeating discrete events does not really exclude inhomogeneity of time, has been nicely mentioned by L. Carroll in *Alice's Adventures in Wonderland* [6, Chap. 7]:

> "...I know I have to beat time when I learn music."
> "Ah! That accounts for it," said the Hatter. "He [Time] won't stand beating. Now, if you only kept on good terms with him, he'd do almost anything you liked to do with the clock..."

Figures 1.6 and 1.7 show those "clock ticks", which we can register, only symbolically. One can interpret them as if there exists some *absolute*, or *cosmic*, inhomogeneous time axis, to which we can compare *individual* homogeneous time, represented by some "clock ticks". Our picture of the individual homogeneous time has the form shown in Fig. 1.6. The cosmic time may be not necessarily flowing equably, like that shown in Fig. 1.7.

To illustrate the idea, let us consider the following situation. Suppose person N has two devices: one is a speedometer, and another one is the clock, which is slowing down, so the interval between the two subsequent ticks is double comparing to the interval between the previous ticks (see Fig. 1.7). Person N reads the velocity values indicated by the speedometer at each encountered "second", without knowing that the clock is, in fact, slowing down.

Using these two series of data, namely the recorded sequence of values of speed, and the sequence of the counted "seconds", person N can estimate the distance which he has passed.

For simplicity, let us suppose that the first "second" of the time shown by the clocks is equal to the absolute time "second". The results of observations in this hypothetical experiment are given in Table 1.

Person N will compute the distance he has passed as

$$S_N = 10 \cdot 1 + 11 \cdot 1 + 12 \cdot 1 + 13 \cdot 1 + 12 \cdot 1 + 11 \cdot 1 + 10 \cdot 1 = 79.$$

However, if there would be an independent observer O, knowing about the slowing-down clock, then such an observer would obtain a notably different result for the distance

Table 1.1: Recording speed using slowing-down clocks

Person N individual "seconds"	Recorded values of velocity [m/s]	Observer O absolute (cosmic) "seconds"
0	10	0
1	11	1
2	12	3
3	13	7
4	12	15
5	11	31
6	10	63
7	9	127

passed by person N:

$$S_O = 10 \cdot 1 + 11 \cdot 2 + 12 \cdot 4 + 13 \cdot 8 + 12 \cdot 16 + 11 \cdot 32 + 10 \cdot 64 = 1368.$$

Below we use this idea for giving a new mechanical interpretation of the Stieltjes integral.

1.4 Physical interpretation of the Stieltjes integral

Imagine a car equipped with two devices for measurements: the speedometer recording the velocity $v(\tau)$, and the clock which should show the time τ. The clock, however, shows the time incorrectly; let us suppose that the relationship between the wrong time τ, which is shown by the clock and which the driver considers as the correct time, on one hand, and the true time T, on the other, is described by the function $T = g(\tau)$. This means that where the driver "measures" the time interval $d\tau$, the real time interval is given by $dT = dg(\tau)$.

The driver A, who do not know about wrong operation of the clock, will compute the passed distance as the classical integral:

$$S_A(t) = \int_0^t v(\tau) d\tau \,. \tag{1.16}$$

However, the observer O knowing about the wrong clock and having the function $g(\tau)$, which restores the correct values of time from the driver's wrong time τ, will compute

the really passed distance as

$$S_O(t) = \int_0^t v(\tau) dg(\tau). \tag{1.17}$$

This example shows that the Stieltjes integral (1.17) can be interpreted as the real distance passed by a moving object, for which we have recorded correct values of speed and incorrect values of time; the relationship between the wrongly recorded time τ and the correct time T is given by a known function $T = g(\tau)$.

1.5 Physical interpretation of fractional integration: Shadows of the past

Now let us consider the left-sided Riemann–Liouville fractional integral

$$S_O(t) = \int_0^t v(\tau) dg_t(\tau) = {}_0I_t^{\alpha} v(t), \tag{1.18}$$

where $g_t(\tau)$ is given by (1.3).

The fractional integral $S_O(t)$ of the function $v(\tau)$ can be interpreted as the real distance passed by a moving object, for which we have recorded the local values of its speed $v(\tau)$ (individual speed) and the local values of its time τ (individual time); the relationship between the locally recorded time τ (which is considered as flowing equably) and the cosmic time (which flows non-equably) is given by a known function $g_t(\tau)$.

The function $g_t(\tau)$ describes the inhomogeneous time scale, which depends not only on τ, but also on the parameter t representing the last measured value of the individual time of the moving object. When t changes, the entire preceding cosmic time interval changes as well. This is in agreement with the current views in physics. Indeed, B. N. Ivanov [12, p. 33] mentioned that time intervals depend on gravitational fields. Similarly, S. Hawking [11, p. 32–33] wrote that:

> "... time should appear to run slower near a massive body like the earth."
> [...]
> "... there is no unique absolute time, but instead each individual has his own personal measure of time that depends on where he is and how he is moving."

When a moving body changes its position in space–time, the gravitational field in the entire space–time also changes due to this movement. As a consequence, the cosmic time interval, which corresponds to the history of the movement of the moving object, changes. This affects the calculation (using formula (1.18)) of the real distance $S_O(t)$ passed by such a moving object.

In other words, the left-sided Riemann–Liouville fractional integral of the individual speed $v(\tau)$ of a moving object, for which the relationship between its individual time τ and the cosmic time T at each individual time instance t is given by the known function $T = g_t(\tau)$ described by the equation (1.3), represents the real distance $S_O(t)$ passed by that object.

1.6 Physical interpretation of the Riemann-Liouville fractional derivative

On the other hand, we can use the properties of fractional differentiation and integration [20, 24] and express $v(t)$ from the equation (1.18) as a left-sided Riemann–Liouville fractional derivative of $S_O(t)$:

$$v(t) = {}_0D_t^{\alpha} S_O(t) \tag{1.19}$$

where ${}_0D_t^{\alpha}$ denotes the Riemann–Liouville fractional derivative [20, 24], which is for $0 < \alpha < 1$ defined by

$${}_0D_t^{\alpha} f(t) = \frac{1}{\Gamma(1-\alpha)} \frac{d}{dt} \int_0^t \frac{f(\tau)d\tau}{(t-\tau)^{\alpha}}. \tag{1.20}$$

This shows that the left-sided Riemann–Liouville fractional derivative of the real distance $S_O(t)$ passed by a moving object, for which the relationship between its individual time τ and the cosmic time T at each individual time instance t is given by the known function $T = g_t(\tau)$ described by equation (1.3), is equal to the individual speed $v(\tau)$ of that object.

On the other hand, we can differentiate the relationship (1.18) with respect to the cosmic time variable t, which gives the relationship between the velocity $v_O(t) = S'_O(t)$ of the movement from the viewpoint of the independent observer O and the individual velocity $v(t)$:

$$v_O(t) = \frac{d}{dt}\, {}_0I_t^{\alpha} v(t) = {}_0D_t^{1-\alpha} v(t), \tag{1.21}$$

Therefore, the $(1-\alpha)$-th–order Riemann–Liouville derivative of the individual velocity $v(t)$ is equal to the velocity $v_O(t)$ from the viewpoint of the independent observer, if the individual time τ and the cosmic time T are related by the function $T = g_t(\tau)$ described by equation (1.3). For $\alpha = 1$, when there is no dynamic deformation of the time scale, both velocities coincide: $v_O(t) = v(t)$.

1.7 Physical interpretation of the Caputo fractional derivative

Applying fractional integration of order $\beta = 1-\alpha$ to both parts of the relationship (1.21) gives:

$$v(t) = {}_0I_t^{1-\alpha} v_O(t) = {}_0I_t^{1-\alpha} S'_O(t) = {}_0^C D_t^{\alpha} S_O(t), \tag{1.22}$$

where ${}_0^C D_t^{\alpha}$ denotes the Caputo fractional derivative [4, 5, 20], which is for $0 < \alpha < 1$ defined by

$${}_0^C D_t^{\alpha} f(t) = \frac{1}{\Gamma(1-\alpha)} \int_0^t \frac{f'(\tau)d\tau}{(t-\tau)^{\alpha}}. \tag{1.23}$$

The relationship (1.22) is similar to (1.19). Therefore, the Caputo fractional derivative has the same physical interpretation as the Riemann–Liouville fractional derivative (see Section 1.6). This coincidence becomes more obvious, if we recall [20] that if $f(0) = 0$, then the Riemann–Liouville derivative and the Caputo derivative of order α $(0 < \alpha < 1)$, coincide: ${}_0^C D_t^{\alpha} f(t) = {}_0D_t^{\alpha} f(t)$.

1.8 Two kinds of time – II

The suggested physical interpretation of fractional integration and fractional differentiation is based on using two kinds of time: the cosmic time and the individual time.

As mentioned above, due to the history of the development of mathematics and physics, we are taught to think about the time, in fact, geometrically. The real roots of this go even far back to Euclid [26]:

> "Euclid considered space as the primary concept of science and relegated time to poor second."

The entire integral and differential calculus is based on using mathematical (homogeneous, equably flowing) time. There is no chance to change this state, and there is nothing to suggest instead of the classical calculus. Moreover, there is probably even no need for this. We can just realize that the classical calculus provides tools for describing the dynamic properties of the cosmic time, which – according to physicists – is inhomogeneous (flowing non-equably). Indeed [11, p. 33–34],

> "The old idea of an essentually unchanging universe that could have existed, and could continue to exist, forever was replaced by the notion of a dynamic, expanding universe that seemed to have begun a finite time ago..."

Clearly, the expansion of the universe implies that neither spatial scale nor time scale remains homogeneous; they both are dynamic. For describing the inhomogeneous time, the ideal homogeneous time scale can be used. This approach is not new; it has already

been used in the theory of relativity for describing shortening of time intervals. This means that in fact two time scales are considered simultaneously: the ideal, equably flowing homogeneous time, and the cosmic (inhomogeneous) time. The change of scale of the cosmic time is described using the homogeneous time scale as a reference scale. In other words, the homogeneous time scale is just an ideal notion, which is necessary for developing mathematical models describing inhomogeneous cosmic time and its change. In this respect we can, without discussing other views on this subject, recall the remark made by A. Daigneault and A. Sangalli in their essay [8] about I. E. Segal and his two-time cosmology ("chronometric cosmology", or CC) [25] – note that "*perhaps!*":

> "According to CC, Einstein's model is the correct one to understand the universe as a whole (i.e., global space–time), except that there are two kinds of time: a cosmic or Einstein's time t, and a local or Minkowski's time x_0, which is (perhaps!) the time measured by existing techniques. [...] Simply put, Einstein's cosmic time is the "real" one, whereas Minkowski's time is only an approximation of t."

So, the ideal model of equably flowing homogeneous time can be considered as a rough approximation of the cosmic time.

1.9 Geometric and physical interpretation of the Volterra convolution integral

It should be mentioned that we can also provide a geometric and physical interpretation for more general integrals.

The Riemann–Liouville fractional integral is a particular case of convolution integrals of the Volterra type:

$$K * f(t) = \int_0^t f(\tau)k(t-\tau)d\tau \tag{1.24}$$

Assuming that $k(t) = K'(t)$, we can write this integral in the form

$$K * f(t) = \int_0^t f(\tau)dq_t(\tau), \tag{1.25}$$

$$q_t(\tau) = K(t) - K(t-\tau). \tag{1.26}$$

The geometric and physical interpretation of the Volterra convolution integral is then similar to the suggested interpretations for fractional integrals. The function $q_t(\tau)$ determines the changing shape of the "live fence" (in the case of the geometric interpretation, see Figs. 1.1 and 1.2) and the relationship between the individual time and the cosmic time of a moving object (in the case of the physical interpretation).

1.10 Chapter summary

Possible geometric and physical interpretations of fractional-order operators have been presented in this chapter. They are based on using two time axes – one is idea mathematical axis, and the other is a "live" and changing axis. Fractional-order operators then have interpretations in the form of some dynamic shape (in the case of geometric interpretation) and of some past history of the process that is continuously re-weighted (in the case of physical interpretation).

1.11 References

[1] F. Ben Adda. Geometric interpretation of the fractional derivative. *Journal of Fractional Calculus,* vol. 11, May 1997, pp. 21–52.

[2] F. Ben Adda. Interprétation géometrique de la différentiabilité et du gradient d'ordre réel. *C. R. Acad. Sci. Paris*, t. 326, Série I, 1998, pp. 931–934.

[3] G. L. Bullock. A geometric interpretation of the Riemann–Stieltjes integral. *American Mathematical Monthly*, vol. 95, no. 5, May 1988, pp. 448–455.

[4] M. Caputo. Linear model of dissipation whose Q is almost frequency independent – II. *Geophys. J. R. Astr. Soc.*, vol. 13, 1967, pp. 529–539.

[5] M. Caputo. *Elasticità e Dissipazione.* Zanichelli, Bologna, 1969.

[6] L. Carroll. *Alice's Adventures in Wonderland.* Progress, Moscow, 1979 (or any other edition).

[7] G. Clemence. Time and its measurement. *The American Scientist*, vol. 40, 1952, pp. 260–269.

[8] A. Daigneault and A. Sangalli. Einstein's static universe: An idea whose time has come back? *Notices of the AMS*, vol. 48, no. 1, 2001, pp. 9–16.

[9] Fu-Yao Ren, Zu-Guo Yu, Feng Su. Fractional integral associated to the self-similar set of the generalised self-similar set and its physical interpretation. *Phys. Lett. A*, vol. 219, 1996, pp. 59–68.

[10] R. Gorenflo. Afterthoughts on interpretation of fractional derivatives and integrals. In: P. Rusev, I. Dimovski, V. Kiryakova (eds.), *Transform Methods and Special Functions, Varna'96,* Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Sofia, 1998, ISBN 954-8986-05-1, pp. 589–591.

[11] S. W. Hawking. *A Brief History of Time.* Bantam Press, London, 1988.

[12] B. N. Ivanov. *Printsipy sovremennoi fiziki (Principles of Modern Physics).* Nauka, Moscow, 1973 (in Russian).

[13] V. Kiryakova. A long standing conjecture failed? In: P. Rusev, I. Dimovski, V. Kiryakova (eds.), *Transform Methods and Special Functions, Varna'96,* Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Sofia, 1998, ISBN 954-8986-05-1, pp. 579–588.

[14] F. Mainardi. Considerations on fractional calculus: interpretations and applications. In: P. Rusev, I. Dimovski, V. Kiryakova (eds.), *Transform Methods and Special Functions, Varna'96,* Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Sofia, 1998, ISBN 954-8986-05-1, pp. 594–597.

[15] A. McBride, G. Roach (eds). *Fractional Calculus.* Research Notes in Mathematics, vol. 138, Pitman, Boston–London–Melbourne, 1985.

[16] M. Monsrefi-Torbati and J.K. Hammond. Physical and geometrical interpretation of fractional operators. *J. Franklin Inst.*, vol. 335B, no. 6, 1998, pp. 1077–1086.

[17] I. Newton. *Mathematical Principles* (trans. A. Motte, ed. F. Cajori), Berkeley, 1934.

[18] R. R. Nigmatullin. A fractional integral and its physical interpretation. *Theoret. and Math. Phys.*, vol. 90, no. 3, 1992, pp. 242–251.

[19] K. Nishimoto (ed.). *Fractional Calculus and Its Applications.* Nihon University, Koriyama, 1990.

[20] I. Podlubny. *Fractional Differential Equations.* Academic Press, San Diego, 1999.

[21] B. Ross (ed.). *Fractional Calculus and Its Applications,* Lecture Notes in Mathematics, vol. 457, Springer-Verlag, New York, 1975.

[22] R. S. Rutman. On the paper by R.R. Nigmatullin "A fractional integral and its physical interpretation". *Theoret. and Math. Phys.*, vol. 100, no. 3, 1994, pp. 1154–1156.

[23] R. S. Rutman. On physical interpretations of fractional integration and differentiation. *Theoret. and Math. Phys.*, vol. 105, no. 3, 1995, pp. 1509–1519.

[24] S. G. Samko, A. A. Kilbas, O. I. Marichev. *Integraly i proizvodnye drobnogo poryadka i nekotorye ich prilozheniya (Integrals and Derivatives of the Fractional Order and Some of Their Applications).* Nauka i Tekhnika, Minsk, 1987 (English translation: S.G. Samko, A.A. Kilbas, and O.I. Marichev, *Fractional Integrals and Derivatives*, Gordon and Breach, Amsterdam, 1993).

[25] I. E. Segal. *Mathematical Cosmology and Extragalactic Astronomy*, Academic Press, New York, 1976.

[26] G. J. Whitrow. *The Natural Philosophy of Time.* Thomas Nelson and Sons Ltd., London and Edinburgh, 1961.

[27] Zu-Guo Yu, Fu-Yao Ren and Ji Zhou. Fractional integral associated to generalized cookie-cutter set and its physical interpretation. *J. Phys.A: Math. Gen.*, vol. 30, 1997, pp. 5569–5577.

Chapter 2

Physical interpretation of initial conditions for fractional differential equations with Riemann-Liouville fractional derivatives

2.1 Introduction

Many physical phenomena lead to their description in terms of non integer order differential equations. Formulations of non integer order derivatives, generally called fractional derivatives, fall into two main classes: Riemann-Liouville derivatives and Grünwald-Letnikov derivatives, on one hand, defined as (Podlubny 1999, Samko et al. 1993)

$$ {}_0D_t^\alpha f(t) = \frac{1}{\Gamma(n-\alpha)} \left(\frac{d}{dt}\right)^n \int_0^t \frac{f(\tau)\, d\tau}{(t-\tau)^{\alpha-n+1}}, \tag{2.1} $$

or the Caputo derivative on the other, defined as (Caputo and Mainardi 1971)

$$ {}_0^C D_t^\alpha f(t) = \frac{1}{\Gamma(n-\alpha)} \int_0^t \frac{f^{(n)}(\tau)\, d\tau}{(t-\tau)^{\alpha-n+1}}, \tag{2.2} $$

where $n-1 \leq \alpha < n$.

In this article we deal only with the Riemann-Liouville fractional derivatives. Fractional differential equations in terms of the Riemann-Liouville derivatives require initial

conditions expressed in terms of initial values of fractional derivatives of the unknown function (Podlubny 1999, Samko et al. 1993), like, for example, in the following initial value problem (where $n-1 < \alpha < n$):

$$ {}_0D_t^{\alpha} f(t) + af(t) = h(t); \qquad (t > 0) \tag{2.3} $$

$$ \left[{}_0D_t^{\alpha-k} f(t) \right]_{t\to 0} = b_k, \qquad (k = 1, 2, \ldots, n). \tag{2.4} $$

On the contrary, initial conditions for the Caputo derivatives are expressed in terms of initial values of integer order derivatives. It is known that for zero initial conditions the Riemann-Liouville, Grünwald-Letnikov and Caputo fractional derivatives coincide (Podlubny 1999). This allows a numerical solution of initial value problems for differential equations of non integer order independently of the chosen definition of the fractional derivative. For this reason, many authors either resort to Caputo derivatives, or use the Riemann-Liouville derivatives but avoid the problem of initial values of fractional derivatives by treating only the case of zero initial conditions.

It is frequently stated that the physical meaning of initial conditions expressed in terms of fractional derivatives is unclear or even non existent. The old and ubiquitous requirement for physical interpretation of such initial conditions was most clearly formulated recently by Diethelm et al. (2005):

> "A typical feature of differential equations (both classical and fractional) is the need to specify additional conditions in order to produce a unique solution. For the case of Caputo FDEs, these additional conditions are just the static initial conditions ..., which are akin to those of classical ODEs, and *are therefore familiar to us*. In contrast, for Riemann-Liouville FDEs, these additional conditions constitute certain fractional derivatives (and/or integrals) of the unknown solution at the initial point $x = 0$ ..., which are functions of x. These initial conditions *are not physical*; furthermore, it is not clear how such quantities are to be measured from experiment, say, so that they can be appropriately assigned in an analysis."

(Emphasis has been added). This quotation highlights the utmost importance of the interpretation of initial conditions in terms of fractional derivatives for further applications in various fields of science. The physical and geometric interpretations of operations of fractional integration and differentiation were suggested recently by Podlubny (2002). However, the problem of interpretation of initial conditions still remained open.

In this chapter we shall show that initial conditions for fractional differential equations with Riemann-Liouville derivatives expressed in terms of fractional derivatives *have physical meaning*, and that the corresponding quantities *can be obtained from measurements*. We shall also demonstrate that in many instances of practical significance zero initial conditions, which are used so frequently in practice, appear in a natural way.

2.2 Number of initial conditions, past history and memory

When a physical process can be described in terms of a differential equation of integer order n, it is well known that n conditions are required to solve the system. In this chapter only initial conditions are considered. It is also known (Podlubny 1999, Samko et al. 1993) that fractional differential equations of order α require $\alpha*$ initial conditions, where $\alpha*$ is the lowest integer greater than α. This means that if $\alpha < 1$ as is the case in viscoelasticity when inertial effects are negligible, a single initial condition is sufficient. However, one of the reasons for the success encountered in describing viscoelasticity by means of differential equations of non integer order is their ability to describe real behaviour, including memory effects such as are observed in polymers, using only a restricted number of material parameters. Such memory effects may continue to affect the material response long after the cause has disappeared, as observed in stress relaxation after a non monotonous loading programme (Heymans and Kitagawa 2004). In such a case a single initial condition would appear insufficient to predict material response.

Here we shall consider only the response of a system starting at $t = 0$ from a state of absolute rest. As a further simplification, we shall consider only response to ideal loading programs, such as step or impulse response. The effects of a finite loading time, of the details of the loading program, and of past history will be accounted for separately in a sequel.

It has been shown (Beris and Edwards 1993) that thermodynamically valid constitutive equations for viscoelasticity are completely equivalent to analog models containing only elements (springs and dashpots) with positive coefficients. A suitable hierarchical arrangement of springs and dashpots gives rise to spring-pot behaviour (described below), either exactly at all timescales for an infinite tree (Heymans and Bauwens 1994), or in the long-term (or low-frequency) limit for an infinite ladder or infinite Sierpinski gasket (Heymans and Bauwens 1994, Schiessel and Blumen 1993, 1995). In the latter case short-term behaviour is similar to a Maxwell model with one element replaced by a spring-pot. Therefore the equivalence demonstrated by Beris and Edwards can be generalized to models including spring-pots (Heymans 1996), hence discussion here will be limited to such models.

2.3 Spring-pot model

We shall start with a spring-pot alone, which is a linear viscoelastic element whose behaviour is intermediate between that of an elastic element (spring) and a viscous element (dashpot). The term "spring-pot" was introduced by Koeller (1984), although the concept of an element with intermediate properties had been introduced some time

earlier. The constitutive equation of a spring-pot is:

$$\sigma(t) = K\,{}_0D_t^{\alpha}\epsilon(t) \qquad \text{or} \qquad \epsilon(t) = \frac{1}{K}\,{}_0D_t^{-\alpha}\sigma(t) \tag{2.5}$$

where σ is stress, ϵ is strain and K is the model constant. The spring-pot is the viscoelastic version of Westerlund's "simplest model" (Westerlund 2002). If $\alpha = 0$ the element is linear elastic (Hookean spring) whereas if $\alpha = 1$ it is purely viscous (Newtonian dashpot). Insight can be gained from the response of a spring-pot in a few simple cases, using the general relationship (Podlubny 1999, Samko et al. 1993)

$${}_0D_t^{\alpha}(at^p) = a\,\frac{\Gamma(1+p)}{\Gamma(1+p-\alpha)}t^{p-\alpha}. \tag{2.6}$$

2.3.1 Creep or general finite load

In the case of creep, a stress step σ_0 is applied at initial time $t = 0$. The strain response is hence $\epsilon(t) = (\sigma_0/K\Gamma(1+\alpha))t^{\alpha}$. The initial value of the strain vanishes, i.e. there is no instantaneous (elastic) strain, only an anelastic (retarded) response. However, the first ordinary derivative of strain is unbounded, so that a finite though undefined strain can be reached in an arbitrarily small time interval.

The change of $\epsilon(t)$ is described by the fractional differential equation

$${}_0D_t^{\alpha}\epsilon(t) = \frac{\sigma_0}{K} \tag{2.7}$$

In accordance with the theory of fractional differential equations in terms of Riemann-Liouville derivatives, an initial condition involving ${}_0D_t^{\alpha-1}\epsilon(t)$ is required. This condition can be found by taking the first-order integral of the constitutive equation as

$$\left[{}_0D_t^{\alpha-1}\epsilon(t)\right]_{t\to 0} = \left[{}_0D_t^{-1}(\sigma_0/K)\right]_{t\to 0}.$$

In the case under consideration stress is finite at all times, hence $\left[{}_0D_t^{-1}\sigma_0\right]_{t\to 0} = 0$, which leads to zero initial condition for ${}_0D_t^{\alpha-1}\epsilon(t)$, namely

$$\left[{}_0D_t^{\alpha-1}\epsilon(t)\right]_{t\to 0} = 0. \tag{2.8}$$

The same considerations apply to a general finite load $\sigma(t)$. In the latter case the equation to be solved is

$${}_0D_t^{\alpha}\epsilon(t) = \frac{\sigma(t)}{K}, \tag{2.9}$$

and the initial condition to be attached to this equation is the zero initial condition (2.8).

2.3.2 Stress relaxation or general deformation

The stress response to a strain step ϵ_0 is $\sigma(t) = (\epsilon_0 K/\Gamma(1-\alpha))t^{-\alpha}$. The initial stress is unbounded reflecting the fact that a spring-pot (just like a dashpot) cannot respond immediately to a bounded stress: it has an infinite initial modulus or a vanishing initial compliance. However, relaxation to a finite though undefined stress occurs in an arbitrarily small time interval.

The change of $\sigma(t)$ is described by the fractional differential equation

$$ {}_0D_t^{-\alpha}\sigma(t) = K\epsilon_0. \tag{2.10}$$

From the known value of ϵ_0 we can obtain the initial value (as t approaches zero) of ${}_0D_t^{-\alpha}\sigma(t)$. Clearly, if ${}_0D_t^{-\alpha}\sigma(t)$ is to be finite although it is defined over a vanishingly small time interval, $\sigma(t)$ must be unbounded. On the contrary, the initial value of ${}_0D_t^{-\alpha}\sigma(t)$ is well defined and finite, and that of ${}_0D_t^{-\alpha-1}\sigma(t)$ is zero. Thus, contrary to the idea expressed by some authors (e.g., Glöckle and Nonnenmacher 1991), initial value problems expressed in terms of fractional integrals are not better posed than those expressed in terms of fractional derivatives.

If strain increases linearly with time, stress increases as $t^{1-\alpha}$. Stress is bounded, but the initial values of its integer order derivatives are unbounded. The known strain rate allows us to define the initial value of ${}_0D_t^{1-\alpha}\sigma(t)$. In fact, in this case, zero initial conditions are found both for ${}_0D_t^{-\alpha}\sigma(t)$ and ${}_0D_t^{-\alpha-1}\sigma(t)$.

For any general finite strain $\epsilon(t)$, following the same reasoning, again zero initial conditions are found.

In all three examples given here, initial conditions expressed in terms of fractional derivatives or integrals arise naturally when taking measurable quantities into account.

2.3.3 Impulse response

The impulse response is seldom used in viscoelasticity except as a mathematical convenience, because it is even more problematic to apply a homogeneous impulse of stress or strain on a sample than it is to apply a step. However, we shall investigate the impulse response following the same reasoning as for the step response above.

Consider an impulse of stress defined as $B\delta(t)$ applied to the spring-pot at time $t = 0$. After that, the stress remains zero. The strain response is $\epsilon(t) = (B/K\Gamma(\alpha))t^{\alpha-1}$. The initial stress singularity gives rise to a lower-order strain singularity, since a spring-pot cannot deform immediately.

The strain $\epsilon(t)$ for $t > 0$ is the solution to the fractional differential equation

$$ {}_0D_t^{\alpha}\epsilon(t) = 0. \tag{2.11}$$

In accordance with the theory of fractional differential equations with Riemann-Liouville derivatives, an initial condition involving $\left[{}_0D_t^{\alpha-1}\epsilon(t)\right]_{t\to 0}$ is required. This can be found through integration of the constitutive equation, as

$$\left[{}_0D_t^{\alpha-1}\epsilon(t)\right]_{t\to 0} = \left[{}_0D_t^{-1}\sigma(t)/K\right]_{t\to 0} = B/K,$$

which gives the following initial condition to equation (2.11):

$$\left[{}_0D_t^{\alpha-1}\epsilon(t)\right]_{t\to 0} = B/K. \tag{2.12}$$

In this problem in terms of Riemann-Liouville derivatives B is the initial impulse of stress $\sigma(t)$, $\epsilon(t)$ is the strain after application of this impulse, and the known impulse of stress yields a non-zero initial condition (2.12) involving a fractional derivative of strain. This fractional derivative is non zero, well defined, and bounded. Note that both strain and its integer-order derivatives are unbounded, and its first order integral is zero, so that a meaningful initial condition expressing the loading conditions cannot be obtained using integral-order derivatives.

The physically unrealistic stress response to a prescribed strain impulse will not be considered here. In fact, the analytical solution has a strong $t^{-(1+\alpha)}$ divergence, reflecting the fact that a strain impulse cannot be applied to a spring-pot.

2.4 The key: look for inseparable twins

Now, after introducing the above simple example, let us formulate our general approach to interpretation of initial conditions involving the Riemann-Liouville fractional derivatives.

In a general case, when we consider some fractional differential equation for, say, $U(t)$, we have to consider also some function $V(t)$, for which some *dual relation* exists between $U(t)$ and $V(t)$. For example, in viscoelasticity we have to consider the pair of stress $\sigma(t)$ and strain $\epsilon(t)$; in electrical circuits the pair of current $i(t)$ and voltage $v(t)$; in heat conduction the pair of the temperature difference $T(t)$ and the heat flux $q(t)$; etc. Functions $U(t)$ and $V(t)$ are normally related by some basic physical law for the particular field of science. In each scientific field there are such pairs of functions like the aforementioned, which are as *inseparable as Siamese twins*: the left-hand side of the initial condition involves one of them, whereas the evaluation of the right-hand side is related to the other.

This concept is not restricted to the spring-pot treated above, but is further applied in the subsequent sections to more elaborate models of viscoelastic behaviour. Indeed, a spring-pot is a particularly crude model, which has several unrealistic and unphysical characteristics. As pointed out above, it has a vanishing initial compliance or an infinite

initial modulus. Viscoelastic solids, on the contrary, have a well-defined instantaneous modulus. (Note that an unbounded initial modulus is no more of a problem when describing a viscoelastic fluid than it is when describing Newtonian viscosity: if a step strain is applied to a dashpot, the initial stress is also unbounded). At long times there is no limit to anelastic strain of a spring-pot: creep continues indefinitely. Also, stress relaxes to vanishingly small values. The increase of stress in constant strain-rate conditions means that if attempting to describe a viscoelastic fluid, steady state flow is never attained. When describing a viscoelastic solid, again we find an unbounded modulus at $t = 0$. In spite of these limitations, the spring-pot can give an approximate description of polymer viscoelasticity in the intermediate time range. Several slightly more elaborate models, which alleviate some oversimplifications of a single spring-pot, will be investigated below.

2.5 The fractional order Voigt model

The fractional Voigt model (a spring and a spring-pot in parallel) is nowadays generally understood as a long-term approximation to the fractional Zener model and it might seem irrelevant to express concern over initial value problems for the Voigt model. However, as the purpose of this note is mainly to examine how initial conditions endowed with physical meaning may be expressed in systems whose constitutive equations contain fractional derivatives, we shall continue to examine the fractional Voigt model.

The constitutive equation of this model is

$$\sigma(t) = E\epsilon(t) + K\,{}_0D_t^{\alpha}\epsilon(t). \tag{2.13}$$

The Voigt element or associations thereof are considered in viscoelasticity modelling to be appropriate to obtain the strain response to a prescribed stress program, so we shall investigate only such cases here.

Assume a stress impulse $B\delta(t)$ is applied to a Voigt element at time $t = 0$. Then the fractional order equation we need to solve for $\epsilon(t)$ $(t > 0)$ is

$$E\epsilon(t) + K\,{}_0D_t^{\alpha}\epsilon(t) = 0. \tag{2.14}$$

In agreement with the theory of fractional differential equations in terms of Riemann-Liouville derivatives, we need an initial condition, which will involve the value of ${}_0D_t^{\alpha-1}\epsilon(t)$ for $t \to 0$. This condition can be obtained by integration of the constitutive equation as

$$\Big[E\,{}_0D_t^{-1}\epsilon(t) + K\,{}_0D_t^{\alpha-1}\epsilon(t) = {}_0D_t^{-1}\sigma(t)\Big]_{t\to 0}. \tag{2.15}$$

The limit of the right hand side is the magnitude B of the stress impulse. On physical grounds, the spring-pot cannot deform instantaneously under a finite stress, and, as is

the case for a spring-pot alone, any singularity of $\epsilon(t)$ must be weaker than that of the stress impulse, thus

$$\left[{}_0D_t^{-1}\epsilon(t)\right]_{t\to 0} = 0.$$

This can also be found from examination of the behaviour of the left hand side of the relationship (2.15): if $\left[{}_0D_t^{-1}\epsilon(t)\right]_{t\to 0}$ is non zero, then $\left[{}_0D_t^{\alpha-1}\epsilon(t)\right]_{t\to 0}$ is unbounded and equation (2.15) cannot be fulfilled. Hence the initial condition finally takes on the form of

$$\left[K\,{}_0D_t^{\alpha-1}\epsilon(t)\right]_{t\to 0} = B. \tag{2.16}$$

This condition expresses the initial value of the fractional derivative of strain, ${}_0D_t^{\alpha-1}\epsilon(t)$, in terms of the stress impulse. We see that the initial condition is obtained expressing a fractional derivative of the unknown strain in terms of a measurable, physically meaningful value of its "inseparable Siamese twin"– the stress. The obtained initial condition is, in fact, identical to the initial condition of the spring-pot alone. This reflects the known fact that the spring in the Voigt model only affects long-term behaviour.

Now let us consider creep, i. e. the response to a stress step σ_0 applied at $t = 0$. The equation to be solved for the strain $\epsilon(t)$ is

$$E\epsilon(t) + K\,{}_0D_t^{\alpha}\epsilon(t) = \sigma_0, \tag{2.17}$$

and the initial condition for (2.17) can be found from

$$\left[E\,{}_0D_t^{-1}\epsilon(t) + K\,{}_0D_t^{\alpha-1}\epsilon(t) = {}_0D_t^{-1}\sigma(t)\right]_{t\to 0},$$

where the limit of the right hand side is zero. A bounded stress can produce only a bounded strain, so the limit of the first-order ordinary integral of strain in the left hand side is also zero. Thus the initial condition has the form:

$$\left[{}_0D_t^{\alpha-1}\epsilon(t)\right]_{t\to 0} = 0. \tag{2.18}$$

Once more, knowledge of a measurable quantity (σ_0) leads to an initial condition expressed in terms of a fractional order derivative of the unknown ($\epsilon(t)$), its inseparable Siamese twin.

The case of a general finite stress program is similar to that of creep. The equation to be solved is now

$$E\epsilon(t) + K\,{}_0D_t^{\alpha}\epsilon(t) = \sigma(t), \tag{2.19}$$

and the initial condition is identical to the condition (2.18) obtained in creep.

2.6 The fractional order Maxwell model

To keep to a simple model while describing realistic behaviour for a viscoelastic solid, a spring expressing instantaneous elasticity must be associated in series with the spring-pot. This eliminates the unbounded initial stress in description of relaxation. The Maxwell element or associations thereof are considered in viscoelasticity modelling to be appropriate to obtain the stress response to a prescribed strain program, so we shall investigate only such cases here.

The constitutive equation of the Maxwell model is

$$\epsilon(t) = \frac{1}{E}\,\sigma(t) + \frac{1}{K}\,({}_0D_t^{-\alpha}\sigma(t)),$$

or

$$\frac{1}{E}\,{}_0D_t^{\alpha}\sigma(t) + \frac{1}{K}\,\sigma(t) = {}_0D_t^{\alpha}\epsilon(t). \tag{2.20}$$

In stress relaxation, a step strain ϵ_0 is applied at $t = 0$. Then the equation to be solved is

$$\frac{1}{E}\,{}_0D_t^{\alpha}\sigma(t) + \frac{1}{K}\,\sigma(t) = \frac{\epsilon_0 t^{-\alpha}}{\Gamma(1-\alpha)}. \tag{2.21}$$

An initial condition is required, involving the value of $\left[{}_0D_t^{\alpha-1}\sigma(t)\right]_{t\to 0}$. Integrating the constitutive equation (2.20) and considering the limit as t approaches zero, we have

$$\left[\frac{1}{E}\,{}_0D_t^{\alpha-1}\sigma(t) + \frac{1}{K}\,{}_0D_t^{-1}\sigma(t) = {}_0D_t^{\alpha-1}\epsilon(t)\right]_{t\to 0}. \tag{2.22}$$

Since strain remains bounded during loading, and $\alpha < 1$, the right hand side inside brackets is bounded and vanishes when $t \to 0$. Since the left hand side is a linear combination of positive functions with positive coefficients, it can only vanish if each term vanishes. This means that stress remains bounded during loading, and hence that we obtain the following initial condition:

$$\left[{}_0D_t^{\alpha-1}\sigma(t)\right]_{t\to 0} = 0. \tag{2.23}$$

Here again we observe that the initial condition on the unknown stress arises naturally from its Siamese twin, the known strain.

Now we shall consider the strain impulse response. A strain impulse of magnitude $A\delta(t)$ is applied at time t=0. Thereafter, the equation to be solved is

$$\frac{1}{E}\,\sigma(t) + \frac{1}{K}\,{}_0D_t^{-\alpha}\sigma(t) = 0. \tag{2.24}$$

The required initial condition is obtained as above by integrating the constitutive equation (2.20) and considering the limit as t approaches zero:

$$\left[\frac{1}{E}\,{}_0D_t^{-1}\sigma(t) + \frac{1}{K}\,{}_0D_t^{-\alpha-1}\sigma(t) = {}_0D_t^{-1}\epsilon(t)\right]_{t\to 0}. \tag{2.25}$$

The limit of the right hand side of (2.25) is A. Hence the limit of the left hand side must also be bounded. This means that the limit of the first-order integral of stress must be bounded, and the $\alpha+1$ integral must vanish, and the initial condition is finally

$$\left[\frac{1}{E}\,{}_0D_t^{-1}\sigma(t)\right]_{t\to 0} = A. \tag{2.26}$$

The singularity in the stress response to a strain impulse is now of the same order as that of the strain impulse itself: adding a spring in series with the spring-pot has weakened the singularity.

The strain response to a stress impulse is identical to that of a spring-pot alone since the impulse response of the spring vanishes.

2.7 The fractional order Zener model

Among the fractional order models of viscoelasticity considered in this article, the most general is the Zener model. Its constitutive equation is

$$\sigma(t) + \nu\,{}_0D_t^{\alpha}\sigma(t) = \lambda\,\epsilon(t) + \mu\,{}_0D_t^{\alpha}\epsilon(t), \tag{2.27}$$

where $\lambda = E_\infty$ is the long-term modulus, $\mu = K(E_0 - E_\infty)/E_0$, $\nu = \mu/E_0$ and E_0 is the instantaneous modulus.

Let us first investigate the response to a stress impulse $B\delta(t)$ applied to the Zener element at time $t = 0$. Then the fractional differential equation we need to solve for $\epsilon(t)$ $(t > 0)$ is:

$$\lambda\epsilon(t) + \mu\,{}_0D_t^{\alpha}\epsilon(t) = 0. \tag{2.28}$$

In accordance with the theory of fractional differential equations, we need an initial condition involving the initial value of ${}_0D_t^{\alpha-1}\epsilon(t)$. Integration of the constitutive equation gives:

$${}_0D_t^{-1}\sigma(t) + \nu\,{}_0D_t^{\alpha-1}\sigma(t) = \lambda\,{}_0D_t^{-1}\epsilon(t) + \mu\,{}_0D_t^{\alpha-1}\epsilon(t). \tag{2.29}$$

The initial condition can be found by considering equation (2.29) as $t \to 0$:

$$\left[{}_0D_t^{-1}\sigma(t) + \nu\,{}_0D_t^{\alpha-1}\sigma(t) = \lambda\,{}_0D_t^{-1}\epsilon(t) + \mu\,{}_0D_t^{\alpha-1}\epsilon(t)\right]_{t\to 0}. \tag{2.30}$$

Using considerations similar to those in case of the Voigt model under stress impulse, we obtain the initial condition in the form:

$$\left[\mu\,{}_0D_t^{\alpha-1}\epsilon(t)\right]_{t\leftrightarrow 0} = B \tag{2.31}$$

As in case of the Voigt model, this condition gives the initial value of the fractional derivative of unknown strain, ${}_0D_t^{\alpha-1}\epsilon(t)$, in terms of its "inseparable twin" – the stress.

The right and left hand sides of equations (2.28) and (2.30) are formally identical, hence following the same reasoning as above the response to a strain impulse $A\delta(t)$ applied to the Zener element at time $t = 0$ will be the solution to the equation

$$\sigma(t) + \nu\, {}_0D_t^{\alpha}\sigma(t) = 0 \tag{2.32}$$

with the initial condition

$$\left[\nu\, {}_0D_t^{\alpha-1}\sigma(t)\right]_{t\leftrightarrow 0} = A \tag{2.33}$$

This formal equivalence between response to a stress or strain impulse reflects the well known fact that the Zener model is the simplest model capable of describing response to a stress or strain program equally well.

In case of creep, i.e. a step-stress $\sigma(t) = \sigma_0$ for $\sigma > 0$, we have the equation for $\epsilon(t)$:

$$\lambda\epsilon(t) + \mu\, {}_0D_t^{\alpha}\epsilon(t) = \sigma_0 + \nu\,\sigma_0\,\frac{t^{-\alpha}}{\Gamma(1-\alpha)}. \tag{2.34}$$

The initial condition can also be found by considering equation (2.29) as $t \to 0$.

Following a similar reasoning to that given above for the Maxwell model in stress relaxation, we find a zero initial condition to accompany equation (2.34):

$$\left[{}_0D_t^{\alpha-1}\epsilon(t)\right]_{t\to 0} = 0. \tag{2.35}$$

This initial condition in terms of fractional derivative of $\epsilon(t)$ appeared from consideration of its "inseparable twin" $\sigma(t)$.

Similarly, for stress relaxation, $\epsilon(t) = \epsilon_0$, we obtain the equation for $\sigma(t)$:

$$\sigma(t) + \nu\, {}_0D_t^{\alpha}\sigma(t) = \lambda\epsilon_0 + \mu\,\epsilon_0\,\frac{t^{-\alpha}}{\Gamma(1-\alpha)}. \tag{2.36}$$

The initial condition for the unknown stress $\sigma(t)$,

$$\left[{}_0D_t^{\alpha-1}\sigma(t)\right]_{t\to 0} = 0, \tag{2.37}$$

appears naturally from consideration of the initial value of strain.

Let us now consider the case of general load $\sigma(t) = \sigma_*(t)$. The equation to be solved for $\epsilon(t)$ is

$$\lambda\epsilon(t) + \mu\, {}_0D_t^{\alpha}\epsilon(t) = \sigma_*(t) + \nu\, {}_0D_t^{\alpha}\sigma_*(t) \tag{2.38}$$

The corresponding initial condition can be obtained using the following procedure. Consider some small $t = a$. Starting at $t = 0$, stress $\sigma(t)$ must be recorded until $t = a$, and based on the recorded values the left hand side of the relationship (2.29) must be evaluated. The obtained quantity provides an approximation of the initial value for the expression in the right hand side of (2.29).

In some cases it is possible to find the limit of such approximation as $a \to 0$. For example, for a physically realisable continuous load $\sigma_*(t)$ we obtain a zero initial condition in the form:

$$\left[{}_0D_t^{\alpha-1}\epsilon(t)\right]_{t\to 0} = 0. \tag{2.39}$$

It is worth mentioning that this procedure amounts, in fact, to the same as measuring the initial value of, for example, the first derivative in the case of classical differential equations of integer order. From the examples given above, it can be seen that for any physically realistic model, zero initial conditions will be found for a continuous loading program or even in the case of a step discontinuity. Non-zero conditions will only be found in the case of an impulse.

2.8 Chapter summary

Examples presented in this chapter demonstrate that it is possible to attribute physical meaning to initial conditions expressed in terms of Riemann-Liouville fractional derivatives.

To summarize, expressing initial conditions in terms of fractional derivatives of a function $U(t)$ is not a problem, because it does not require a direct experimental evaluation of these fractional derivatives. Instead, one should consider its "inseparable twin" $V(t)$ related to $U(t)$ via a basic physical law, and measure (or consider) its initial values.

It is worth noting that the only case where non zero initial conditions appeared in our considerations, is the case of impulse response. In other cases (including the Zener model under physically realisable load program), the initial conditions are zero, and in such cases the use of the Riemann-Liouville derivatives, the Grünwald-Letnikov derivatives, and the Caputo derivatives is equivalent.

2.9 References

[1] Beris AN, Edwards BJ (1993) On the admissibility criteria for linear viscoelastic kernels. Rheol Acta 32:505–510

[2] Caputo M, Mainardi F. (1971) Linear models of dissipation in anelastic solids. Riv. Nuovo Cimento (Ser. II) 1:161-198

[3] Diethelm K, Ford NJ, Freed AD, Luchko Yu (2005) Algorithms for the fractional calculus: A selection of numerical methods. Comput Methods Appl Mech Engrg 194:743–773

[4] Glöckle WG, Nonnenmacher TF (1991) Fractional integral operators and Fox functions in the theory of viscoelasticity. Macromolecules 24:6426–6434

[5] Heymans N, Bauwens J-C (1994) Fractal rheological models and fractional differential equations for viscoelastic behavior. Rheol Acta 33:210–219

[6] Heymans N (1996) Hierarchical models for viscoelasticity: dynamic behaviour in the linear range. Rheol Acta 35:508–519

[7] Heymans N, Kitagawa M (2004) Modelling "unusual" behaviour after strain reversal with hierarchical fractional models. Rheol Acta 43:383–389

[8] Koeller RC (1984) Applications of Fractional Calculus to the Theory of Viscoelasticity. J Appl Mech 51:299–307

[9] Podlubny I (1999) Fractional Differential Equations. Academic Press, San Diego

[10] Podlubny I (2002) Geometric and physical interpretation of fractional integration and fractional differentiation. Fract Calc Appl Anal 5:367–386

[11] Samko SG, Kilbas AA, Marichev OI (1993) Fractional Integrals and Derivatives: Theory and Applications. Gordon and Breach, Amsterdam.

[12] Schiessel H, Blumen A (1993) Hierarchical Analogues to Fractional Relaxation Equations. J Phys A: Math Gen 26:5057-5069

[13] Schiessel H, Blumen A (1995) Mesoscopic Pictures of the Sol-Gel Transition: Ladder Models and Fractal Networks. Macromolecules 28:4013-4019

[14] Westerlund S (2002) Dead Matter has Memory! Causal Consulting, Kalmar, Sweden, 2002

Chapter 3

Matrix approach for ordinary fractional differential equations

3.1 Introduction

There are several well-known approaches to unification of notions of differentiation and integration, and their extension to non-integer orders [13].

The approach, which is described in this chapter, unifies *numerical* differentiation of integer order and n-fold integration, using so-called triangular strip matrices [1, 8, 14]. Applied to numerical solution of differential equations, it also unifies solution of ordinary integer- and fractional-order differential equations, and fractional integral equations.

Triangular strip matrices already appeared in some studies on fractional integral equations [2, 3, 6, 7, 9, 10, 11], but until today their usefulness for approximating fractional derivatives and solving fractional differential equations has not been recognized.

The general structure of this chapter is the following. First of all, triangular strip matrices, and operations on them, are introduced. Then discrete forms of integer-order differentiation and n-fold integration are considered using triangular strip matrices, and a generalisation for the case of an arbitrary (non-integer) order of differentiation and integration is presented. The advantages of the use of triangular strip matrices for numerical solution of fractional integral and differential equations of some important types are described and illustrated with four examples.

3.2 Triangular strip matrices

In this and subsequent chapters we will deal with matrices of a specific structure, which are called *triangular strip matrices* [14, p. 20], and which have been mentioned in [1, 8]. We will consider lower triangular strip matrices,

$$L_N = \begin{bmatrix} \omega_0 & 0 & 0 & 0 & \cdots & 0 \\ \omega_1 & \omega_0 & 0 & 0 & \cdots & 0 \\ \omega_2 & \omega_1 & \omega_0 & 0 & \cdots & 0 \\ \ddots & \ddots & \ddots & \ddots & \ldots & \ldots \\ \omega_{N-1} & \ddots & \omega_2 & \omega_1 & \omega_0 & 0 \\ \omega_N & \omega_{N-1} & \ddots & \omega_2 & \omega_1 & \omega_0 \end{bmatrix}, \tag{3.1}$$

and upper triangular strip matrices,

$$U_N = \begin{bmatrix} \omega_0 & \omega_1 & \omega_2 & \ddots & \omega_{N-1} & \omega_N \\ 0 & \omega_0 & \omega_1 & \ddots & \ddots & \omega_{N-1} \\ 0 & 0 & \omega_0 & \ddots & \omega_2 & \ddots \\ 0 & 0 & 0 & \ddots & \omega_1 & \omega_2 \\ \cdots & \cdots & \cdots & \cdots & \omega_0 & \omega_1 \\ 0 & 0 & 0 & \cdots & 0 & \omega_0 \end{bmatrix}, \tag{3.2}$$

A lower (upper) triangular strip matrix is completely described by its first column (row). Because of this, it may be convenient in the future to use a compact notation of the form

$$L_N = ||\omega_0, \omega_1, \ldots, \omega_N||^T,$$
$$U_N = ||\omega_0, \omega_1, \ldots, \omega_N||,$$

where $||\cdot||^T$ denotes matrix transposition. However, in this chapter we prefer to use full matrix notation for clarity.

Obviously, if matrices C and D are both lower (upper) triangular strip matrices, then they commute:

$$CD = D\,C \tag{3.3}$$

Denoting

$$\Omega_N = L_N - \omega_0 E, \qquad \Psi_N = U_N - \omega_0 E, \tag{3.4}$$

where E is the unit matrix, we can write

$$L_N = \omega_0 E + \Omega_N, \qquad U_N = \omega_0 E + \Psi_N \tag{3.5}$$

We can also consider $(N+1)\times(N+1)$ matrices E_p^+, $p=1,\dots N$, with ones on p-th diagonal above the main diagonal and zeroes elsewhere, and matrices E_p^-, $p=1,\dots N$, with ones on p-th diagonal below the main diagonal and zeroes elsewhere. We will also denote $E_0^\pm \equiv E$ the unit matrix.

It can be shown that

$$E_p^\pm E_q^\pm = \begin{cases} E_{p+q}^\pm, & (p+q \le N), \\ O, & (p+q > N), \end{cases} \tag{3.6}$$

from which follows that for integer k

$$(E_p^\pm)^k = \begin{cases} E_{pk}^\pm, & (pk \le N), \\ O, & (pk > N), \end{cases} \qquad (E_1^\pm)^{N+1} = O. \tag{3.7}$$

Noting that

$$\Omega_N = \sum_{k=1}^{N} \omega_k E_k^-, \qquad \Psi_N = \sum_{k=1}^{N} \omega_k E_k^+, \tag{3.8}$$

it can be shown that $(N+1)$-th power of Ω_N and of Ψ_N gives the zero matrix:

$$\Omega_N^{N+1} = O, \qquad \Psi_N^{N+1} = O. \tag{3.9}$$

Using (3.9) it is easy to check that the inverse matrices $(L_N)^{-1}$ and $(U_N)^{-1}$ are given by the following explicit expressions [5, p. 62]:

$$(L_N)^{-1} = \omega_0^{-1}E - \omega_0^{-2}\Omega_N + \omega_0^{-3}\Omega_N^2 + \ldots + (-1)^N\omega_0^{-N-1}\Omega_N^N, \tag{3.10}$$

$$(U_N)^{-1} = \omega_0^{-1}E - \omega_0^{-2}\Psi_N + \omega_0^{-3}\Psi_N^2 + \ldots + (-1)^N\omega_0^{-N-1}\Psi_N^N. \tag{3.11}$$

There is a link between matrix polynomials and triangular strip matrices. Namely, if we introduce the polynomial $\varrho_N(z)$,

$$\varrho_N(z) = \omega_0 + \omega_1 z + \omega_2 z^2 + \ldots + \omega_N z^N, \tag{3.12}$$

and take into account the relationship (3.7), then we can write:

$$\varrho_N(E_1^-) = \omega_0 E + \omega_1 E_1^- + \omega_2 (E_1^-)^2 + \ldots + \omega_N (E_1^-)^N = L_N, \tag{3.13}$$

$$\varrho_N(E_1^+) = \omega_0 E + \omega_1 E_1^+ + \omega_2 (E_1^+)^2 + \ldots + \omega_N (E_1^+)^N = U_N. \tag{3.14}$$

where L_N and U_N defined by relationships (3.1) and (3.2).

If we define the truncation operation, $\mathrm{trunc}_N(\cdot)$, which truncates (in a general case) the power series $\varrho(z)$,

$$\varrho(z) = \sum_{k=0}^{\infty} \omega_k z^k \tag{3.15}$$

to the polynomial $\varrho_N(z)$,

$$\operatorname{trunc}_N(\varrho(z)) \stackrel{\text{def}}{=} \sum_{k=0}^{N} \omega_k z^k = \varrho_N(z), \tag{3.16}$$

then we can consider the function $\varrho(z)$ as a generating series for the set of lower (or upper) triangular matrices L_N (or U_N), $N = 1, 2, \ldots$.

We will need the following properties of the truncation operation:

$$\begin{aligned} \operatorname{trunc}_N(\gamma\lambda(z)) &= \gamma \operatorname{trunc}_N(\lambda(z)), && (3.17)\\ \operatorname{trunc}_N(\lambda(z)+\mu(z)) &= \operatorname{trunc}_N(\lambda(z)) + \operatorname{trunc}_N(\mu(z)), && (3.18)\\ \operatorname{trunc}_N(\lambda(z)\mu(z)) &= \operatorname{trunc}_N\Big(\operatorname{trunc}_N(\lambda(z))\ \operatorname{trunc}_N(\mu(z))\Big). && (3.19) \end{aligned}$$

3.3 Operations with triangular strip matrices

Due to special structure of triangular strip matrices, operations with them, such as addition, subtraction, multiplication, and inversion, can be expressed in the form of operations with their generating series (3.15).

Let us consider two $(N+1)\times(N+1)$ lower triangular strip matrices: matrix A_N with elements a_k, $k = 0, 1, \ldots, N$ in its first column, and matrix B_N with elements b_k, $k = 0, 1, \ldots, N$ in its first column. Denoting $\lambda(z)$ and $\mu(z)$ the generating series of A_N and B_N respectively, and using the representation (3.13), we can write:

$$A_N = \sum_{k=0}^{N} a_k (E_1^-)^k = \lambda_N(E_1^-), \qquad B_N = \sum_{k=0}^{N} b_k (E_1^-)^k = \mu_N(E_1^-), \tag{3.20}$$

where $\lambda_N(z) = \operatorname{trunc}_N(\lambda(z))$, $\mu_N = \operatorname{trunc}_N(\mu(z))$, and therefore

$$A_N \pm B_N = \sum_{k=0}^{N} (a_k \pm b_k)(E_1^-)^k. \tag{3.21}$$

In symbolic form, using the generating series $\lambda(z)$ and $\mu(z)$ and the properties of truncation operation (3.17) and (3.18), this can be written as

$$\begin{aligned} A_N \pm B_N \quad &\longleftrightarrow \quad \operatorname{trunc}_N(\lambda(z)\pm\mu(z)) = \lambda_N(z) \pm \mu_N(z) = \\ &= \quad \sum_{k=0}^{N} (a_k \pm b_k) z^k \longleftrightarrow \sum_{k=0}^{N} (a_k \pm b_k)(E_1^-)^k. \end{aligned} \tag{3.22}$$

This means that the coefficient on k-th diagonal of the sum of two lower triangular strip matrices is equal to the sum of k-th coefficients of the generating series of those matrices. Therefore, summation of lower triangular strip matrices is equivalent to summation of their respective generating series with a subsequent truncation.

Multiplication by a constant γ is simple:

$$\gamma A_N \longleftrightarrow \mathrm{trunc}_N\left(\gamma\lambda(z)\right) = \gamma\lambda_N(z) = \sum_{k=0}^{N} \gamma a_k z^k \longleftrightarrow \sum_{k=0}^{N} \gamma a_k (E_1^-)^k, \tag{3.23}$$

and it is equivalent to multiplication of the generating series by γ followed by truncation.

Taking into account the property (3.7) of the matrix E_1^-, we obtain the product of A_N and B_N:

$$\begin{aligned} A_N B_N &= \left(\sum_{k=0}^{N} a_k (E_1^-)^k\right)\left(\sum_{k=0}^{N} b_k (E_1^-)^k\right) && (3.24)\\ &= \sum_{k=0}^{N}\left(\sum_{i=0}^{k} a_i b_{k-i}\right)(E_1^-)^k. && (3.25) \end{aligned}$$

Using the truncation operation, the product of the matrices A_N and B_N can also be expressed in terms of their generating series:

$$\begin{aligned} A_N B_N \longleftrightarrow\ & \mathrm{trunc}_N\left(\lambda(z)\mu(z)\right) = \mathrm{trunc}_N\left(\sum_{k=0}^{\infty}\left(\sum_{i=0}^{k} a_i b_{k-i}\right) z^k\right) = \\ =\ & \sum_{k=0}^{N}\left(\sum_{i=0}^{k} a_i b_{k-i}\right) z^k \longleftrightarrow \sum_{k=0}^{N}\left(\sum_{i=0}^{k} a_i b_{k-i}\right)(E_1^-)^k. \end{aligned} \tag{3.26}$$

In other words, the product of two lower triangular matrices A_N and B_N is equivalent to the truncated product of their generating series.

The use of the generating series is especially convenient for inverting the lower triangular strip matrices.

If A_N is a lower triangular strip matrix with a generating function $\lambda(z)$, then the generating function for the inverse matrix $(A_N)^{-1}$ is simply $\mathrm{y}(z) = \lambda^{-1}(z)$. Indeed,

$$A_N\,(A_N)^{-1} \longleftrightarrow \mathrm{trunc}_N\left(\lambda(z)\,\lambda^{-1}(z)\right) = 1 \longleftrightarrow E. \tag{3.27}$$

This means that the coefficients on the first column of the inverse matrix $(A_N)^{-1}$ are the coefficients of the polynomial

$$\mathrm{y}_N(z) = \mathrm{trunc}_N\left(\lambda^{-1}(z)\right), \tag{3.28}$$

which is the truncation of the generating series for the inverse matrix. This method of inversion of triangular strip matrices is even simpler than the formulas (3.10) and (3.11).

All the above rules involving generating functions can also be used for upper triangular strip matrices.

3.4 Integer-order differentiation

Let us consider equidistant nodes with the step h: $t_k = kh$, $(k = 0, \dots, N)$, in the interval $[a, b]$, where $t_0 = a$ and $t_N = b$.

3.4.1 Backward differences

For a function $f(t)$, differentiable in $[a, b]$, we can consider first-order approximation of its derivative $f'(t)$ at the points t_k, $k = 1, \dots, N$, using first-order backward differences:

$$f'(t_k) \approx \frac{1}{h} \nabla f(t_k) = \frac{1}{h} \left(f_k - f_{k-1} \right), \quad k = 1, \dots, N. \tag{3.29}$$

All N formulas (3.29) can be written simultaneously in the matrix form:

$$\begin{bmatrix} h^{-1} f_0 \\ h^{-1} \nabla f(t_1) \\ h^{-1} \nabla f(t_2) \\ \vdots \\ h^{-1} \nabla f(t_{N-1}) \\ h^{-1} \nabla f(t_N) \end{bmatrix} = \frac{1}{h} \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 0 & 0 & \cdots & 0 \\ 0 & -1 & 1 & 0 & \cdots & 0 \\ \dots & \dots & \dots & \ddots & \dots & \dots \\ 0 & \cdots & 0 & -1 & 1 & 0 \\ 0 & 0 & \cdots & 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} f_0 \\ f_1 \\ f_2 \\ \vdots \\ f_{N-1} \\ f_N \end{bmatrix} \tag{3.30}$$

In the formula (3.30) the column vector of function values f_k $(k = 0, \dots, N)$ is multiplied by the matrix

$$B_N^1 = \frac{1}{h} \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 0 & 0 & \cdots & 0 \\ 0 & -1 & 1 & 0 & \cdots & 0 \\ \dots & \dots & \dots & \ddots & \dots & \dots \\ 0 & \cdots & 0 & -1 & 1 & 0 \\ 0 & 0 & \cdots & 0 & -1 & 1 \end{bmatrix}, \tag{3.31}$$

and the result is the column vector of approximated values of $f'(t_k)$, $k = 1, \dots, N$, with the exception of the first element, depending on the value of the function $f(t)$ at the initial point, namely $h^{-1} f_0 = h^{-1} f(a)$. We can look at the matrix B_N^1 as at a discrete analog of first-order differentiation. The generating function for the matrix B_N^1 is

$$\beta_1(z) = h^{-1}(1 - z). \tag{3.32}$$

Similarly, we can consider the approximation of the second-order derivative using second-order backward differences:

$$f''(t_k) \approx \frac{1}{h^2} \nabla^2 f(t_k) = \frac{1}{h^2} \left(f_k - 2 f_{k-1} + f_{k-2} \right), \quad k = 2, \dots, N, \tag{3.33}$$

which in the matrix form corresponds to the relationship

$$\begin{bmatrix} h^{-2}\, f_0 \\ h^{-2}\,(-2f_0+f_1) \\ h^{-2}\,\nabla^2 f(t_2) \\ \vdots \\ h^{-2}\,\nabla^2 f(t_{N-1}) \\ h^{-2}\,\nabla^2 f(t_N) \end{bmatrix} = \frac{1}{h^2} \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ -2 & 1 & 0 & 0 & \cdots & 0 \\ 1 & -2 & 1 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \ddots & \cdots & \cdots \\ \cdots & 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & \cdots & 1 & -2 & 1 \end{bmatrix} \begin{bmatrix} f_0 \\ f_1 \\ f_2 \\ \vdots \\ f_{N-1} \\ f_N \end{bmatrix} \tag{3.34}$$

In the formula (3.34) the column vector of function values f_k $(k = 0, \ldots, N)$ is multiplied by the matrix

$$B_N^2 = \frac{1}{h^2} \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ -2 & 1 & 0 & 0 & \cdots & 0 \\ 1 & -2 & 1 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \ddots & \cdots & \cdots \\ \cdots & 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & \cdots & 1 & -2 & 1 \end{bmatrix} \tag{3.35}$$

and the result is the column vector of approximations of $f''(t_k)$, $k = 2, 3, \ldots, N$, with the exception of the first two elements, namely $h^{-2}f_0$ and $h^{-2}(-2f_0 + f_1)$. We can look at the matrix B_N^2 as at a discrete analog of second-order differentiation. The generating function for the matrix B_N^2 is

$$\beta_2(z) = h^{-2}(1 - 2z + z^2) = h^{-2}(1 - z)^2. \tag{3.36}$$

Further, we can consider a matrix B_N^p, where p is a positive integer:

$$B_N^p = \frac{1}{h^p} \begin{bmatrix} \omega_0 & 0 & \ldots & 0 & 0 & 0 & 0 & 0 \\ \omega_1 & \omega_0 & 0 & \ldots & 0 & 0 & 0 & 0 \\ \omega_2 & \omega_1 & \omega_0 & 0 & \ldots & 0 & 0 & 0 \\ \ldots & \ldots & \ldots & \ddots & \ldots & \ldots & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ddots & \ldots & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ddots & 0 & \ldots \\ 0 & \ldots & 0 & \omega_p & \omega_{p-1} & \ldots & \omega_0 & 0 \\ 0 & 0 & \ldots & 0 & \omega_p & \omega_{p-1} & \ldots & \omega_0 \end{bmatrix} \tag{3.37}$$

$$\omega_j = (-1)^j \binom{p}{j}, \qquad j = 0,\ 1,\ 2,\ \ldots\ ,\ p. \tag{3.38}$$

The matrix B_N^p is a discrete analog of differentiation of p-th order, if backward differences of the p-th order are used. The generating function for the matrix B_N^p is

$$\beta_p(z) = h^{-p}(1-z)^p. \tag{3.39}$$

For the generating functions of the form $\beta_p(z)$ we have:

$$\begin{aligned} \beta_2(z) &= \beta_1(z)\beta_1(z) \\ \beta_p(z) &= \underbrace{\beta_1(z)\dots\beta_1(z)}_{p} \\ \beta_{p+q}(z) &= \beta_p(z)\beta_q(z) = \beta_q(z)\beta_p(z), \end{aligned}$$

from which in view of (3.26) follows that

$$B_N^2 = B_N^1\, B_N^1, \tag{3.40}$$

$$B_N^p = \underbrace{B_N^1\, B_N^1 \dots B_N^1}_{p}, \tag{3.41}$$

$$B_N^{p+q} = B_N^p\, B_N^q = B_N^q\, B_N^p, \tag{3.42}$$

where p and q are positive integers.

3.4.2 Forward differences

Similarly to the previous section, we obtain that the matrix F_N^p, where p is a positive integer,

$$F_N^p = \frac{1}{h^p}\begin{bmatrix} \omega_0 & \dots & \omega_{p-1} & \omega_p & 0 & \dots & 0 & 0 \\ 0 & \omega_0 & \dots & \omega_{p-1} & \omega_p & 0 & \dots & 0 \\ \dots & 0 & \ddots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & \ddots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \dots & \ddots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 0 & \omega_0 & \omega_1 & \omega_2 \\ 0 & 0 & 0 & 0 & \dots & 0 & \omega_0 & \omega_1 \\ 0 & 0 & 0 & 0 & 0 & \dots & 0 & \omega_0 \end{bmatrix} \tag{3.43}$$

$$\omega_j = (-1)^j \binom{p}{j}, \qquad j = 0,\, 1,\, 2,\, \dots,\, p. \tag{3.44}$$

is a discrete analog of differentiation of p-th order, namely of $(-1)^p f^{(p)}(t)$, if forward differences of the p-th order are used. The generating function for F_N^p is the same as for B_N^p: $\beta_p(z) = h^{-p}(1-z)^p$.

Since the generating functions are the same as in case of the matrices B_N^p, we have for F_N^p the similar properties:

$$F_N^2 = F_N^1\, F_N^1, \tag{3.45}$$

$$F_N^p = \underbrace{F_N^1\, F_N^1 \ldots F_N^1}_{p}, \tag{3.46}$$

$$F_N^{p+q} = F_N^p\, F_N^q = F_N^q\, F_N^p, \tag{3.47}$$

where p and q are positive integers.

It also should be noted that transposition of the matrix B_N^p, representing the backward difference operation, gives the matrix F_N^p, which corresponds to forward differencing:

$$\left(B_N^p\right)^T = F_N^p, \qquad \left(F_N^p\right)^T = B_N^p. \tag{3.48}$$

3.5 n-fold integration

Now let us turn to integration. To deal with operations, which are inverse to differentiation, we have to consider definite integrals with one limit fixed and another moving.

3.5.1 Moving upper limit of integration

Let us take a function $f(t)$, integrable in $[a, b]$, and consider integrals with fixed lower limit and moving upper limit:

$$g_1(t) = \int_a^t f(t)dt, \tag{3.49}$$

for which we have $g_1'(t) = f(t)$ in (a, b).

Let us consider equidistant nodes with the step h: $t_k = kh$, $(k = 0, \ldots, N)$, in the interval $[a, b]$, where $t_0 = a$ and $t_N = b$. We can use the left rectangular quadrature rule for approximating the integral (3.49) at the points t_k, $k = 1, \ldots, N$:

$$g_1(t_k) \approx h \sum_{i=0}^{k-1} f_i, \qquad k = 1, \ldots, N. \tag{3.50}$$

All N formulas (3.50) can be written simultaneously in the matrix form:

$$\begin{bmatrix} g_1(t_1) \\ g_1(t_2) \\ g_1(t_3) \\ \vdots \\ g_1(t_N) \\ g_1(t_N + h) \end{bmatrix} = h \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 1 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \ddots & \cdots & \cdots \\ 1 & \cdots & 1 & 1 & 1 & 0 \\ 1 & 1 & \cdots & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} f_0 \\ f_1 \\ f_2 \\ \vdots \\ f_{N-1} \\ f_N \end{bmatrix} \tag{3.51}$$

We see that the column vector of function values f_k $(k = 0, \ldots, N)$ is multiplied by the matrix

$$I_N^1 = h \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 1 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \ddots & \cdots & \cdots \\ 1 & \cdots & 1 & 1 & 1 & 0 \\ 1 & 1 & \cdots & 1 & 1 & 1 \end{bmatrix}, \tag{3.52}$$

and the result is the column vector of approximated values of the integral (3.49), namely $g_1(t_k)$, $k = 1, \ldots, N$, with the exception of the last element, which corresponds to the node lying outside of the considered interval $[a, b]$. We can look at the matrix I_N^1 as at a discrete analog of left rectangular quadrature rule for evaluating the integral (3.49). The generating function for I_N^1 is

$$\varphi_1(z) = h(1 - z)^{-1}. \tag{3.53}$$

It must be noted here that the matrix I_N^1 is inverse to the matrix B_N^1, which corresponds to backward difference approximation of the first derivative. We have:

$$B_N^1 I_N^1 = I_N^1 B_N^1 \longleftrightarrow \mathrm{trunc}_N\left(\beta_1(z)\,\varphi_1(z)\right) = 1 \longleftrightarrow E. \tag{3.54}$$

Therefore, having one of these matrices, we can immediately obtain another by matrix inversion.

Similarly, we can consider the two-fold integral with moving upper boundary:

$$g_2(t) = \int_a^t dt \int_a^t f(t)dt, \tag{3.55}$$

for which we have $g_2''(t) = g_1'(t) = f(t)$ in (a, b).

Using the left rectangular quadrature rule twice for approximating $g_2(t_k)$ and taking into account that $g_1(t_0) = 0$, we have:

$$\begin{aligned} g_2(t_k) &= h\sum_{i=0}^{k-1} g_1(t_i) = h\sum_{i=1}^{k-1} g_1(t_i) = h\sum_{i=1}^{k-1} h\sum_{j=0}^{i-1} f_j = \\ &= h^2 \sum_{i=1}^{k-1}\sum_{j=0}^{i-1} f_j = h^2 \sum_{j=0}^{k-2}(k-j-1)\, f_j = \\ &= h^2\Big((k-1)f_0 + (k-2)f_1 + \ldots + 2f_{k-3} + f_{k-2}\Big), \\ & \qquad k = 2, 3, \ldots, N. \end{aligned} \tag{3.56}$$

The equations (3.56) can be written simultaneously in the matrix form:

$$\begin{bmatrix} g_2(t_2) \\ g_2(t_3) \\ \vdots \\ g_2(t_N) \\ g_2(t_N+h) \\ g_2(t_N+2h) \end{bmatrix} = h^2 \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ 2 & 1 & 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \ddots & \cdots & \cdots & \cdots \\ \cdots & 3 & 2 & 1 & 0 & 0 \\ N & \cdots & 3 & 2 & 1 & 0 \\ N+1 & N & \cdots & 3 & 2 & 1 \end{bmatrix} \begin{bmatrix} f_0 \\ f_1 \\ \vdots \\ f_{N-2} \\ f_{N-1} \\ f_N \end{bmatrix} \tag{3.57}$$

We see that the column vector of function values f_k $(k = 0, \ldots, N)$ is multiplied by the matrix

$$I_N^2 = h^2 \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ 2 & 1 & 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \ddots & \cdots & \cdots & \cdots \\ \cdots & 3 & 2 & 1 & 0 & 0 \\ N & \cdots & 3 & 2 & 1 & 0 \\ N+1 & N & \cdots & 3 & 2 & 1 \end{bmatrix}, \tag{3.58}$$

and the result is the column vector of approximated values of the integral (3.55), namely $g_2(t_k)$, $k = 2, \ldots, N$, with the exception of the last two elements, which correspond to the nodes lying outside of the considered interval $[a, b]$. We can look at the matrix I_N^2 as at a discrete analog of left rectangular quadrature rule for evaluating the two-fold integral (3.55). The generating function for I_N^2 is

$$\varphi_2(z) = h^2(1-z)^{-2}. \tag{3.59}$$

It must be mentioned here that the matrix I_N^2 is inverse to the matrix B_N^2, which corresponds to backward difference approximation of the second derivative. We have:

$$B_N^2 I_N^2 = I_N^2 B_N^2 \longleftrightarrow \mathrm{trunc}_N\left(\beta_2(z)\,\varphi_2(z)\right) = 1 \longleftrightarrow E. \tag{3.60}$$

Therefore, having one of these matrices, we can immediately obtain another by matrix inversion.

If we consider p-fold integration with moving upper limit,

$$g_p(t) = \int_a^t d\tau_p \int_a^{\tau_p} d\tau_{p-1} \ldots \int_a^{\tau_2} f(\tau_1) d\tau_1, \tag{3.61}$$

and apply the left rectangular quadrature rule p times, then we arrive at the following

relationship in the matrix form:

$$\begin{bmatrix} g_p(t_p) \\ g_p(t_{p+1}) \\ \vdots \\ g_p(t_N) \\ \vdots \\ g_p(t_N+h) \\ g_p(t_N+ph) \end{bmatrix} = h^p \begin{bmatrix} \gamma_0 & 0 & 0 & 0 & \cdots & \cdots & 0 \\ \gamma_1 & \gamma_0 & 0 & 0 & \cdots & \cdots & 0 \\ \cdots & \cdots & \ddots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \gamma_2 & \gamma_1 & \gamma_0 & 0 & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \ddots & \cdots & \cdots \\ \gamma_{N-1} & \cdots & \cdots & \gamma_2 & \gamma_1 & \gamma_0 & 0 \\ \gamma_N & \gamma_{N-1} & \cdots & \cdots & \gamma_2 & \gamma_1 & \gamma_0 \end{bmatrix} \begin{bmatrix} f_0 \\ f_1 \\ \vdots \\ f_p \\ \vdots \\ f_{N-1} \\ f_N \end{bmatrix}, \quad (3.62)$$

involving the lower triangular strip matrix I_N^p with the generating function $\varphi_p(z) = h^p(1-z)^{-p}$,

$$I_N^p = h^p \begin{bmatrix} \gamma_0 & 0 & 0 & 0 & \cdots & \cdots & 0 \\ \gamma_1 & \gamma_0 & 0 & 0 & \cdots & \cdots & 0 \\ \cdots & \cdots & \ddots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \gamma_2 & \gamma_1 & \gamma_0 & 0 & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \ddots & \cdots & \cdots \\ \gamma_{N-1} & \cdots & \cdots & \gamma_2 & \gamma_1 & \gamma_0 & 0 \\ \gamma_N & \gamma_{N-1} & \cdots & \cdots & \gamma_2 & \gamma_1 & \gamma_0 \end{bmatrix}, \quad (3.63)$$

which is inverse to the matrix B_N^p, corresponding to backward difference approximation of the p-th derivative:

$$B_N^p I_N^p = I_N^p B_N^p \longleftrightarrow \mathrm{trunc}_N\left(\beta_p(z)\,\varphi_p(z)\right) = 1 \longleftrightarrow E. \quad (3.64)$$

In view of (3.26) it follows from the properties of the generating functions $\varphi_p(z) = h^p(1-z)^{-p}$ that

$$I_N^2 = I_N^1\, I_N^1, \quad (3.65)$$

$$I_N^p = \underbrace{I_N^1\, I_N^1 \ldots I_N^1}_{p}, \quad (3.66)$$

$$I_N^{p+q} = I_N^p\, I_N^q = I_N^q\, I_N^p, \quad (3.67)$$

where p and q are positive integers. Moreover, matrices I_N^p commute also with matrices B_N^p.

3.5.2 Moving lower limit of integration

If we consider p-fold integration with moving lower limit,

$$y_p(t) = \int_t^b d\tau_p \int_{\tau_p}^b d\tau_{p-1} \ldots \int_{\tau_2}^b f(\tau_1) d\tau_1, \quad (3.68)$$

then its discrete analog is represented by the upper triangular strip matrix J_N^p with the generating function $\varphi_p(z) = h^p(1-z)^{-p}$:

$$J_N^p = h^p \begin{bmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \cdots & \cdots & \gamma_{N-1} & \gamma_N \\ 0 & \gamma_0 & \gamma_1 & \gamma_2 & \cdots & \cdots & \gamma_{N-1} \\ \cdots & \cdots & \ddots & \cdots & \cdots & \cdots & \cdots \\ \cdots & 0 & 0 & \gamma_0 & \gamma_1 & \gamma_2 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \ddots & \cdots & \cdots \\ 0 & \cdots & \cdots & 0 & 0 & \gamma_0 & \gamma_1 \\ 0 & 0 & \cdots & \cdots & 0 & 0 & \gamma_0 \end{bmatrix}, \tag{3.69}$$

The matrix J_N^p is inverse to the matrix F_N^p, corresponding to backward difference approximation of the p-th derivative:

$$F_N^p J_N^p = J_N^p F_N^p \longleftrightarrow \mathrm{trunc}_N\left(\beta_p(z)\,\varphi_p(z)\right) = 1 \longleftrightarrow E. \tag{3.70}$$

In view of (3.26) it follows from the properties of the generating functions $\varphi_p(z) = h^p(1-z)^{-p}$ that

$$J_N^2 = J_N^1\, J_N^1, \tag{3.71}$$

$$J_N^p = \underbrace{J_N^1\, J_N^1 \ldots J_N^1}_{p}, \tag{3.72}$$

$$J_N^{p+q} = J_N^p\, J_N^q = J_N^q\, J_N^p, \tag{3.73}$$

where p and q are positive integers. Moreover, matrices J_N^p commute also with matrices F_N^p.

It also should be noted that transposition of the matrix I_N^p, representing the integration with moving upper limit, gives the matrix J_N^p, which corresponds to integration with moving lower limit:

$$\left(I_N^p\right)^T = J_N^p, \qquad \left(J_N^p\right)^T = I_N^p. \tag{3.74}$$

3.6 Fractional differentiation

The triangular strip matrices can also be used for fractional derivatives. In this case, we arrive at lower (upper) triangular matrices, which have no zeros below (above) the main diagonal.

3.6.1 Left-sided fractional derivatives

Let us consider a function $f(t)$, defined in $[a, b]$, such that $f(t) \equiv 0$ for $t < a$. (Functions satisfying this condition are often called causal functions.) We assume that the function

$f(t)$ is good enough for considering its left-sided fractional derivative of real order α, $(n-1 \le \alpha < n)$,

$$ {}_aD_t^{\alpha} f(t) = \frac{1}{\Gamma(n-\alpha)} \left(\frac{d}{dt}\right)^n \int_a^t \frac{f(\tau)d\tau}{(t-\tau)^{\alpha-n+1}}, \qquad (a < t < b). \tag{3.75} $$

Let us take equidistant nodes with the step h: $t_k = kh$ $(k = 0, 1, \ldots, N)$, in the interval $[a, b]$, where $t_0 = a$ and $t_N = b$. Using the backward fractional difference approximation for the α-th derivative at the points t_k, $k = 0, 1, \ldots, N$, we have:

$$ {}_aD_{t_k}^{\alpha} f(t) \approx \frac{\nabla^{\alpha} f(t_k)}{h^{\alpha}} = h^{-\alpha} \sum_{j=0}^{k} (-1)^j \binom{\alpha}{j} f_{k-j}, \qquad k = 0, 1, \ldots, N. \tag{3.76} $$

All $N+1$ formulas (3.76) can be written simultaneously in the matrix form:

$$ \begin{bmatrix} h^{-\alpha}\nabla^{\alpha} f(t_0) \\ h^{-\alpha}\nabla^{\alpha} f(t_1) \\ h^{-\alpha}\nabla^{\alpha} f(t_2) \\ \vdots \\ h^{-\alpha}\nabla^{\alpha} f(t_{N-1}) \\ h^{-\alpha}\nabla^{\alpha} f(t_N) \end{bmatrix} = \frac{1}{h^{\alpha}} \begin{bmatrix} \omega_0^{(\alpha)} & 0 & 0 & 0 & \cdots & 0 \\ \omega_1^{(\alpha)} & \omega_0^{(\alpha)} & 0 & 0 & \cdots & 0 \\ \omega_2^{(\alpha)} & \omega_1^{(\alpha)} & \omega_0^{(\alpha)} & 0 & \cdots & 0 \\ \ddots & \ddots & \ddots & \ddots & \cdots & \cdots \\ \omega_{N-1}^{(\alpha)} & \ddots & \omega_2^{(\alpha)} & \omega_1^{(\alpha)} & \omega_0^{(\alpha)} & 0 \\ \omega_N^{(\alpha)} & \omega_{N-1}^{(\alpha)} & \ddots & \omega_2^{(\alpha)} & \omega_1^{(\alpha)} & \omega_0^{(\alpha)} \end{bmatrix} \begin{bmatrix} f_0 \\ f_1 \\ f_2 \\ \vdots \\ f_{N-1} \\ f_N \end{bmatrix} \tag{3.77} $$

$$ \omega_j^{(\alpha)} = (-1)^j \binom{\alpha}{j}, \qquad j = 0, 1, \ldots, N. \tag{3.78} $$

In the formula (3.77) the column vector of function values f_k $(k = 0, \ldots, N)$ is multiplied by the matrix

$$ B_N^{\alpha} = \frac{1}{h^{\alpha}} \begin{bmatrix} \omega_0^{(\alpha)} & 0 & 0 & 0 & \cdots & 0 \\ \omega_1^{(\alpha)} & \omega_0^{(\alpha)} & 0 & 0 & \cdots & 0 \\ \omega_2^{(\alpha)} & \omega_1^{(\alpha)} & \omega_0^{(\alpha)} & 0 & \cdots & 0 \\ \ddots & \ddots & \ddots & \ddots & \cdots & \cdots \\ \omega_{N-1}^{(\alpha)} & \ddots & \omega_2^{(\alpha)} & \omega_1^{(\alpha)} & \omega_0^{(\alpha)} & 0 \\ \omega_N^{(\alpha)} & \omega_{N-1}^{(\alpha)} & \ddots & \omega_2^{(\alpha)} & \omega_1^{(\alpha)} & \omega_0^{(\alpha)} \end{bmatrix} \tag{3.79} $$

and the result is the column vector of approximated values of the fractional derivative ${}_aD_{t_k}^{\alpha} f(t)$, $k = 0, 1, \ldots, N$. We can look at the matrix B_N^{α} as at a discrete analog of left-sided fractional differentiation of order α.

The generating function for the matrix B_N^{α} is

$$ \beta_{\alpha}(z) = h^{-\alpha}(1-z)^{\alpha}. \tag{3.80} $$

Since for lower triangular matrices B_N^α and B_N^β we always have

$$B_N^\alpha B_N^\beta = B_N^\beta B_N^\alpha = B_N^{\alpha+\beta},$$

we can consider such matrices as discrete analogs of the corresponding left-sided fractional derivatives ${}_aD_t^\alpha$ and ${}_aD_t^\beta$, where $n-1 \le \alpha < n$ and $m-1 \le \alpha < m$, only if

$${}_aD_t^\alpha\Big({}_aD_t^\beta f(t)\Big) = {}_aD_t^\beta\Big({}_aD_t^\alpha f(t)\Big) = {}_aD_t^{\alpha+\beta} f(t),$$

which holds if

$$f^{(k)}(a) = 0, \quad k = 1, 2, \ldots, r-1, \tag{3.81}$$

where $r = \max\{n, m\}$.

This means that if left-sided fractional derivatives of a function $f(t)$ of orders less than some integer r are considered, than they can all be replaced with their corresponding discrete analogs, if the function $f(t)$ satisfies the conditions (3.81).

3.6.2 Right-sided fractional derivatives

Let us consider a function $f(t)$, defined in $[a, b]$, such that $f(t) \equiv 0$ for $t > b$. We assume that the function $f(t)$ is good enough for considering its right-sided fractional derivative of real order α, $(n-1 \le \alpha < n)$,

$${}_tD_b^\alpha f(t) = \frac{(-1)^n}{\Gamma(n-\alpha)}\left(\frac{d}{dt}\right)^n \int_t^b \frac{f(\tau)d\tau}{(\tau - t)^{\alpha-n+1}}, \qquad (a < t < b). \tag{3.82}$$

Similarly to the previous section, we can obtain the discrete analog of the right-sided fractional differentiation on the net of equidistant nodes with the step h: $t_k = kh$ $(k = 0, 1, \ldots, N)$, in the interval $[a, b]$, where $t_0 = a$ and $t_N = b$, which is represented by the matrix

$$F_N^\alpha = \frac{1}{h^\alpha}\begin{bmatrix} \omega_0^{(\alpha)} & \omega_1^{(\alpha)} & \ddots & \ddots & \omega_{N-1}^{(\alpha)} & \omega_N^{(\alpha)} \\ 0 & \omega_0^{(\alpha)} & \omega_1^{(\alpha)} & \ddots & \ddots & \omega_{N-1}^{(\alpha)} \\ 0 & 0 & \omega_0^{(\alpha)} & \omega_1^{(\alpha)} & \ddots & \ddots \\ \cdots & \cdots & \cdots & \ddots & \ddots & \ddots \\ 0 & \cdots & 0 & 0 & \omega_0^{(\alpha)} & \omega_1^{(\alpha)} \\ 0 & 0 & \cdots & 0 & 0 & \omega_0^{(\alpha)} \end{bmatrix} \tag{3.83}$$

The generating function for the matrix F_N^α is the same as for the matrix B_N^α: $\beta_\alpha(z) = h^{-\alpha}(1-z)^\alpha$.

It should also be mentioned that transposition of the matrix B_N^α, corresponding to left-sided fractional differentiation, gives the matrix F_N^p, which corresponds to right-sided

differentiation:

$$\left(B_N^\alpha\right)^T = F_N^\alpha, \qquad \left(F_N^\alpha\right)^T = B_N^\alpha. \tag{3.84}$$

Similarly to the previous section, if right-sided fractional derivatives of a function $f(t)$ of orders less than some integer r are considered, than they can all be replaced with their corresponding discrete analogs, if the function $f(t)$ satisfies the conditions

$$f^{(k)}(b) = 0, \quad k = 1, 2, \ldots, r-1. \tag{3.85}$$

3.6.3 Sequential fractional derivatives

For left-sided sequential fractional derivatives, in which all derivatives in the sequence can be arbitrarily interchanged,

$$ {}_a\mathcal{D}_t^{\vec{\alpha}} f(t) = {}_aD_t^{\alpha_1}\ {}_aD_t^{\alpha_2}\ \ldots\ {}_aD_t^{\alpha_n} f(t), \tag{3.86}$$

$$\vec{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_n),$$

(and the same equidistant nodes as above) the discrete analog $\mathcal{B}_N^{\vec{\alpha}}$ has the form of the product of matrices $B_N^{\alpha_k}$, corresponding to operators ${}_aD_t^{\alpha_k}$, $k = 1, 2, \ldots, n$:

$$\mathcal{B}_N^{\vec{\alpha}} = B_N^{\alpha_1} B_N^{\alpha_2} \ldots B_N^{\alpha_n} = \prod_{k=1}^{n} B_N^{\alpha_k}. \tag{3.87}$$

Similarly, for right-sided fractional derivatives, in which all derivatives in the sequence can be arbitrarily interchanged,

$$ {}_t\mathcal{D}_b^{\vec{\alpha}} f(t) = {}_tD_b^{\alpha_1}\ {}_tD_b^{\alpha_2}\ \ldots\ {}_tD_b^{\alpha_n} f(t), \tag{3.88}$$

the discrete analog $\mathcal{F}_N^{\vec{\alpha}}$ is

$$\mathcal{F}_N^{\vec{\alpha}} = F_N^{\alpha_1} F_N^{\alpha_2} \ldots F_N^{\alpha_n} = \prod_{k=1}^{n} F_N^{\alpha_k}. \tag{3.89}$$

3.7 Fractional integration

Discrete analogs of left- and right-sided fractional integrals can be obtained by inversion of the discrete analogs of the corresponding fractional derivatives.

3.7.1 Left-sided fractional integration

To obtain the matrix I_N^α, corresponding to the discrete analog of the left-sided fractional integration ($\alpha > 0$),

$$ {}_aD_t^{-\alpha} f(t) = \frac{1}{\Gamma(\alpha)} \int_a^t (t-\tau)^{\alpha-1} f(\tau) d\tau, \qquad (a < t < b), \tag{3.90}$$

we simply invert the matrix B_N^α, corresponding to the left-sided fractional differentiation:

$$ I_N^\alpha = \left(B_N^\alpha \right)^{-1}. \tag{3.91}$$

If $\varphi_\alpha(z)$ denotes the generation function for I_N^α and $\beta_\alpha(z)$ is the generating function for B_N^α, then, taking into account the rule (3.28), we can write:

$$ I_N^\alpha \longleftrightarrow \varphi_N(z) = \mathrm{trunc}_N \left(\beta_\alpha^{-1}(z) \right) = \mathrm{trunc}_N \left(h^\alpha (1-z)^{-\alpha} \right). $$

Therefore, the matrix I_N^α has the following form:

$$ I_N^\alpha = h^\alpha \begin{bmatrix} \omega_0^{(-\alpha)} & 0 & 0 & 0 & \cdots & 0 \\ \omega_1^{(-\alpha)} & \omega_0^{(-\alpha)} & 0 & 0 & \cdots & 0 \\ \omega_2^{(-\alpha)} & \omega_1^{(-\alpha)} & \omega_0^{(-\alpha)} & 0 & \cdots & 0 \\ \ddots & \ddots & \ddots & \ddots & \cdots & \cdots \\ \omega_{N-1}^{(-\alpha)} & \ddots & \omega_2^{(-\alpha)} & \omega_1^{(-\alpha)} & \omega_0^{(-\alpha)} & 0 \\ \omega_N^{(-\alpha)} & \omega_{N-1}^{(-\alpha)} & \ddots & \omega_2^{(-\alpha)} & \omega_1^{(-\alpha)} & \omega_0^{(-\alpha)} \end{bmatrix} \tag{3.92}$$

3.7.2 Right-sided fractional integration

Similarly, inversion of the matrix F_N^α, corresponding to the right-sided fractional differentiation, gives the matrix J_N^α,

$$ J_N^\alpha = h^\alpha \begin{bmatrix} \omega_0^{(-\alpha)} & \omega_1^{(-\alpha)} & \ddots & \ddots & \omega_{N-1}^{(-\alpha)} & \omega_N^{(-\alpha)} \\ 0 & \omega_0^{(-\alpha)} & \omega_1^{(-\alpha)} & \ddots & \ddots & \omega_{N-1}^{(-\alpha)} \\ 0 & 0 & \omega_0^{(-\alpha)} & \omega_1^{(-\alpha)} & \ddots & \ddots \\ \cdots & \cdots & \cdots & \ddots & \ddots & \ddots \\ 0 & \cdots & 0 & 0 & \omega_0^{(-\alpha)} & \omega_1^{(-\alpha)} \\ 0 & 0 & \cdots & 0 & 0 & \omega_0^{(-\alpha)} \end{bmatrix} \tag{3.93}$$

which is the discrete analog of the right-sided fractional integration ($\alpha > 0$):

$$ {}_tD_b^{-\alpha} f(t) = \frac{1}{\Gamma(\alpha)} \int_t^b (\tau - t)^{\alpha-1} f(\tau) d\tau, \qquad (a < t < b). \tag{3.94}$$

3.7.3 Feller and Riesz potentials on a finite interval

Let us consider the Riesz potential $R^\alpha f(t)$ and the modified Riesz potential $M^\alpha f(t)$ on a finite interval [13, Chap. 3]:

$$R^\alpha f(t) = \frac{1}{2\Gamma(\alpha)\cos(\alpha\pi/2)} \int_a^b \frac{\varphi(\tau)d\tau}{|t-\tau|^{1-\alpha}}, \qquad (a < t < b), \tag{3.95}$$

$$M^\alpha f(t) = \frac{1}{2\Gamma(\alpha)\sin(\alpha\pi/2)} \int_a^b \frac{\operatorname{sign}(t-\tau)\varphi(\tau)d\tau}{|t-\tau|^{1-\alpha}}, \qquad (a < t < b). \tag{3.96}$$

Obviously, both these operators are linear combinations of the left-sided and right-sided fractional integrals:

$$R^\alpha f(t) = \frac{1}{2\cos(\alpha\pi/2)} \left({}_aD_t^{-\alpha} f(t) + {}_tD_b^{-\alpha} f(t) \right), \tag{3.97}$$

$$M^\alpha f(t) = \frac{1}{2\sin(\alpha\pi/2)} \left({}_aD_t^{-\alpha} f(t) - {}_tD_b^{-\alpha} f(t) \right). \tag{3.98}$$

The matrices $\mathcal{R}_N^\alpha$ and $\mathcal{M}_N^\alpha$ of their discrete analogs are linear combinations of the matrix I_N^α, corresponding to the left-sided fractional integral, and the matrix J_N^α, corresponding to the right-sided fractional integral:

$$\mathcal{R}_N^\alpha = \frac{1}{2\cos(\alpha\pi/2)} \left(I_N^\alpha + J_N^\alpha \right), \tag{3.99}$$

$$\mathcal{M}_N^\alpha = \frac{1}{2\sin(\alpha\pi/2)} \left(I_N^\alpha - J_N^\alpha \right). \tag{3.100}$$

The Feller potential operator $\Phi^\alpha f(t)$ is also a linear combination of left- and right-sided fractional integrals, but with general constant coefficients u, v [13, Chap. 3]:

$$\Phi^\alpha f(t) = u\, {}_aD_t^{-\alpha} + v\, {}_tD_b^{-\alpha} f(t), \tag{3.101}$$

and the matrix of its discrete analog is

$$\Phi_N^\alpha = uI_N^\alpha + vJ_N^\alpha. \tag{3.102}$$

Numerical inversion of the Riesz and Feller potential operators (3.95), (3.96), and (3.101), reduces to inversion of the corresponding square matrices $\mathcal{R}_N^\alpha$, $\mathcal{M}_N^\alpha$, and Φ_N^α.

Example 1. Let us consider the fractional integral equation with the Riesz kernel:

$$\frac{1}{\Gamma(1-\alpha)} \int_{-1}^{1} \frac{y(\tau)\,d\tau}{|t-\tau|^\alpha} = 1, \quad (-1 < t < 1), \tag{3.103}$$

which has the solution [12, eq. 6.116]

$$y(t) = \pi^{-1}\Gamma(1-\alpha)\cos\left(\frac{\alpha\pi}{2}\right)(1-t^2)^{(\alpha-1)/2}. \tag{3.104}$$

Writing the equation (3.103) in the form

$${}_{-1}D_t^{-(1-\alpha)}y(t) + {}_tD_1^{-(1-\alpha)}y(t) = 1,$$

and replacing the fractional derivatives with their discrete analogs, we obtain the system of linear algebraic equations

$$\left(B_N^{-(1-\alpha)} + F_N^{-(1-\alpha)}\right)Y_N = F_N, \tag{3.105}$$

where $F_N = \Big(1, 1, \ldots, 1\Big)^T$, for determination of $Y_N = \Big(y(t_0), y(t_1), \ldots, y(t_N)\Big)^T$, which represents the approximate solution of the equations (3.103).

The numerical solution of the equation (3.103) for $\alpha = 0.8$ is shown in Fig. 3.1. In this figure, and also in all subsequent figures, only a subset of the points of the obtained numerical solution is shown; otherwise it would be difficult to depict the analytical solution and the numerical one, which are very close each other.

3.8 Numerical solution of fractional differential equations

The use of triangular strip matrices significantly simplifies numerical solution of fractional differential equations. Instead of writing unwieldy recurrence relationships for determination of the values of the unknown function in equidistant discretization nodes, one can immediately write a system of algebraic equations for those values.

For convenience, let us introduce a certain type of matrices, which are obtained from the $N \times N$ unit matrix E by keeping only some of its rows and omitting all other rows: S_1 is obtained by omitting only the first row of E; S_2 is obtained by omitting only the second row; $S_{1,2}$ is obtained by omitting only the first and the second row of E; and, in general, $S_{r_1,r_2,\ldots,r_k}$ is obtained by omitting the rows with the numbers $r_1, r_2, \ldots, r_k$.

If A is a square $N \times N$ matrix, then the product $S_{r_1,r_2,\ldots,r_k}A$ contains only rows of A with the numbers different from $r_1, r_2, \ldots, r_k$. Similarly, the product $AS^T_{r_1,r_2,\ldots,r_k}$ contains only columns of A with the numbers different from $r_1, r_2, \ldots, r_k$. Because of this property, the matrix $S_{r_1,r_2,\ldots,r_k}$ is called an *eliminator*. In case of infinite matrices, similar matrices appeared in [4].

The following simple example illustrates the main property of eliminators:

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}; \qquad S_1 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}; \qquad S_1\,A = \begin{bmatrix} a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix};$$

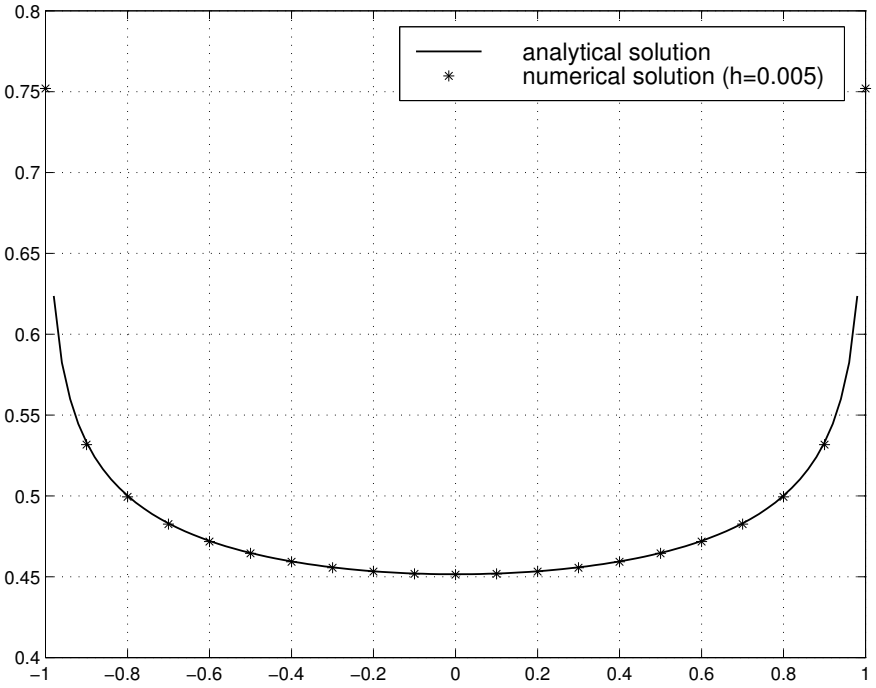


Figure 3.1: Solution of equation $\frac{1}{\Gamma(1-\alpha)} \int_{-1}^{1} |t-\tau|^{-\alpha} y(\tau)\, d\tau = 1$.

$$A\,S_1^T = \begin{bmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix}; \qquad S_1\,A\,S_1^T = \begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix}.$$

Considering $(N+1)\times(N+1)$ lower triangular matrices L_N (3.1) and upper triangular matrices U_N (3.2), and numbering rows and columns from 0 to N, we have the following useful relationships:

$$S_0 \left\{ \begin{array}{c} L_N \\ U_N \end{array} \right\} S_0^T = \left\{ \begin{array}{c} L_{N-1} \\ U_{N-1} \end{array} \right\}, \tag{3.106}$$

$$S_N \left\{ \begin{array}{c} L_N \\ U_N \end{array} \right\} S_N^T = \left\{ \begin{array}{c} L_{N-1} \\ U_{N-1} \end{array} \right\}, \tag{3.107}$$

$$S_{0,1,\ldots,k} \left\{ \begin{array}{c} L_N \\ U_N \end{array} \right\} S_{0,1,\ldots,k}^T = \left\{ \begin{array}{c} L_{N-k-1} \\ U_{N-k-1} \end{array} \right\}, \tag{3.108}$$

$$S_{N-k,N-k+1,\ldots,N} \left\{ \begin{array}{c} L_N \\ U_N \end{array} \right\} S_{N-k,N-k+1,\ldots,N}^T = \left\{ \begin{array}{c} L_{N-k-1} \\ U_{N-k-1} \end{array} \right\}. \tag{3.109}$$

In other words, simultaneous multiplication of a triangular strip matrix by the eliminator $S_{0,1,\ldots,k}$ (or by $S_{N-k,N-k+1,\ldots,N}$) on the left and $S_{0,1,\ldots,k}^T$ (respectively, by $S_{N-k,N-k+1,\ldots,N}^T$) on the right preserves the type and the structure of the triangular strip matrix, and only reduces its size by $k+1$ rows and $k+1$ columns.

3.8.1 Initial value problems for FDEs

The general procedure of numerical solution of fractional differential equations consists of two steps.

First, initial conditions are used to reduce a given initial-value problem to a problem with zero initial conditions. At this stage, instead of a given equation a modified equation, incorporating initial values, is obtained.

Then the system of algebraic equations is obtained by replacing all derivatives (of fractional and integer orders) in the obtained modified equation by the corresponding matrices (B_N^α for left-sided derivatices, F_N^α for right-sided derivatives) for their discrete analogs.

We will consider an m-term linear fractional differential equation with non-constant coefficients of the following form:

$$\sum_{k=1}^{m} p_k(t) D^{\alpha_k} y(t) = f(t), \tag{3.110}$$

$$0 \le \alpha_1 < \alpha_2 < \ldots < \alpha_m, \qquad n-1 < \alpha_m < n,$$

where D^{α_k} denotes either Riemann-Liouville or Caputo left-sided fractional derivative of order α_k.

Let us denote

$$P_N^{(k)} = \operatorname{diag}\Big(p_k(t_0), p_k(t_1), \ldots, p_k(t_N)\Big) = \begin{bmatrix} p_k(t_0) & 0 & \ldots & 0 \\ 0 & p_k(t_1) & 0 & \ldots \\ 0 & \ldots & \ddots & 0 \\ 0 & \ldots & 0 & p_k(t_N) \end{bmatrix}, \tag{3.111}$$

$$Y_N = \Big(y(t_0), y(t_1), \ldots, y(t_N)\Big)^T, \qquad F_N = \Big(f(t_0), f(t_1), \ldots, f(t_N)\Big)^T. \tag{3.112}$$

Using these notations and taking into account that the discrete analog of the left-sided fractional derivative D^{α_k} is $B_N^{\alpha_k}$, we can write a discrete analog of the fractional differential equation (3.110):

$$\sum_{k=1}^{m} P_N^{(k)} B_N^{\alpha_k} Y_N = F_N. \tag{3.113}$$

3.8.2 Zero initial conditions

If $n-1 < \alpha_m < n$, then the Riemann-Liouville and the Caputo formulations of the equation (3.110) are equivalent under the assumption of zero initial values of the function $y(t)$ and its $(n-1)$ derivatives [12]:

$$y^{(k)}(t_0) = 0, \qquad k = 0, 1, \ldots, n-1. \tag{3.114}$$

Approximating the derivatives in the initial conditions (3.114) by backward differences, we immediately obtain:

$$y(t_0) = y(t_1) = \ldots = y(t_{n-1}) = 0. \tag{3.115}$$

The linear algebraic system for determination of $y_n, \ldots, y_N$ is obtained from the system (3.113) by omitting its first n rows and substituting the zero starting values (3.115) into the remaining equations. This can be *symbolically* written with the help of eliminator:

$$\left\{ S_{0,1,\ldots,n-1} \left\{ \sum_{k=1}^{m} P_N^{(k)} B_N^{\alpha_k} \right\} S_{0,1,\ldots,n-1}^T \right\} \{ S_{0,1,\ldots,n-1} Y_N \} = S_{0,1,\ldots,n-1} F_N. \tag{3.116}$$

Solution of the linear algebraic system (3.116) along with the starting values (3.115) gives the numerical solution of the fractional differential equation (3.110) under zero initial conditions (3.114).

If the coefficients $p_k(t)$ are constant, i.e. $p_k(t) \equiv p_k$, then the system (3.116) takes on the simplest form:

$$\sum_{k=1}^{m} p_k B_{N-n}^{\alpha_k} \{S_{0,1,\ldots,n-1} Y_N\} = S_{0,1,\ldots,n-1} F_N. \tag{3.117}$$

EXAMPLE 2. Let us consider the following two-term fractional differential equation under zero initial conditions:

$$y^{(\alpha)}(t) + y(t) = 1, \tag{3.118}$$

$$y(0) = 0, \quad y'(0) = 0, \tag{3.119}$$

which has the analytical solution

$$y(t) = t^{\alpha} E_{\alpha,\alpha+1}(-t^{\alpha}). \tag{3.120}$$

The numerical solution of the problem (3.118)–(3.119) can be found from the system (3.117), where we have $m = 2$, $\alpha_1 = \alpha$, $\alpha_2 = 0$, $n = 2$, $p_1 = p_2 = 1$, $B_{N-n}^{\alpha_1} = B_{N-2}^{\alpha}$, $B_{N-n}^{\alpha_2} = E_{N-2}$, $F_N = (\underbrace{1, 1, \ldots, 1}_{N})^T$. For these vales, the system of algebraic equations for determining y_k, $k = 2, 3, \ldots, N$ takes on the form:

$$\{B_{N-2}^{\alpha} + E_{N-2}\} \{S_{0,1} Y_N\} = S_{0,1} F_N. \tag{3.121}$$

It should be also added that from the initial conditions we have $y_0 = y_1 = 0$.

The numerical solution of the problem (3.118)–(3.119) for $\alpha = 1.8$ is shown in Fig. 3.2.

3.8.3 Initial conditions in terms of integer-order derivatives

If fractional derivatives in the equation (3.110), where $n - 1 < \alpha_m < n$, are Caputo derivatives, then the initial conditions are expressed in terms of classical integer-order derivatives and can be non-zero:

$$y^{(k)}(t_0) = c_k, \qquad k = 0, 1, \ldots, n-1. \tag{3.122}$$

The solution of the initial-value problem problem (3.110)–(3.122) can be written in the form

$$y(t) = y_*(t) + z(t), \tag{3.123}$$

where $y_*(t)$ is some known function, satisfying the conditions $y^{(k)}(t_0) = c_k$, $k = 0, 1, \ldots, n-1$, and $z(t)$ is a new unknown function.

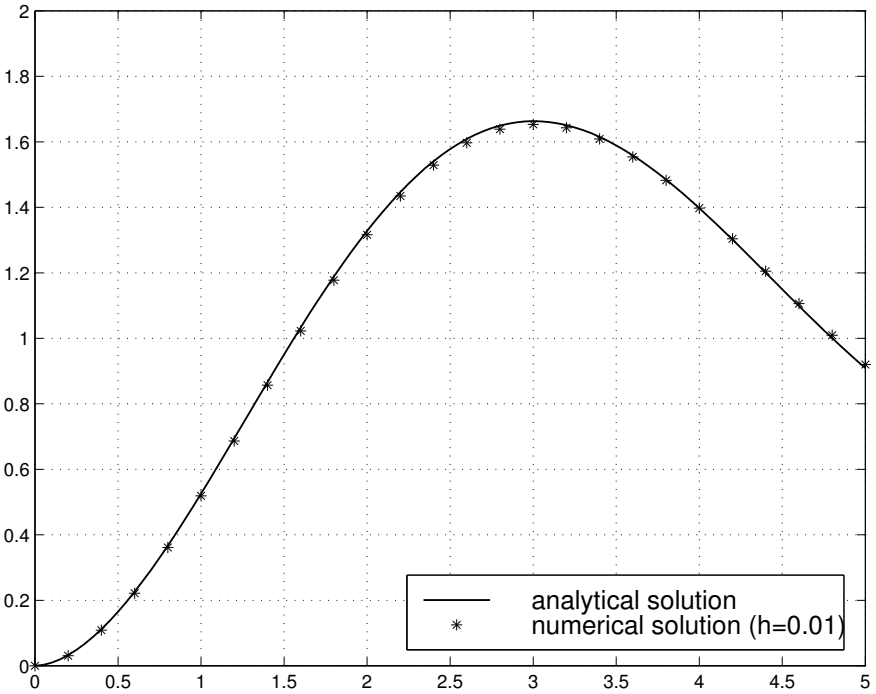


Figure 3.2: Solution of the problem $y^{(1.8)}(t) + y(t) = 1, y(0) = 0, y'(0) = 0$

Substituting (3.123) into the equation (3.110) and the initial conditions (3.122), we obtain for the function $z(t)$ an initial-value problem with zero initial conditions, which can be solved as described in Section 3.8.2.

EXAMPLE 3. Let us consider the following two-term fractional differential equation under non-zero initial conditions:

$$y^{(\alpha)}(t) + y(t) = 1, \tag{3.124}$$

$$y(0) = c_0, \quad y'(0) = c_1. \tag{3.125}$$

The analytical solution, obtained with the help of the Laplace transform of the Caputo fractional derivatives [12], is given by expression

$$y(t) = c_0 E_{\alpha,1}(-t^\alpha) + c_1 t E_{\alpha,2}(-t^\alpha) + t^\alpha E_{\alpha,\alpha+1}(-t^\alpha). \tag{3.126}$$

To obtain numerical solution, we have first to transform the problem (3.124)–(3.125) to the problem with zero initial conditions. For this, let us introduce an auxiliary function $z(t)$, such that

$$y(t) = c_0 + c_1 t + z(t).$$

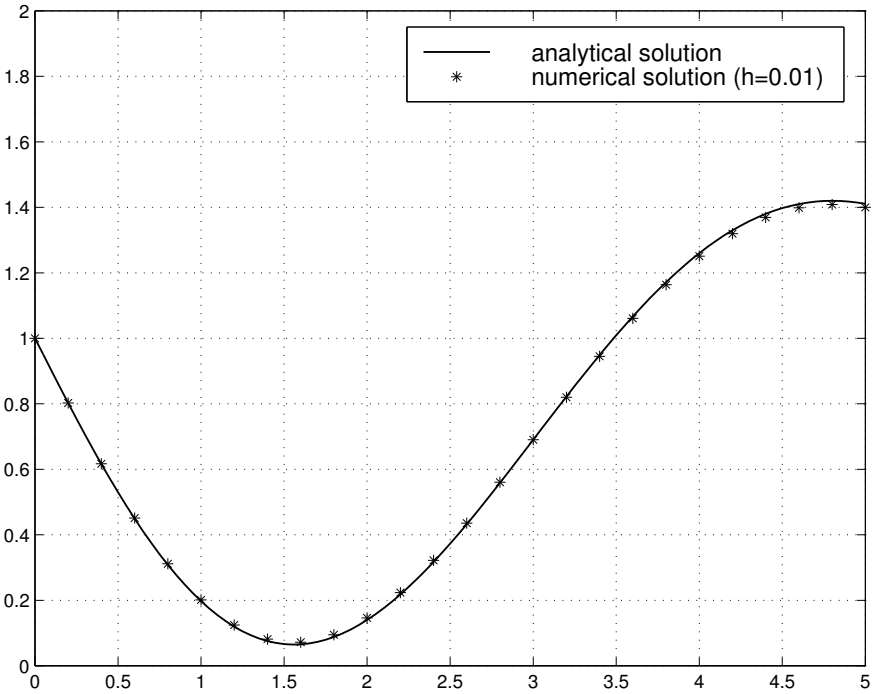


Figure 3.3: Solution of the problem $y^{(1.8)}(t) + y(t) = 1, y(0) = 1, y'(0) = -1$

Substituting this expression into the equation (3.124) and in the initial conditions (3.125), we obtain the problem for finding $z(t)$:

$$z^{(\alpha)}(t) + z(t) = 1 - c_0 - c_1 t, \tag{3.127}$$

$$z(0) = 0, \quad z'(0) = 0. \tag{3.128}$$

The numerical solution of this problem can be found as described in Section 3.8.2, and the numerical solution $y(t)$ of the problem (3.124)–(3.125) is obtained using the relationship $y(t) = c_0 + c_1 t + z(t)$.

The numerical solution of the problem (3.124)–(3.125) for $\alpha = 1.8$, $c_0 = 1$, $c_1 = -1$ is shown in Fig. 3.3.

3.8.4 Initial conditions in terms of Riemann-Liouville fractional derivatives

Initial value problems for fractional differential equations with non-zero initial conditions in terms of Riemann-Liouville derivatives, namely

$${}_aD_t^{\alpha-k-1} y(t)\Big|_{t\to a} = c_k, \qquad k = 0, 1, \ldots, n-1, \tag{3.129}$$

can be also transformed to initial-value problems with zero initial condition. Such a transformation allows us to circumvent the difficulty consisting in the fact that there is still no known approximation for such initial conditions.

EXAMPLE 4. Let us consider the following two-term fractional differential equation under non-zero initial conditions:

$$y^{(\alpha)}(t) + y(t) = 1, \tag{3.130}$$

$$y^{(\alpha-1)}(0) = c_0, \quad y^{(\alpha-2)}(0) = c_1. \tag{3.131}$$

The analytical solution, obtained with the help of the Laplace transform of the Riemann–Liouville fractional derivative [12], is given by expression

$$y(t) = c_0 t^{\alpha-1} E_{\alpha,\alpha}(-t^\alpha) + c_1 t^{\alpha-2} E_{\alpha,\alpha-1}(-t^\alpha) + t^\alpha E_{\alpha,\alpha+1}(-t^\alpha). \tag{3.132}$$

To obtain numerical solution, we have first to transform the problem (3.130)–(3.131) to the problem with zero initial conditions. For this, let us introduce an auxiliary function $z(t)$, such that

$$y(t) = c_0 t^{\alpha-1} + c_1 t^{\alpha-2} + z(t).$$

Substituting this expression into the equation (3.130) and into the initial conditions (3.131), we obtain the problem for finding $z(t)$:

$$z^{(\alpha)}(t) + z(t) = 1 - c_0 t^{\alpha-1} - c_1 t^{\alpha-2}, \tag{3.133}$$

$$z(0) = 0, \quad z'(0) = 0. \tag{3.134}$$

The numerical solution of this problem can be found as described in Section 3.8.2, and the numerical solution $y(t)$ of the problem (3.130)–(3.131) is obtained using the relationship $y(t) = c_0 t^{\alpha-1} + c_1 t^{\alpha-2} + z(t)$.

The numerical solution of the problem (3.130)–(3.131) for $\alpha = 1.8$, $c_0 = 1$, $c_1 = -1$ is shown in Fig. 3.4.

3.8.5 Nonlinear FDEs

Triangular strip matrices can be useful also for solving fractional differential equations of a general form. Let us write, for example, an equation with left-sided fractional derivatives $y^{(\alpha_i)}(t) = {}_aD_t^{\alpha_i} y(t)$:

$$y^{(\alpha_1)}(t) = f(t, y^{(\alpha_2)}(t), y^{(\alpha_3)}(t), \ldots, y^{(\alpha_k)}(t)), \tag{3.135}$$

$$(0 < \alpha_1 < \alpha_2 < \ldots < \alpha_k \leq n.)$$

assuming that the initial conditions are already transformed to zero initial conditions.

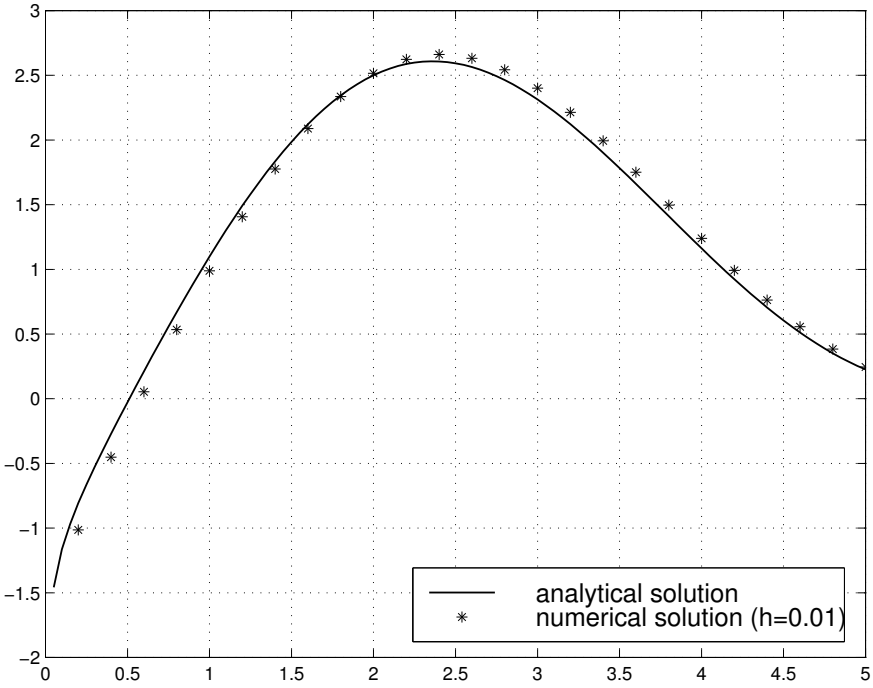


Figure 3.4: Solution of the problem $y^{(1.8)}(t) + y(t) = 1; y^{(0.8)}(0) = 1; y^{(-0.2)}(0) = -1$.

Replacing all fractional derivatives in the equation (3.135) with their discrete analogs and utilizing zero initial conditions, we obtain a nonlinear algebraic system

$$B_N^{\alpha_1} Y_N = f(Et_N, B_N^{\alpha_2} Y_N, B_N^{\alpha_3} Y_N, \ldots, B_N^{\alpha_k} Y_N), \tag{3.136}$$

$$y_j = 0, \quad j = 1, 2, \ldots, n-1,$$

where $Y_N = (y_0, y_1, \ldots, y_N)^T$, $t_N = (t_0, t_1, \ldots, t_N)^T$, $y_j = y(t_j)$, $t_j = jh$, $(j = 0, 1, \ldots, N)$, and E is $(N+1) \times (N+1)$ unit matrix.

3.9 Chapter summary

The suggested approach, using the triangular strip matrices, provides:

- a uniform approach to discretization of derivatives of arbitrary real order, including classical integer-order derivatives, and various types of fractional derivatives, including left- and right-sided derivatives, and sequential fractional derivatives;
- a uniform approach to numerical solution of differential equations of integer order and of fractional order;
- a convenient language for discretization of differential equations of arbitrary real order;

- a method for numerical solution of initial value problems and boundary value problems for ordinary differential equations of arbitrary real order;
- a possible method for numerical solution of non-linear differential equations of arbitrary real order.

The triangular matrix approach can also be used for obtaining new quadrature formulas for fractional integrals. For this, any approximation of fractional derivatives should be written in the form of a triangular strip matrix, inversion of which gives the corresponding quadrature formula for fractional integrals.

Similarly, new approximations of fractional derivatives can be obtained by inverting the triangular strip matrices, corresponding to quadrature formulas for fractional integrals.

3.10 References

[1] B. V. Bulgakov. *Kolebaniya (Vibrations).* Gostekhizdat, Moscow, 1954.

[2] R. F. Cameron and S. McKee. High accuracy convergent product integration methods for the generalized Abel equation. *Journal of Integral Equations*, 7:103–125, 1984.

[3] R. F. Cameron and S. McKee. The analysis of product integration methods for Abel's equation using discrete fractional differentiation. *IMA Journal of Numerical Analysis*, 5:339–353, 1985.

[4] R. G. Cooke. *Infinite Matrices and Sequence Spaces.* Fizmatgiz, Moscow, 1960.

[5] B. P. Demidovich. *Lektsii po matematicheskoy teorii ustoychivosti (Lectures on the Mathematical Theory of Stability).* Nauka, Moscow, 1967.

[6] L. J. Derr. Triangular matrices with the isoclinal property. *Pacific Journal of Mathematics*, 37(1):41–43, 1971.

[7] P. P. B. Eggermont. A new analysis of the trapezoidal-discretization method for the numerical solution of Abel-type integral equations. *Journal of Integral Equations*, 3:317–332, 1981.

[8] F. R. Gantmakher. *Teoriya matrits (Theory of Matrices).* Nauka, Moscow, 1988.

[9] P.A.W.Holyhead and S.McKee. Stability and convergence of multistep methods for linear Volterra integral equations of the first kind. *SIAM J. Numer. Anal.*, 13(2):269–292, April 1976.

[10] P. A. W. Holyhead, S. McKee, and P. J. Taylor. Multistep methods for solving linear Volterra integral equations of the first kind. *SIAM J. Numer. Anal.*, 12(5):698–711, October 1975.

[11] S. McKee. Discretization methods and block isoclinal matrices. *IMA Journal of Numerical Analysis*, 3:467–491, 1983.

[12] I. Podlubny. *Fractional Differential Equations.* Academic Press, San Diego, 1999.

[13] S. G. Samko, A. A. Kilbas, and O. I. Marichev. *Integraly i proizvodnye drobnogo poryadka i nekotorye ich prilozheniya (Integrals and Derivatives of the Fractional Order and Some of Their Applications).* Nauka i Tekhnika, Minsk, 1987 (English translation: S. G. Samko, A. A. Kilbas, and O. I. Marichev, *Fractional Integrals and Derivatives*, Gordon and Breach, Amsterdam, 1993).

[14] D. A. Suprunenko and R. I. Tyshkevich. *Perestanovochnye matritsy.* Nauka i Tekhnika, Minsk, 1966 (English translation: D. A.Suprunenko and R. I. Tyshkevich, *Commutative Matrices*, Academic Press, New York, 1968).

Chapter 4

Matrix approach for partial fractional differential equations

4.1 Introduction

Recently, kinetic equations of the diffusion, diffusion–advection, and Fokker–Planck type with partial fractional derivatives were recognized as a useful approach for the description of transport dynamics in complex systems whose temporal evolution deviates from the standard laws, that is, from exponential Debye or Gaussian laws, and from fast decaying correlations. Examples include systems exhibiting Hamiltonian chaos, disordered medium, plasma and fluid turbulence, underground water pollution, dynamics of protein molecules, motions under the influence of optical tweezers, reactions in complex systems, and more (see reviews on fractional kinetics [6, 44, 56, 68, 45], the recent multi-author book [28], and references therein). These fractional equations are derived asymptotically from basic random walk models, the generalized master and Langevin equations. The advantage of the fractional models lies in the straightforward way of including external force terms and of calculating boundary value problems. Also, the consideration of transport in the phase space spanned by both position and velocity coordinates is possible within the fractional approach. However, because of complicated integro-differential structure of fractional kinetic equations the analytical solutions are presently known only in a very few relatively simple cases. Therefore, the development of numerical methods is of current importance.

Let us recall briefly how the kinetic equations with integer partial derivatives can be "fractionalized". There are two generic types of fractionalization, which can be explained by taking as an example the parabolic diffusion equation for the particles density $u(x,t)$ in a one-dimensional space,

$$\frac{\partial u}{\partial t} = \chi \frac{\partial^2 u}{\partial x^2}, \quad (\, t > 0, \;\; a < x < b \,) \tag{4.1}$$

where constant χ is diffusion coefficient. The first type of fractionalization leads to a *time fractional diffusion equation* by means of replacing the first order time derivative by afractional derivative of order α less than 1,

$$ {}_0^C D_t^\alpha u = \chi \frac{\partial^2 u}{\partial x^2}, \quad (\, t > 0, \;\; a < x < b \,) \tag{4.2}$$

Here, ${}_0^C D_t^\alpha$ is the Caputo fractional derivative [2], which is defined as

$$ {}_a^C D_x^\mu \phi(x) = \frac{1}{\Gamma(m-\mu)} \int_a^x \frac{\phi^{(m)}(\xi) d\xi}{(x-\xi)^{\mu-m+1}}, \qquad (m-1 < \mu \le m) \tag{4.3}$$

Taking $\alpha = 1$ in (4.2) gives the classical diffusion equation (4.1).

Other two forms of a time fractional diffusion equation that appears in the literature use the Riemann-Liouville fractional derivative instead of the Caputo one [44]. Although recently, in addition to a geometric and physical interpretation of fractional integration and fractional differentiation [52], a physical interpretation for the initial conditions in terms of the Riemann-Liouville fractional derivatives of the unknown function has been suggested [27], the use of Caputo derivative in physical problems is perhaps more convenient since it allows using initial conditions expressed in terms of values of the unknown function and its integer-order derivatives [50]. However, all three forms of "time-fractionalization" are equivalent if zero initial conditions are posed. In what follows we use the form with the Caputo derivative, equation (4.2), since some of the illustrating examples use non-zero initial conditions.

In the second type of fractionalization, the second order spatial derivative is replaced by the fractional derivative of the order β between 1 and 2, thus leading to spatial fractional diffusion equation,

$$\frac{\partial u}{\partial t} = \chi \frac{\partial^\beta u}{\partial |x|^\beta}, \quad (\, t > 0, \;\; a < x < b \,) \tag{4.4}$$

where $\partial^\beta / \partial |x|^\beta$ (we adopt here the notation introduced in [54]) is a partial (with respect to spatial variable) symmetric Riesz derivative, which is defined as a half-sum of the left- and right-sided Riemann-Liouville derivatives [50, 51]:

$$\frac{d^\beta \phi(x)}{d|x|^\beta} = D_R^\beta \phi(x) = \frac{1}{2}\Big({}_a D_x^\beta \phi(x) + {}_x D_b^\beta \phi(x) \Big), \tag{4.5}$$

where the left- and right-sided Riemann-Liouville derivatives are defined by

$$ {}_aD_x^\mu \phi(x) = \frac{1}{\Gamma(m-\mu)} \left(\frac{d}{dx}\right)^m \int_a^x \frac{\phi(\xi)d\xi}{(x-\xi)^{\mu-m+1}}, \quad (m-1 < \mu \leq m), \tag{4.6} $$

$$ {}_xD_b^\mu \phi(x) = \frac{1}{\Gamma(m-\mu)} \left(-\frac{d}{dx}\right)^m \int_x^b \frac{\phi(\xi)d\xi}{(\xi-x)^{\mu-m+1}}, \quad (m-1 < \mu \leq m), \tag{4.7} $$

For $\beta = 2$ the equation (4.4) becomes the classical diffusion equation (4.1).

Other forms of asymmetric space fractional generalizations use the left-side Riemann - Liouville derivative instead of the symmetric Riesz derivative [12, 40], or asymmetric derivative with different asymmetry factors at the left- and right-side derivatives [13, 11, 42]. In terms of random walk schemes, the symmetric derivative corresponds to a symmetric jump probability distribution of a diffusing particle, whereas any asymmetry in space derivative accounts for inherent force-free preferable direction of jumps which may occur, e.g., in heterogeneous porous media or magnetically confined fusion plasmas. In this chapter we restrict ourselves to symmetric case, equation (4.4).

Of course, there are different generalizations of time and space fractional diffusion equations, including: multidimensional fractional diffusion and kinetic equations [5, 17], both time and space fractional generalizations [36], different regular forces in space and time fractional Fokker-Planck equations [43, 4, 14, 57, 26, 64], variable transport coefficients [65], equations with fractional derivatives of distributed and variable orders [7, 8, 3, 37] etc. The realm of fractional kinetics is growing, and therefore it is desirable to have at hand a method for numerical solution which would be relatively simple and at the same time general enough to deal effectively with different forms of fractional kinetic equations. However, while different numerical tools for ordinary fractional equations exist and a basic framework of their numerical solution is already established, relatively few numerical methods exist to solve fractional equations with partial derivatives, and the development of effective numerical schemes is now on the agenda. We recall briefly the different approaches used in the literature.

The numerical methods differ essentially in the way in which normal and fractional derivatives are discretized. In [35] to solve diffusion-reaction equation with the left Riemann-Liouville derivative between 1 and 2, the L2 discretization method was used taken from [47], together with its modification, L2C (both L2 and L2C methods are based on numerical approximation of a fractional integral that appears in the definition of the Riemann-Liouville fractional derivative). It was shown that the former is the most accurate for orders larger than 1.5, whereas the latter is the most accurate for orders less than 1.5. For the first order time derivative, the explicit forward Euler formula and semi-implicit scheme were used.

Langlands and Henry [29] used L1 scheme from [47] to discretize the Riemann-Liouville fractional time derivative of order between 1 and 2.

Yuste [67] considered a Grünwald-Letnikov approximation for the Riemann-Liouville time derivative and used a weighted average for the second-order space derivative.

Scherer et al. [55] introduced very recently a modification of the Grünwald-Letnikov approximation for the case of the Caputo derivative of a function which is not zero in the starting point of the considered time interval, and applied that approximation for the numerical solution of fractional diffusion equations with the Caputo time derivative and non-zero initial conditions.

To solve the one-dimensional space fractional advection-dispersion equation with left-side Riemann-Liouville derivative and variable coefficients the shifted Grünwald-Letnikov approximation was proposed by Meerschaert and Tadjeran [40]. For two-sided space-fractional partial differential equations the shifted Grünwald-Letnikov formula was proposed and discussed in [41]. The fractional Crank-Nicholson method based on the shifted formula was elaborated, giving temporally and spatially second-order numerical estimates [61]. The generalizations of the shifted formula and of the fractional Crank-Nicholson method in the two-dimensional case were discussed in [39] and [60], respectively.

Another method to solve the space-fractional Fokker-Planck equation with constant coefficient on the fractional derivative term was pursued by Liu et al. [32]. They transform the partial differential equation into a system of ordinary differential equations, which is solved by a method of lines.

Ervin and Roop [15, 16] presented a theoretical framework for the Galerkin finite element approximation to the steady state fractional advection-diffusion equation, and extended this approach to multidimensional partial differential equations with constant coefficients on the fractional derivative terms.

Valko and Abate [62] solved the time-fractional diffusion equation on a semi-infinite domain by numerical inversion of the two-dimensional Laplace transform. To solve the time-fractional diffusion equation in a bounded domain, Lin and Xu [31] proposed the method based on a finite difference scheme in time and Legendre spectral method in space.

Liang and Chen [30] used a combination of symbolic computations and numerical inversion of the Laplace transform for solving a time-fractional diffusion-wave equation with the time derivative of order between 1 and 2.

We also mention that in order to approximate shifted Caputo time derivative appearing in hydrodynamic equations for heterogeneous porous media the modification of Yuan and Agrawal's method [66] was used to transform a fractional derivative into an infinite integral over auxiliary internal variables [34].

Another approach for the solution of fractional kinetic equations employs the methods

of Monte Carlo type (random walk based methods). A set of random walk schemes applied to fractional diffusion equations based on the Grünwald-Letnikov approximation was developed in the papers by Gorenflo, Mainardi and co-workers. They were applied to solve (i) symmetric space-fractional diffusion equation [20, 22]; (ii) asymmetric space-fractional diffusion equation in the Lèvy–Feller form [21]; (iii) time-fractional diffusion equation with Caputo derivative [25]; (iv) time-space fractional diffusion equation [24, 23]. Chechkin et al. [9] generalized the approach on a double-order time fractional diffusion equation. Gorenflo and Abdel-Rehim [19] proposed discrete approximations to time-fractional diffusion process with non-homogeneous drift towards the origin by generalization of Ehrenfest's urn model. The Lèvy–Feller diffusion-advection process with a constant drift was approximated by random walk and finite difference method by Liu et al. [33]. The random walk particle tracking approach to solve one-dimensional space-fractional diffusion-advection equation with space dependent coefficients was employed by Meerschaert and co-authors [65]. The method based on numerical solution of a coupled stochastic differential equations driven by Lèvy symmetric stable processes was proposed in [58] to solve a non-linear evolution problem involving the fractional Laplacian operator.

All aforementioned works indicate that numerical solution of partial fractional differential equations plays an important and increasing role in the applications of the methods and models of non-integer order.

In the present chapter we propose a general approach to the numerical solution of partial fractional differential equations, which is based on the matrix form representation of discretized fractional operators introduced in [51]. This approach unifies the numerical differentiation of arbitrary (including integer) order and the n-fold integration, using the so-called triangular matrices. Applied to numerical solution of differential equations, it also unifies the solution of integer- and fractional-order partial differential equations. The suggested approach leads to significant simplification of the numerical solution of partial differential equations, and it is general enough to deal with different types of partial fractional differential equations, even with delays.

4.2 The idea of the suggested method

The method that we suggest is based on triangular strip matrix approach [51] to discretization of operators of differentiation and integration of arbitrary real order.

In contrast with generally used numerical methods, where the solution is obtained step-by-step by moving from the previous time layer to the next one, let us consider the whole time interval of interest at once. This allows us to create a net of discretization nodes. In the simplest case of one spatial dimension this step gives a 2D net of nodes. An example of such discretization is shown in Fig. 4.1. The values of the unknown function in inner nodes (shaded area in Fig. 4.1) are to be found. The values at the boundaries

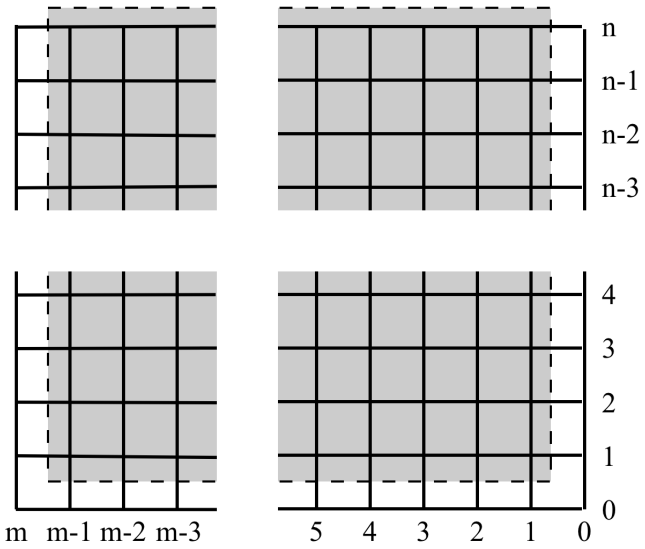


Figure 4.1: Nodes and their right-to-left, and bottom-to-top numbering.

are known: they are used later in constructing the system of algebraic equations.

The system of algebraic equations is obtained by approximating the equation in all inner nodes simultaneously (this gives the left-hand side of the resulting system of algebraic equations) and then utilizing the initial and boundary conditions (the values of which appear in the right-hand side of the resulting system).

The discretization nodes in Fig. 4.1 are numbered from right to left in each time level, and the time levels are numbered from bottom to top. We use such numbering in this article for the clarity of presentation of our approach, although standard numberings work equally well.

In the following sections we recall the basic tools that are necessary for the method: the triangular strip matrices, the Kronecker product, the eliminators, and the shifters. Then we show how they are used for approximating partial derivatives of arbitrary real order and the equation, and how the resulting system of algebraic equations appears.

4.3 Triangular strip matrices

In this chapter we use matrices of a specific structure, which are called *triangular strip matrices* [51, 59], and which have been also mentioned in [1, 18]. We will need lower

triangular strip matrices,

$$L_N = \begin{bmatrix} \omega_0 & 0 & 0 & 0 & \cdots & 0 \\ \omega_1 & \omega_0 & 0 & 0 & \cdots & 0 \\ \omega_2 & \omega_1 & \omega_0 & 0 & \cdots & 0 \\ \ddots & \ddots & \ddots & \ddots & \ldots & \ldots \\ \omega_{N-1} & \ddots & \omega_2 & \omega_1 & \omega_0 & 0 \\ \omega_N & \omega_{N-1} & \ddots & \omega_2 & \omega_1 & \omega_0 \end{bmatrix}, \tag{4.8}$$

and upper triangular strip matrices,

$$U_N = \begin{bmatrix} \omega_0 & \omega_1 & \omega_2 & \ddots & \omega_{N-1} & \omega_N \\ 0 & \omega_0 & \omega_1 & \ddots & \ddots & \omega_{N-1} \\ 0 & 0 & \omega_0 & \ddots & \omega_2 & \ddots \\ 0 & 0 & 0 & \ddots & \omega_1 & \omega_2 \\ \cdots & \cdots & \cdots & \cdots & \omega_0 & \omega_1 \\ 0 & 0 & 0 & \cdots & 0 & \omega_0 \end{bmatrix}, \tag{4.9}$$

A lower (upper) triangular strip matrix is completely described by its first column (row). Therefore, if we define the truncation operation, $\mathrm{trunc}_N(\cdot)$, which truncates (in a general case) the power series $\varrho(z)$,

$$\varrho(z) = \sum_{k=0}^{\infty} \omega_k z^k \tag{4.10}$$

to the polynomial $\varrho_N(z)$,

$$\mathrm{trunc}_N(\varrho(z)) \stackrel{\text{def}}{=} \sum_{k=0}^{N} \omega_k z^k = \varrho_N(z), \tag{4.11}$$

then we can consider the function $\varrho(z)$ as a generating series for the set of lower (or upper) triangular matrices L_N (or U_N), $N = 1, 2, \ldots$

It was shown in [51] that operations with triangular strip matrices, such as addition, subtraction, multiplication, and inversion, can be expressed in the form of operations with their generating series (4.10).

Among properties of triangular strip matrices it should be noticed that if matrices C and D are both lower (upper) triangular strip matrices, then they commute:

$$C\,D = D\,C. \tag{4.12}$$

4.4 Kronecker matrix product

The Kronecker product $A \otimes B$ of the $n \times m$ matrix A and the $p \times q$ matrix B

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{bmatrix}, \qquad B = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1q} \\ b_{21} & b_{22} & \dots & b_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ b_{p1} & b_{p2} & \dots & b_{pq} \end{bmatrix}, \tag{4.13}$$

is the $np \times mq$ matrix having the following block structure:

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \dots & a_{1m}B \\ a_{21}B & a_{22}B & \dots & a_{2m}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1}B & a_{n2}B & \dots & a_{nm}B \end{bmatrix}. \tag{4.14}$$

For example, if

$$A = \begin{bmatrix} 1 & 2 \\ 0 & -3 \end{bmatrix}, \qquad B = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}, \tag{4.15}$$

then

$$A \otimes B = \begin{bmatrix} 1 & 2 & 3 & 2 & 4 & 6 \\ 4 & 5 & 6 & 8 & 10 & 12 \\ 0 & 0 & 0 & -3 & -6 & -9 \\ 0 & 0 & 0 & -12 & -15 & -18 \end{bmatrix}. \tag{4.16}$$

Among many known interesting properties of the Kronecker product we would like to recall those that are important for the subsequent sections. Namely [63],

- if A and B are band matrices, then $A \otimes B$ is also a band matrix,
- if A and B are lower (upper) triangular, then $A \otimes B$ is also lower (upper) triangular.

We will also need two specific Kronecker products, namely the products $E_n \otimes A$ and $A \otimes E_m$, where E_n is an $n \times n$ identity matrix. For example, if A is a 2×3 matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \tag{4.17}$$

then

$$E_2 \otimes A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & 0 & 0 & 0 \\ a_{21} & a_{22} & a_{23} & 0 & 0 & 0 \\ 0 & 0 & 0 & a_{11} & a_{12} & a_{13} \\ 0 & 0 & 0 & a_{21} & a_{22} & a_{23} \end{bmatrix} \tag{4.18}$$

$$A \otimes E_3 = \begin{bmatrix} a_{11} & 0 & 0 & a_{12} & 0 & 0 & a_{13} & 0 & 0 \\ 0 & a_{11} & 0 & 0 & a_{12} & 0 & 0 & a_{13} & 0 \\ 0 & 0 & a_{11} & 0 & 0 & a_{12} & 0 & 0 & a_{13} \\ a_{21} & 0 & 0 & a_{22} & 0 & 0 & a_{23} & 0 & 0 \\ 0 & a_{21} & 0 & 0 & a_{22} & 0 & 0 & a_{23} & 0 \\ 0 & 0 & a_{21} & 0 & 0 & a_{22} & 0 & 0 & a_{23} \end{bmatrix} \tag{4.19}$$

This illustrates that left multiplication of $A_{n\times m}$ by E_n creates an $n \times n$ block diagonal matrix by repeating the matrix A on the diagonal, and that right multiplication of $A_{n\times m}$ by E_m creates a sparse matrix made of $n \times m$ diagonal blocks.

4.5 Eliminators

The suggested method requires also the use of a certain type of matrices called *eliminators* [51], which are obtained from the $N \times N$ unit matrix E by keeping only some of its rows and omitting all other rows: S_1 is obtained by omitting only the first row of E; S_2 is obtained by omitting only the second row; $S_{1,2}$ is obtained by omitting only the first and the second row of E; and, in general, $S_{r_1,r_2,\ldots,r_k}$ is obtained by omitting the rows with the numbers $r_1, r_2, \ldots, r_k$. In case of infinite matrices, similar matrices appeared in [10].

If A is a square $N \times N$ matrix, then the product $S_{r_1,r_2,\ldots,r_k} A$ contains only rows of A with the numbers different from $r_1, r_2, \ldots, r_k$. Similarly, the product $AS^T_{r_1,r_2,\ldots,r_k}$ contains only columns of A with the numbers different from $r_1, r_2, \ldots, r_k$.

The following simple example illustrates the main property of eliminators:

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}; \qquad S_1 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}; \qquad S_1 A = \begin{bmatrix} a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix};$$

$$A S_1^T = \begin{bmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix}; \qquad S_1 A S_1^T = \begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix}.$$

4.6 Shifters

For some types of approximation of differential operators (for example, one of the approximations of the symmetric Riesz derivative below in this article) and especially for

numerical solution of differential equations of arbitrary order (integer or fractional) with delays, it is convenient to introduce another special kind of matrices – *shifters* –, which will represent discrete shifts, like, for example, delays.

Shifters (although without using this term) were used in [51] for a simple generation of triangular strip matrices. There are shifters of two kinds: $(N+1)\times(N+1)$ matrices $E^{+}_{N,p}$, $p = 1, \dots N$, with ones on p-th diagonal above the main diagonal and zeroes elsewhere, and matrices $E^{-}_{N,p}$, $p = 1, \dots N$, with ones on p-th diagonal below the main diagonal and zeroes elsewhere. We also denote $E^{\pm}_{N,0} \equiv E_N$ the unit matrix.

The shift of all the coefficients in the triangular strip matrix U_N in the south-west (bottom-left) direction can be easily written if we start with U_{N+1} and then use shifters and eliminators:

$$ {}_{-1}U_N = S_1 \, E^{-}_{N+1,1} \, U_{N+1} \, E^{-}_{N+1,1} \, S^{T}_{N+1} \tag{4.20} $$

Similarly, the shift of all the coefficients in the triangular strip matrix U_N in the north-east (top-right) direction can be easily obtained as:

$$ {}_{+1}U_N = S_{N+1} \, E^{+}_{N+1,1} \, U_{N+1} \, E^{+}_{N+1,1} \, S^{T}_{1} \tag{4.21} $$

4.7 Discretization of ordinary fractional derivatives

It follows from [51], that the left-sided Riemann-Liouville or Caputo fractional derivative $v^{(\alpha)}(t) = {}_0D^{\alpha}_t v(t)$ can be approximated in all nodes of the equidistant discretization net $t = j\tau$ $(j = 0, 1, \dots, n)$ simultaneously with the help of the upper triangular strip matrix $B^{(\alpha)}_n$ as [1] :

$$ \left[v^{(\alpha)}_n \;\; v^{(\alpha)}_{n-1} \;\; \dots \;\; v^{(\alpha)}_1 \;\; v^{(\alpha)}_0 \right]^T = B^{(\alpha)}_n \left[v_n \;\; v_{n-1} \;\; \dots \;\; v_1 \;\; v_0 \right]^T \tag{4.22} $$

[1] In this article due to the use of the descending numbering of discretization nodes the roles of the matrices $B^{(\alpha)}_n$ (originally for backward fractional differences) and $F^{(\alpha)}_n$ (originally for forward fractional differences) are swapped in comparison with [51], where these matrices were introduced for the first time. However, we would like to preserve the notation $B^{(\alpha)}_n$ for the case of the backward fractional differences approximation and $F^{(\alpha)}_n$ for the case of the forward fractional differences approximation.

where

$$B_n^{(\alpha)} = \frac{1}{\tau^\alpha} \begin{bmatrix} \omega_0^{(\alpha)} & \omega_1^{(\alpha)} & \ddots & \ddots & \omega_{n-1}^{(\alpha)} & \omega_n^{(\alpha)} \\ 0 & \omega_0^{(\alpha)} & \omega_1^{(\alpha)} & \ddots & \ddots & \omega_{n-1}^{(\alpha)} \\ 0 & 0 & \omega_0^{(\alpha)} & \omega_1^{(\alpha)} & \ddots & \ddots \\ \cdots & \cdots & \cdots & \ddots & \ddots & \ddots \\ 0 & \cdots & 0 & 0 & \omega_0^{(\alpha)} & \omega_1^{(\alpha)} \\ 0 & 0 & \cdots & 0 & 0 & \omega_0^{(\alpha)} \end{bmatrix} \tag{4.23}$$

$$\omega_j^{(\alpha)} = (-1)^j \binom{\alpha}{j}, \qquad j = 0, 1, \ldots, n. \tag{4.24}$$

Similarly, the right-sided Riemann-Liouville or Caputo fractional derivative $v^{(\alpha)}(t) = {}_tD_b^\alpha v(t)$ can be approximated in all nodes of the equidistant discretization net $t = j\tau$ $(j = 0, 1, \ldots, n)$ simultaneously with the help of the lower triangular strip matrix $F_n^{(\alpha)}$:

$$\begin{bmatrix} v_n^{(\alpha)} & v_{n-1}^{(\alpha)} & \ldots & v_1^{(\alpha)} & v_0^{(\alpha)} \end{bmatrix}^T = F_n^{(\alpha)} \begin{bmatrix} v_n & v_{n-1} & \ldots & v_1 & v_0 \end{bmatrix}^T \tag{4.25}$$

$$F_n^{(\alpha)} = \frac{1}{\tau^\alpha} \begin{bmatrix} \omega_0^{(\alpha)} & 0 & 0 & 0 & \cdots & 0 \\ \omega_1^{(\alpha)} & \omega_0^{(\alpha)} & 0 & 0 & \cdots & 0 \\ \omega_2^{(\alpha)} & \omega_1^{(\alpha)} & \omega_0^{(\alpha)} & 0 & \cdots & 0 \\ \ddots & \ddots & \ddots & \ddots & \ldots & \ldots \\ \omega_{n-1}^{(\alpha)} & \ddots & \omega_2^{(\alpha)} & \omega_1^{(\alpha)} & \omega_0^{(\alpha)} & 0 \\ \omega_n^{(\alpha)} & \omega_{n-1}^{(\alpha)} & \ddots & \omega_2^{(\alpha)} & \omega_1^{(\alpha)} & \omega_0^{(\alpha)} \end{bmatrix} \tag{4.26}$$

The symmetric Riesz derivative of order β can be approximated based on its definition (4.5) as a combination of the approximations (4.22) and (4.25) for the left- and right-sided Riemann-Liouville derivatives, or using the centred fractional differences approximation of the symmetric Riesz derivative suggested recently by Ortigueira [48, 49]. The general formula is the same:

$$\begin{bmatrix} v_m^{(\beta)} & v_{m-1}^{(\beta)} & \ldots & v_1^{(\beta)} & v_0^{(\beta)} \end{bmatrix}^T = R_m^{(\beta)} \begin{bmatrix} v_m & v_{m-1} & \ldots & v_1 & v_0 \end{bmatrix}^T \tag{4.27}$$

In the first case, the approximation for the left-sided Caputo derivative is taken one step ahead, and the approximation for the right-sided Caputo derivatve is taken one step back. This leads to the matrix

$$R_m^{(\beta)} = \frac{h^{-\alpha}}{2}\Big[{}_{-1}U_m + {}_{+1}U_m\Big] \tag{4.28}$$

In the second case (Ortigueira's definition [48]) we have the following symmetric matix:

$$R_m^{(\beta)} = h^{-\beta}\begin{bmatrix} \omega_0^{(\beta)} & \omega_1^{(\beta)} & \omega_2^{(\beta)} & \omega_3^{(\beta)} & \cdots & \omega_m^{(\beta)} \\ \omega_1^{(\beta)} & \omega_0^{(\beta)} & \omega_1^{(\beta)} & \omega_2^{(\beta)} & \cdots & \omega_{m-1}^{(\beta)} \\ \omega_2^{(\beta)} & \omega_1^{(\beta)} & \omega_0^{(\beta)} & \omega_1^{(\beta)} & \cdots & \omega_{m-2}^{(\beta)} \\ \ddots & \ddots & \ddots & \ddots & \cdots & \cdots \\ \omega_{m-1}^{(\beta)} & \ddots & \omega_2^{(\beta)} & \omega_1^{(\beta)} & \omega_0^{(\beta)} & \omega_1^{(\beta)} \\ \omega_m^{(\beta)} & \omega_{m-1}^{(\beta)} & \ddots & \omega_2^{(\beta)} & \omega_1^{(\beta)} & \omega_0^{(\beta)} \end{bmatrix} \tag{4.29}$$

$$\omega_k^{(\beta)} = \frac{(-1)^k\,\Gamma(\beta+1)\,\cos(\beta\pi/2)}{\Gamma(\beta/2-k+1)\,\Gamma(\beta/2+k+1)}, \qquad k = 0, 1, \ldots, m \tag{4.30}$$

Both these approximations of symmetric Riesz derivatives give practically the same numerical results and in case of numerical solution of partial fractional differential equations lead to a well-posed matrix of the resulting algebraic system.

4.8 Discretization of partial derivatives in time and space

The simplest implicit discretization scheme for the classical diffusion equation is shown in Fig. 4.2, where the two nodes in time direction are used for approximating the time derivative, and the three points in spatial direction are used for the symmetric approximation of the the spatial derivative. The stencil in Fig. 4.2 involves therefore only two time layers. If we consider fractional-order time derivative, then we have to involve all time levels starting from the very beginning. This is shown in Fig. 4.3 for the case of five time layers.

Similarly, if in addition to fractional-order time derivative we also consider symmetric fractional-order spatial derivatives, then we have to use all nodes at the considered time layer from the leftmost to the rightmost spatial discretization node. This most general situation is shown in Fig. 4.4.

Let us consider the nodes $(ih, j\tau)$, $j = 0, 1, 2, \ldots, n$, corresponding to all time layers at i-th spatial discretization node. It has been shown in [51] that all values of α-th order time derivative of $u(x,t)$ at these nodes are approximated using the discrete analogue of differentiation of arbitrary order:

$$\Big[u_{i,n}^{(\alpha)}\ u_{i,n-1}^{(\alpha)}\ \ldots\ u_{i,2}^{(\alpha)}\ u_{i,1}^{(\alpha)}\ u_{i,0}^{(\alpha)}\Big] = B_n^{(\alpha)}\Big[u_{i,n}\ u_{i,n-1}\ \ldots\ u_{i,2}\ u_{i,1}\ u_{i,0}\Big]^T \tag{4.31}$$

In order to obtain a simultaneous approximation of α-th order time derivative of $u(x,t)$ in all nodes shown in Fig. 4.1, we need to arrange all function values u_{ij} at the discretization nodes to the form of a column vector:

$$\begin{aligned} u_{nm} = \Big[& u_{m,n}\ u_{m-1,n}\ \ldots\ u_{1,n}\ u_{0,n} \\ & u_{m,n-1}\ u_{m-1,n-1}\ \ldots\ u_{1,n-1}\ u_{0,n-1} \\ & \ldots\ldots\ldots \\ & u_{m,1}\ u_{m-1,1}\ \ldots\ u_{1,1}\ u_{0,1} \\ & u_{m,0}\ u_{m-1,0}\ \ldots\ u_{1,0}\ u_{0,0} \Big]^T \end{aligned} \tag{4.32}$$

In visual terms of Fig. 4.1, we first take the nodes of n-th time layer, then the nodes of $(n-1)$-th time layer, and so forth, and put them in this order in a vertical column stack.

The matrix that transforms the vector U_{nm} to the vector $U_t^{(\alpha)}$ of the partial fractional derivative of order α with respect to time variable can be obtained as a Kronecker product of the matrix $B_n^{(\alpha)}$, which corresponds to the fractional ordinary derivative of order α (recall that n is the number of time steps), and the unit matrix E_m (recall that m is the number of spatial discretization nodes):

$$T_{mn}^{(\alpha)} = B_n^{(\alpha)} \otimes E_m \tag{4.33}$$

This is illustrated in Fig. 4.5, where the nodes denoted as white and gray are used to approximate the fractional-oder time derivative at the node shown in gray.

Similarly, the matrix that transforms the vector U to the vector $U_x^{(\beta)}$ of the partial fractional derivative of order β with respect to spatial variable can be obtained as a Kronecker product of the unit matrix E_n (recall that n is the number of spatial discretization nodes), and the matrix $R_m^{(\beta)}$, which corresponds to a symmetric Riesz ordinary derivative of order β [48, 49] (recall that m is the number of time steps):

$$S_{mn}^{(\alpha)} = E_n \otimes R_n^{(\beta)} \tag{4.34}$$

This is also illustrated in Fig. 4.5, where the nodes denoted as black and gray (corresponding to all discretization nodes from the leftmost to the rightmost one) are used to approximate the symmetric fractional-order Riesz derivative at the same node shown in gray.

Having these approximations for partial fractional derivatives with respect to both variables, we can immediately discretize the general form of the fractional diffusion equation

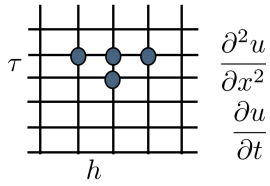


Figure 4.2: A stencil for integer-order derivatives.

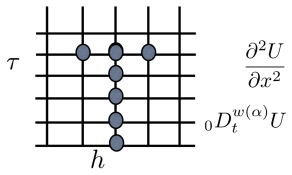


Figure 4.3: A stencil in case of fractional time derivative.

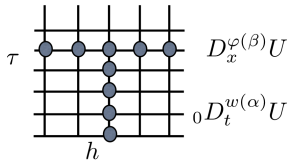


Figure 4.4: A stencil in case of fractional time and spatial derivatives.

by simply replacing the derivatives with their discrete analogs (Fig. 4.6). Namely, the equation

$$ {}_0^C D_t^\alpha u - \chi \frac{\partial^\beta u}{\partial |x|^\beta} = f(x,t) \tag{4.35} $$

is discretized as

$$ \left\{ B_n^{(\alpha)} \otimes E_m - \chi\, E_n \otimes R_m^{(\beta)} \right\} u_{nm} = f_{nm}, \tag{4.36} $$

and the matrix of this system has the structure shown in Fig. 4.7.

4.9 Initial and boundary conditions

Initial and boundary conditions must be equal to zero. If it is not so, then an auxiliary unknown function must be introduced, which satisfies the zero initial and boundary conditions. In this way, the non-zero initial and boundary conditions moves to the right-hand side of the equation for the new unknown function.

4.10 Implementation in MATLAB

We provide a set of MATLAB routines for implementing the suggested method [53]. The function BCRECUR returns the values of the coefficients that appear in the fractional difference approximations of fractional derivatives. The function BAN returns the matrix for the backward difference approximation of the left-sided ordinary fractional derivative, the function FAN returns the matrix for approximating the right-sided ordinary fractional derivative, and the functions RANSYM and RANORT return the matrices for approximating the symmetric Riesz using the formulas (4.28) and (4.29), respectively. The function ELIMINATOR returns the eliminator matrix, and the function SHIFT implements the operations (4.20) and (4.21).

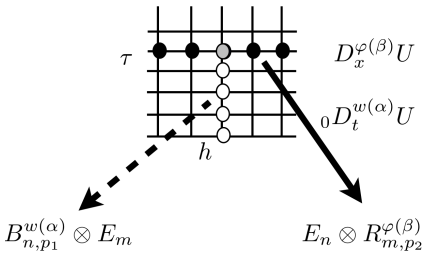


Figure 4.5: Discretization of partial derivatives.

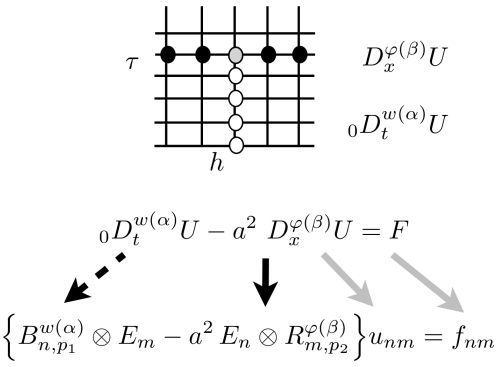


Figure 4.6: Discretization of partial derivatives and of the equation

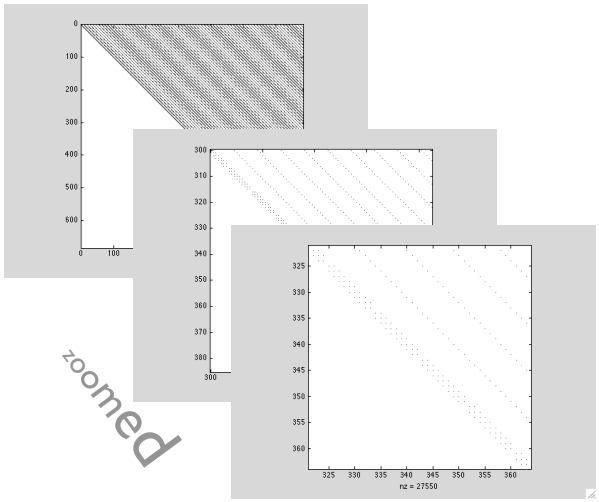


Figure 4.7: The structure of the matrix of the resulting algebraic system.

The use of these routines is illustrated by the demo functions FRACDIFFDEMOU, which implements Examples 1 and 2 below, FRACDIFFDEMOY, which implements Examples 3 and 4, and FRACDIFFDEMOYDELAY, which implements Example 5.

4.11 Examples

In this section we introduce several examples illustrating the use of the suggested method.

First, we demonstrate that for the classical integer-order diffusion equation our method gives proper results, which are in agreement with the analytical and numerical solutions provided in [46].

Second, we obtain the numerical solution of a time-fractional diffusion equation. This solution is in perfect agreement with the numerical solution obtained in the very recent work [55] by a different approach.

Then we consider fractional diffusion equation with spatial fractional derivative. The fractional derivative with respect to the spatial variable is considered as a Riesz fractional derivative.

After that, we show the results of numerical solution of a general fractional diffusion equation, where time and spatial derivatives are both of fractional order – the time fractional derivative is a left-sided Riemann–Liouville derivative, and the spatial fractional derivative is a Riesz fractional derivatie.

Finally, we demonstrate that consideration of partial differential equations with fractional derivatives and delays is equally simple in the framework of the suggested general approach.

In all examples, the spatial interval is finite.

4.11.1 Example 1: Classical diffusion equation

Let us start with the classical problem [46]:

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} \tag{4.37}$$

$$u(0,t) = 0, \quad u(1,t) = 0 \tag{4.38}$$

$$u(x,0) = 4x(1-x) \tag{4.39}$$

To reduce this problem to a problem with zero initial conditions (the boundary conditions are already zero), let us introduce an auxiliary function

$$y(x,t) = u(x,t) - u(x,0) \tag{4.40}$$

It follows from (4.40) and (4.37)–(4.39) that the function $y(x,t)$ must satisfy

$$\frac{\partial y}{\partial t} - \frac{\partial^2 y}{\partial x^2} = f(x,t), \qquad (\text{with } f(x,t) \equiv 8) \tag{4.41}$$

$$y(0,t) = 0, \quad y(1,t) = 0; \qquad y(x,0) = 0. \tag{4.42}$$

The problem (4.41)–(4.42) can be discretized using the described method (see Fig. 4.6), which gives

$$\left\{ B_n^{(1)} \otimes E_m - E_n \otimes R_m^{(2)} \right\} y_{nm} = f_{nm} \tag{4.43}$$

where m is the number of spatial discretization intervals and n is the number of time steps.

To obtain the system for finding the unknown values of y_{nm} for the inner nodes of the discretization net, we have to use the initial and boundary conditions. Since they all are zero, it is sufficient to delete the corresponding rows and columns in the system (4.43), which is easily done with the help of *eliminators*.

The result of computation of $y(x,t)$ for the spatial step $h = 0.1$ and the time step $\tau = h^2/6$ is shown in Fig. 4.8 (on the left) for $n = 37$ time steps. These values were chosen for the purpose of comparison with the results from [46]. Using (4.40), we can compute $u(x,t)$, and the result is shown in Fig. 4.8 (on the right). The values of $u(x,t)$ are in perfect agreement with the values given in [46] for the same values of h, τ, and n.

4.11.2 Example 2: Diffusion equation with time fractional derivative

Now let us consider the equation with the Caputo fractional-order time derivative:

$$ {}_0^C D_t^\alpha u = \frac{\partial^2 u}{\partial x^2} \tag{4.44}$$

$$u(0,t) = 0, \quad u(1,t) = 0 \tag{4.45}$$

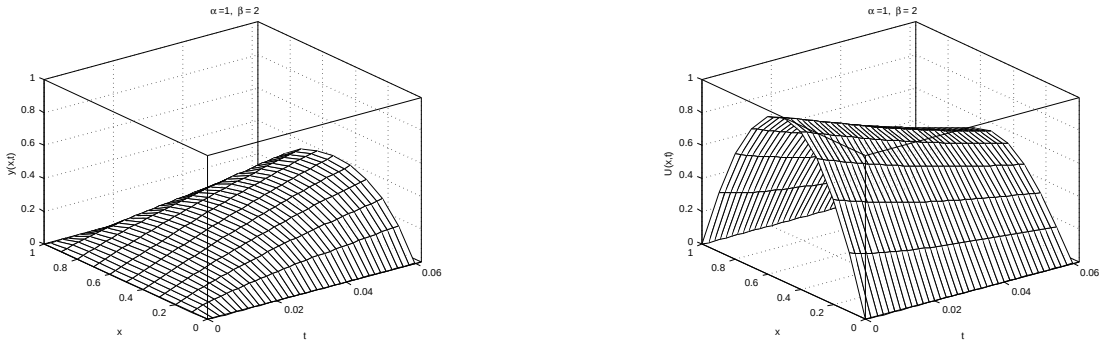


Figure 4.8: Solutions $y(x,t)$ (left) and $u(x,t)$ (right) of Example 1, with the same values of parameters as in [46].

$$u(x,0) = 4x(1-x) \tag{4.46}$$

Since the Caputo derivative of a constant is zero [2, 50], for the auxiliary function $y(x,t)$ defined by equation (4.40) we obtain a problem with zero initial and boundary conditions similar to (4.41)–(4.42):

$$ {}_0^C D_t^\alpha y - \frac{\partial^2 y}{\partial x^2} = f(x,t), \qquad (\text{with } f(x,t) \equiv 8) \tag{4.47}$$

$$y(0,t) = 0, \quad y(1,t) = 0; \qquad y(x,0) = 0 \tag{4.48}$$

This problem can be discretized in the same manner as the previous one (refer to Fig. 4.6), with the only difference that instead of the first-order time derivative we have now a derivative of order α:

$$\left\{ B_n^{(\alpha)} \otimes E_m - E_n \otimes R_m^{(2)} \right\} y_{nm} = f_{nm} \tag{4.49}$$

where m is the number of spatial discretization intervals and n is the number of time steps.

As above, the use of the zero initial conditions means that the corresponding rows and columns in the system (4.49) are removed with the help of *eliminators*.

The results of computations of $y(x,t)$ and then $u(x,t)$ for $\alpha = 1$, $\alpha = 0.7$, $\alpha = 0.5$ with $h = 0.05$ and $\tau = h^2/6$ are shown in Fig. 4.9. The structure of the matrix is the same as shown in Fig. 4.7.

Obviously, for $\alpha = 1$ we have the classical case and the same plots as in Fig. 4.8, and therefore Example 1 is a particular case of Example 2. As α goes to zero, the

function $y(x,t)$ slowly tends to $u(x,0) = 4x(1-x)$ for all t. This is also not a surprize, because, indeed, for $\alpha = 0$ the function $y(x,t)$ does not depend on t and therefore must satisfy

$$y''(x) + 8 = 0, \quad y(0) = y(1) = 0,$$

which has the solution $y(x) = 4x(1-x)$.

It should be noted that almost the same problem as (4.44)–(4.46) was numerically solved in [55] using a very different approach. The initial condition in [55] was $u(x,0) = x(1-x)$. Scaling the plots in figures 1 and 2 in [55] by the factor of 4, we obtain the plots which are practically identical with our results for $u(x,t)$ shown in Fig. 4.9. For this comparison we considered the shorter interval $0 \le t \le 0.02$ used in [55].

4.11.3 Example 3: Diffusion equation with spatial fractional derivative

Let us now focus on the role of *spatial* fractional derivative. For clarity, let us directly write the following analog of the problem (4.41)–(4.42) for determining the function $y(x,t)$:

$$\frac{\partial y}{\partial t} - \frac{\partial^\beta y}{\partial |x|^\beta} = f(x,t), \qquad (\text{with } f(x,t) \equiv 8) \tag{4.50}$$

$$y(0,t) = 0, \quad y(1,t) = 0; \qquad y(x,0) = 0. \tag{4.51}$$

where $1 < \beta \le 2$. The right-hand side is the same as in (4.41), but instead of second order spatial derivative we deal with the Riesz-Caputo fractional derivative.

The problem (4.50)–(4.51) can be discretized using the described method (see Fig. 4.6), which gives

$$\left\{ B_n^{(1)} \otimes E_m - E_n \otimes R_m^{(\beta)} \right\} y_{nm} = f_{nm} \tag{4.52}$$

where m is the number of spatial discretization intervals and n is the number of time steps, and the corresponding rows and columns in the system (4.52) are removed with the help of eliminators.

The results of computations for four different values of β are shown in Fig. 4.10.

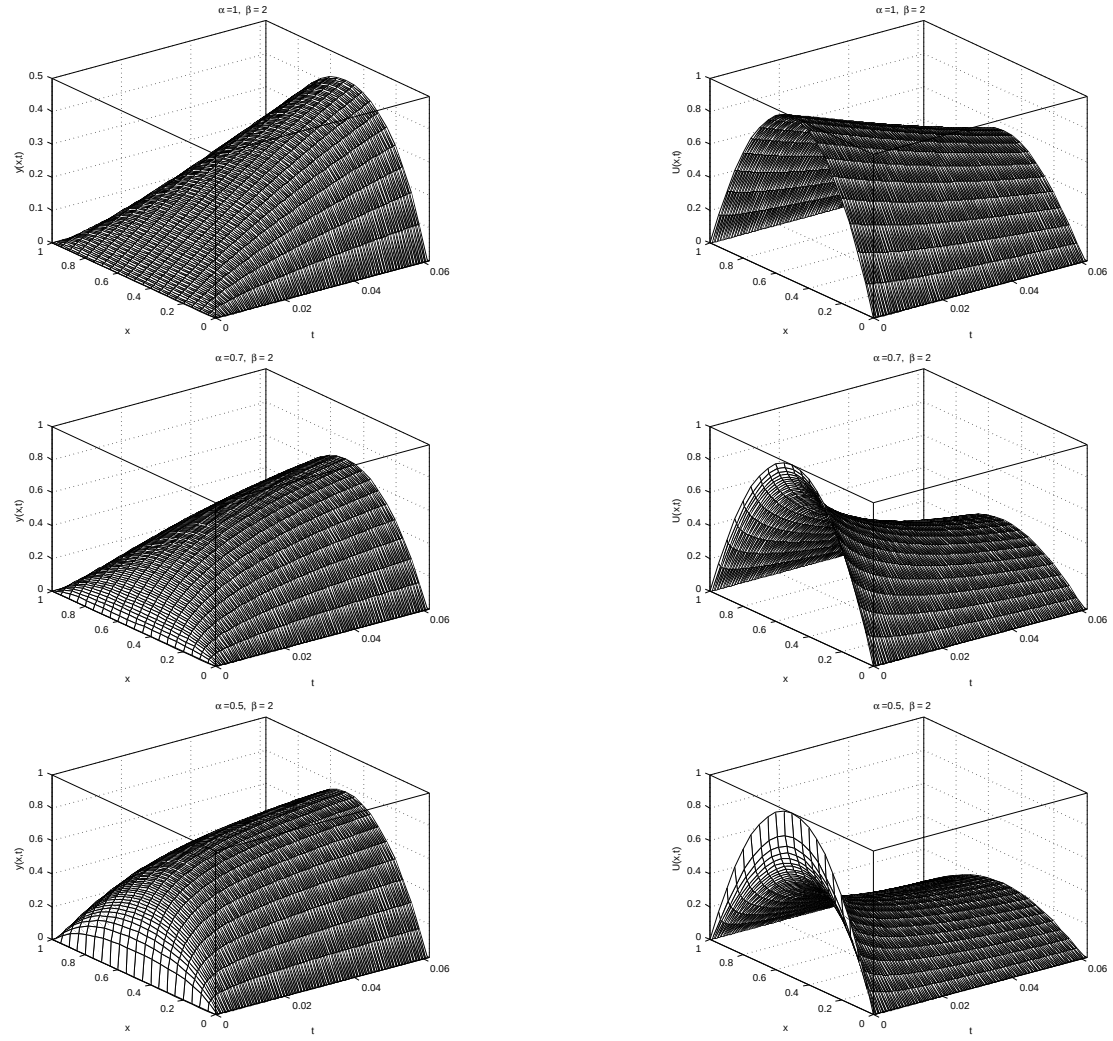


Figure 4.9: Solutions $y(x,t)$ (left column) and $u(x,t)$ (right column) of Example 2, for $\alpha = 1$ (top), $\alpha = 0.7$ (middle) and $\alpha = 0.5$ (bottom), with spatial step $h = 0.05$ and time step $\tau = h^2/6$.

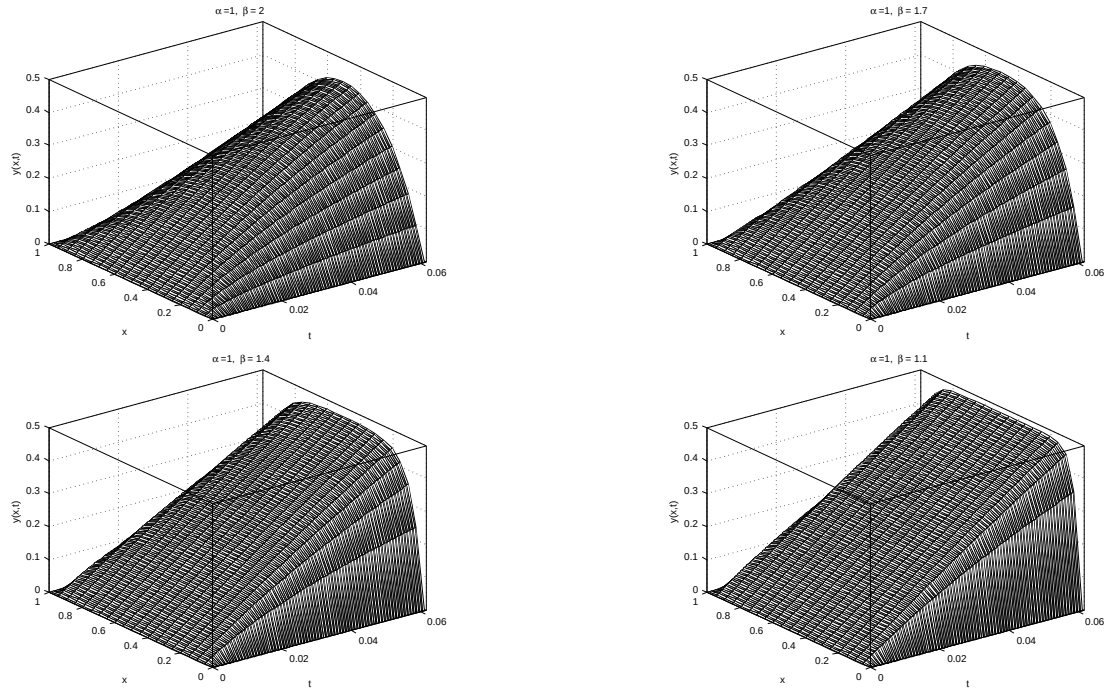


Figure 4.10: Solutions $y(x,t)$ (left column) of Example 3, for $\beta = 2$ (top-left), $\beta = 1.7$ (top-right), $\beta = 1.4$ (bottom-left), and $\beta = 1.1$ (bottom-right), with spatial step $h = 0.05$ and time step $\tau = h^2/6$.

4.11.4 Example 4: General fractional diffusion equation

Now we can illustrate that the method works also in the case when both derivatives are of fractional order. Let us consider the most general situation:

$$ {}_0^C D_t^\alpha y - \frac{\partial^\beta y}{\partial |x|^\beta} = f(x,t), \qquad (\text{with } f(x,t) \equiv 8) \tag{4.53} $$

$$ y(0,t) = 0, \quad y(1,t) = 0; \qquad y(x,0) = 0. \tag{4.54} $$

The right-hand side is the same as in (4.41) and (4.50), but now *both* derivatives are allowed to be of non-integer order.

The problem (4.53)–(4.54) can be discretized using the described method (see Fig. 4.6), which gives

$$ \left\{ B_n^{(\alpha)} \otimes E_m - E_n \otimes R_m^{(\beta)} \right\} y_{nm} = f_{nm} \tag{4.55} $$

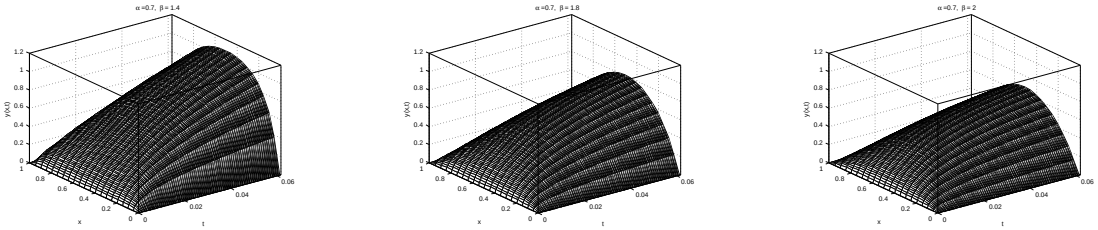


Figure 4.11: Solutions $y(x,t)$ of Example 4, for $\alpha = 0.7$ and $\beta = 1.4$ (left), $\alpha = 0.7$ and $\beta = 1.8$ (middle), $\alpha = 0.7$ $\beta = 2$ (right), with spatial step $h = 0.05$ and time step $\tau = h^2/6$.

where m is the number of spatial discretization intervals and n is the number of time steps, and the corresponding rows and columns in the system (4.55) are as in all previous examples removed with the help of eliminators.

The results of computations for some sample combinations of non-integer orders α and different values of β are shown in Fig. 4.11.

4.11.5 Example 5: Fractional diffusion equation with delay

Finally, let us consider the equation with two Caputo fractional-order time derivatives, of which one is with delay δ (we do not go into the physical interpretation of this equation, because physical interpretation of a delayed fractional derivative is not known so far, but use it for demonstrating how broad can be the field of application of our approach):

$$\frac{1}{2}\left\{ {}_0^C D_t^{\alpha} y + {}_0^C D_{t-\delta}^{\gamma} y \right\} - \frac{\partial^{\beta} y}{\partial |x|^{\beta}} = f(x,t) \qquad (\text{with } f(x,t) \equiv 8) \tag{4.56}$$

$$y(0,t) = 0, \quad y(1,t) = 0 \quad y(x,0) = 0 \tag{4.57}$$

Obviously, for $\gamma = \alpha$ and $\delta = 0$ we have the equation considered in Example 2. Let us select the discretization step so that δ is a multiple of the time step τ: $\delta = k\tau$. Then the problem (4.56)–(4.57) can be discretized using the described method (see Fig. 4.6 and the equation (4.21)), which gives:

$$\left\{ \frac{1}{2}\left(B_n^{(\alpha)} \otimes E_m + {}_{+k}B_n^{(\gamma)} \otimes E_m \right) - E_n \otimes R_m^{(\beta)} \right\} y_{nm} = f_{nm} \tag{4.58}$$

$${}_{+k}B_n^{(\gamma)} = S_{n+1,\ldots,n+k}\, E_{n+k,k}^{+}\, B_{n+k}^{(\gamma)}\, E_{n+k,k}^{+}\, S_{1,\ldots,k}^{T}$$

where, as above, m is the number of spatial discretization intervals and n is the number of time steps, k is the number of time steps corresponding to the delay δ, and the appropriate rows and columns in the system (4.58) are as in all previous examples are to removed with the help of eliminators.

The results of computations for a sample combination of non-integer orders α, β and γ and some delays δ represented by the parameter k are shown in Fig. 4.12.

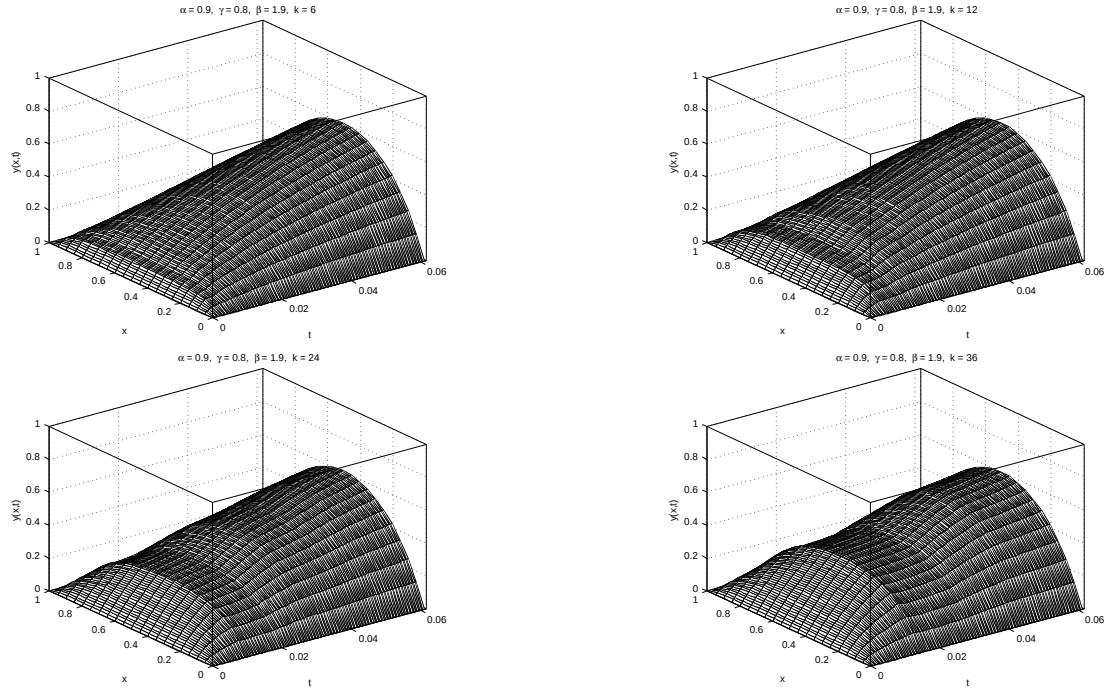


Figure 4.12: Solutions $y(x,t)$ (left column) of Example 5, for $\alpha = 0.9$, $\gamma = 0.8$, $\beta = 1.9$, for delays $\delta_k = k\tau$, k = 6, 12, 24, 36.

4.12 Chapter summary

The suggested method represents a unifying approach to numerical solution of partial differential equations of both integer and non-integer order, including equations with delays.

For the sake of clarity, in this article we considered the case of one spatial variable. However, the suggested method can be easily extended to the case of two and three spatial variables by repeatedly applying the triangular strip matrix representations of fractional-order operators in combination with the Kronecker matrix product.

The problems considered in this article are linear, so the resulting systems of algebraic

equations are linear as well. However, the suggested approach can be extended to the case of nonlinear problems, too.

The suggested method can be used also for solving partial fractional differential equations appearing from the Laplace equation by replacing second order spatial derivatives with fractional Riesz derivatives.

The suggested method can be used also for partial fractional FDEs of variable and distributed order(s) and for equations with delays.

Appendix: sample evaluation of the symmetric Riesz fractional derivative

For $\phi(x) = x(1-x)$ and the order of differentiation $1 < \beta < 2$ the left-sided Riemann-Liouville fractional derivative (4.6) of the function $\phi(x)$ is

$$ {}_0D_x^\beta \phi(x) = \frac{x^{1-\beta}}{\Gamma(2-\beta)} - \frac{2\,x^{2-\beta}}{\Gamma(3-\beta)}. \tag{4.59} $$

Similarly, the right-sided Riemann-Liouville derivative (4.7) of $\phi(x)$ is

$$ {}_xD_1^\beta \phi(x) = \frac{(1-x)^{1-\beta}}{\Gamma(2-\beta)} - \frac{2\,(1-x)^{2-\beta}}{\Gamma(3-\beta)}. \tag{4.60} $$

Therefore, the symmetric Riesz fractional derivative (4.5) of the function $\phi(x)$ is:

$$ \frac{d^\beta \phi}{d|x|^\beta} = \frac{1}{2}\left\{ {}_0D_x^\beta \phi(x) + {}_xD_1^\beta \phi(x) \right\} \tag{4.61} $$

$$ = \frac{x^{1-\beta} + (1-x)^\beta}{2\,\Gamma(2-\beta)} - \frac{x^{2-\beta} + (1-x)^{2-\beta}}{\Gamma(3-\beta)}. \tag{4.62} $$

4.13 References

[1] B. V. Bulgakov, Kolebaniya (Vibrations), Gostekhizdat, Moscow, 1954.

[2] M. Caputo, Elasticità e dissipazione, Zanichelli, Bologna, 1969.

[3] A. Chechkin, V. Gonchar, R. Gorenflo, N. Korabel, I. Sokolov, Generalized fractional diffusion equations for accelerating subdiffusion and truncated levy flights, Physical Review E 78 (2008) Article 021111.

[4] A. Chechkin, V. Gonchar, J. Klafter, R. Metzler, L. Tanatarov, Stationary states of non-linear oscillators driven by Lévy noise, Chemical Physics 284 (1–2) (2002) 233—251.

[5] A. Chechkin, V. Gonchar, M. Szydłowsky, Fractional kinetics for relaxation and superdiffusion in magnetic field, Physics of Plasmas 9 (1) (2002) 78–88.

[6] A. Chechkin, V. Y. Gonchar, J. Klafter, R. Metzler, Fundamentals of Lévy flight processes., Advances in Chemical Physics 133 (2006) 439–496.

[7] A. Chechkin, R. Gorenflo, I. Sokolov, Retarding subdiffusion and accelerating superdiffusion governed by distributed order fractional diffusion equations, Physical Review E 66 (4, art 046129) (2002) 1—7.

[8] A. Chechkin, R. Gorenflo, I. Sokolov, Fractional diffusion in inhomogeneous media, Journal of Physics A: Mathematical and General 38 (42) (2005) L679—L684.

[9] A. Chechkin, R. Gorenflo, I. Sokolov, V. Gonchar, Distributed order time fractional diffusion equation, Fractional Calculus and Applied Analysis 6 (3) (2003) 259—279.

[10] R. G. Cooke, Infinite Matrices and Sequence Spaces, Fizmatgiz, Moscow, 1960.

[11] D. del Castillo-Negrete, Fractional diffusion models of nonlocal transport, Physics of Plasmas 13 (8) (2006) Article 082308.

[12] D. del Castillo-Negrete, B. A. Carreras, V. E. Lynch, Front dynamics in reaction-diffusion systems with Lévy flights: A fractional diffusion approach, Phys. Rev. Lett. 91 (1) (2003) Article 018302.

[13] D. del Castillo-Negrete, B. A. Carreras, V. E. Lynch, Nondiffusive transport in plasma turbulence: A fractional diffusion approach, Phys. Rev. Lett. 94 (2005) Article 065003.

[14] D. del Castillo-Negrete, V. Gonchar, A. Chechkin, Fluctuation-driven directed transport in presence of Lévy flights, Physica A 27 (2008) 6693–6704.

[15] V. J. Ervin, J. P. Roop, Variational formulation for the stationary fractional advection dispersion equation, Numerical Methods for Partial Differential Equations 22 (3) (2005) 558–576.

[16] V. J. Ervin, J. P. Roop, Variational solution of fractional advection dispersion equations on bounded domains in r^d, Numerical Methods for Partial Differential Equations 23 (2) (2006) 256–281.

[17] R. Friedrich, Statistics of Lagrangian velocities in turbulent flows, Physical Review Letters 90 (8) (2003) Article 084501.

[18] F. R. Gantmakher, Theory of Matrices, Nauka, Moscow, 1988.

[19] R. Gorenflo, E. Abdel-Rehim, Discrete models of time-fractional diffusion in a potential well, Fractional Calculus and Applied Analysis 8 (2) (2005) 173–200.

[20] R. Gorenflo, G. De Fabritiis, F. Mainardi, Discrete random walk models for symmetric Lévy–Feller diffusion processes, Physica A 269 (1999) 79–89.

[21] R. Gorenflo, F. Mainardi, Random walk models for space-fractional diffusion processes, Fractional Calculus and Applied Analysis 1 (2) (1998) 167–192.

[22] R. Gorenflo, F. Mainardi, Random walk models approximating symmetric space fractional diffusion processes, Problems in Mathematical Physics 121 (2001) 120–145.

[23] R. Gorenflo, F. Mainardi, D. Moretti, G. Pagnini, P. Paradisi, Discrete random walk models for space-time fractional diffusion, Chemical Physics 284 (1–2) (2002) 521–541.

[24] R. Gorenflo, F. Mainardi, D. Moretti, G. Pagnini, P. Paradisi, Fractional diffusion: probability distributions and random walk models., Physica A 305 (1–2) (2002) 106–112.

[25] R. Gorenflo, F. Mainardi, D. Moretti, P. Paradisi, Time fractional diffusion: A discrete random walk approach., Nonlinear Dynamics 29 (2002) 129–143.

[26] E. Heinsalu, M. Patriarca, I. Goychuk, H. P., Use and abuse of a fractional Fokker-Planck dynamics for time-dependent driving, Physical Review Letters 99 (2007) Article 120602.

[27] N. Heymans, I. Podlubny, Physical interpretation of initial conditions for fractional differential equations with Riemann-Liouville fractional derivatives, Rheologica Acta 45 (5) (2006) 765–771.

[28] R. Klages, G. Radons, I. M. Sokolov (eds.), Anomalous Transport, Wiley-VCH, Berlin, 2008.

[29] T. Langlands, B. Henry, The accuracy and stability of an implicit solution method for the fractional diffusion equation, Journal of Computational Physics 205 (2) (2005) 719–736.

[30] J. Liang, Y.-Q. Chen, Hybrid symbolic and numerical simulation studies of time-fractional order wave-diffusion systems, International Journal of Control 79 (11) (2006) 1462–1470.

[31] Y. Lin, C. Xu, Finite difference/spectral approximations for the time-fractional diffusion equation, Journal of Computational Physics 225 (2) (2007) 1533–1552.

[32] F. Liu, V. Anh, I. Turner, Numerical solution of the space fractional Fokker–Planck equation, Journal of Computational and Applied Mathematics 166 (1) (2004) 209–219.

[33] Q. Liu, F. Liu, I. Turner, V. Anh, Approximation of the Levy–Feller advection-dispersion process by random walk and finite difference method, Journal of Computational Physics 222 (1) (2007) 57–70.

[34] J.-F. Lu, A. Hanyga, Wave field simulation for heterogeneous porous media with singular memory drag force, Journal of Computational Physics 208 (2) (2005) 651–674.

[35] V. E. Lynch, B. A. Carreras, D. del Castillo-Negrete, K. Ferreira-Mejias, Numerical methods for the solution of partial differential equations of fractional order, Journal of Computational Physics 192 (2003) 406–421.

[36] F. Mainardi, Y. Luchko, G. Pagnini, The fundamental solution of the space-time fractional diffusion equation, Fractional Calculus and Applied Analysis 4 (2) (2001) 153–192.

[37] F. Mainardi, A. Mura, G. Pagnini, R. Gorenflo, Time-fractional diffusion of distributed order, Journal of Vibration and Control 14 (9–10) (2008) 1267–1290.

[38] M. Meerschaert, D. Benson, B. Bäumer, Multidimensional advection and fractional dispersion, Physical Review E 59 (5) (1999) 5026–5028.

[39] M. Meerschaert, H.-P. Scheffler, C. Tadjeran, Finite difference methods for two-dimensional fractional dispersion equation, Journal of Computational Physics 211 (2006) 249–261.

[40] M. Meerschaert, C. Tadjeran, Finite difference approximations for fractional advection-dispersion equations, Journal of Computational and Applied Mathematics 172 (1) (2004) 65–77.

[41] M. Meerschaert, C. Tadjeran, Finite difference approximations for two-sided space-fractional partial differential equations, Applied Numerical Mathematics 56 (2006) 80–90.

[42] M. M. Meerschaert, H.-P. Scheffler, C. Tadjeran, Finite difference methods for two-dimensional fractional dispersion equation, Journal of Computational Physics 211 (1) (2006) 249–261.

[43] R. Metzler, E. Barkai, J. Klafter, Deriving fractional Fokker–Planck equations from a generalized master equation, Europhysics Letters 46 (4) (1999) 431–436.

[44] R. Metzler, J. Klafter, The random walk's guide to anomalous diffusion: a fractional dynamics approach, Phys. Reports. 339 (1) (2000) 1–77.

[45] R. Metzler, J. Klafter, The restaurant at the end of the random walk: recent developments in the description of anomalous transport by fractional dynamics, Journ. Phys. A: Math. Gen. 37 (2) (2004) R161–R208.

[46] W. E. Milne, Numerical Solution of Differential Equations, Wiley, New York, 1953.

[47] K. B. Oldham, J. Spanier, The Fractional Calculus, Academic Press, New York, 1974.

[48] M. D. Ortigueira, Riesz potential operators and inverses via fractional centred derivatives, International Journal of Mathematics and Mathematical Sciences Article ID 48391 (2006) 1–12.

[49] M. D. Ortigueira, A. G. Batista, On the relation between the fractional Brownian motion and the fractional derivatives, Physics Letters A 372 (2008) 958–968.

[50] I. Podlubny, Fractional Differential Equations, Academic Press, San Diego, 1999.

[51] I. Podlubny, Matrix approach to discrete fractional calculus, Fractional Calculus and Applied Analysis 3 (4) (2000) 359–386.

[52] I. Podlubny, Geometric and physical interpretation of fractional integration and fractional differentiation, Fractional Calculus and Applied Analysis 5 (4) (2002) 367–386.

[53] I. Podlubny, A. Chechkin, T. Skovranek, Y. Chen, B. Vinagre Jara, Matrix approach to discretization of ODEs and PDEs of arbitrary real order, http://www.mathworks.com/matlabcentral/fileexchange/22071 (November 12, 2008).

[54] A. Saichev, G. Zaslavsky, Fractional kinetic equations: solutions and applications, Chaos 7 No4 (1997) 753–764.

[55] R. Scherer, S. L. Kalla, L. Boyadjiev, B. Al-Saqabi, Numerical treatment of fractional heat equations, Applied Numerical Mathematics 58 (8) (2008) 1212–1223.

[56] I. Sokolov, J. Klafter, A. Blumen, Fractional kinetics, Physics Today 55 (2002) 48–54.

[57] I. M. Sokolov, J. Klafter, Field-induced dispersion in subdiffusion, Physical Review Letters 97 (2006) Article 140602.

[58] D. Stanescu, D. Kim, W. A. Woyczynski, Numerical study of interacting particles approximation for integro-differential equations, Journal of Computational Physics 206 (2) (2005) 706–726.

[59] D. A. Suprunenko, R. I. Tyshkevich, Commutative Matrices, Nauka i Tekhnika, Minsk, 1966.

[60] C. Tadjeran, M. M. Meerschaert, A second-order accurate numerical method for the two-dimensional fractional diffusion equation, Journal of Computational Physics 220 (2) (2007) 813–823.

[61] C. Tadjeran, M. M. Meerschaert, H.-P. Scheffler, A second-order accurate numerical approximation for the fractional diffusion equation, Journal of Computational Physics 213 (1) (2006) 205–213.

[62] P. P. Valko, J. Abate, Numerical inversion of 2-d Laplace transforms applied to fractional diffusion equation, Applied Numerical Mathematics 53 (2005) 73–88.

[63] C. F. van Loan, The ubiquitous Kronecker product, Journal of Computational and Applied Mathematics 123 (2000) 85–100.

[64] A. Weron, M. Magdziarz, K. Weron, Modeling of subdiffusion in space-time-dependent force fields beyond the fractional Fokker-Planck equation, Physical Review E 77 (2008) Article 036704.

[65] Z. Yong, D. A. Benson, M. M. Meerschaert, H.-P. Scheffler, On using random walks to solve the space-fractional advection-dispersion equations, Journal of Statistical Physics 123 No.1 (2006) 89–110.

[66] L. Yuan, O. P. Agrawal, A numerical scheme for dynamic systems containing fractional derivatives, Journal of Vibration and Acoustics 124 (2) (2002) 321–324.

[67] S. Yuste, Weighted average finite difference methods for fractional diffusion equations, Journal of Computational Physics 216 (1) (2006) 264–274.

[68] G. Zaslavsky, Chaos, fractional kinetics, and anomalous transport, Phys. Rep. 371 (1) (2002) 461–580.

Chapter 5

Matrix approach for differential equations of distributed order

5.1 Introduction

In this chapter we present a general approach to numerical solution to discretization of distributed-order derivatives and integrals, and to numerical solution of ordinary and partial differential equations of distributed order.

This approach is based on the matrix form representation of discretized fractional operators of constant order introduced for the first time in [54] and extended further in the works [60, 58, 65, 56].

This approach unifies the numerical differentiation of arbitrary (including integer) order and the n-fold integration, using the so-called triangular matrices. Applied to numerical solution of differential equations, it also unifies the solution of integer- and fractional-order partial differential equations. The matrix approach lead to significant simplification of the numerical solution of partial differential equations as well, and it is general enough to deal with different types of partial fractional differential equations.

In this chapter we extend the range of applicability of the matrix approach to discretization of distributed-order derivatives and integrals, and to numerical solution of distributed-order differential equations (both ordinary and partial).

Since the distributed-order operators are represented by integrals of weighted constant-order operators, we necessarily first introduce the matrix approach to discretization of constant order and then demonstrate how this method can be extended to allow numerical solution of distributed-order differential equations.

5.2 Triangular Strip Matrices

We use matrices of a specific structure, which are called *triangular strip matrices* [54, 68], and which have been also mentioned in [1, 22]. We will need lower triangular strip matrices ,

$$L_N = \begin{bmatrix} \omega_0 & 0 & 0 & 0 & \cdots & 0 \\ \omega_1 & \omega_0 & 0 & 0 & \cdots & 0 \\ \omega_2 & \omega_1 & \omega_0 & 0 & \cdots & 0 \\ \ddots & \ddots & \ddots & \ddots & \ldots & \ldots \\ \omega_{N-1} & \ddots & \omega_2 & \omega_1 & \omega_0 & 0 \\ \omega_N & \omega_{N-1} & \ddots & \omega_2 & \omega_1 & \omega_0 \end{bmatrix}, \tag{5.1}$$

and upper triangular strip matrices ,

$$U_N = \begin{bmatrix} \omega_0 & \omega_1 & \omega_2 & \ddots & \omega_{N-1} & \omega_N \\ 0 & \omega_0 & \omega_1 & \ddots & \ddots & \omega_{N-1} \\ 0 & 0 & \omega_0 & \ddots & \omega_2 & \ddots \\ 0 & 0 & 0 & \ddots & \omega_1 & \omega_2 \\ \cdots & \cdots & \cdots & \cdots & \omega_0 & \omega_1 \\ 0 & 0 & 0 & \cdots & 0 & \omega_0 \end{bmatrix}, \tag{5.2}$$

A lower (upper) triangular strip matrix is completely described by its first column (row). Therefore, if we define the truncation operation, $\mathrm{trunc}_N(\cdot)$, which truncates (in a general case) the power series $\varrho(z)$,

$$\varrho(z) = \sum_{k=0}^{\infty} \omega_k z^k \tag{5.3}$$

to the polynomial $\varrho_N(z)$,

$$\mathrm{trunc}_N(\varrho(z)) \stackrel{\mathrm{def}}{=} \sum_{k=0}^{N} \omega_k z^k = \varrho_N(z), \tag{5.4}$$

then we can consider the function $\varrho(z)$ as a generating series for the set of lower (or upper) triangular matrices L_N (or U_N), $N = 1, 2, \ldots$

It was shown in [54] that operations with triangular strip matrices, such as addition, subtraction, multiplication, and inversion, can be expressed in the form of operations with their generating series (5.3).

Among properties of triangular strip matrices it should be noticed that if matrices C and D are both lower (upper) triangular strip matrices, then they commute:

$$C\,D = D\,C. \tag{5.5}$$

5.3 Kronecker Matrix Product

The Kronecker product $A \otimes B$ of the $n \times m$ matrix A and the $p \times q$ matrix B

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{bmatrix}, \qquad B = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1q} \\ b_{21} & b_{22} & \dots & b_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ b_{p1} & b_{p2} & \dots & b_{pq} \end{bmatrix}, \tag{5.6}$$

is the $np \times mq$ matrix having the following block structure:

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \dots & a_{1m}B \\ a_{21}B & a_{22}B & \dots & a_{2m}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1}B & a_{n2}B & \dots & a_{nm}B \end{bmatrix}. \tag{5.7}$$

For example, if

$$A = \begin{bmatrix} 1 & 2 \\ 0 & -3 \end{bmatrix}, \qquad B = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}, \tag{5.8}$$

then

$$A \otimes B = \begin{bmatrix} 1 & 2 & 3 & 2 & 4 & 6 \\ 4 & 5 & 6 & 8 & 10 & 12 \\ 0 & 0 & 0 & -3 & -6 & -9 \\ 0 & 0 & 0 & -12 & -15 & -18 \end{bmatrix}. \tag{5.9}$$

Among many known interesting properties of the Kronecker product we would like to recall those that are important for the subsequent sections. Namely [72],

- if A and B are band matrices, then $A \otimes B$ is also a band matrix,
- if A and B are lower (upper) triangular, then $A \otimes B$ is also lower (upper) triangular.

We will also need two specific Kronecker products, namely the products $E_n \otimes A$ and $A \otimes E_m$, where E_n is an $n \times n$ identity matrix. For example, if A is a 2×3 matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \tag{5.10}$$

then

$$E_2 \otimes A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & 0 & 0 & 0 \\ a_{21} & a_{22} & a_{23} & 0 & 0 & 0 \\ 0 & 0 & 0 & a_{11} & a_{12} & a_{13} \\ 0 & 0 & 0 & a_{21} & a_{22} & a_{23} \end{bmatrix}, \tag{5.11}$$

$$A \otimes E_3 = \begin{bmatrix} a_{11} & 0 & 0 & a_{12} & 0 & 0 & a_{13} & 0 & 0 \\ 0 & a_{11} & 0 & 0 & a_{12} & 0 & 0 & a_{13} & 0 \\ 0 & 0 & a_{11} & 0 & 0 & a_{12} & 0 & 0 & a_{13} \\ a_{21} & 0 & 0 & a_{22} & 0 & 0 & a_{23} & 0 & 0 \\ 0 & a_{21} & 0 & 0 & a_{22} & 0 & 0 & a_{23} & 0 \\ 0 & 0 & a_{21} & 0 & 0 & a_{22} & 0 & 0 & a_{23} \end{bmatrix}. \tag{5.12}$$

This illustrates that left multiplication of $A_{n\times m}$ by E_n creates an $n \times n$ block diagonal matrix by repeating the matrix A on the diagonal, and that right multiplication of $A_{n\times m}$ by E_m creates a sparse matrix made of $n \times m$ diagonal blocks.

5.4 Discretization of Ordinary Fractional Derivatives of Constant Order

It follows from [54], that the left-sided Riemann-Liouville or Caputo fractional derivative $v^{(\alpha)}(t) = {}_0\mathrm{D}_t^\alpha v(t)$ can be approximated in all nodes of the equidistant discretization net $t = j\tau$ $(j = 0, 1, \ldots, n)$ simultaneously with the help of the upper triangular strip matrix $B_n^{(\alpha)}$ as [1]

$$\begin{bmatrix} v_n^{(\alpha)} & v_{n-1}^{(\alpha)} & \ldots & v_1^{(\alpha)} & v_0^{(\alpha)} \end{bmatrix}^T = B_n^{(\alpha)} \begin{bmatrix} v_n & v_{n-1} & \ldots & v_1 & v_0 \end{bmatrix}^T \tag{5.13}$$

where

$$B_n^{(\alpha)} = \frac{1}{\tau^\alpha} \begin{bmatrix} \omega_0^{(\alpha)} & \omega_1^{(\alpha)} & \ddots & \ddots & \omega_{n-1}^{(\alpha)} & \omega_n^{(\alpha)} \\ 0 & \omega_0^{(\alpha)} & \omega_1^{(\alpha)} & \ddots & \ddots & \omega_{n-1}^{(\alpha)} \\ 0 & 0 & \omega_0^{(\alpha)} & \omega_1^{(\alpha)} & \ddots & \ddots \\ \cdots & \cdots & \cdots & \ddots & \ddots & \ddots \\ 0 & \cdots & 0 & 0 & \omega_0^{(\alpha)} & \omega_1^{(\alpha)} \\ 0 & 0 & \cdots & 0 & 0 & \omega_0^{(\alpha)} \end{bmatrix}, \tag{5.14}$$

$$\omega_j^{(\alpha)} = (-1)^j \binom{\alpha}{j}, \qquad j = 0, 1, \ldots, n. \tag{5.15}$$

[1] Here due to the use of the descending numbering of discretization nodes the roles of the matrices $B_n^{(\alpha)}$ (originally for backward fractional differences) and $F_n^{(\alpha)}$ (originally for forward fractional differences) are swapped in comparison with [54], where these matrices were introduced for the first time. However, we would like to preserve the notation $B_n^{(\alpha)}$ for the case of the backward fractional differences approximation and $F_n^{(\alpha)}$ for the case of the forward fractional differences approximation.

Similarly, the right-sided Riemann-Liouville or Caputo fractional derivative $v^{(\alpha)}(t) = {}_t\mathrm{D}_b^{\alpha} v(t)$ can be approximated in all nodes of the equidistant discretization net $t = j\tau$ $(j = 0, 1, \ldots, n)$ simultaneously with the help of the lower triangular strip matrix $F_n^{(\alpha)}$:

$$\left[v_n^{(\alpha)} \;\; v_{n-1}^{(\alpha)} \;\; \ldots \; v_1^{(\alpha)} \;\; v_0^{(\alpha)}\right]^T = F_n^{(\alpha)} \left[v_n \;\; v_{n-1} \;\; \ldots \; v_1 \;\; v_0\right]^T, \tag{5.16}$$

$$F_n^{(\alpha)} = \frac{1}{\tau^{\alpha}} \begin{bmatrix} \omega_0^{(\alpha)} & 0 & 0 & 0 & \cdots & 0 \\ \omega_1^{(\alpha)} & \omega_0^{(\alpha)} & 0 & 0 & \cdots & 0 \\ \omega_2^{(\alpha)} & \omega_1^{(\alpha)} & \omega_0^{(\alpha)} & 0 & \cdots & 0 \\ \ddots & \ddots & \ddots & \ddots & \ldots & \ldots \\ \omega_{n-1}^{(\alpha)} & \ddots & \omega_2^{(\alpha)} & \omega_1^{(\alpha)} & \omega_0^{(\alpha)} & 0 \\ \omega_n^{(\alpha)} & \omega_{n-1}^{(\alpha)} & \ddots & \omega_2^{(\alpha)} & \omega_1^{(\alpha)} & \omega_0^{(\alpha)} \end{bmatrix}. \tag{5.17}$$

The symmetric Riesz derivative of order β can be approximated based on its definition as a half-sum of the approximations (5.13) and (5.16) for the left-and right-sided Riemann-Liouville derivatives. We, however, prefer using the centered fractional differences approximation of the symmetric Riesz derivative suggested recently by Ortiguieira [51, 52], which gives

$$\left[v_m^{(\beta)} \;\; v_{m-1}^{(\beta)} \;\; \ldots \; v_1^{(\beta)} \;\; v_0^{(\beta)}\right]^T = R_m^{(\beta)} \left[v_m \;\; v_{m-1} \;\; \ldots \; v_1 \;\; v_0\right]^T \tag{5.18}$$

with the following symmetric matrix:

$$R_m^{(\beta)} = h^{-\beta} \begin{bmatrix} \omega_0^{(\beta)} & \omega_1^{(\beta)} & \omega_2^{(\beta)} & \omega_3^{(\beta)} & \cdots & \omega_m^{(\beta)} \\ \omega_1^{(\beta)} & \omega_0^{(\beta)} & \omega_1^{(\beta)} & \omega_2^{(\beta)} & \cdots & \omega_{m-1}^{(\beta)} \\ \omega_2^{(\beta)} & \omega_1^{(\beta)} & \omega_0^{(\beta)} & \omega_1^{(\beta)} & \cdots & \omega_{m-2}^{(\beta)} \\ \ddots & \ddots & \ddots & \ddots & \ldots & \ldots \\ \omega_{m-1}^{(\beta)} & \ddots & \omega_2^{(\beta)} & \omega_1^{(\beta)} & \omega_0^{(\beta)} & \omega_1^{(\beta)} \\ \omega_m^{(\beta)} & \omega_{m-1}^{(\beta)} & \ddots & \omega_2^{(\beta)} & \omega_1^{(\beta)} & \omega_0^{(\beta)} \end{bmatrix}, \tag{5.19}$$

$$\omega_k^{(\beta)} = \frac{(-1)^k \, \Gamma(\beta+1) \, \cos(\beta\pi/2)}{\Gamma(\beta/2-k+1) \, \Gamma(\beta/2+k+1)}, \qquad k = 0, 1, \ldots, m. \tag{5.20}$$

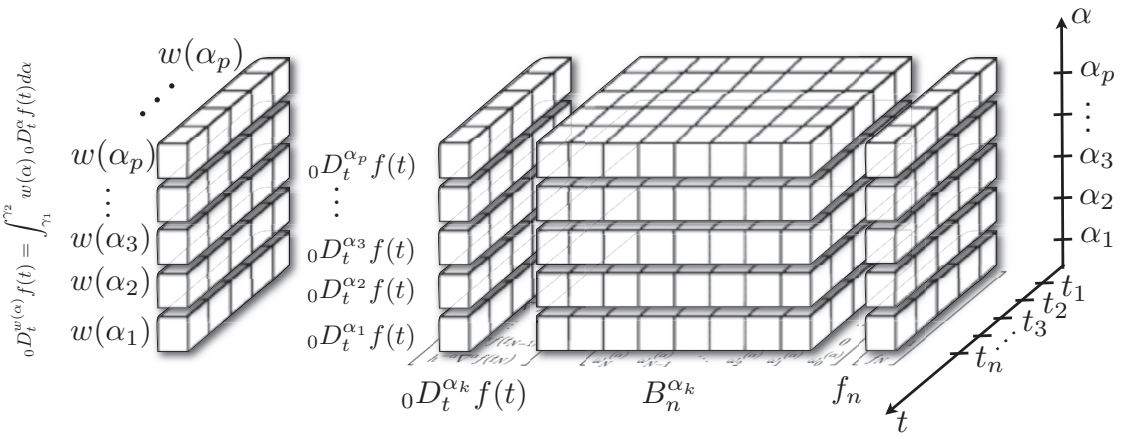


Figure 5.1: Visualization of numerical evaluation of distributed-order derivatives with the help of the matrix approach.

5.5 Discretization of Ordinary Derivatives of Distributed Order

Using the matrix approach, the discretization of a derivative of distributed order is very easy. Let us discretize the interval $[a, b]$, in which the order α is changing, using the grid with the steps $\Delta\alpha_k$, not necessarily equidistant. Then we have

$$
\begin{aligned}
{}_0\mathrm{D}_t^{w(\alpha)} f(t) &= \int_{\gamma_1}^{\gamma_2} w(\alpha)\ {}_0\mathrm{D}_t^{\alpha} f(t)\, d\alpha \approx \sum_{k=1}^{p} w(\alpha_k) \Big({}_0\mathrm{D}_t^{\alpha_k} f(t) \Big) \Delta\alpha_k && (5.21) \\
&\approx \sum_{k=1}^{p} w(\alpha_k) \Big(B_n^{\alpha_k} f_n \Big) \Delta\alpha_k = \Big(\sum_{k=1}^{p} B_n^{\alpha_k}\, w(\alpha_k)\, \Delta\alpha_k \Big) f_n. && (5.22)
\end{aligned}
$$

In other words, the discrete analog of distributed-order differentiation is given by the matrix that we will further denote as $B_{n,p}^{w(\alpha)}$,

$$
B_{n,p}^{w(\alpha)} = \sum_{k=1}^{p} B_n^{\alpha_k}\, w(\alpha_k)\, \Delta\alpha_k, \tag{5.23}
$$

and we can obtain the values of the distributed-order derivative at all points t_j $(j = 1, \ldots, n)$ at once using the following relationship:

$$
{}_0\mathrm{D}_t^{w(\alpha)} f(t) \approx B_{n,p}^{w(\alpha)} f_n. \tag{5.24}
$$

In the notation $B_{n,p}^{w(\alpha)}$ the order $w(\alpha)$ means the function describing the distribution of orders α in the interval $[\gamma_1, \gamma_2]$, and the second index p is the number of discretization

steps for α; the first index n, as above, is the number of discretization steps with respect to the variable t.

The visualization of the formula (5.22) is shown in Fig. 5.1. On each k-th layer of the shown "cake" the input vector f_n of the values of the function $f(t)$ at the nodes t_j is multiplied by the matrix $B_n^{\alpha_k}$ and gives the output vector of the values of the fractional derivative ${}_0\mathrm{D}_t^{\alpha_k}$ at the same nodes t_j. Those vectors ${}_0\mathrm{D}_t^{\alpha_k}$ are then multiplied by weights $w(\alpha_k)$ and discretization steps $\Delta\alpha_k$, and the final summation with respect to k ("summation across the layers of orders") gives the vector of the distributed-order derivative ${}_0\mathrm{D}_t^{w(\alpha)}$ evaluated at the nodes t_j $(j = 1, \ldots, n)$.

5.6 Discretization of Partial Derivatives of Distributed Order

In contrast with generally used numerical methods, where the solution is obtained step-by-step by moving from the previous time layer to the next one, let us consider the whole time interval of interest at once. This allows us to create a net of discretization nodes. In the simplest case of one spatial dimension this step gives a 2D net of nodes. An example of such discretization is shown in Fig. 5.2. The values of the unknown function in inner nodes (shaded area in Fig. 5.2) are to be found. The values at the boundaries are known, they are used later in constructing the system of algebraic equations.

The system of algebraic equations is obtained by approximating the equation in all inner nodes simultaneously (this gives the left-hand side of the resulting system of algebraic equations) and then utilizing the initial and boundary conditions (the values of which appears in the right-hand side of the resulting system).

The discretization nodes in Fig. 5.2 are numbered from right to left in each time level, and the time levels are numbered from bottom to top. We use such numbering for the clarity of presentation of our approach, although standard numberings work equally well.

The simplest implicit discretization scheme used for numerical solution of partial differential equations, like diffusion equation, is shown in Fig. 5.3, where the two nodes in time direction are used for approximating the time derivative, and the three points in spatial direction are used for the symmetric approximation of the the spatial derivative. The stencil in Fig. 5.3 involves therefore only two time layers. If we consider fractional-order time derivative or distributed-order derivative, then we have to involve all time levels starting from the very beginning. This is shown in Fig. 5.4 for the case of five time layers.

Similarly, if in addition to time derivative of distributed order or of fractional order we also consider symmetric spatial derivatives of distributed order or fractional order, then

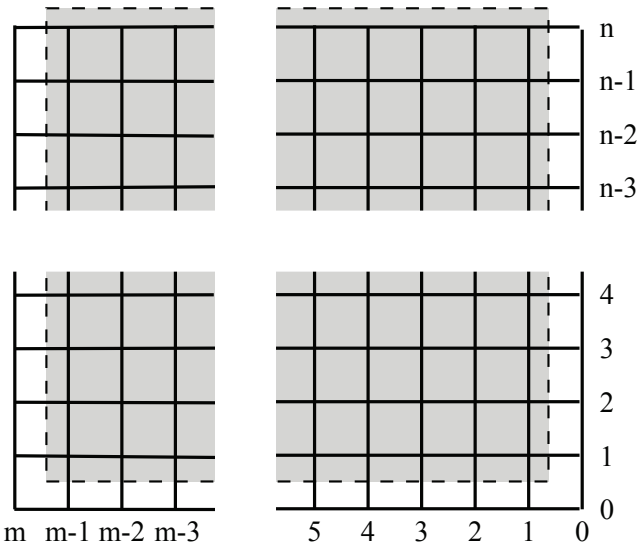


Figure 5.2: Nodes and their right-to-left, and bottom-to-top numbering.

we have to use all nodes at the considered time layer. This most general situation is shown in Fig. 5.5.

Let us consider the nodes $(ih, j\tau)$, $j = 0, 1, 2, \ldots, n$, corresponding to all time layers at i-th spatial discretization node. Similarly to the case of constant fractional orders [54], all values of α-th order time derivative of $u(x,t)$ at these nodes are approximated using the discrete analogue of distributed-order differentiation:

$$\left[u_{i,n}^{(w(\alpha))}\ u_{i,n-1}^{(w(\alpha))}\ \ldots\ u_{i,2}^{(w(\alpha))}\ u_{i,1}^{(w(\alpha))}\ u_{i,0}^{(w(\alpha))}\right] = B_{n,p}^{w(\alpha)}\left[u_{i,n}\ u_{i,n-1}\ \ldots\ u_{i,2}\ u_{i,1}\ u_{i,0}\right]^T. \tag{5.25}$$

In order to obtain a simultaneous approximation of α-th order time derivative of $u(x,t)$ in all nodes shown in Fig. 5.2, we need to arrange all function values u_{ij} at the discretization nodes to the form of a column vector:

$$\begin{aligned} U_{nm} = \Big[& u_{m,n}\ u_{m-1,n}\ \ldots\ u_{1,n}\ u_{0,n} \\ & u_{m,n-1}\ u_{m-1,n-1}\ \ldots\ u_{1,n-1}\ u_{0,n-1} \\ & \ldots\ldots\ldots \\ & u_{m,1}\ u_{m-1,1}\ \ldots\ u_{1,1}\ u_{0,1} \\ & u_{m,0}\ u_{m-1,0}\ \ldots\ u_{1,0}\ u_{0,0}\ \Big]^T. \end{aligned} \tag{5.26}$$

In visual terms of Fig. 5.2, we first take the nodes of n-th time layer, then the nodes of $(n-1)$-th time layer, and so forth, and put them in this order in a vertical column stack.

The matrix that transforms the vector U_{nm} to the vector $U_t^{w(\alpha)}$ of the partial derivative of distributed order $w(\alpha)$ with respect to time variable can be obtained as a Kronecker product of the matrix $B_{n,p}^{w(\alpha)}$, which corresponds to the ordinary derivative of distributed order $w(\alpha)$ (recall that n is the number of time steps), and the unit matrix E_m (recall that m is the number of spatial discretization steps):

$$T_{mn}^{w(\alpha)} = B_{n,p}^{w(\alpha)} \otimes E_m. \tag{5.27}$$

This is illustrated in Fig. 5.6, where the nodes denoted as white and grey are used to approximate the fractional-order time derivative at the node shown in grey.

Similarly, the matrix that transforms the vector U to the vector $U_x^{\varphi(\beta)}$ of the derivative of distributed order $\varphi(\beta)$ with respect to spatial variable can be obtained as a Kronecker product of the unit matrix E_n (recall that n is the number of spatial discretization nodes), and the matrix $R_{m,p}^{\varphi(\beta)}$, which corresponds to symmetric Riesz ordinary derivative of distributed order $\varphi(\beta)$ (recall that m is the number of time steps):

$$S_{mn}^{\varphi(\beta)} = E_n \otimes R_{n,p}^{\varphi(\beta)}. \tag{5.28}$$

This is also illustrated in Fig. 5.6, where the nodes denoted as black and grey are used to approximate the symmetric fractional-order Riesz derivative at the same node shown in grey.

Having these approximations for partial fractional derivatives with respect to both variables, we can immediately discretize, for example, the diffusion equation in terms of time- and space-derivatives of distributed order by simply replacing the derivatives with their discrete analogs (Fig. 5.7). Namely, the equation

$${}_0\mathrm{D}_t^{w(\alpha)} u - \chi \frac{\partial^{\varphi(\beta)} u}{\partial |x|^{\varphi(\beta)}} = f(x,t) \tag{5.29}$$

is discretized as

$$\left\{ B_{n,p}^{w(\alpha)} \otimes E_m - \chi\, E_n \otimes R_{m,p}^{\varphi(\beta)} \right\} u_{nm} = f_{nm}. \tag{5.30}$$

5.7 Initial and Boundary Conditions for Using the Matrix Approach

It is always emphasized in case of the matrix approach to solution of differential equations that initial and boundary conditions must be equal to zero. If it is not so, then an

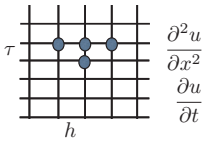


Figure 5.3: A stencil for integer-order derivatives.

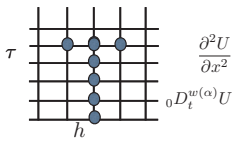


Figure 5.4: A stencil in case of distributed-order and fractional time derivative and second-order spatial derivative.

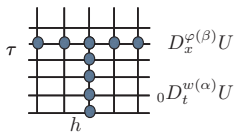


Figure 5.5: A stencil in case of distributed-order and fractional-order time and spatial derivatives.

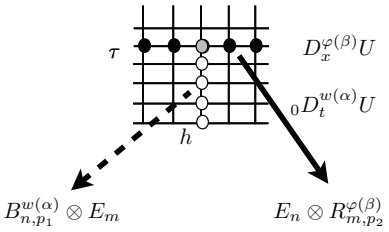


Figure 5.6: Discretization of partial derivaives.

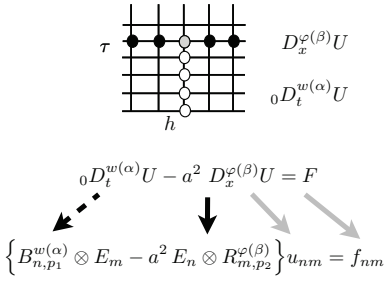


Figure 5.7: Discretization of partial derivaives and of the equation

auxiliary unknown function must be introduced, which satisfies the zero initial and boundary conditions. In this way, the non-zero initial and boundary conditions moves to the right-hand side of the equation for the new unknown function. After obtaining the solution for the auxiliary function, the backward substitution gives the solution of the original equation.

5.8 Implementation in MATLAB

A set of MATLAB routines implementing the described method is provided for download [55]. Those routines require the previously published toolbox for numerical solution of differential equations of arbitrary (fractional) constant order [57].

The function DOBAN returns the matrix for the backward difference approximation of the left-sided distributed-order derivative, the function DOFAN returns the matrix for approximating the right-sided distributed-order derivative, DORANORT return the matrix for approximating the symmetric Riesz distributed-order derivative.

The use of these routines is illustrated by the demo functions included in the toolbox.

5.9 Numerical Examples

The use of the matrix approach for numerical solution of differential equations with derivatives of distributed orders is illustrated below on three examples that generalize the standard frequently used models of the applied fractional calculus. The relaxation, oscillation, and diffusion equations play extremely important role in numerous fields of science and engineering, and, because of their importance, they are also often used for benchmarking new methods and algorithms.

To the knowledge of the authors, these are the first examples of numerical solution of such distributed-order problems.

It is worth mentioning that existence and uniqueness of solutions of such types of distributed-order differential equations in the particular case of $w(\alpha) = 1$ were investigated by Pskhu [61].

5.9.1 Example 1: Distributed-order relaxation

Let us consider the following initial value problem for the distributed-order relaxation equation:

$$ {}_0\mathrm{D}_t^{w(\alpha)} x(t) + bx(t) = f(t), \tag{5.31} $$

$$ x(0) = 1, \tag{5.32} $$

where the distribution of the orders α is given by the function $w(\alpha) = 6\,\alpha\,(1-\alpha)$, $(0 \le \alpha \le 1)$. To be able to use the matrix approach, we need zero initial condition. Introducing an auxiliary function $u(t)$,

$$ x(t) = u(t) + 1 $$

gives the following initial value problem for the new unknown $u(t)$:

$$ {}_0\mathrm{D}_t^{w(\alpha)} u(t) + bu(t) = f(t) - b, \tag{5.33} $$

$$ u(t) = 0. \tag{5.34} $$

The discretization of equation (5.33) gives the following system of algebraic equations in the matrix form:

$$ \left(B_{n,p}^{w(\alpha)} + bE_n\right) U_n = F_n, \tag{5.35} $$

where U_n is the vector of the values of $u(t)$ at the discretization nodes, and F_n is the vector of the values of the right-hand side, $f(t) - B$, at the same nodes; E_n is the identity matrix. The Matlab code for solving Example 1 is in the Appendix, and the results of computations are shown in Fig. 5.8.

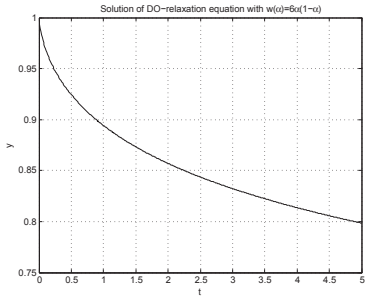


Figure 5.8: Solution of the distributed-order relaxation equation with $w(\alpha) = 6\alpha(1-\alpha)$.

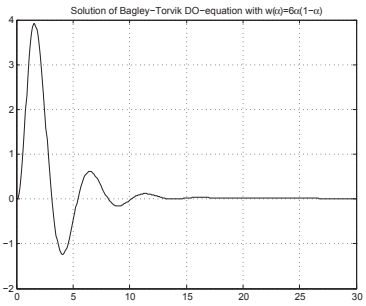


Figure 5.9: Solution of the distributed-order oscillator equation (Bagley-Torvik equation) with $w(\alpha) = 6\alpha(1-\alpha)$.

5.9.2 Example 2: Distributed-order oscillator

Let us consider the Bagley-Torvik equation with a damping term described by a distributed-order derivative. When the damping term is of constant (integer or non-integer) order, this equation is also called the fractional oscillator equation.

$$ay''(t) + by^{(w(\alpha))}(t) + cy(t) = f(t), \qquad f(t) = \begin{cases} 8, & (0 \le t \le 1) \\ 0, & (t > 1) \end{cases}, \tag{5.36}$$

$$y(0) = y'(0) = 0. \tag{5.37}$$

Similarly to Example 1, we just replace continuous operators with their corresponding discrete analogs in the form of matrices, and the known and unknown function by the vectors of their values in the discretization nodes. This gives the following algebraic system in the matrix form:

$$\Big(a\,B_n^2 + b\,B_{n,p}^{w(\alpha)} + c\Big)Y_n = F_n. \tag{5.38}$$

The Matlab code for solving Example 2 is in the Appendix, and the results of computations are shown in Fig. 5.9.

5.9.3 Example 3: Distributed-order diffusion

The last example that is provided in this section is an initial value problem for a partial differential equation with a derivative of distributed-order $w(\alpha)$ with respect to time variable t and with a constant-order symmetric fractional derivative of order β with respect to the spatial variable x:

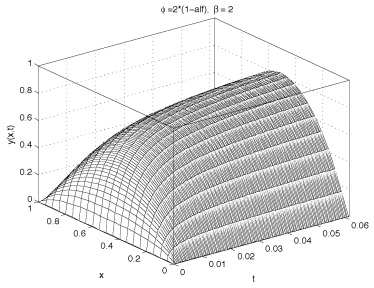


Figure 5.10: Solution of distributed-order diffusion equation with $w_1(\alpha) = 2(1-\alpha)$.

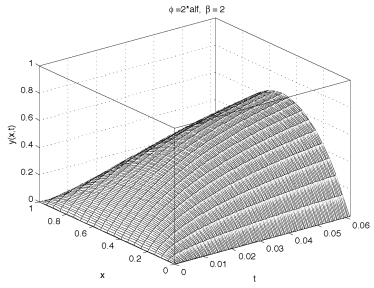


Figure 5.11: Solution of distributed-order diffusion equation with $w_2(\alpha) = 2\alpha$.

$$ {}_0\mathrm{D}_t^{w(\alpha)} y - \frac{\partial^\beta y}{\partial |x|^\beta} = f(x,t) \tag{5.39} $$

$$ y(0,t) = 0, \quad y(1,t) = 0; \qquad y(x,0) = 0. \tag{5.40} $$

In order to be able to check for a "backward compatibility" of the obtained solution with the solution of the classical diffusion equation and with the solution of the constant-order fractional diffusion equation, we take $f(x,t) = 8$.

Again, we just replace continuous operators with their corresponding discrete analogs in the form of matrices, and the known and unknown function by the vectors of their values in the discretization nodes. This gives the following algebraic system in the matrix form:

$$ \left(B_{n,p}^{w(\alpha)} \otimes E_m - E_n \otimes R_m^\beta \right) Y_{nm} = F_{nm}. \tag{5.41} $$

Here we demonstrate that distributed-order derivatives, integer-order derivatives and fractional-order derivatives can appear in the same equations and are treated in the same manner for the purpose of numerical solution using the matrix approach.

The Matlab code for solving Example 3 is also provided in the Appendix. The results of computations are shown in Fig. 5.10 for the case $w_1(\alpha) = 2(1-\alpha)$, and in Fig. 5.11 for the case $w_2(\alpha) = 2\alpha$. Although we have

$$ \int_0^1 w_1(\alpha)\,\mathrm{d}\alpha = \int_0^1 w_2(\alpha)\,\mathrm{d}\alpha = 1, $$

the obtained solutions are different because of different weights assigned by the functions $w_1(\alpha)$ and $w_2(\alpha)$ to the fractional derivatives of orders close to 0 and 1.

5.10 Chapter Summary

In this chapter we introduced the extension the matrix approach to the case of distributed-order derivatives. The matrix approach provides an extremely convenient language and framework for discretization of differentiation of any order – integer, fractional, and distributed order. Using discrete analogs of all those forms of differentiation, one can easily discretize differential equations with all possible combinations of derivatives – classical integer-order derivatives, left- and right-sided fractional order derivatives, symmetric fractional derivatives, and left-sided, right-sided, and symmetric distributed-order derivatives.

We have provided examples of solution of the three important types of problems that are important for applications and appear in many fields of science and engineering. The Matlab code provided in the Appendix demonstrates how easy it is using the matrix approach.

5.11 References

[1] B. V. Bulgakov. *Kolebaniya (Vibrations).* Gostekhizdat, Moscow, 1954.

[2] R. F. Cameron and S. McKee. High accuracy convergent product integration methods for the generalized Abel equation. *Journal of Integral Equations*, 7:103–125, 1984.

[3] R. F. Cameron and S. McKee. The analysis of product integration methods for Abel's equation using discrete fractional differentiation. *IMA Journal of Numerical Analysis*, 5:339–353, 1985.

[4] M. Caputo. *Elasticità e dissipazione.* Zanichelli, Bologna, 1969.

[5] A.V. Chechkin, V. Yu. Gonchar, J. Klafter, and R. Metzler. Fundamentals of Lévy flight processes. *Advances in Chemical Physics*, 133:439–496, 2006.

[6] A.V. Chechkin, V.Yu. Gonchar, J. Klafter, R. Metzler, and L.V. Tanatarov. Stationary states of non-linear oscillators driven by Lévy noise. *Chemical Physics*, 284(1–2):233—251, 2002.

[7] A.V. Chechkin, V.Yu. Gonchar, and M. Szydłowsky. Fractional kinetics for relaxation and superdiffusion in magnetic field. *Physics of Plasmas*, 9(1):78–88, 2002.

[8] A.V. Chechkin, R. Gorenflo, and I. Sokolov. Fractional diffusion in inhomogeneous media. *Journal of Physics A: Mathematical and General*, 38(42):L679—L684, 2005.

[9] A.V. Chechkin, R. Gorenflo, and I.M. Sokolov. Retarding subdiffusion and accelerating superdiffusion governed by distributed order fractional diffusion equations. *Physical Review E*, 66(4, art 046129):1—7, 2002.

[10] A.V. Chechkin, R. Gorenflo, I.M. Sokolov, and V.Yu. Gonchar. Distributed order time fractional diffusion equation. *Fractional Calculus and Applied Analysis*, 6(3):259—279, 2003.

[11] R. G. Cooke. *Infinite Matrices and Sequence Spaces*. Fizmatgiz, Moscow, 1960.

[12] del Castillo-Negrete. Fractional diffusion models of nonlocal transport. *Physics of Plasmas*, 13(8):Article 082308, 2006.

[13] D. del Castillo-Negrete, B. A. Carreras, and V. E. Lynch. Front dynamics in reaction-diffusion systems with Lévy flights: A fractional diffusion approach. *Phys. Rev. Lett.*, 91(1):Article 018302, 2003.

[14] D. del Castillo-Negrete, B. A. Carreras, and V. E. Lynch. Nondiffusive transport in plasma turbulence: A fractional diffusion approach. *Phys. Rev. Lett.*, 94:Article 065003, 2005.

[15] B. P. Demidovich. *Lectures on the Mathematical Theory of Stability*. Nauka, Moscow, 1967.

[16] L. J. Derr. Triangular matrices with the isoclinal property. *Pacific Journal of Mathematics*, 37(1):41–43, 1971.

[17] P. P. B. Eggermont. A new analysis of the trapezoidal-discretization method for the numerical solution of Abel-type integral equations. *Journal of Integral Equations*, 3:317–332, 1981.

[18] Vincent J. Ervin and John Paul Roop. Variational formulation for the stationary fractional advection dispersion equation. *Numerical Methods for Partial Differential Equations*, 22(3):558–576, 2005.

[19] Vincent J. Ervin and John Paul Roop. Variational solution of fractional advection dispersion equations on bounded domains in r^d. *Numerical Methods for Partial Differential Equations*, 23(2):256–281, 2006.

[20] G. Pagnini F. Mainardi, Yu. Luchko. The fundamental solution of the space-time fractional diffusion equation. *Fractional Calculus and Applied Analysis*, 4(2):153–192, 2001.

[21] R. Friedrich. Statistics of Lagrangian velocities in turbulent flows. *Physical Review Letters*, 90(8, 084501), 2003.

[22] F. R. Gantmakher. *Theory of Matrices*. Nauka, Moscow, 1988.

[23] R. Gorenflo and E.A. Abdel-Rehim. Discrete models of time-fractional diffusion in a potential well. *Fractional Calculus and Applied Analysis*, 8(2):173–200, 2005.

[24] R. Gorenflo and F. Mainardi. Random walk models for space-fractional diffusion processes. *Fractional Calculus and Applied Analysis*, 1(2):167–192, 1998.

[25] R. Gorenflo and F. Mainardi. Random walk models approximating symmetric space fractional diffusion processes. *Problems in Mathematical Physics*, 121:120–145, 2001.

[26] R. Gorenflo, F. Mainardi, D. Moretti, G. Pagnini, and P. Paradisi. Discrete random walk models for space-time fractional diffusion. *Chemical Physics*, 284(1–2):521–541, 2002.

[27] R. Gorenflo, F. Mainardi, D. Moretti, G. Pagnini, and P. Paradisi. Fractional diffusion: probability distributions and random walk models. *Physica A*, 305(1–2):106–112, 2002.

[28] R. Gorenflo, F. Mainardi, D. Moretti, and P. Paradisi. Time fractional diffusion: A discrete random walk approach. *Nonlinear Dynamics*, 29:129–143, 2002.

[29] Rudolf Gorenflo, Gianni De Fabritiis, and Francesco Mainardi. Discrete random walk models for symmetric Lévy–Feller diffusion processes. *Physica A*, 269:79–89, 1999.

[30] Nicole Heymans and Igor Podlubny. Physical interpretation of initial conditions for fractional differential equations with Riemann-Liouville fractional derivatives. *Rheologica Acta*, 45(5):765–771, 2006.

[31] P. A. W. Holyhead and S. McKee. Stability and convergence of multistep methods for linear Volterra integral equations of the first kind. *SIAM J. Numer. Anal.*, 13(2):269–292, April 1976.

[32] P. A. W. Holyhead, S. McKee, and P. J. Taylor. Multistep methods for solving linear Volterra integral equations of the first kind. *SIAM J. Numer. Anal.*, 12(5):698–711, October 1975.

[33] T.A.M. Langlands and B.I. Henry. The accuracy and stability of an implicit solution method for the fractional diffusion equation. *Journal of Computational Physics*, 205(2):719–736, May 2005.

[34] J. Liang and Yang-Quan Chen. Hybrid symbolic and numerical simulation studies of time-fractional order wave-diffusion systems. *International Journal of Control*, 79(11):1462–1470, 2006.

[35] Yumin Lin and Chuanju Xu. Finite difference/spectral approximations for the time-fractional diffusion equation. *Journal of Computational Physics*, 225(2):1533–1552, 2007.

[36] F. Liu, V. Anh, and I. Turner. Numerical solution of the space fractional Fokker–Planck equation. *Journal of Computational and Applied Mathematics*, 166(1):209–219, 2004.

[37] Q. Liu, F. Liu, I. Turner, and V. Anh. Approximation of the Levy–Feller advection-dispersion process by random walk and finite difference method. *Journal of Computational Physics*, 222(1):57–70, March 2007.

[38] Jian-Fei Lu and Andrzej Hanyga. Wave field simulation for heterogeneous porous media with singular memory drag force. *Journal of Computational Physics*, 208(2):651–674, September 2005.

[39] V. E. Lynch, B. A. Carreras, D. del Castillo-Negrete, and K.M. Ferreira-Mejias. Numerical methods for the solution of partial differential equations of fractional order. *Journal of Computational Physics*, 192:406–421, 2003.

[40] S. McKee. Discretization methods and block isoclinal matrices. *IMA Journal of Numerical Analysis*, 3:467–491, 1983.

[41] M. Meerschaert and Ch. Tadjeran. Finite difference approximations for fractional advection-dispersion equations. *Journal of Computational and Applied Mathematics*, 172(1):65–77, 2004.

[42] M. M. Meerschaert, Hans-Peter Scheffler, and Charles Tadjeran. Finite difference methods for two-dimensional fractional dispersion equation. *Journal of Computational Physics*, 211(1):249–261, January 2006.

[43] M.M. Meerschaert, D.A. Benson, and B. Bäumer. Multidimensional advection and fractional dispersion. *Physical Review E*, 59(5):5026–5028, 1999.

[44] M.M. Meerschaert, H.-P. Scheffler, and Ch. Tadjeran. Finite difference methods for two-dimensional fractional dispersion equation. *Journal of Computational Physics*, 211:249–261, 2006.

[45] M.M. Meerschaert and Charles Tadjeran. Finite difference approximations for two-sided space-fractional partial differential equations. *Applied Numerical Mathematics*, 56:80–90, 2006.

[46] R. Metzler, E. Barkai, and J. Klafter. Deriving fractional Fokker–Planck equations from a generalized master equation. *Europhysics Letters*, 46(4):431–436, 1999.

[47] R. Metzler and J. Klafter. The random walk's guide to anomalous diffusion: a fractional dynamics approach. *Phys. Reports.*, 339(1):1–77, 2000.

[48] R. Metzler and J. Klafter. The restaurant at the end of the random walk: recent developments in the description of anomalous transport by fractional dynamics. *Journ. Phys. A: Math. Gen.*, 37(2):R161–R208, 2004.

[49] W. E. Milne. *Numerical Solution of Differential Equations.* Wiley, New York, 1953.

[50] Keith B. Oldham and Jerome Spanier. *The Fractional Calculus.* Academic Press, New York, 1974.

[51] Manuel Duarte Ortigueira. Riesz potential operators and inverses via fractional centred derivatives. *International Journal of Mathematics and Mathematical Sciences*, Article ID 48391:1–12, 2006.

[52] Manuel Duarte Ortigueira and Arnaldo Guimaraes Batista. On the relation between the fractional Brownian motion and the fractional derivatives. *Physics Letters A*, 372:958–968, 2008.

[53] I. Podlubny. *Fractional Differential Equations*. Academic Press, San Diego, 1999.

[54] I. Podlubny. Matrix approach to discrete fractional calculus. *Fractional Calculus and Applied Analysis*, 3(4):359–386, 2000.

[55] I. Podlubny. *Matrix approach to distributed-order ODEs and PDEs*. http://www.mathworks.com/matlabcentral/fileexchange/, 2011.

[56] I. Podlubny, T. Skovranek, V. Verbickij, B. M. Vinagre, Y. Q. Chen, and I. Petras. Discrete fractional calculus: Non-equidistant grids and variable step length. In *Proceedings of the ASME 2011 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference IDETC/CIE 2011 August 28- August 31, 2011, Washington, D.C., USA, article DETC2011-47623*, 2011.

[57] I. Podlubny, T. Skovranek, and B. M. Vinagre. *Matrix approach to discretization of ODEs and PDEs of arbitrary real order*. http://www.mathworks.com/matlabcentral/fileexchange/ 22071, 2008.

[58] I. Podlubny, T. Skovranek, and B. M. Vinagre. Matrix approach to discretization of ordinary and partial differential equations of arbitrary real order: the matlab toolbox. In *Proceedings of the ASME 2009 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference IDETC/CIE 2009 August 30-September 2, 2009, San Diego, USA, article DETC2009-86944*, 2009.

[59] Igor Podlubny. Geometric and physical interpretation of fractional integration and fractional differentiation. *Fractional Calculus and Applied Analysis*, 5(4):367–386, 2002.

[60] Igor Podlubny, Aleksei Chechkin, Tomas Skovranek, Yang Quan Chen, and Blas M. Vinarge. Matrix approach to discrete fractional calculus ii: Partial fractional differential equations. *Journal of Computational Physics*, 228(8):3137 – 3153, 2009.

[61] A. V. Pskhu. Differencial'nye uravneniya drobnogo poryadka (partial differential equations of fractional order), 2005.

[62] A.I. Saichev and G.M. Zaslavsky. Fractional kinetic equations: solutions and applications. *Chaos*, 7 No4:753–764, 1997.

[63] S. G. Samko, A. A. Kilbas, and O. I. Maritchev. *Integrals and Derivatives of the Fractional Order and Some of Their Applications*. Nauka i Tekhnika, Minsk, 1987.

[64] R. Scherer, S. L. Kalla, L. Boyadjiev, and B. Al-Saqabi. Numerical treatment of fractional heat equations. *Applied Numerical Mathematics*, 58(8):1212–1223, 2008.

[65] T. Skovranek, V. Verbickij, Yashodhan Tarte, and I. Podlubny. Discretization of fractional-order operators and fractional differential equations on a non-equidistant mesh. In *Proceedings of the The 4TH IFAC Workshop on Fractional Differentiation and Its Applications, Badajoz, Spain, October 18-20, 2010, article FDA10-157*, 2010.

[66] I.M. Sokolov, J. Klafter, and A. Blumen. Fractional kinetics. *Physics Today*, 55:48–54, 2002.

[67] Dan Stanescu, Dongjin Kim, and Wojbor A. Woyczynski. Numerical study of interacting particles approximation for integro-differential equations. *Journal of Computational Physics*, 206(2):706–726, July 2005.

[68] D. A. Suprunenko and R. I. Tyshkevich. *Commutative Matrices*. Nauka i Tekhnika, Minsk, 1966.

[69] Charles Tadjeran and Mark M. Meerschaert. A second-order accurate numerical method for the two-dimensional fractional diffusion equation. *Journal of Computational Physics*, 220(2):813–823, January 2007.

[70] Charles Tadjeran, Mark M. Meerschaert, and Hans-Peter Scheffler. A second-order accurate numerical approximation for the fractional diffusion equation. *Journal of Computational Physics*, 213(1):205–213, March 2006.

[71] Peter P. Valko and Joseph Abate. Numerical inversion of 2-d Laplace transforms applied to fractional diffusion equation. *Applied Numerical Mathematics*, 53:73–88, 2005.

[72] Ch. F. van Loan. The ubiquitous Kronecker product. *Journal of Computational and Applied Mathematics*, 123:85–100, 2000.

[73] Zhang Yong, David A. Benson, Mark M. Meerschaert, and Hans-Peter Scheffler. On using random walks to solve the space-fractional advection-dispersion equations. *Journal of Statistical Physics*, 123 No.1:89–110, 2006.

[74] Lixia Yuan and Om P. Agrawal. A numerical scheme for dynamic systems containing fractional derivatives. *Journal of Vibration and Acoustics*, 124(2):321–324, 2002.

[75] S.B. Yuste. Weighted average finite difference methods for fractional diffusion equations. *Journal of Computational Physics*, 216(1):264–274, July 2006.

[76] G.M. Zaslavsky. Chaos, fractional kinetics, and anomalous transport. *Phys. Rep.*, 371(1):461–580, 2002.

Chapter 6

Matrix approach for non-equidistant grids and variable step length

6.1 Introduction

The results that we present in this article were motivated by two important challenges in applied numerical methods of fractional calculus.

First, until recent times, the fractional derivatives were discretized using equidistant nodes. This has roots in the famous Grünwald-Letnikov definition of fractional-order differentiation, which is based on generalization of finite differences defined on an equidistant grid, and which gives the simplest and very efficient approximation for numerical evaluation of fractional derivatives. This Grünwald-Letnikov-based approach to discretization of fractional derivatives had so strong impact on the way of thinking in the fractional calculus, that even fractional integrals were routinely discretized on equidistant grids, too. However, it is clear that for fractional integrals it was not a necessity at all. On the other hand, it was unclear what one could do with approximation of fractional integrals on non-equidistant grids, if one wants to have a uniform approach to discretization of both fractional derivatives and fractional integrals.

Second, solution of fractional differential equations in large time intervals is a computational problem due to the fact, that the number of points taken into account in computations grows with the growing value of the time variable. This is caused by the non-local character of fractional-order differentiation. So far, the only aid in this respect was the "short memory principle" [10]. Methods known from classical numerical solutions of integer-order differential equations, especially variable step length techniques, were not available for fractional differential equations.

The systematic and continuous development of the "matrix approach" [11] allowed us to find some answers to both challenges, and in this article we present them as two mutually related methods for solving problems of discrete fractional calculus on non-uniformly spaced discretization grids. Moreover, we extend this approach to distributed-order operators and distributed-order differential equations.

We start with demonstrating how the matrix approach can be extended to numerical evaluation of fractional-order integrals and derivatives on non-equidistant grids, and how fractional differential equations with constant-order derivatives can be solved on such grids. This finally unifies the discretization of fractional derivatives and fractional integrals on arbitrary (equidistant and non-equidistant) grids.

After that we make the next step and extend the matrix approach to discretization of distributed-order operators and to numerical solution of distributed-order differential equations.

Then we move the focus on using the variable step length for solving fractional differential equations. In this article we for the first time present the method that we call "the method of large steps". We provide the general framework and illustrate this method by a numerical example that, for verification purposes, allows easy exact solution as well.

Since each "large step" consists of a set of "small steps", it can be done using the matrix approach, and the "small steps" can be equidistant or non-equidistant. This is illustrated by included little pieces of Matlab code using our publicly available toolbox [13, 15].

The methods presented in this article finally allow fractional-order differentiation and integration of non-uniformly sampled signals, and the development of variable step length techniques for solving fractional differential equations (ordinary and partial).

The standard basic notation and basic definitions of fractional derivatives and fractional integrals can be found in [10, 17, 7].

6.2 Fractional-order integration and differentiation on non-equidistant grids

Although equidistant grids are used in application frequently, in many situations the use of non-equidistant grids brings notable advantages. For example, many numerical methods for solving differential equations use variable time step technique, so the time step can increase or decrease depending on how rapidly the resulting solution is changing.

Up to now, the fractional derivatives were discretized using equidistant nodes. This was, of course, due to the famous Grünwald-Letnikov definition of fractional-order differen-

tiaton, which is based on generalization of finite differences defined on an equidistant grid.

The matrix approach to discretization of integrals and derivatives of arbitrary real order, developed by Podlubny [11, 14], allows us to generalize discretization of fractional-order integrals and derivatives to non-equidistant grids.

The idea is to create first a discretization matrix I^α for integration of order α. After the matrix I^α for discrete fractional integration on non-equidistant grid is obtained, we can easily derive the matrix D^α for discretization of fractional order derivatives by matrix inversion:

$$D^\alpha = (I^\alpha)^{-1}\,.$$

In the simplest case, the function under differentiation can be approximated by a piecewise constant function, and for the non-equidistant discretization nodes t_k $(k = 1, \ldots, N)$, the coefficients of the lower triangular matrix I^α can be evaluated as

$$I_{k,j} = \frac{(t_k - t_{j-1})^\alpha - (t_k - t_j)^\alpha}{\Gamma(\alpha + 1)}, \tag{6.1}$$

$$j = 1, \ldots, k; \quad k = 1, \ldots, N. \tag{6.2}$$

Other formulas for numerical integration will give other expressions for the coefficients $I_{k,j}$; however, as we demonstrate below, even this simple approximation works well.

In the case of non-equidistant nodes, however, the matrices I^α and D^α are not strip matrices, although for one-sided fractional integrals and derivatives they are still triangular.

In the examples given below we use non-equidistant nodes generated with random steps. We generate N random points between 0 and 1, sort them in ascending order, and then scale to the considered interval of length L. After that, we replace the first and the last randomly generated node with the exact left and right bounds of the considered interval.

6.2.1 Example 1: Evaluation of integer-order integrals and derivatives

Let us consider the function $y = \sin(t)$. In Fig. 6.1, its exact first-order derivative (dashed line) and its exact (dotted ine) one-fold integral (dash-dotted line) are depicted.

The results of numerical differentiation (solid line) and numerical integration (dash-dotted line) of the same function using the matrix approach on a non-equidistant grid of 200 randomly generated nodes in the interval $[0, 2]$ are also shown in Fig. 6.1.

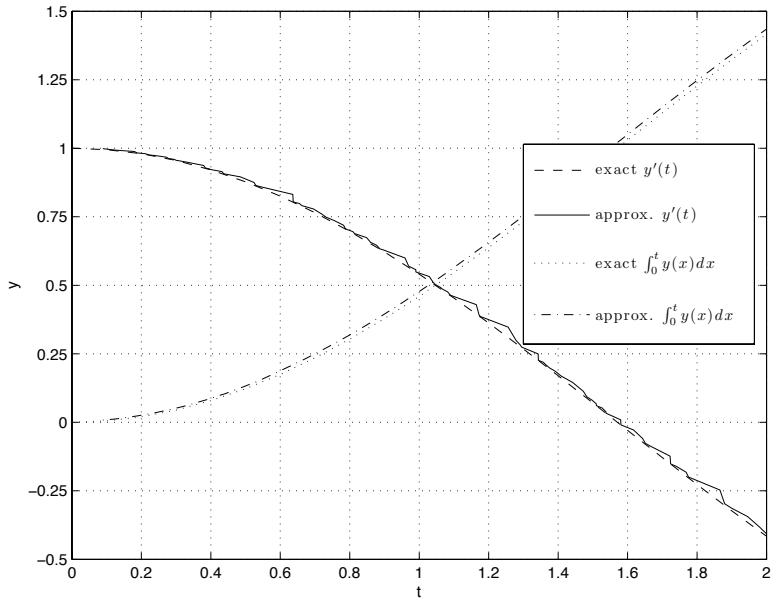


Figure 6.1: Exact and approximate evaluation of the first order derivative and the one-fold integral of $y = \sin(t)$ on the grid of 200 random non-equidistant nodes.

6.2.2 Example 2: Evaluation of fractional-order integrals and derivatives

The proposed approach is suitable also for evaluating fractional-order integrals and derivatives. In Fig. 6.2 and Fig. 6.3 fractional order integrals and derivatives of orders $\alpha = 0.1,\ \alpha = 0.3,\ \alpha = 0.5$, and $\alpha = 0.7$ of function $y = \sin(t)$ are plotted. Each derivative obtained using non-equidistant step is compared with solution obtained using the "matrix approach" with equidistant step. The match shows good agreement of the results.

6.3 Solution of fractional differential equations on non-equidistant grids

Using discrete analogs of fractional-order derivatives on a non-equidistant grid, we can easily and conveniently perform discretization and numerical solution of fractional differential equations on such grids. We illustrate the developed approach on two classical benchmark problems: two-term fractional relaxation equation and the Bagley-Torvik equation.

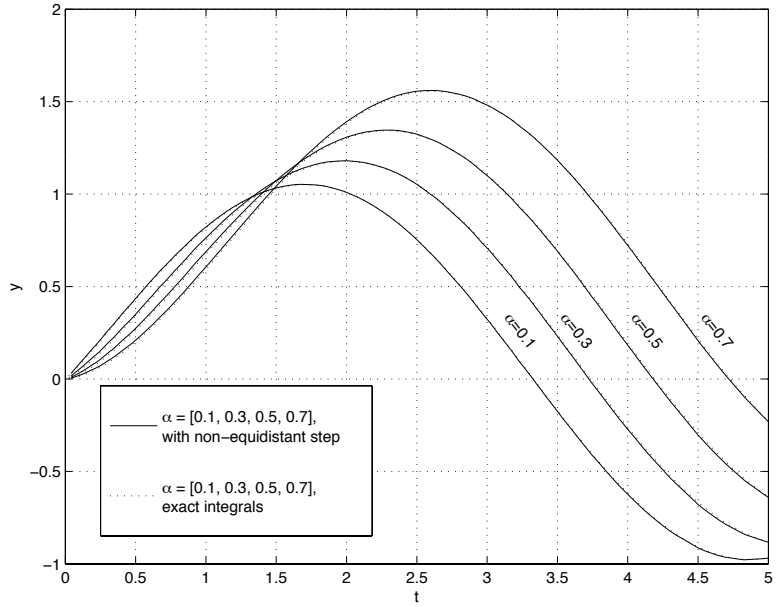


Figure 6.2: Exact and approximate evaluation of α-th order integrals of $y = \sin(t)$ on the grid of 200 random non-equidistant nodes for $\alpha = 0.1, 0.3, 0.5, 0.7$.

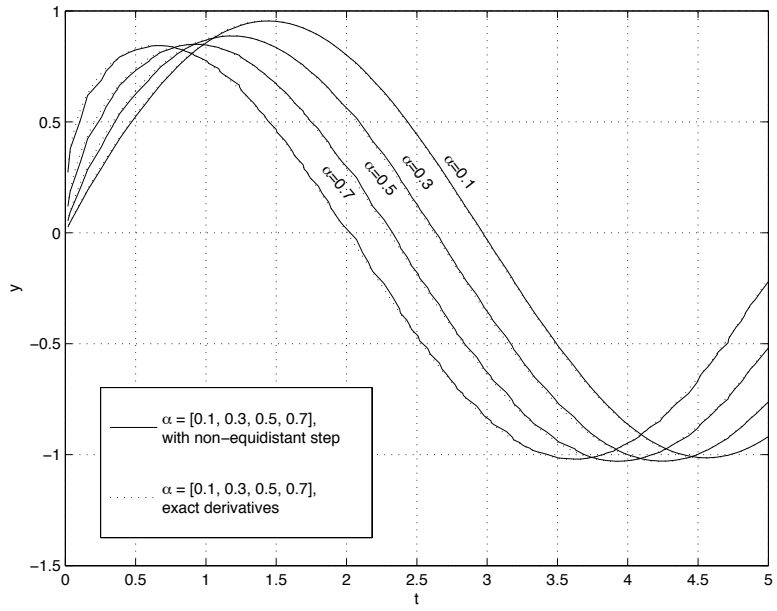


Figure 6.3: Exact and approximate evaluation of α-th order derivatives of $y = \sin(t)$ on the grid of 200 random non-equidistant nodes for $\alpha = 0.1, 0.3, 0.5, 0.7$.

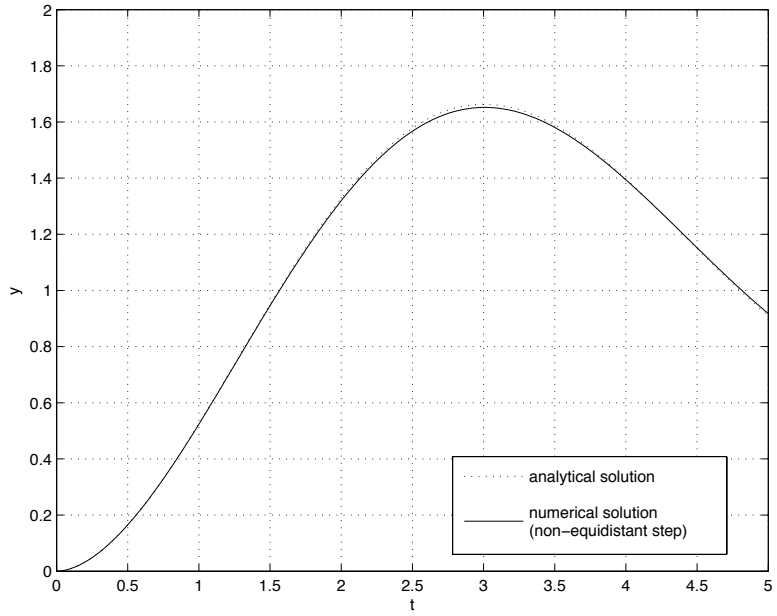


Figure 6.4: Analytical and numerical solution of problem (6.3)

6.3.1 Example 3: Fractional relaxation equation

In the first work on the matrix approach to discrete fractional calculus [11], the following sample two-term fractional differential equation in terms of the Caputo derivatives [10] under zero initial conditions was considered:

$$y^{(\alpha)}(t) + y(t) = 1, \tag{6.3}$$

$$y(0) = 0, \quad y'(0) = 0.$$

The exact analytical solution of this problem can be expressed using the Mittag-Leffler function:

$$y(t) = t^{\alpha} E_{\alpha,\alpha+1}(-t^{\alpha}). \tag{6.4}$$

In Fig. 6.4 the comparison of the exact analytical solution (dotted line), and numerical solution obtained with the help of developed approach (solid line) using non-equidistant nodes (with $N = 500$) for the case of $\alpha = 1.8$ is shown.

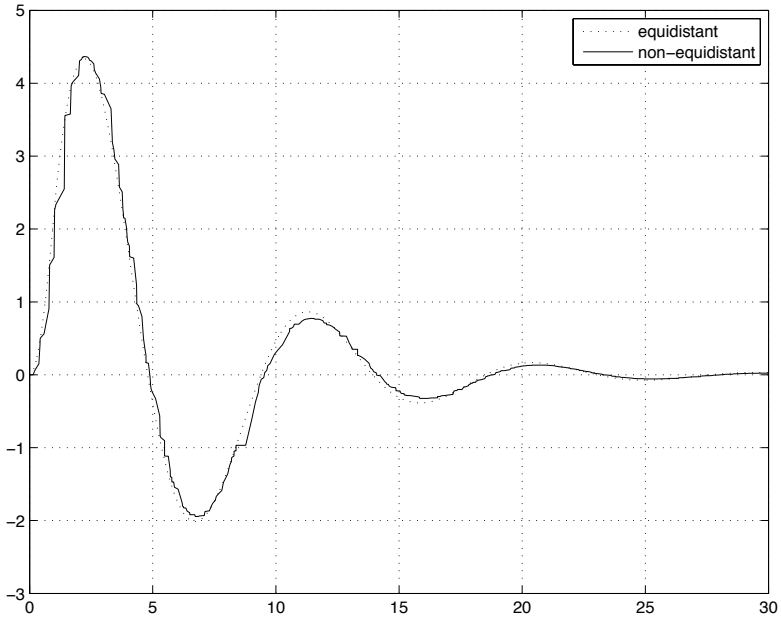


Figure 6.5: Solution of the Bagley-Torvik equation for $A = 1$, $B = 1$, $C = 1$

6.3.2 Example 4: Fractional oscillator equation

As mentioned above, the proposed approach allows easy solution of ordinary differential equations with derivatives of arbitrary real order (integer and non-integer). Let us consider the following classical initial value problem for the Bagley-Torvik equation (also known as damped fractional oscillator equations):

$$Ay''(t) + By^{(3/2)}(t) + Cy(t) = F(t), \tag{6.5}$$

$$F(t) = \begin{cases} 8, & (0 \le t \le 1) \\ 0, & (t > 1) \end{cases}, \quad y(0) = y'(0) = 0.$$

The solutions of the Bagley-Torvik equation for $A = 1$, $B = 1$, $C = 1$ in the time interval $[0; 30]$, obtained using two approaches, are shown in Fig. 6.5, and the solutions of the same problem are shown for $A = 1$, $B = 0.5$, $C = 0.5$ are depicted in Fig. 6.6. In both cases the number of discretizaton steps is 400. The dotted lines represent the numerical solutions obtained using equidistant steps (with $h = 0.075$), and the solid lines represent the numerical solutions with the same number of randomly generated non-equidistant steps.

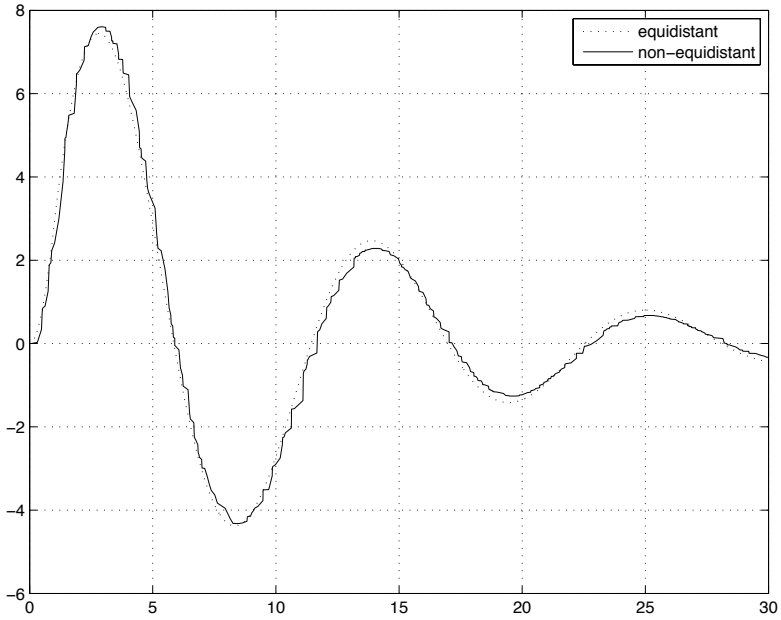


Figure 6.6: Solution of the Bagley-Torvik equation for $A = 1$, $B = 0.5$, $C = 0.5$

6.4 Solution of distributed-order differential equations on non-equidistant grids

The presented extension of the matrix approach to discretization of non-integer order integrals and derivatives and to numerical solution of equations with such operators on non-equidistant grids can be used for solving distributed-order differential equations, too. In this chapter, we use the following definition of distributed-order differential/integral operators [6]:

$$ {}_0D_t^{w(\alpha)} f(t) = \int_{\gamma_1}^{\gamma_2} w(\alpha)\, {}_0D_t^{\alpha} f(t)\, d\alpha, \tag{6.6} $$

where $w(\alpha)$ denotes the weight function of distribution of order $\alpha \in [\gamma_1, \gamma_2]$. The weight function $w(\alpha)$ is normalized, so

$$ \int_{\gamma_1}^{\gamma_2} w(\alpha)\, d\alpha = 1. \tag{6.7} $$

The idea of distributed-order differential equations was first introduced most probably by Caputo [2, 3].

As Jiao, Chen, and Podlubny showed recently [6], distributed-order derivatives can be discretized as

$$ {}_0D_t^{w(\alpha)} f(t) \approx B^{w(\alpha)} f_n, \tag{6.8} $$

$$ B^{w(\alpha)} = \sum_{k=1}^{p} B^{\alpha_k} \, w(\alpha_k) \, \Delta\alpha_k, \tag{6.9} $$

where matrices B^{α_k} are discrete analogs of fractional derivatives of orders α_k on a given grid – in our case, on a non-equidistant grid. The matrices B^{α_k} for discrete differentiation of order α_k on a non-equidistand grid are obtained as in described above, and then the matrix $B^{w(\alpha)}$ for discrete distributed-order differentiation on the same non-equidistand grid is computed using relationship (6.9).

6.4.1 Example 5: Distributed-order relaxation equation

Let us consider the following initial value problem for the distributed-order relaxation equation:

$$ {}_0D_t^{w(\alpha)} x(t) + bx(t) = f(t), \tag{6.10} $$

$$ x(0) = 1, \tag{6.11} $$

where the distribution of the orders α is given by the function $w(\alpha) = 6\,\alpha\,(1-\alpha)$, $(0 \le \alpha \le 1)$. To be able to use the matrix approach, we need zero initial condition. Introducing an auxiliary function $u(t)$,

$$ x(t) = u(t) + 1 $$

gives the following initial value problem for the new unknown $u(t)$:

$$ {}_0D_t^{w(\alpha)} u(t) + bu(t) = f(t) - b, \tag{6.12} $$

$$ u(t) = 0. \tag{6.13} $$

The discretization of equation (6.12) gives, as usual in the matrix approach, the following system of algebraic equations in the matrix form:

$$ \Big(B^{w(\alpha)} + bE_n\Big) U_n = F_n, \tag{6.14} $$

where U_n is the vector of the values of $u(t)$ at the discretization nodes, and F_n is the vector of the values of the right-hand side, $f(t) - b$, at the same nodes; E_n is the identity matrix.

The results of computations for the case of 500 randomly generated nodes in the interval $[0, 5]$ for $b = 0.1$ are shown in Fig. 6.7 by solid line. They are in agreement with the numerical solution obtained for same number of uniformly spaces nodes (dotted line in Fig. 6.7).

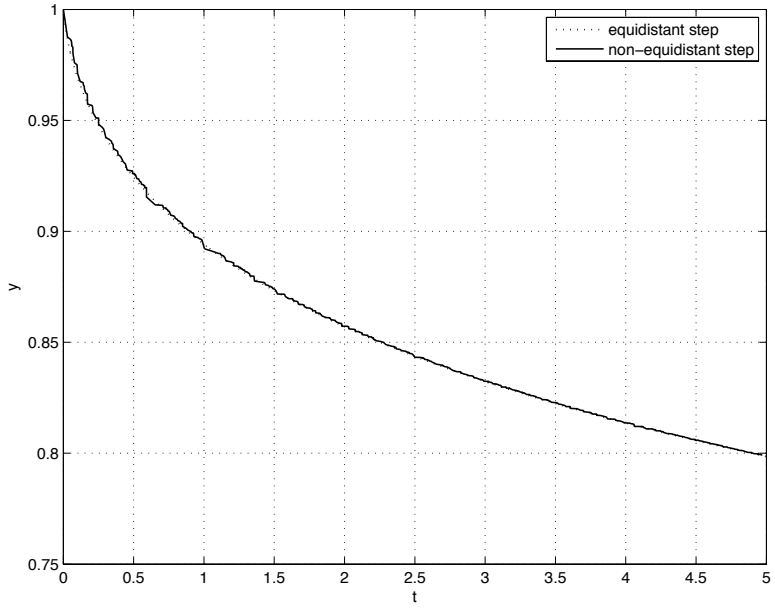


Figure 6.7: Solution of the distributed-order relaxation equation on equidistant and on non-equdistant grids.

6.4.2 Example 6: Distributed-order oscillator

Let us consider an initial value problem for the Bagley-Torvik equation, where the damping term is expressed in terms of distributed-order derivatives:

$$ay''(t) + b\ {}_0D_t^{w(\alpha)}y(t) + cy(t) = f(t), \qquad f(t) = \begin{cases} 8, & (0 \le t \le 1) \\ 0, & (t > 1) \end{cases}, \tag{6.15}$$

$$y(0) = y'(0) = 0. \tag{6.16}$$

Similarly to Example 5, we just replace continuous operators with their corresponding matrix-form discrete analogs on the considered non-equidistant grids, and the known and unknown function by the vectors of their values in the discretization nodes. This gives the following algebraic system in the matrix form:

$$\Big(a\,B_n^2 + b\,B^{w(\alpha)} + c\Big)Y_n = F_n. \tag{6.17}$$

The results of computations for the case of 400 randomly generated nodes in the interval $[0, 30]$ for $A = 1$, $B = 1$, $C = 1$, $\varphi(\alpha) = 6\alpha(1-\alpha)$, $\alpha \in [0, 1]$, are shown in Fig. 6.8 by solid line. They are in obvious agreement with the numerical solution obtained for same number of uniformly spaces nodes (dotted line in Fig. 6.8).

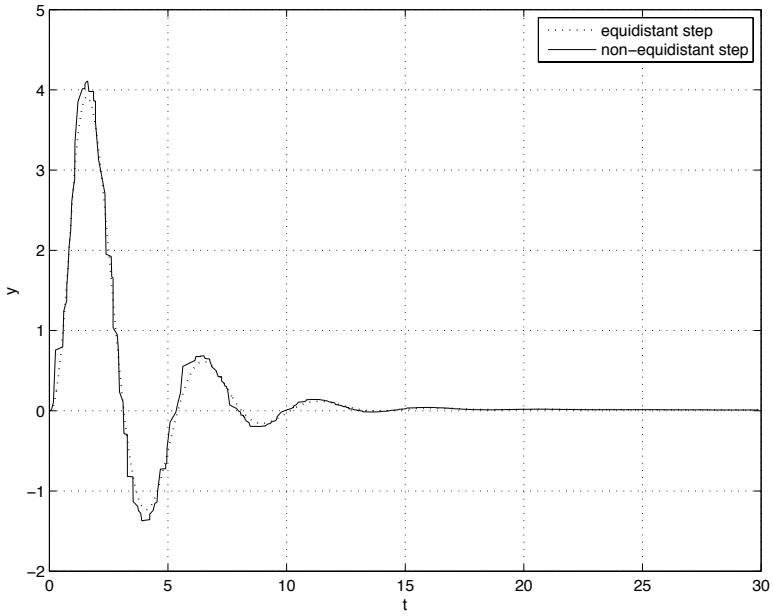


Figure 6.8: Solution of the distributed-order Bagley-Torvik equation for $A = 1$, $B = 1$, $C = 1$, $\varphi(\alpha) = 6\alpha(1-\alpha)$, $\alpha \in [0, 1]$, on equidistant and on non-equdistant grids.

The Matlab implementation of the matrix approach to discretization of distributed-order operators and numerical solution of distributed-order differential equations can be found in the update to the available toolbox [12].

6.5 Method of "large steps"

We will illustrate the idea of the proposed "method of large steps" on an easy-to-follow and sample problem, which allows exact analytical solution. In this section, we prefer using small snippets of the Matlab code in order to illustrate the simplicity of the procedure and the idea of how "large steps" are performed.

6.5.1 Sample problem in the interval (0, 2)

It is worth reminding that we use the Caputo derivatives [10]. Let us consider the following sample problem for large-steps method.

$$ {}_0^C D_t^{1/2} y(t) + y(t) = \frac{t^{0.5}}{\Gamma(1.5)} + t, \quad (t > 0), \tag{6.18} $$

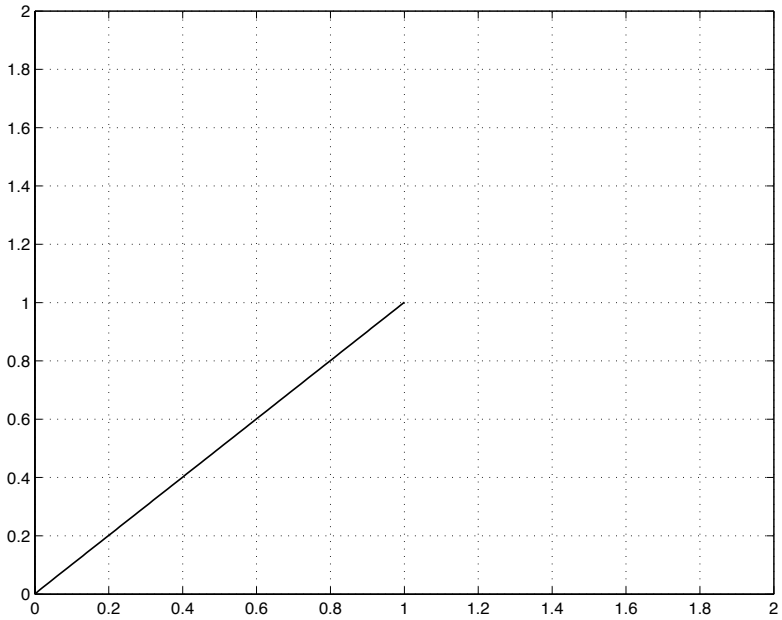


Figure 6.9: Solution of (6.18)–(6.19) after the first "large step"

$$y(0) = 0. \tag{6.19}$$

One can easily verify that the exact solution of this problem is $y(t) = t$.

6.5.2 First "large step": interval (0,1)

We can solve the problem (6.18)–(6.19) numerically in the interval (0,1) using the recently developed matrix approach [11, 14] and the corresponding MATLAB toolbox [11].

```
clear all
h = 0.01;
t = 0:h:1;
N = 1/h + 1;
M = zeros(N,N);
M = ban(0.5, N, h) + eye(N,N);
F = (t.^(0.5)/gamma(1.5) + t)';
M = eliminator(N,[1])*M*eliminator(N,[1])';
F = eliminator(N,[1])*F;
Y = M\F;
Y0 = [0; Y];
plot (t,Y0,'b')
set(gca, 'xlim', [0 2], 'ylim', [0 2] )
grid on, hold on
```

So, we have solved the previous problem in (0,1) and we know $y(t)$ for t in (0,1). How can we continue from the point to which we have arrived?

6.5.3 Second "large step": interval (1,2)

Taking into account that for $t > 1$ (we recall that we use the Caputo derivatives here)

$$ {}_0^C D_t^{1/2} y(t) = {}_1^C D_t^{1/2} y(t) + \frac{1}{\Gamma(0.5)} \int_0^1 \frac{y'(\tau) d\tau}{(t-\tau)^{1/2}}, \quad (t > 1) \tag{6.20} $$

and that we already have $y(t) = t$ in the interval (0,1), the problem (6.18)–(6.19) can be written as

$$ {}_1^C D_t^{1/2} y(t) + y(t) = \frac{t^{0.5}}{\Gamma(1.5)} + t - \frac{1}{\Gamma(0.5)} \int_0^1 \frac{d\tau}{(t-\tau)^{1/2}} \quad (t > 1). \tag{6.21} $$

The integral in the last term can be easily evaluated as

$$ \int_0^1 \frac{d\tau}{(t-\tau)^{1/2}} = 2t^{0.5} - 2(t-1)^{0.5}, \quad (t > 1). \tag{6.22} $$

Now we are ready to make the second "large step", i.e. solution in the interval $(1, 2)$. In the interval $(1, 2)$ (second step) we have to solve the following problem:

$$ {}_1^C D_t^{\alpha} y(t) + y(t) = \frac{t^{0.5}}{\Gamma(1.5)} + t - \frac{2t^{0.5}}{\Gamma(0.5)} + \frac{2(t-1)^{0.5}}{\Gamma(0.5)}; \quad (t > 1) \tag{6.23} $$

$$ y(1) = 1. \tag{6.24} $$

To solve this problem using the matrix approach [11, 14], we need to obtain zero initial conditions. For this, we make substitution

$$ y(t) = u(t) + 1, \tag{6.25} $$

and for the auxiliary function $u(t)$ we have the desired initial value problem with zero initial condition:

$$ {}_1^C D_t^{\alpha} u(t) + u(t) = \frac{t^{0.5}}{\Gamma(1.5)} + t - \frac{2t^{0.5}}{\Gamma(0.5)} + \frac{2(t-1)^{0.5}}{\Gamma(0.5)} - 1; \quad (t > 1) \tag{6.26} $$

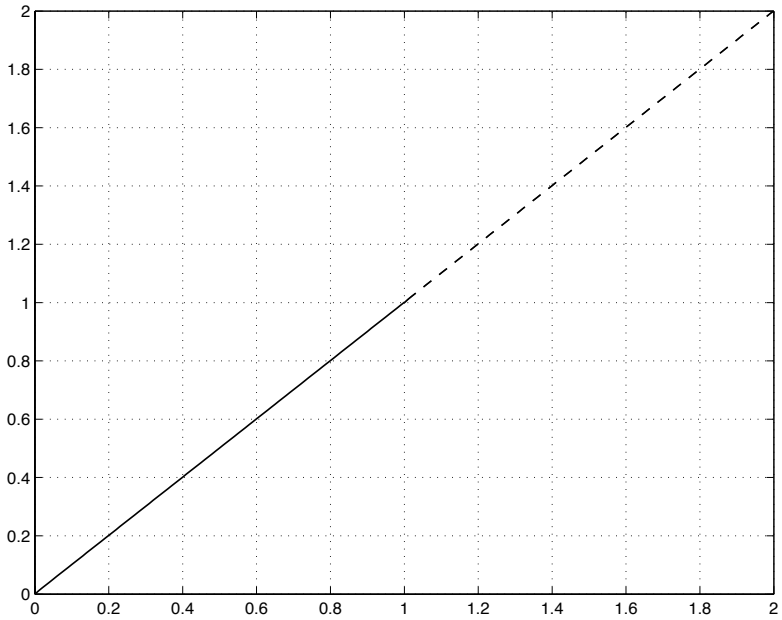


Figure 6.10: Solution of (6.23)–(6.24) after the second "large step"

$$u(1) = 0. \tag{6.27}$$

Now we solve the problem for $u(t)$ using the same matrix approach toolbox, and plot the solution.

```
clear all
h = 0.01;
t = 1:h:2;
N = 1/h + 1;
M = zeros(N,N);
M = ban(0.5, N, h) + eye(N,N);
F = (t.^(0.5)/gamma(1.5) + t - 2*t.^(0.5)/gamma(0.5) ...
       + 2*(t-1).^(0.5)/gamma(0.5) - 1)';
M = eliminator(N,[1])*M*eliminator(N,[1])';
F = eliminator(N,[1])*F;
U = M \F;
U0 = [0; U];
Y0 = U0 + 1;
plot(t, Y0, 'g')
```

We see that finally we obtained the solution of the original problem in the interval (0,2) using two "large steps": the first step was numerical solution in (0,1), and the second step was numerical solution in (1,2). In the right-hand side of the equation for the interval (1,2) two additional terms appeared as the result of considering fractional differentiation with a different lower terminal, namely ${}_1^C D_t^{1/2} y(t)$.

6.6 Method of "large steps": general scheme

In general, if we have considered the problem ($0 < \alpha < 1$)

$$ {}_0^C D_t^\alpha y(t) = f(y(t), t), \quad (t > 0), \tag{6.28} $$

$$ y(0) = 0, \tag{6.29} $$

and obtained its solution in the interval $(0, a)$ (and the final value y_a at $t = a$), then we can use this for transforming the problem to

$$ {}_a^C D_t^\alpha y(t) = f(y(t), t) - {}_0P_a^\alpha y(t), \quad (t > a), \tag{6.30} $$

$$ y(a) = y_a, \tag{6.31} $$

where

$$ {}_0P_a^\alpha y(t) = \frac{1}{\Gamma(1-\alpha)} \int_0^a (t-\tau)^{\alpha-1} y'(\tau)\, d\tau, \quad (t > a) \tag{6.32} $$

is the contribution of the "past" of the process $y(t)$ in the interval $[0, a]$ to the differential equations describing its current state in the interval $[a, b]$.

It is useful to note here that ${}_0P_a^\alpha y(t)$ can be evaluated as a fractional derivative of the function $y^*(t) = (1 - H(t-a))y(t)$, where $H(t)$ is the Heaviside unit-step function:

$$ {}_0P_a^\alpha y(t) = {}_0^C D_t^\alpha \Big((1 - H(t-a))y(t) \Big) \tag{6.33} $$

Introducing an auxiliary function $y(t) = u(t) + y_a$, we arrive at the problem with zero initial condition for the function $u(t)$, which can be solved numerically:

$$ {}_a^C D_t^\alpha u(t) = f(u(t) + y_a, t) - {}_0P_a^\alpha y(t) - y_a, \quad (t > a), \tag{6.34} $$

$$ u(a) = 0. \tag{6.35} $$

This process of making "large steps" can be continued as long as necessary.

6.7 Linear fractional differential equations

Let us consider a linear fractional differential equation with constant coefficients in the interval $(0, b)$,

$$\sum_{k=1}^{m} p_k \ {}_0^C D_t^{\alpha_k} y(t) + p_0 \, y(t) = f(t), \quad (0 < t < b), \tag{6.36}$$

If we assume that $a_k < n$, $(k = 1, \ldots, m)$ and $n - 1 < \max\limits_k a_k < n$, then we have to add n initial conditions, for example,

$$y^{(k)}(0) = 0, \quad k = 0, \ldots, n-1. \tag{6.37}$$

The equation for the second "large step" in the interval (a, b) will be

$$\sum_{k=1}^{m} p_k \ {}_a^C D_t^{\alpha_k} y(t) + p_0 y(t) = f(t) - \sum_{k=1}^{m} p_k \ {}_0 P_a^{\alpha_k} y(t), \quad (a < t < b) \tag{6.38}$$

and the initial conditions for the second "large step" will have the values of the final values $y_a^{(k)}$ of the solution in the first interval:

$$y^{(k)}(a) = y_a^{(k)}, \quad k = 0, \ldots, n-1. \tag{6.39}$$

The initial conditions should be, as usual, transformed to zero initial conditions. For this we have to introduce the auxiliary function $u(t)$

$$y(t) = u(t) + \sum_{k=0}^{n-1} y_a^{(k)} \frac{t^k}{k!}. \tag{6.40}$$

This process of making "large steps" can be continued as long as necessary.

6.8 Method of "large steps" and the problem of initialization of fractional derivatives

Lorenzo and Hartley [8, 9] raised the question about initialization of fractional derivatives. Their motivation was to use or recover the information about the process $y(t)$ in the interval $(0, a)$, if we consider fractional derivatives of $y(t)$ in (a, b).

It is worth noting that in the second "large step" in the considered sample problem we used, in fact, the proper initialization of the fractional derivative in the interval $(1, 2)$ based on the known behavior of $y(t)$ in $(0, 1)$.

In other words, proper initialization in the interval (a, b) can be done only when we know all values of $y(t)$ in the preceding interval $(0, a)$.

6.9 Chapter summary

The presented extension of the matrix approach to discretization of non-integer order derivatives and integrals of constant and distributed order allows numerical solution of differential equations with such derivatives on non-equidistant grids.

The matrix approach proves to be a very easy, algorithmic, modular, and convenient method, that unifies numerical solution of many types of problems in various settings. In the examples in this chapter we, for simplicity, used only ordinary differential equations with generalized derivatives; partial differential equations on non-equidistant grids can be solved using the technique published earlier [14].

Of course, the proposed method can be used for equations containing any mixture of left-sided, right-sided, and two-sided derivatives of integer orders, constant non-integer orders, and distributed orders, on equidistant (uniform) or non-equidistant (non-uniform) grids. In all cases, those derivatives are simply replaced with their discrete analogs in the form of easily computable matrices.

The methods presented in this article finally allow fractional-order differentiation and integration, and also distributed-order differentiation and integration, of non-uniformly sampled signals.

The proposed method of "large steps" allows to avoid in many situations the limitations of the "short memory principle" and to obtain numerical solutions in large intervals with higher accuracy. At the same time, the method of "large steps" finally allows development of variable and adaptive step length techniques for solving differential equations of non-integer order (ordinary and partial).

We would like to mention that other kind of efforts towards using non-equidistant grids for numerical solution of fractional differential equations can be found in [5, 18], and that some existing methods, like the collocation method [1, 4] or explicit numerical methods [16], can be re-considered in terms of non-equidistant grids as well.

6.10 References

[1] L. Blank. Numerical treatment of differential equations of fractional order. Numerical Analysis Report 287, Manchester Center for Computational Mathematics,

1996.

[2] M. Caputo. *Elasticità e dissipazione*. Zanichelli, Bologna, 1969.

[3] M. Caputo. Mean fractional-order-derivatives differential equations and filters. *Annali dell'Universita di Ferrara*, 41(1)(1):73–84, 1995.

[4] F. Dubois and S. Mengué. Mixed collocation for fractional differential equations. *Numerical Algorithms*, 34(2):303–311, 2003.

[5] N. Ford and A. Simpson. The numerical solution of fractional differential equations: speed versus accuracy. *Numerical Algorithms*, 26(4):333–346, 2001.

[6] Z. Jiao, Y.Q. Chen, and I. Podlubny. *Distributed-Order Dynamic Systems*. Springer, 2012.

[7] A. A. Kilbas, H. M. Srivastava, and J. J. Trujillo. *Theory and Applications of Fractional Differential Equations*, volume 204 of *North-Holland Mathematics Studies*. Elsevier, 2006.

[8] C. F. Lorenzo and T. Hartley. Initialized fractional calculus. Technical Report NASA/TP-2000-209943, 2000.

[9] C. F. Lorenzo and T. Hartley. Initialization, conceptualization, and application in the generalized (fractional) calculus. *Critical Reviews in Biomedical Engineering*, 35(6):447–553, 2007.

[10] I. Podlubny. *Fractional Differential Equations*. Academic Press, San Diego, 1999.

[11] I. Podlubny. Matrix approach to discrete fractional calculus. *Fractional Calculus and Applied Analysis*, 3(4):359–386, 2000. ISSN 0021-9991.

[12] I. Podlubny. *Matrix approach to distributed-order ODEs and PDEs*. http://www.mathworks.com/matlabcentral/fileexchange/, 2012.

[13] I. Podlubny, T. Skovranek, and B. M. Vinagre Jara. Matrix approach to discretization of ODEs and PDEs of arbitrary real order (a MATLAB toolbox), 2008. URL http://www.mathworks.com/matlabcentral/fileexchange/22071.

[14] I. Podlubny, A. Chechkin, T. Skovranek, YQ. Chen, and B. M. Vinagre Jara. Matrix approach to discrete fractional calculus II: Partial fractional differential equations. *Journal of Computational Physics*, 228:3137–3153, 2009a. ISSN 0021-9991. doi: 10.1016/j.jcp.2009.01.014.

[15] I. Podlubny, T. Skovranek, and B. M. Vinagre Jara. Matrix approach to discretization of ordinary and partial differential equations of arbitrary real order: the matlab toolbox. In *Proc. of the ASME 2009 IDETC/CIE 2009, August 30 – September 2, 2009, San Diego*. article DETC2009-86944. ISBN 9780791838563, 2009b.

[16] J. Quintana Murillo and Santos Bravo Yuste. An explicit difference method for solving fractional diffusion and diffusion-wave equations in the Caputo form. *Journal of Computational and Nonlinear Dynamics*, 6(021014), 2011.

[17] S. Samko, A. Kilbas, and O. Marichev. *Fractional Integrals and derivatives: Theory and Applications*. Gordon and Breach Science Publishers, Yverdon, 1993.

[18] B. P. Sprouse, Ch. L. MacDonald, and G. A. Silva. Computational efficiency of fractional diffusion using adaptive time step memory. In *Proc. of 4th IFAC Workshop Fractional Differentiation and its Applications. Badajoz, Spain, October 18-20, 2010*. article no. FDA10-051, 2010.

Chapter 7

Fitting of experimental data using Mittag-Leffler function

7.1 Introduction

Fitting experimental data is, undoubtedly, the most important step on the way to good models of considered processes for modeling and control purposes.

Formulation of basic laws of physics, chemistry, electrotechnics, and other fields of science, normally starts from experimental observations. An experiment is set up, data are collected from the experiment, and a researcher postulates a hypothesis based on his assessment of those experimental data. Such assessment can be intuitive or exact; nowadays, researchers usually pre-process the collected data in order to remove artifacts, outliers, noise, or other disturbances, and then use some simple function $y(t) = f(t, params)$, where t is the independent variable and $params$ are model parameters, in order to describe the results analytically. The main problem is to find such a set of parameters $params$, which gives satisfactory agreement between the experimental data and the fitting function $y(t)$.

A good choice of the fitting function $y(t)$ with appropriate number of parameters (having less parameters is better) can serve as a general model for a wide class of objects or processes.

Let us recall a historical example of the discovery of Hooke's law in the theory of elasticity. Experimenting with elongation of various kinds of elastic materials led Robert Hooke to the observation that for small deformations the stress σ in the material due to its deformation is approximately proportional to the deformation ε:

$$\sigma \sim \varepsilon$$

It is well known that he formulated this observation in the form of an anagram, which after decoding sounds as "Ut tensio, sic vis" ("As the extension, so the force"). In other words, he suggested the model which we now call the Hooke's law:

$$\sigma = E\,\varepsilon,$$

where E (which later got the name of the modulus of elasticity or the Young modulus) is a constant depending on a particular material; taking different materials, we obtain different values of the parameter E.

7.2 Recalling Basic Notions of the Fractional-Order Calculus

The standard notation for denoting the left-sided fractional-order differentiation of a function $f(t)$ defined in the interval $[a, b]$ is ${}_aD_t^\alpha f(t)$, with $\alpha \in \mathrm{R}$. Sometimes a simplified notation $f^{(\alpha)}(t)$ or $d^\alpha f(t)/dt^\alpha$ is used. In some applications also right-sided fractional derivatives ${}_tD_b^\alpha f(t)$ are used, but in the present article we will use only left-sided fractional derivatives. Even from the notation one can see that evaluation of the left-sided fractional-order operators require the values of the function $f(t)$ in the interval $[a, t]$. When α becomes an integer number, this interval shrinks to the vicinity of the point t, and we obtain the classical integer-order derivatives as particular cases.

There are several definitions of the fractional derivatives and integrals, of which we need only the following two.

The Caputo definition of fractional differentiation can be written as [1]:

$$ {}_a^C D_t^\alpha f(t) = \frac{1}{\Gamma(n-\alpha)} \int_a^t \frac{f^{(n)}(\tau)}{(t-\tau)^{\alpha-n+1}} d\tau, \qquad (7.1) $$

$$ (n-1 \le \alpha < n) $$

where $\Gamma(z)$ is Euler's gamma function.

Above Caputo definition is extremely useful in the time domain studies, because the initial conditions for the fractional-order differential equations with the Caputo derivatives can be given in the same form as for the integer-order differential equations. This is an advantage in applied problems, which require the use of initial conditions containing starting values of the function and its integer-order derivatives $f(a)$, $f'(a)$, $f''(a)$, ..., $f^{(n-1)}(a)$.

The formula for the Laplace transform of the Caputo fractional derivative (7.1) has the form [1]:

$$ \int_0^\infty e^{-st}\, {}_0^C D_t^\alpha f(t)\, dt = s^\alpha F(s) - \sum_{k=0}^{n-1} s^{\alpha-k-1}\, f^{(k)}(0), \qquad (7.2) $$

$$(n-1 \leq \alpha < n).$$

If the process $f(t)$ is considered from the state of absolute rest, so $f(t)$ and its integer-order derivatives are equal to zero at the starting time $t = 0$, then the Laplace transform of the α-th derivative of $f(t)$ is simply $s^{\alpha}F(s)$.

7.3 The Mittag-Leffler function

As it is obvious from its name, the Mittag-Leffler function was introduced by G. M. Mittag-Leffler. This function is a generalization of exponential function, and it plays in the fractional-order calculus the same fundamental role as the exponential function plays in the classical integer-order calculus and integer-order differential equations. Many known functions, which we used to consider as different, are, in fact, just particular cases of the Mittag-Leffler function.

7.4 Data fitting using the Mittag-Leffler function

In order to provide a tool for quick and easy creation of models of arbitrary real (integer and non-integer) order, we have developed a new approach to data fitting, which is based on using the Mittag-Leffler function.

The idea of our method is based on the following. When it comes to obtaining a mathematical models from measurements or observations, it is a common practice in many fields of science and engineering to choose the type of the fitting curve and identify its parameters using some criterion (usually a least squares method). We would like to point out that choosing a particular type of a curve means that, in fact, the process is modeled by a differential equation, for which that curve is a solution.

For example, fitting data using the equation $y(t) = at + b$ (known as linear regression model) means that the process is modeled by the solution of a simple second-order differential equation under two initial conditions:

$$y'' = 0, \quad y(0) = b, \quad y'(0) = a. \tag{7.3}$$

Similarly, the fitting function in the form $y = a\sin(\omega t) + b\cos(\omega t)$ means that the process is modeled by the solution of the initial value problem of the form

$$y'' + \omega^2 y = 0, \quad y(0) = b, \quad y'(0) = a\omega. \tag{7.4}$$

Choosing the fitting function in another frequently used form, $y = ae^{bt}$, means that the process is modeled by the solution of the initial value problem

$$y' - b\,y = 0, \quad y(0) = a. \tag{7.5}$$

Thinking in this way, we conclude that instead of postulating the shape of the fitting curve it is possible to postulate the form of the initial-value problem and identify the parameters appearing in the differential equation and in the initial conditions. For the first time this method was suggested and used in [1, Chapter 10]. We would like to emphasize that obtaining a fitting function $y(t)$ for measurements of a dynamic process immediately means that that process is described by an initial-value problem of which $y(t)$ is the solution.

Suppose that the measured data are fitted by

$$y = y_0\, E_{\alpha,1}(a\, t^{\alpha}) \tag{7.6}$$

where $E_{\alpha,\beta}(z)$ is the Mittag-Leffler function defined as [1]

$$E_{\alpha,\beta}(z) = \sum_{k=0}^{\infty} \frac{z^k}{\Gamma(\alpha k + \beta)}. \tag{7.7}$$

The parameters to be identified are α, a, and y_0.

For example, If the data are fitted by the function (7.6), then this means that they are modeled by the solution of the following initial-value problem for a two-term fractional-order differential equation containing the Caputo fractional derivative of order α:

$${}_0^C D_t^{\alpha} y(t) - a\, y(t) = 0, \quad y(0) = y_0. \tag{7.8}$$

7.5 Examples

The proposed method of fitting is illustrated below on several examples, which includes "restoration" of the Mittag-Leffler function from its noised values, fitting a complimentary error function, fitting a sine wave, and fitting damped oscillations.

7.5.1 Fitting back the noised Mittag-Leffler function:

The first example is the "restoration" of the function $y(x) = 0.8E_{1.5}(-0.2x^{1.5})$. Three series of "measured" data are created by adding noise to the values of the Mittag-Leffler function at the same set of nodes x. Such noisy data are fitted using MLFFIT1.M function.

```
% Define the set of nodes (x)
% and the parameter $\alpha$ of the
% Mittag-Leffler function:
x = 0:0.35:20;
alfa = 1.5;

% Since for computing the one-parameter
% Mittag-Leffler function
% we call the Matlab function for computing
% the two-parameter Mittag-Leffler function,
% the second parameter is equal to 1:
beta = 1;

% Now let us simulate measurements
% by adding noise to the exact values
% of the original function
y1 = 0.8*mlf(alfa, beta, -0.2*x.^alfa, 6) ...
+ (-.05 + .1*rand(size(x)));
y2 = 0.8*mlf(alfa, beta, -0.2*x.^alfa, 6) ...
+ (-.05 + .1*rand(size(x)));
y3 = 0.8*mlf(alfa, beta, -0.2*x.^alfa, 6) ...
+ (-.05 + .1*rand(size(x)));

% and fit these "measurements"
% by calling MLFFIT1:
[c, R2] = mlffit1([x x x], [y1 y2 y3], ...
[0.5; 0.5; 0.5; -0.5], 6);

% Let us check if the coefficients
% of the fitting Mittag-Leffler function
% are close to the original coefficients:
alpha = c(1)
C = c(3)
a = c(4)

% Finally, we can plot the "measurements",
% the original function and the function
% that fits the "measurement":

xfine = x(1):0.01:x(end);
yfit = c(3)*mlf(c(1),c(2),c(4)*xfine.^c(1),6);
yorig = 0.8*mlf(alfa,beta,-0.2*xfine.^alfa,6);

figure(1)
plot(xfine, yorig, 'r', x,y1, '.b', x, y2, '.g', ...
x, y3, '.m', xfine, yfit, 'k')
grid on
legend('original function', 'noised data', ...
'noised data', 'noised data', 'fitting')
```

The output of the above code was

$$\alpha = 1.4949, \quad C = 0.8053, \quad a = -0.2008,$$

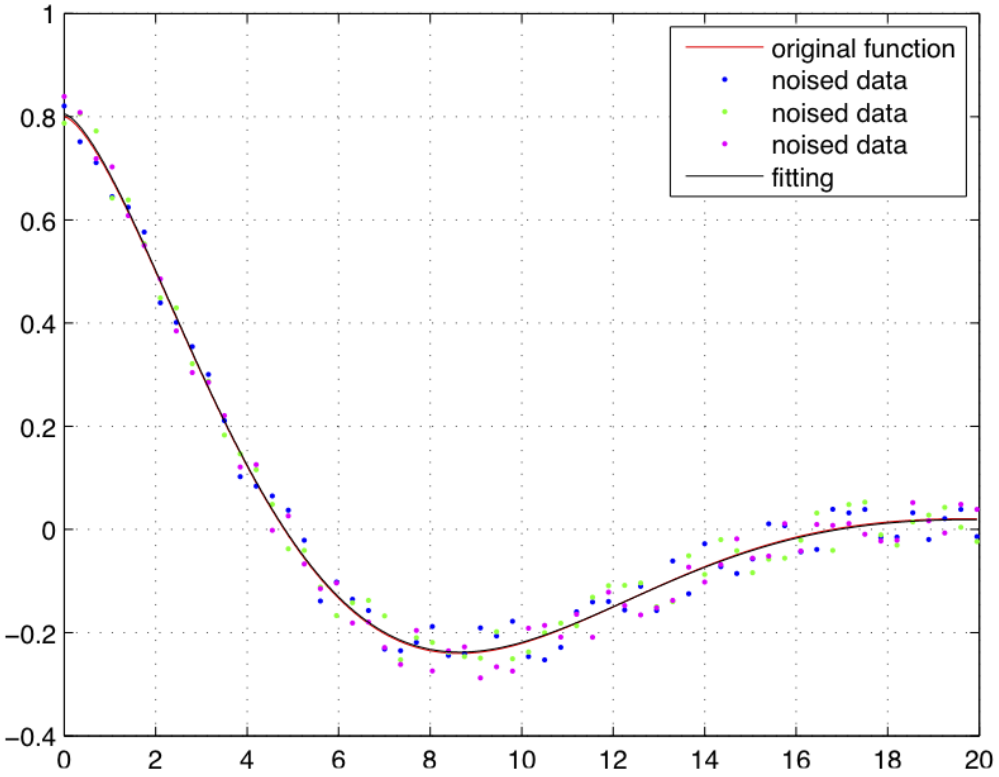


Figure 7.1: Restoring the function $y(x) = 0.8E_{1.5}(-0.2x^{1.5})$

which is very close to the values of these parameters for the original function $y(x) = 0.8E_{1.5}(-0.2x^{1.5})$. The noised data, and the fitting curve are shown in Fig. 7.1.

7.5.2 Fitting the classics (complementary error function):

Let us "restore" the following function: $y(x) = e^x erfc(\sqrt{x})$. it should be mentioned that this function can be written as $y(x) = E_{1/2,1}(-x^{1/2})$.

```
% Define the set of nodes (x)
x = 0:0.35:20;

% Now let us simulate measurements
% by adding noise to  the exact values
% of the original function
y1 = exp(x).*erfc(sqrt(x))  ...
+ (-.02 + .04*rand(size(x)));
y2 = exp(x).*erfc(sqrt(x)) ...
 + (-.02 + .04*rand(size(x)));
y3 = exp(x).*erfc(sqrt(x)) ...
 + (-.02 + .04*rand(size(x)));
```

```
% and fit these "measurements"
% by calling MLFFIT1:
[c, R2] = mlffit1([x x x], [y1 y2 y3], ...
[0.5; 0.5; 0.5; -0.5], 6);

% Let us output the coefficients
% of the fitting Mittag-Leffler function:
alpha = c(1), C = c(3), a = c(4)

% Finally, we can plot the "measurements",
% the original function and the function
% that fits the "measurement":
xfine = x(1):0.01:x(end);
yfit = c(3)*mlf(c(1),c(2),c(4)*xfine.^c(1),6);
yorig = exp(xfine).*erfc(sqrt(xfine));

plot(xfine, yorig,'r',x,y1,'.b',...
x,y2,'.g',x,y3,'.m',xfine,yfit,'k')
grid on
legend('original function','noised data',...
'noised data','noised data','fitting')
```

The output of the above code was

$$\alpha = 0.4966, \quad C = 1.0028, \quad a = -1.0175,$$

which is very close to the values of these parameters for the original function $y(x) = e^x erfc(\sqrt{x}) = E_{1/2,1}(-x^{1/2})$. The noised data, and the fitting curve are shown in Fig. 7.2.

7.5.3 Dumped oscillation fitting:

Let us test if the Mittag-Leffler function is abe to fit dumped oscillations: $y(x) = e^(-\alpha x)\cos(x)$.

```
% Define the dumping coefficient:
alfa = 0.2;

% Define the set of nodes (x)
x = 0:0.35:20;

% Now let us simulate measurements by adding noise to
% the exact values of the original function
y1 = exp(-alfa*x).*cos(x)+(-.05+.1*rand(size(x)));
y2 = exp(-alfa*x).*cos(x)+(-.05+.1*rand(size(x)));
y3 = exp(-alfa*x).*cos(x)+(-.05+.1*rand(size(x)));

% and fit these "measurements" by calling MLFFIT1:
[c, R2] = mlffit1([x x x], [y1 y2 y3], ...
[0.5; 0.5; 0.5; -0.5], 6);

% Let us output if the coefficients
% of the fitting Mittag-Leffler function:
alpha = c(1), C = c(3), a = c(4)

% Finally, we can plot the "measurements",
```

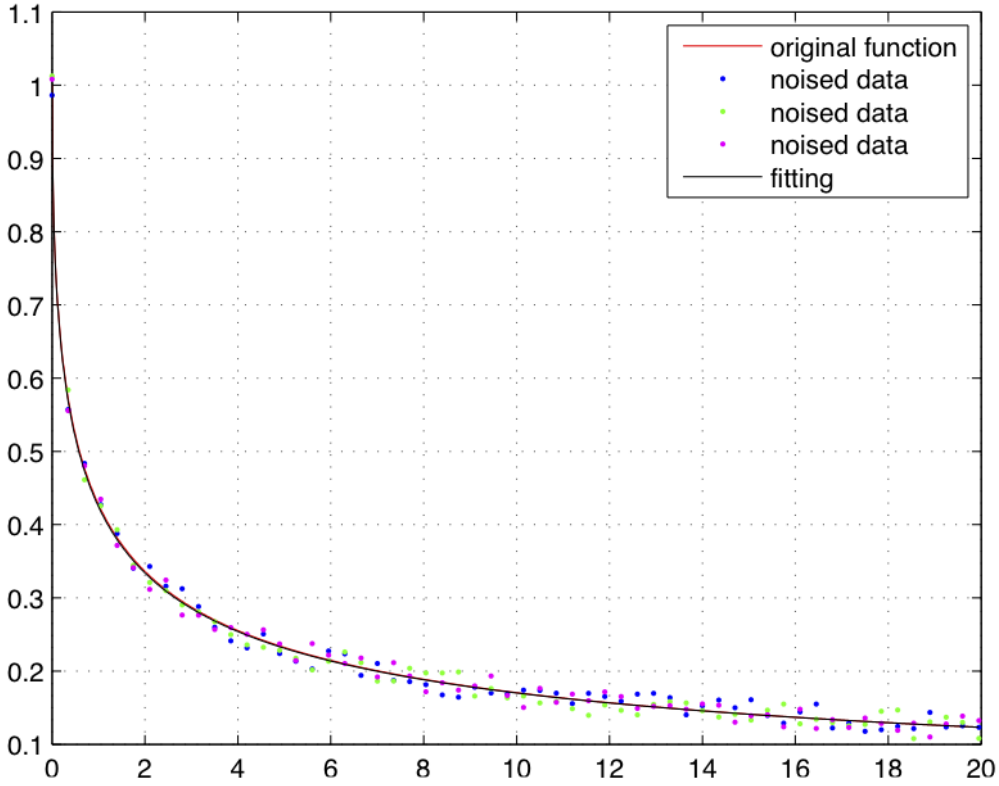


Figure 7.2: Restoring the function $y(x) = e^x erfc(\sqrt{x})$

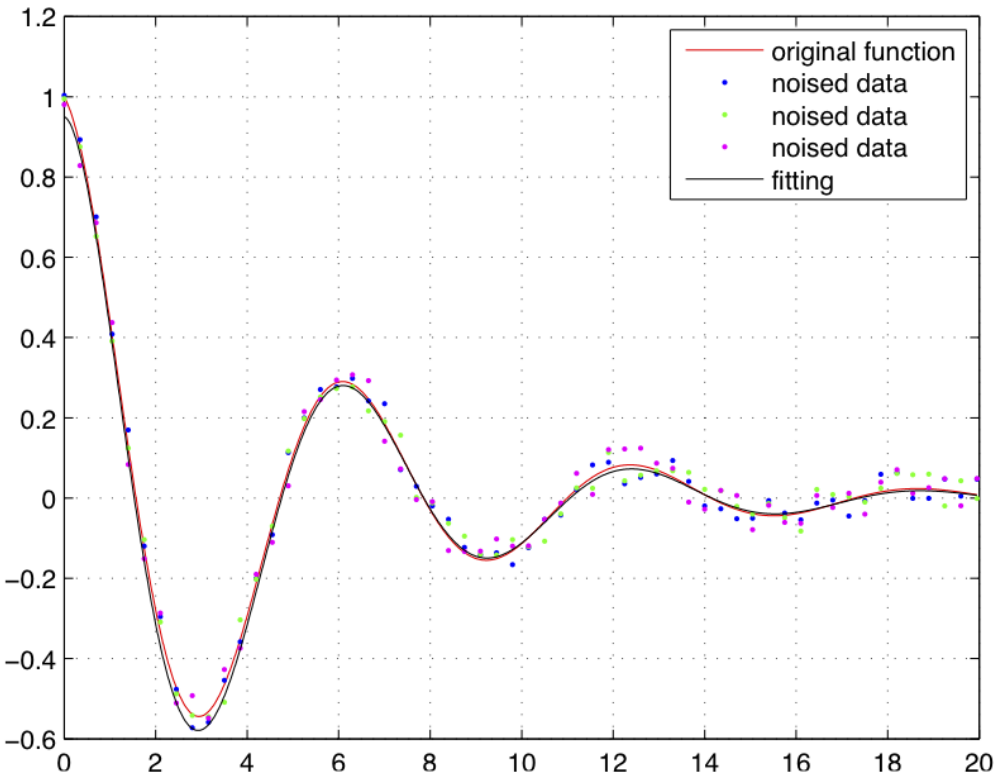


Figure 7.3: Fitting damped oscillations $y(x) = e^{(} - 0.2\,x)\cos(x)$

```
% the original function and the function
% that fits the "measurement":
xfine = x(1):0.01:x(end);
yfit = c(3)*mlf(c(1),c(2),c(4)*xfine.^c(1),6);
yorig = exp(-alfa*xfine).*cos(xfine);

plot(xfine,yorig,'r',x,y1,'.b',x,y2,'.g', ...
x,y3,'.m',xfine,yfit,'k')
legend('original function','noised data',...
'noised data','noised data','fitting')
```

The output of the above code was

$$\alpha = 1.7631, \quad C = 0.9495, \quad a = -1.0340.$$

The noised data, and the fitting curve are shown in Fig. 7.3.

7.6 Chapter summary

The Mittag-Leffler function can be used as a universal fitting function, which is capable of capturing the behavior of various types of processes, including such practically important cases as monotonic processes,

oscillatory behavior, and damped oscillations. There is no need in postulating a narrow type of the fitting function anymore; the MIttag-Leffler function flexibly uncovers the nature of the fitted data. The broad field of potential applicaitons the proposed approach was applied recently for providing the first example of identification of variable-order systems [8].

In addition, as soon as the data are fitted with the help of the Mittag-Leffler function, the process can be described by a two-term fractional-order differential equation. This is an important advantage of the proposed approach to fitting experimental data – it opens a way to creating standard models of elementary processes.

7.7 References

[1] I. Podlubny, *Fractional Differential Equations*, San Diego: Academic Press, 1999.

[2] K. B. Oldham and J. Spanier, *The Fractional Calculus*, New York: Academic Press, 1974.

[3] R. L. Magin, *Fractional Calculus in Bioengineering*, Connecticut: Begell House Publishers, 2006.

[4] C. A. Monje, Y. Q. Chen, B. M. Vinagre, D. Xue and V. Feliu, *Fractional-order Systems and Controls*, Series: Advances in Industrial Control, Springer, 2010.

[5] R. Caponetto, G. Dongola, L. Fortuna, and I. Petráš, *Fractional Order Systems: Modeling and Control Applications*, Singapore: World Scientific, 2010.

[6] I. Podlubny. *Fitting data using the Mittag-Leffler function.* Matlab Central File Exchange, submission #32170, 11 Jul 2011 (Updated 19 Jul 2011)
http://www.mathworks.com/matlabcentral/fileexchange/32170

[7] I. Podlubny. *Mittag-Leffler function.* Matlab Central File Exchange, submission #8738, 17 Oct 2005 (Updated 25 Mar 2009)
http://www.mathworks.com/matlabcentral/fileexchange/8738

[8] D. Sierociuk, I. Podlubny, and I. Petras; , Experimental evidence of variable-order behavior of ladders and nested ladders, EEE Transactions on Control Systems Technology I, 2012, doi: 10.1109/TCST.2012.2185932 (early access).

Chapter 8

Identification of parameters of a variable-order system

8.1 Introduction

The standard notation for denoting the left-sided fractional-order differentiation of a function $f(t)$ defined in the interval $[a, b]$ is ${}_aD_t^\alpha f(t)$, with $\alpha \in \mathrm{R}$. Sometimes a simplified notation $f^{(\alpha)}(t)$ or $d^\alpha f(t)/dt^\alpha$ is used. In some applications also right-sided fractional derivatives ${}_tD_b^\alpha f(t)$ are used, but in the present article we will use only left-sided fractional derivatives. Even from the notation one can see that evaluation of the left-sided fractional-order operators require the values of the function $f(t)$ in the interval $[a, t]$. When α becomes an integer number, this interval shrinks to the vicinity of the point t, and we obtain the classical integer-order derivatives as particular cases.

There are several definitions of the fractional derivatives and integrals, of which we recall only the following two.

The Caputo definition of fractional differentiation can be written as [1]:

$$ {}_a^C D_t^\alpha f(t) = \frac{1}{\Gamma(n-\alpha)} \int_a^t \frac{f^{(n)}(\tau)}{(t-\tau)^{\alpha-n+1}} d\tau, \tag{8.1} $$

$$ (n-1 \leq \alpha < n) $$

where $\Gamma(z)$ is Euler's gamma function.

The Caputo definition is extremely useful in the time domain studies, because the initial conditions for the fractional-order differential equations with the Caputo derivatives can be given in the same form as for the integer-order differential equations. This is an advantage in applied problems, which require the use of initial conditions containing starting values of the function and its integer-order derivatives $f(a)$, $f^{'}(a)$, $f^{''}(a)$, $\ldots$, $f^{(n-1)}(a)$.

The formula for the Laplace transform of the Caputo fractional derivative (8.1) has the form [1]:

$$ \int_0^\infty e^{-st}\, {}_0^C D_t^\alpha f(t)\, dt = s^\alpha F(s) - \sum_{k=0}^{n-1} s^{\alpha-k-1} f^{(k)}(0), \tag{8.2} $$

$$ (n-1 \leq \alpha < n). $$

If the process $f(t)$ is considered from the state of absolute rest, so $f(t)$ and its integer-order derivatives are equal to zero at the starting time $t = 0$, then the Laplace transform of the α-th derivative of $f(t)$ is simply $s^\alpha F(s)$.

The second definition, which we need, is the definition of the left-sided Caputo-Weyl fractional derivative:

$$ {}_{-\infty}^{W}D_t^\alpha f(t) = \frac{1}{\Gamma(n-\alpha)} \int_{-\infty}^{t} \frac{f^{(n)}(\tau)d\tau}{(t-\tau)^{\alpha-n+1}}, \quad (n-1 \leq \alpha < n) \tag{8.3} $$

The Fourier transform of ${}_{-\infty}^{W}D_t^\alpha f(t)$ is simply $(j\omega)^\alpha$. The Caputo-Weyl definition must be used in the frequency domain studies of fractional-order systems. The Caputo-Weyl derivative can be considered as the Caputo derivative with $a \to -\infty$. In other words, the Caputo definition allows the study of the transient effects in fractional-order systems, which were initially at the state of rest, while the Caputo-Weyl definition allows the study of frequency responses of such systems.

Fractional-order models have been already used for modeling of electrical circuits (such as domino ladders, tree structures, etc.) and elements (coils, memristor, etc.). The review of such models can be found in [6, 7, 8].

Let us consider, for instance, a capacitor as a basic element of many circuits. Westerlund and Ekstam in 1994 proposed a new linear capacitor model [9]. It is based on Curie's empirical law of 1889 which states that the current through a capacitor is

$$ i(t) = \frac{u_0}{h_1 t^\alpha}, $$

where h_1 and α are constants, u_0 is the dc voltage applied at $t = 0$, and $0 < \alpha < 1$, $(\alpha \in \mathrm{R})$.

For a general input voltage $u(t)$ the current is

$$ i(t) = C\frac{d^\alpha u(t)}{dt^\alpha} \equiv C\,{}_0D_t^\alpha u(t), \tag{8.4} $$

where C is capacitance of the capacitor. It is related to the kind of dielectric used in the capacitor. The order α is related to the losses of the capacitor. Westerlund and Ekstam provided in their work the table of various capacitor dielectrics with appropriated constants α which have been obtained experimentally by measurements.

The relationship between the current and the voltage in a capacitor is described using fractional-order integration:

$$ u(t) = \frac{1}{C}\int_0^t i(t)dt^\alpha \equiv \frac{1}{C}\,{}_0D_t^{-\alpha} i(t). \tag{8.5} $$

Then the impedance of a fractional capacitor is:

$$ Z_c(s) = \frac{1}{C\,s^\alpha} = \frac{1}{\omega^\alpha C} e^{j(-\alpha\frac{\pi}{2})}, \quad \omega \in (-\infty, \infty). \tag{8.6} $$

Ideal Bode's characteristics of the transfer function for a real capacitor (8.6) are depicted in Fig. 8.1.

General characteristics of the transfer function of a real capacitor (8.6) are [10]:

- Magnitude: constant slope of $-\alpha 20dB/dec.$;
- Crossover frequency: a function of $1/C$;
- Phase: horizontal line of $-\alpha\frac{\pi}{2}$;

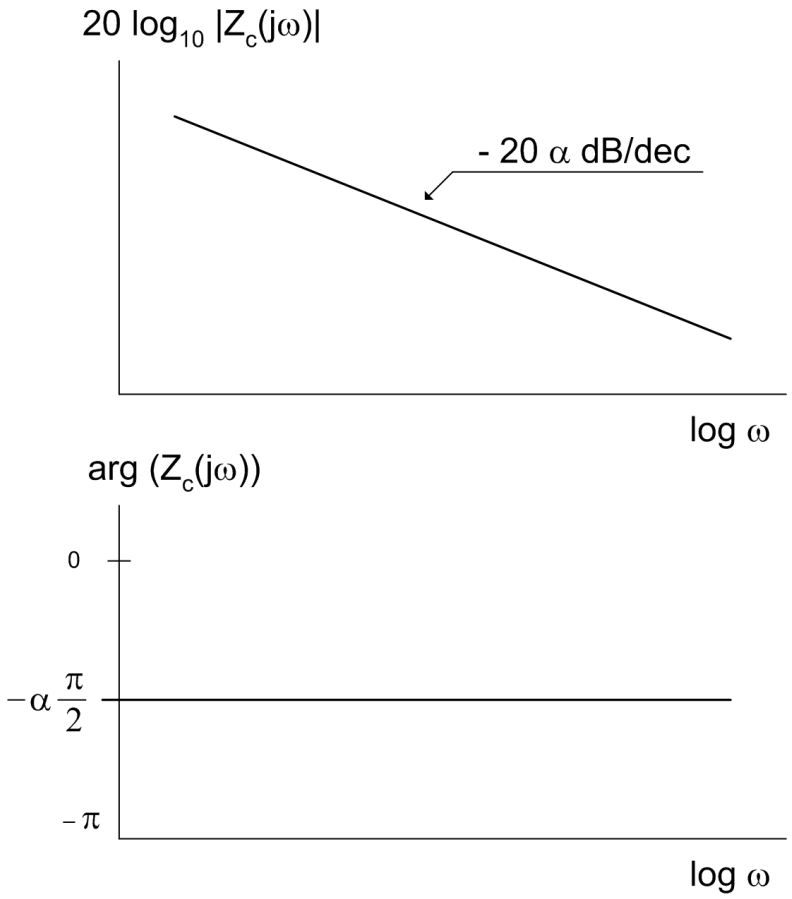


Figure 8.1: Bode plots of real capacitor.

Besides this fractional-order capacitor model, we can mention the new fractional-order models of coils [11], memristive systems [12], ultracapacitors [13, 14], and the element called fractor [15]. Such elements can be combined with classical passive and active elements for creating various types of electrical circuits.

Among the aforementioned fractional-order elements, the fractor is of special interest, because it is known that the order of fractor slowly changes in time with aging of chemical materials of which it is composed [16, Table I]. In other words, fractor is an example of an element of variable non-integer order. Such variable-order behavior of the fractor was experimentally studied in [16]. In this chapter we demonstrate that variable-order behavior can be observed in a wide class of ladder-type circuits composed of standard passive elements.

8.2 Fractional Devices and Fractance

Besides simple elements like a capacitor, electrical circuits of more or less complex structure were studied by many authors. The review of most of the previous efforts can be found in [6]. A circuit that exhibits fractional-order behavior is called a *fractance* [1].

8.2.1 Fractances

The fractance devices have the following characteristics [17]. First, the phase angle is constant independent of the frequency within a wide frequency band. Second, it is possible to construct a filter having a moderate characteristics which can not be realized by using the conventional devices.

Generally speaking, there are three basic types of fractances. The most popular is a domino ladder circuit network [22]. Another type is a tree structure of electrical elements [17], and finally, we can find out also some transmission line circuit (or symmetrical domino ladder [18]).

Design of fractances having given order α can be done easily using any of the rational approximations or a truncated continued fraction expansion (CFE), which also gives a rational approximation [19, 20]. Truncated CFE does not require any further transformation; a rational approximation based on any other methods must be first transformed to the form of a continued fraction; then the values of the electrical elements, which are necessary for building a fractance, are determined from the obtained finite continued fraction. If all coefficients of the obtained finite continued fraction are positive, then the fractance can be made of classical passive elements (resistors and capacitors). If some of the coefficients are negative, then the fractance can be made with the help of negative impedance converters [6, 19].

It is worth mentioning also the constant phase element (CPE), which exhibits the fractional-order behavior as well. It is a metal-insulator-solution or metal-insulator-liquid interface used in electrochemistry. CPE interprets a dipole layer capacitance [21]. The impedance of CPE is expressed as $Z_{CPE}(s) = Qs^{-\alpha}$ and CPE cannot be described by a finite number of passive elements with frequency independent values.

8.2.2 Traditional domino ladder (half-order integrator)

Several different algorithms for approximation the fractional order integrators are currently available [6, 22, 23, 24, 25, 26]. Most of them are based on some form of approximation of irrational transfer functions in the complex domain. The commonly used approaches include the aforementioned CFE method and its modifications, or representation by a quotient of polynomials in s in a factorized form.

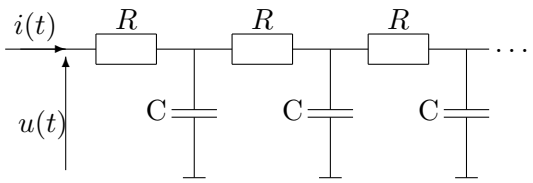


Figure 8.2: Domino ladder scheme.

The main disadvantage of these algorithms is that the values of electrical elements (like resistors and capacitors) needed for the approximation are not the standard values of elements produced by manufacturers.

However, it is still possible to obtain highly accurate and practically usable implementations of a fractional-order integrator using only standard elements with the standard values available in the market. The idea of this practical approach to implementation of fractional-order systems is based on the domino ladder structure.

The domino ladder circuit shown in Fig. 8.2 has the following impedance:

$$G(s) = R + \cfrac{1}{sC + \cfrac{1}{R + \cfrac{1}{sC + \cfrac{1}{R + \cfrac{1}{sC + \cfrac{1}{R + \cfrac{1}{sC + \dots}}}}}}} = \frac{1}{(Ts)^{0.5}}, \tag{8.7}$$

where $T = C/R$. In the ideal case of infinite realization, (8.7) gives a half-order integrator; a truncated realization gives its approximation.

The domino ladder circuit can be also considered as a model of a semi-infinite RC line, which is described by the following partial differential equations [27, 28]:

$$\frac{\partial}{\partial x} u(t,x) = Ri(t,x), \tag{8.8}$$

$$\frac{\partial}{\partial x} i(t,x) = C\frac{\partial}{\partial t} u(t,x), \tag{8.9}$$

where $u(t,x)$ is the voltage and $i(t,x)$ is the current at point x at time instance t.

This can be rewritten as

$$\frac{\partial^2}{\partial x^2} u(t,x) = RC\frac{\partial}{\partial t} u(t,x). \tag{8.10}$$

From this equation a relationship between the current $i(t,0)$ and voltage $u(t,0)$ at the beginning of the semi-infinite RC line can be obtained in terms of half-order integral; in the Laplace domain it has the following form:

$$G(s) = \frac{U(s,0)}{I(s,0)} = \sqrt{\frac{R}{C}}\frac{1}{s^{0.5}}, \tag{8.11}$$

where $I(s,0)$ and $U(s,0)$ are the Laplace transforms of $i(t,0)$ and $u(t,0)$.

8.2.3 Domino ladder with alternating resistors

For building accurate analog approximation of the half-order integrator using easily accessible elements available in the market, the approach presented in Fig. 8.3 can be used.

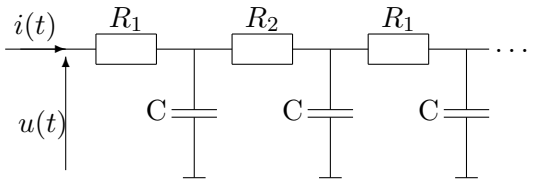


Figure 8.3: Proposed analogue model of half-order integrator.

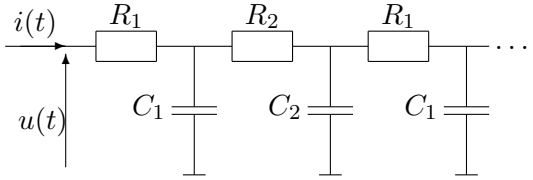


Figure 8.4: Enhanced domino ladder for half-order integration.

Based on the observation made in article [29], we can formulate the following design algorithm:

(a) Choose the values of R_1 and C in order to obtain the required low frequency limit.

(b) Choose value of R_2 in order to satisfy the condition $R_2 \approx 4R_1$. This condition allows to select those values of resistors that are available as manufactured.

(c) Choose the ladder length n (number of steps in the domino ladder) in order to obtain the desired frequency range of approximation.

8.2.4 Enhanced domino ladder for half-order integration

The modified ladder with two alternating values of resistors performs better than the classical domino ladder, but the phase shift is still equal not to 45°, but to approximately 43°–44° (45° achieved only at very short frequency range).

To further improve the accuracy of approximation, let us modify the structure of the domino ladder in such a way that there are not only two values of resistors, but also two values of capacitors are used (Fig. 8.4).

Fig 8.5 presents the experimental results for the enhanced domino ladder for half-order integration with the following parameters of the circuit presented in Fig. 8.4: $R_1 = 2320\Omega$, $R_2 = 8200\Omega$, $C_1 = 330\text{nF}$, $C_2 = 220\text{nF}$, and the number of steps in the ladder is equal to $n = 34$. The results are compared with the realization presented in Fig. 8.3. It is obvious that the phase plot of the enhanced ladder is indeed is much closer to the 45° than in the case of the classical domino ladder.

8.2.5 A new type of fractances: a nested ladder

Based on the above results, we can easily extend them to build a fractional order integrator of order 0.25. This can be done by replacing the capacitors in the scheme in Fig. 8.3 by half-order integrators,

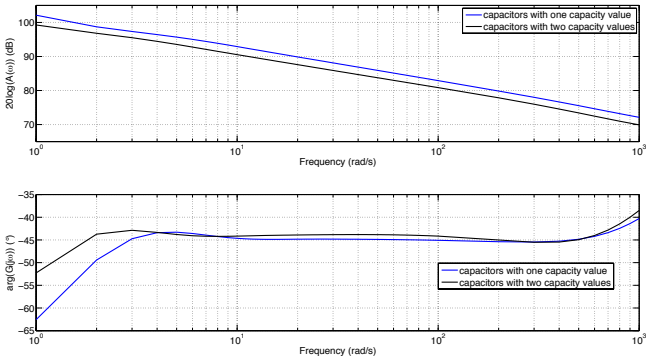


Figure 8.5: Results of modeling of half-order integrator using a modified domino ladder (dotted line: capacitors with one capacity value; solid line: capacitors with two capacity values).

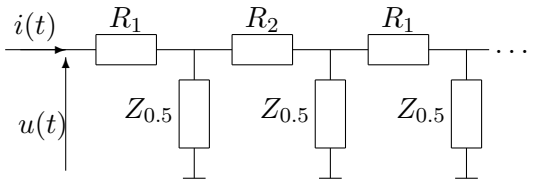


Figure 8.6: Integrator of order $\alpha = 0.25$ in the form of a nested ladder.

which can be either classical domino ladders or enhanced domino ladders. This step can be interpreted as an introduction a half-order dynamics into the equation (8.9). This results in a transfer function of order $\alpha = 0.25$, which corresponds to a quarter-order integrator.

In Fig. 8.6 the scheme of the approximation of a quarter-order integrator is shown; $Z_{0.5}$ are the impedances of modified or enhanced domino ladders implementing half-order integrators.

In the same way (namely, by replacing impedances $Z_{0.5}$ with $Z_{0.25}$) an integrator of order $\alpha = 0.125$ and so forth can be built, but this will need a large number of elements.

We call such a structure of electrical circuit the *nested ladder*. The nested ladder is an example of using the ideas of self-similarity and fractality for creating electrical circuits exhibiting non-integer order behavior.

8.3 Data fitting using the Mittag-Leffler function

In order to obtain a model for the measured data from the considered electrical circuits (ladders and nested ladders), we have developed a new approach to data fitting, which is based on using the Mittag-Leffler function and which, in fact, allows obtaining models of non-integer order.

The idea of our method is based on the following. When it comes to obtaining a mathematical models

from measurements or observations, it is a common practice in many fields of science and engineering to choose the type of the fitting curve and identify its parameters using some criterion (usually a least squares method). We would like to point out that choosing a particular type of a curve means that, in fact, the process is modeled by a differential equation, for which that curve is a solution.

For example, fitting data using the equation $y(t) = at + b$ (known as linear regression model) means that the process is modeled by the solution of a simple second-order differential equation under two initial conditions:

$$y'' = 0, \quad y(0) = b, \quad y'(0) = a. \tag{8.12}$$

Similarly, the fitting function in the form $y = a\sin(\omega t) + b\cos(\omega t)$ means that the process is modeled by the solution of the initial value problem of the form

$$y'' + \omega^2 y = 0, \quad y(0) = b, \quad y'(0) = a\omega. \tag{8.13}$$

Choosing the fitting function in another frequently used form, $y = ae^{bt}$, means that the process is modeled by the solution of the initial value problem

$$y' - b\,y = 0, \quad y(0) = a. \tag{8.14}$$

Thinking in this way, we conclude that instead of postulating the shape of the fitting curve it is possible to postulate the form of the initial-value problem and identify the parameters appearing in the differential equation and in the initial conditions. For the first time this method was suggested and used in [1, Chapter 10]. In this chapter we, however, just emphasize that obtaining a fitting function $y(t)$ for measurements of a dynamic process immediately means that that process is described by an initial-value problem of which $y(t)$ is the solution.

In the present article the measured data are fitted by

$$y = y_0\, E_{\alpha,1}(a\, t^\alpha) \tag{8.15}$$

where $E_{\alpha,\beta}(z)$ is the Mittag-Leffler function defined as [1]

$$E_{\alpha,\beta}(z) = \sum_{k=0}^{\infty} \frac{z^k}{\Gamma(\alpha k + \beta)}. \tag{8.16}$$

The parameters to be identified are α, a, and y_0.

If the data are fitted by the function (8.15), then this means that they are modeled by the solution of the following initial-value problem for a two-term fractional-order differential equation containing the Caputo fractional derivative of order α:

$${}_0^C D_t^\alpha y(t) - a\, y(t) = 0, \quad y(0) = y_0. \tag{8.17}$$

8.4 Experimental Results

8.4.1 Experimental setup

For the experimental verification of the introduced method, the circuits presented in Section 8.2 were built. For measurements, the modified domino ladder circuit and the nested ladder were connected to

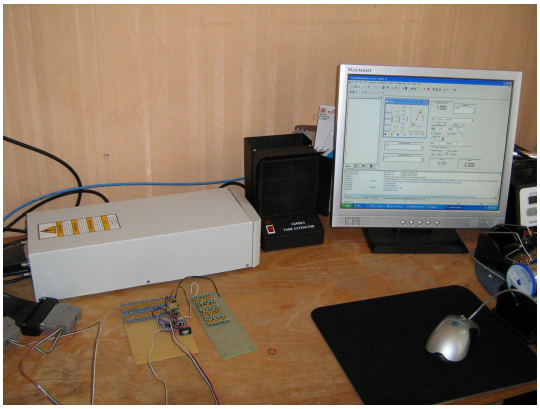


Figure 8.7: Experimental setup used for all measurements: 1 – domino ladder, 2 – nested ladder, 3 – dSpace card, 4 – computer with Matlab/Simulink software.

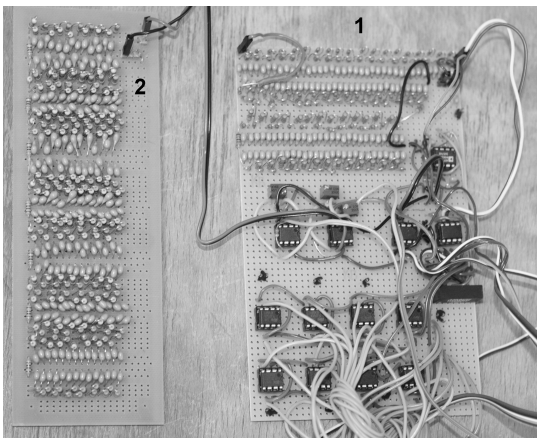


Figure 8.8: Detailed view: 1 – domino ladder, 2 – nested ladder.

the amplifier electronic circuit of the operational amplifiers TL071 and to the dSpace DS1103 DSP card connected to a computer. The real laboratory setup is shown in Fig. 8.7 and detailed view of the ladders is in Fig. 8.8.

The electronic scheme presented in Fig. 8.9 uses two operational amplifiers. The first one is working in the integrator configuration and the second one is working in the inverse unit-gain for compensating the signal inversion of the integrator amplifier. The resistor R_i can be used for changing the gain of the integrator and it was chosen to $R_i = 3.3\text{k}\Omega$ in order to keep a gain equal one. The u_1 is an input and u_2 is an output of the integrator system.

8.4.2 Modified half-order domino ladder measurements

The tested circuit has the following parameters of the circuit presented in Fig. 8.3: $R_1 = 2000\Omega$, $R_2 = 8200\Omega$, $C = 470\text{nF}$, and numbers of steps in the ladders was taken first $n = 60$ and then $n = 130$. The sampling period was $T_s = 0.0001$ s. The manufacturing tolerance of the elements used for making

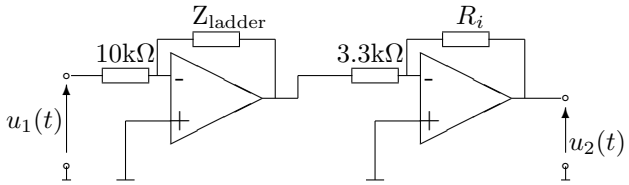


Figure 8.9: Electronic circuit of measurement setup for integrator.

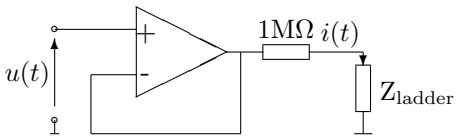


Figure 8.10: Electronic circuit of direct measurement setup for ladders.

such ladders is 1% for resistors and 20% for capacitors. As it can be seen in Fig. 8.11 and Fig. 8.12, the obtained experimental results fully confirm the theoretical considerations and simulations.

8.4.3 Quarter-order domino ladder measurements

The tested circuit has the following parameters of the circuit presented in Fig. 8.3: $R_1 = 2000\Omega$, $R_2 = 8200\Omega$, $C = 470$nF and realization length equal $n = 14 \times 14$, that is 14 sub-ladders with 14 steps each. The sampling period was $T_s = 0.0001$ s. The manufacturing tolerance of the elements used for making such ladders is 1% for resistors and 20% for capacitors. As it can be seen in Fig. 8.13 and Fig. 8.14, the obtained experimental results confirm the theoretical considerations and simulations. A little deviation in the time domain is due to small number of the nested ladder steps, which can be also observed in the frequency domain (only two and half decades approximation).

8.5 Variable-order behavior

If the measurements are obtained for the fixed interval $[0, t]$, then fitting using the Mittag-Leffler function (8.15), described in Section 8.3, immediately gives the model (8.17) of fractional order α.

However, if we consider the changing length of the interval, then the resulting order of the model will be, in general, a function of this changing interval length t: $\alpha = \alpha(t)$. The same holds for other two parameters.

In our experiments we considered the growing number of measurements that are used for fitting the measured data. We increment the length of the time interval by 1 s within first 5 s, and then use the increment of 5 s up to 100 s. This allowed us to better examine the time-domain response of the considered circuits (discharge of both ladders), connected as in Fig. 8.10, near the starting point $t = 0$, and also their time-domain responses in long run, which was in our case the interval up to 100 seconds. Discharges of the 60-steps domino ladder, 130-steps domino ladder, and nested domino ladder are depicted in Figs. 8.15, 8.16, and 8.17, respectively. The sampling period was $T_s = 0.01$ s for all measurements of the discharges used for the computations.

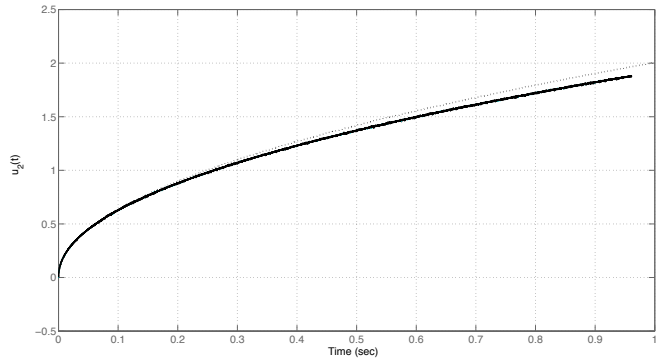


Figure 8.11: Comparison of measured and calculated step responses of half-order integrator with domino ladder of 130 steps: (dotted line) calculated response for $\alpha = 0.5167$ from Table 8.1 for 1 s; (solid line) measured response.

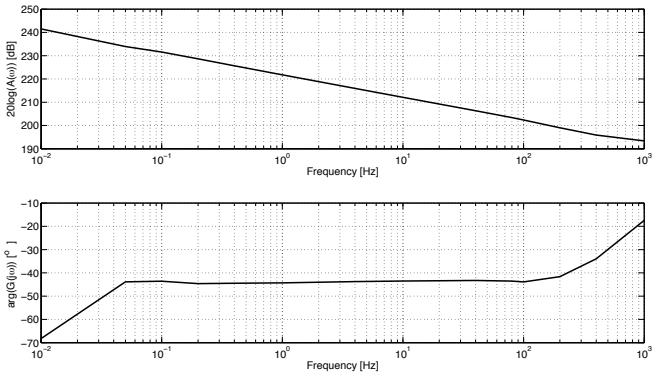


Figure 8.12: Measured Bode plots of half-order integrator with domino ladder of 130 steps.

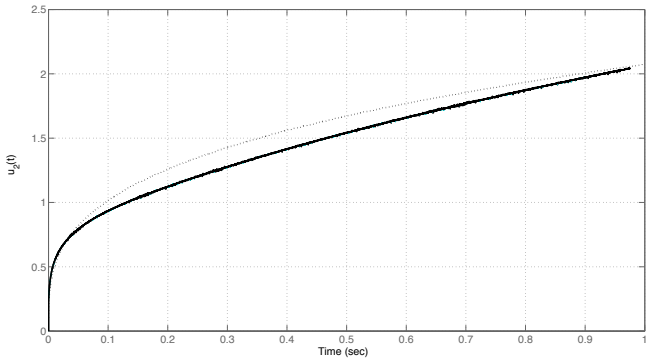


Figure 8.13: Comparison of measured and calculated step responses of quarter-order integrator with nested ladder of size 14×14 steps: (dotted line) calculated response for $\alpha = 0.3126$ from Table 8.1 for 1 s; (solid line) measured response.

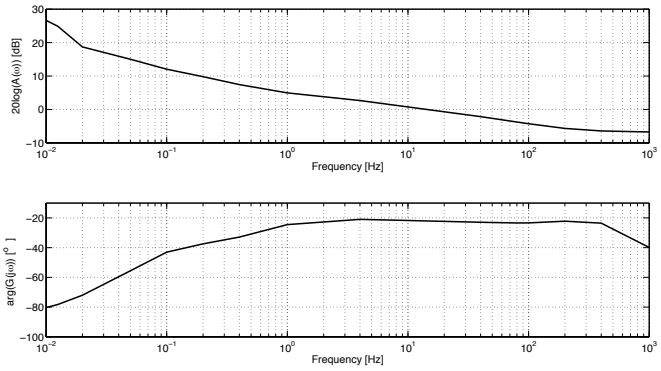


Figure 8.14: Measured Bode plots of quarter-order integrator with nested ladder of size 14×14 steps.

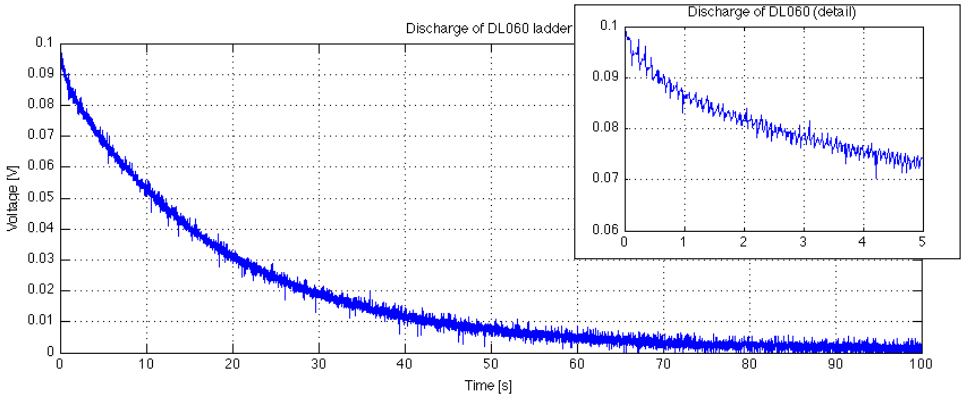


Figure 8.15: Discharge of the 60-steps domino ladder (DL060).

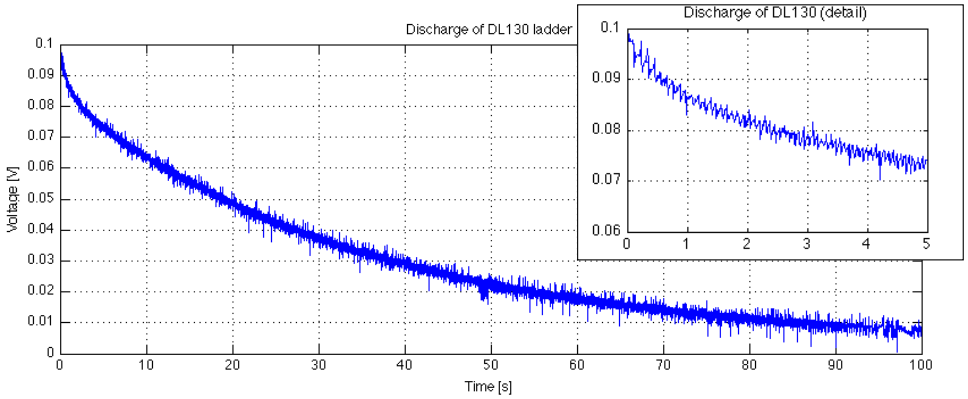


Figure 8.16: Discharge of the 130-steps domino ladder (DL130)).

The results of these computations are presented in Table 8.1 and in the Figs. 8.18 and 8.19.

The method of data fitting using the Mittag-Leffler function is implemented as a Matlab routine [30], and the Mittag-Leffler function is computed also using our Matlab routine [31].

8.6 Discussion

Our main conclusion is that both the domino ladder and the nested ladder exhibit dual behavior in the frequency domain and in the time domain. In some frequency range or in some time interval they behave as fractional-order integrators of (almost) constant order. Outside of that frequency range or outside of that time interval they behave as variable-order integrators; in one case that variable order depends on frequency, in the other case it depends on the time.

The domino ladder behaves as Caputo-Weyl integrator of constant order $\alpha = 0.5$ in a certain frequency range in the frequency domain. This means that in that frequency range it simply shifts the phase by $\alpha\pi/2 = \pi/4$. The frequency range where this behavior is observed can be made larger by increasing the number of steps in the domino ladder. Outside of this frequency range the domino ladder behaves as a

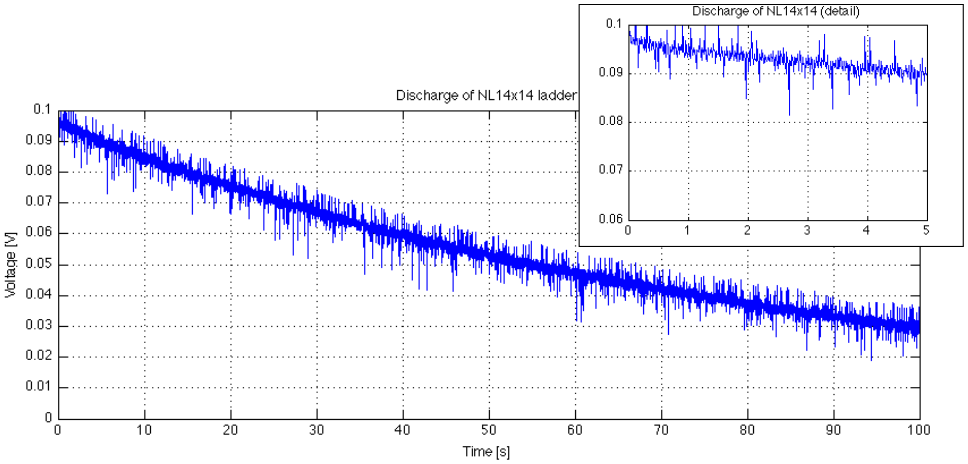


Figure 8.17: Discharge of the nested domino ladder (NL14x14).

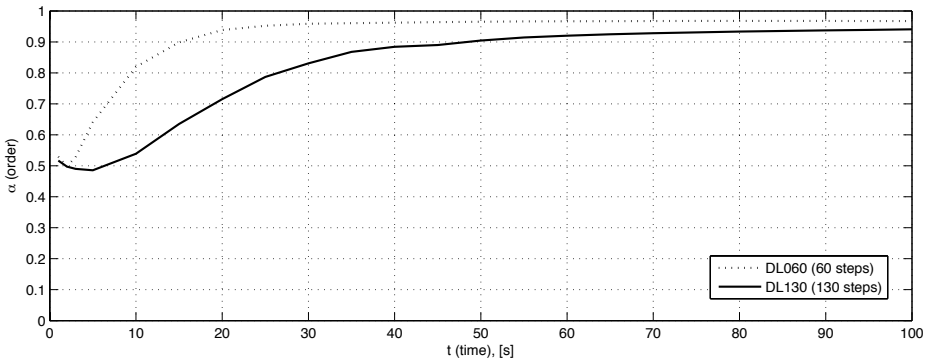


Figure 8.18: Variable order $\alpha(t)$ for the 60-steps domino ladder (DL060), dotted line, and the 130-steps domino ladder (DL130), solid line.

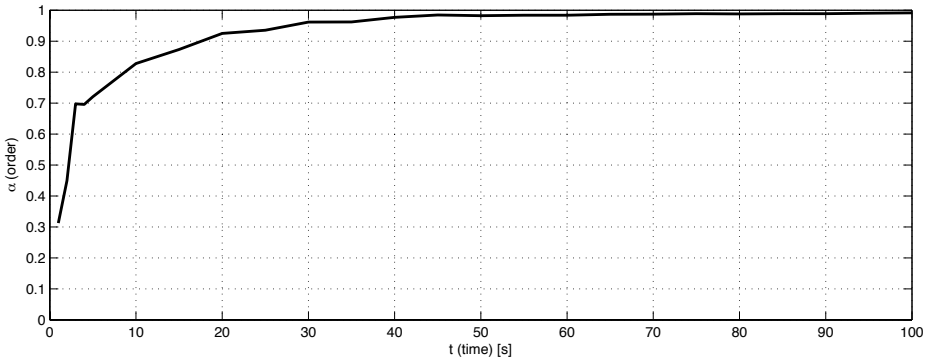


Figure 8.19: Variable order $\alpha(t)$ for the nested domino ladder (NL14x14).

Table 8.1: Variable order $\alpha(t)$ for the 60-steps domino ladder (DL060), the 130-steps domino ladder (DL130), and the nested ladder (NL14x14)

t [s]	$\alpha(t)$		
	DL060	DL130	NL14x14
1	0.5294	0.5167	0.3126
2	0.4984	0.4972	0.4498
3	0.5277	0.4901	0.6978
4	0.5746	0.4821	0.6959
5	0.6408	0.4855	0.7205
10	0.8195	0.5390	0.8278
15	0.8986	0.6326	0.8732
20	0.9385	0.7098	0.9249
25	0.9523	0.7801	0.9354
30	0.9586	0.8227	0.9618
35	0.9604	0.8531	0.9620
40	0.9620	0.8737	0.9770
45	0.9638	0.8900	0.9847
50	0.9651	0.9046	0.9823
55	0.9661	0.9142	0.9837
60	0.9668	0.9201	0.9837
65	0.9670	0.9246	0.9868
70	0.9674	0.9279	0.9872
75	0.9678	0.9307	0.9886
80	0.9677	0.9336	0.9879
85	0.9676	0.9354	0.9888
90	0.9677	0.9373	0.9887
95	0.9672	0.9388	0.9904
100	0.9672	0.9406	0.9915

variable-order system, where the order depends on the frequency, as one can conclude directly from the Bode plots.

At the same time, in the time domain the same domino ladder behaves as an integrator of variable non-integer order, where the order depends on the length of the time interval. Close to the starting time instance $t = 0$, the domino ladder behaves as an integrator of order $\alpha \approx 0.5$, and with growing t the domino ladder behaves closer and closer to the classical integrator of order 1. It should be mentioned that although the domino ladder order, $\alpha(t)$, tends to 1, the order 1 is never reached. However, in many practical applications it is sufficient to neglect the transient effects for some initial time interval and to assume that $\alpha(t) = 1$ for all t.

The similar observations hold for the nested ladder circuit, which has been introduced in this chapter. In some frequency range it behaves as an integrator of order 0.25, and outside of this frequency range it behaves as a variable-order system, where the order depends on the frequency.

In the time domain the nested ladder circuit behaves like a variable-order integrator, with order $\alpha(t)$ starting close to 0.25, and then increasing towards 1; the order 1 is also never reached in the considered

time interval.

A possible explanation of the behavior of both type of fractances in time domain can be the following. For example, in Fig. 8.2, initially the capacitors are not charged and the initial impedance is just that of the first resistor. As the capacitors get charged sequentially, the impedance looks like $2R$, then $3R$ and so on. At very long times, the charged capacitors act as very high impedance, overwhelming the contribution of the resistors. So, the system tends asymptotically toward an exponent of 1.

The frequency range and the time interval, where the order of the nested ladder is close to 0.25, can be extended by increasing the number of levels of the ladders in the nested structure, and by increasing the numbers of steps in those ladders.

8.7 Chapter summary

In this chapter we have presented the experimental study of the two types of electrical circuits made only of passive elements, which exhibit non-integer order behavior. One of them is the domino ladder, which already appeared in the works of other authors on the fractional-order systems. The other one is the circuit that we call the nested ladder and which was introduced in this chapter.

For both these types of circuits we demonstrated that they should be considered not just as non-integer order systems, but as variable-order systems, where the order depends either on the frequency (in the frequency domain) or on the time variable (in the time domain).

While in the frequency domain the frequency-dependent variable order is obvious directly from the Bode plots, providing the evidence of the variable-order behavior of the considered circuits in the time domain required some additional tools. Namely, we suggested a method of data fitting with the help of the Mittag-Leffler function, explained a link between such fitting and fractional-order differential equations, and provided the Matlab routines for such fitting.

The approach to identification of variable-order systems, that we presented in this chapter, can be used for creating variable-order models for many other processes.

8.8 References

[1] I. Podlubny, *Fractional Differential Equations*, San Diego: Academic Press, 1999.

[2] K. B. Oldham and J. Spanier, *The Fractional Calculus*, New York: Academic Press, 1974.

[3] R. L. Magin, *Fractional Calculus in Bioengineering*, Connecticut: Begell House Publishers, 2006.

[4] C. A. Monje, Y. Q. Chen, B. M. Vinagre, D. Xue and V. Feliu, *Fractional-order Systems and Controls*, Series: Advances in Industrial Control, Springer, 2010.

[5] R. Caponetto, G. Dongola, L. Fortuna, and I. Petráš, *Fractional Order Systems: Modeling and Control Applications*, Singapore: World Scientific, 2010.

[6] I. Petráš, I. Podlubny, P. O'Leary, L. Dorčák and B. M. Vinagre, *Analogue Realization of Fractional-Order Controllers*, TU Košice: BERG Faculty, Slovakia, 2002.

[7] I. Petráš, "Fractional-order memristor-based Chua's circuit," *IEEE Transactions on Circuits and Systems-II: Express Briefs*, vol. 57, no. 12, pp. 975–979, 2010.

[8] I. Petráš, *Fractional-Order Nonlinear Systems: Modeling, Analysis and Simulation*, London: Springer and Beijing: HEP, 2011.

[9] S. Westerlund and L. Ekstam, "Capacitor theory," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 1, no. 5, pp. 826–839, 1994.

[10] I. Petráš, Y. Q. Chen and C. Coopmans, "Fractional-order memristive systems," in *Proc. of the 14th IEEE International Conference ETFA'2009*, September 22-26, 2009, Mallorca, Spain.

[11] I. Schafer and K. Kruger, "Modelling of lossy coils using fractional derivatives," *J. Phys. D: Appl. Phys.*, vol. 41, pp. 1–8, 2008.

[12] C. Coopmans, I. Petráš and Y. Q. Chen, "Analogue Fractional-Order Generalized Memristive Devices," in *Proc. of the ASME 2009 Conference*, #86861, San Diego, USA, August 30–September 2, 2009.

[13] A. Dzielinski and D. Sierociuk, "Ultracapacitor modelling and control using discrete fractional order state-space model," *Acta Montanistica Slovaca*, vol. 13, no. 1, pp. 136–145, 2008.

[14] A. Dzielinski, D. Sierociuk and G. Sarwas, "Some applications of fractional order calculus," *Bulletin of The Polish Academy of Sciences-Technical Sciences*, vol. 58, no. 4, pp. 583–592, 2010.

[15] G. Bohannan, "Analog Realization of a Fractional Control Element - Revisited," in *Proc. of the IEEE Int. Conf. on Decision and Control*, 2002, Las Vegas, USA.

[16] H. Sheng, H. G. Sun, C. Coopmans, Y. Q. Chen and G. W. Bohannan, "A Physical experimental study of variable-order fractional integrator and differentiator," *Eur. Phys. J. Special Topics*, vol. 193, pp. 93–104, 2011.

[17] M. Nakagawa and K. Sorimachi, "Basic characteristics of a fractance device," *IEICE Trans. Fundamentals*, vol. E75 - A, no. 12, pp. 1814–1818, 1992.

[18] G. E. Carlson and C. A. Halijak, "Approximation of fractional capacitors $(1/s)^{1/n}$ by a regular Newton process," *IEEE Trans. on Circuit Theory*, vol. 11, no. 2, pp. 210–213, 1964.

[19] I. Podlubny, I. Petráš, B. Vinagre, P. O'Leary and L. Dorčák, "Analogue Realizations of Fractional-order Controllers", *Nonlinear Dynamics*, vol. 29, no. 1–4, pp. 281–296, 2002.

[20] S. Erfani, "Evaluation and realization of continued-fraction expansion revisited," *Computers and Electrical Engineering*, vol. 28, pp. 311–316, 2002.

[21] K. Biswas, S. Sen and P. K. Dutta, "Realization of a Constant Phase Element and Its Performance Study in a Differentiator Circuits," *IEEE Trans. on Circuits and Systems II: Express Brief*, vol. 53, no. 9, pp. 802–806, 2006.

[22] S. D. Roy, "On the Realization of a Constant-Argument Immittance or Fractional Operator," *IEEE Transactions on Circuit Theory*, vol. 14, no. 3, pp. 264–274, September 1967.

[23] S. D. Roy, "Constant argument immittance realization by a distributed RC network," *IEEE Transactions on Circuits and Systems*, vol. 21, no. 5, pp. 655–658, September 1974.

[24] B. T. Krishna and K. V. V. S. Reddy, "Active and Passive Realization of Fractance Device of Order 1/2," *Active and Passive Electronic Components*, vol. 2008, article ID 369421, 5 pages.

[25] P. Yifei, Y. Xiao, L. Ke, Z. Jiliu, Z Ni, Z. Yi and P. Xiaoxian, "Structuring analog fractance circuit for 1/2 order fractional calculus," in *Proc. of The 6th International Conference on ASICON 2005*, vol. 2, 24–30 October, 2005, pp. 1136–1139.

[26] J. C. Wang, "Realizations of generalized Warburg impedance with RC ladder networks and transmission lines," *J. of Electrochem. Soc.*, vol. 134, no. 8, pp. 1915–1920, August 1987.

[27] S. Manabe, "The Non-Integer Integral and its Application to Control Systems," *ETJ of Japan*, vol. 6, no. 3-4, pp. 83–87, 1961.

[28] K. B. Oldham, "Semiintegral Electroanalysis: Analog Implementation," *Analytical Chemistry*, vol. 45, no. 1, pp. 39–47, 1973.

[29] D. Sierociuk and A. Dzielinski, "New method of fractional order integrator analog modeling for orders $\alpha = 0.25$ and $\alpha = 0.5$," in *Proc. of the 16th International Conference on Methods and Models in Automation and Robotics*, 22 – 25 August 2011, Amber Baltic Hotel, Miedzyzdroje, Poland.

[30] I. Podlubny, *Fitting experimental data using the Mittag-Leffler function*, Matlab Central File Exchange, submission #32170. Available at http://www.mathworks.com/matlabcentral/fileexchange/32170.

[31] I. Podlubny and M. Kacenak, *Mittag-Leffler function*, Matlab Central File Exchange, submission #8738. Available at http://www.mathworks.com/matlabcentral/fileexchange/8738.

Chapter 9

Least squares? No! Least circles!

9.1 Introduction

A student came to the final exam on statistical methods in economics. The professor asked him to compute the linear regression of y versus x, and the student successfully computed some a and b of the straight line $y = a + bx$. Then the professor asked the student to compute the linear regression of x versus y, and the student immediately rewrote the previous equation into the form $x = (1/b)y - (a/b)$. The professor was expecting that the student would derive the equation of a conjugate regression line, and evaluated the student's answer as unsatisfactory. But was the student's second line really incorrect? That all depends on how the first line was calculated, which in turns depend on the criterion used for determining a and b in the first line.

9.2 A bit of history

One could hardly name another method which is used as frequently as the method which is known as the least squares method. At the same time, it is difficult to name another method which has been accompanied by such strong and long lasting controversy.

The story of the birth of the least squares method is well covered in the literature and can be summarized as follows. The priority in publication definitely belongs to A. -M. Legendre (Nouvelles méthodes pour la détermination des orbites des comètes, 1805), who also gave the method its famous name. C. F. Gauss (Theoria Motus Corporum Coelestium in Sectionibus Conicis Solem Ambientium, 1809) claimed, however, that he knew and used this method much earlier, about 10 years before Legendre's publications. In a letter to Gauss about his new book Legendre wrote that claims of priority should not be made without a proof by previous publications. Gauss did not have such a publication. Despite this, Gauss was very active in attacking Legendre. We see that his efforts were fruitful enough: in vast majority of today's textbooks the least squares method is attributed to Gauss without any further comments. In fact, Gauss's arguments for his priority were not perfect at all. His diaries with computations claimed to be made by the least squares method were lost. His colleagues did not hurry to acknowledge that he showed him those computations. Indeed, can one imagine that Gauss showed and explained the details of his unpublished computations to his potential competitors? Only many years later did H.W.M. Olbers

(1816) and F.W. Bessel (1832) mention that Gauss showed them something in that sense. But how accurately could they really remember the details of some discussion that happened many years ago? It is also known that H. C. Schumacher suggested repeating Gauss's lost computations claimed to be done by the least squares method. Gauss totally rejected this idea with the words that such attempts would only suggest that he could not be trusted. This, however, has been done by Stigler (1981), who could not reproduce Gauss's results. Later Celmins (1998) also tried to repeat Gauss's computations, including the adjustments suggested by Stigler (1981), and arrived at the same conclusion that Gauss's results cannot be obtained by the least squares method. In other words, it is well known which method Legendre used, and it is not clear at all which method was used by Gauss.

Assuming that it was Gauss who invented the least squares method, it is hard to believe that he did not realize the huge potential of this method and its importance for applications. Knowing Gauss as a prolific mathematician and looking at the present version of the least squares method, one can see a certain contradiction.

9.3 Least squares method

It is difficult to find another method that is both so easy and at the same time so artificial. Figure 9.1 is a version of the picture which appears frequently in the textbooks, slides, and blackboards as a geometric illustration of the least squares method. Let us recall how this figure is created. A set of points (x_k, y_k), $k = 1, \ldots, N$, is approximated by a line $y = a + bx$. The classical least squares fitting consists in minimizing the sum of the squares of the vertical deviations of a set of data points

$$E = \sum_i [y_i - (a + bx_i)]^2 \tag{9.1}$$

from a chosen linear function $y = a+bx$. Each term in (9.1) corresponds to a square in Figure 9.1.

In the opinion of the authors, this picture is ugly. It does not have any sign of mathematical beauty. It could be good for Malevich or Kandinsky, but not for Gauss. The line and the squares are in some visual conflict. This conflict is even more obvious if we assume that the coordinate system is not rectangular. Figure 9.2 gives an illustration. This conflict is absolutely obvious if we would consider, for example, polar or elliptic coordinates. It is difficult to imagine that Gauss would have been happy with such visual interpretation.

The key idea here is that the visualization and, more important, the resulting approximation is dependent on the choice of the coordinate system. The vertical distance is not independent of the coordinate system in which vertical is defined. But some definitions of distance are invariant to the coordinate system!

Shape Recognition and Curve Detection

Nowadays, the distance between two points in k-dimensional space is widely used as an optimal fitting criterion in the field of image processing for industrial and scientific applications, especially in problems of shape recognition and curve detection. An ellipse (a circle) is an ellipse (a circle) in any coordinate system. A parabola is a parabola in any coordinate system, too. Those objects are not defined by equations, but by their general properties, that include the notion of distance. Indeed, everybody knows from the school that a circle is a set of points in a plane that are at the same distance from a given point; an ellipse is a set of point for which the sum of distance from two given points is constant; a parabola is a set of points which are at the same distance from a given point and from a given straight line, and so forth. We sometimes forget that those geometric objects are not related to any particular coordinate

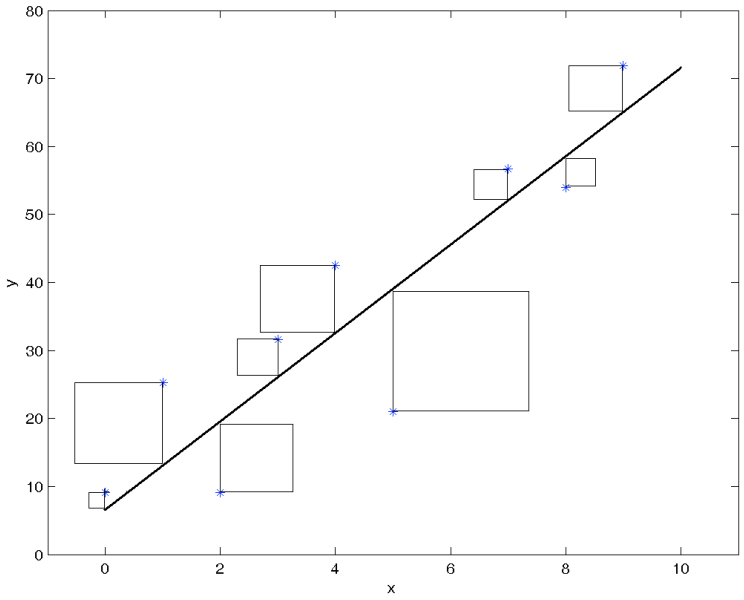


Figure 9.1: Least squares method – a classical illustration

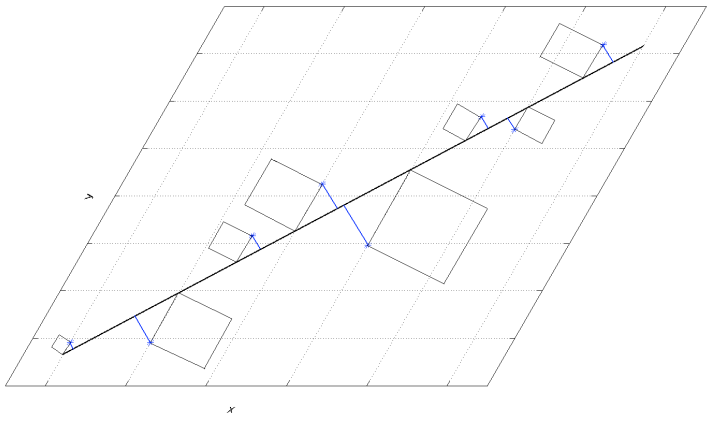


Figure 9.2: Least squares method – non-rectangular coordinates

system, although some coordinate systems are more suitable for describing those objects by equations. Indeed, one can write simpler equations in a suitable coordinate system.

Figure 9.3 illustrates the fitting of an ellipse to a set of points in two-dimensional space. What must one do to fit a set of points by a circle, or an ellipse, or another geometric shape? Simply speaking, one has to draw a sample curve, measure the distance from each point of a set under consideration to the curve, and consider the sum of these distances as a criterion that has to be minimized. Distance in shape recognition and pattern detection is usually a function of squared (or absolute) deviations from the point to the nearest point on the object. For doing this algorithmically using a computer, it is necessary to set the whole picture into some coordinate system. The most common coordinate system we use is the Cartesian rectangular system.

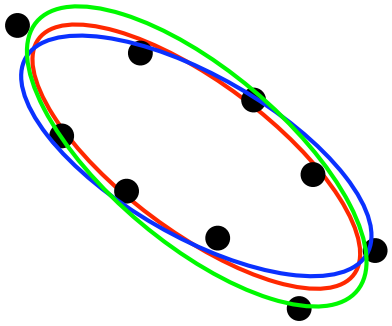


Figure 9.3: Fitting a set of points with an ellipse. The "best" ellipse should be the one for which the sum of distances from the points to the ellipse is minimal

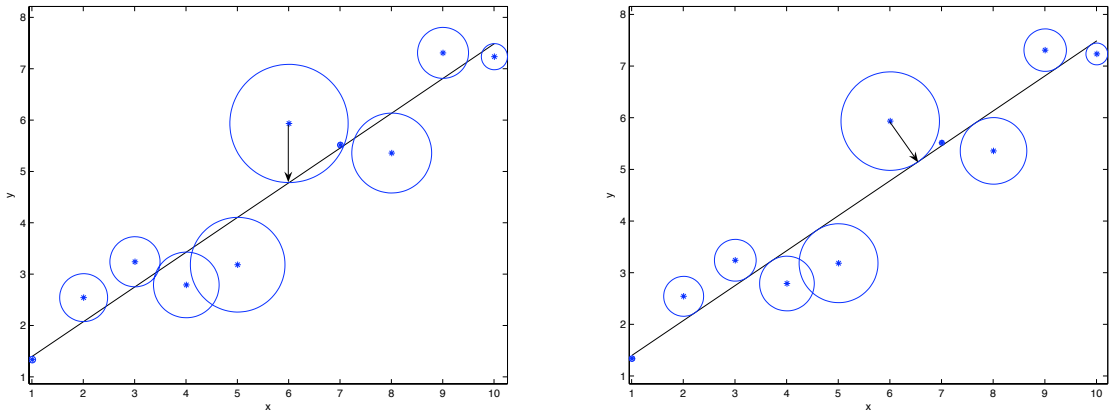
9.4 The Least Squares Method, revisited

Legendre published an idea on how to circumvent the computational problems that arise in the case of trying to minimize the sum of orthogonal distances from data points to a straight line. The idea was to replace the computational problem with a problem in calculus. Put the whole set of the objects (data point and a line) in Cartesian coordinates. Instead of shortest distances from points to a line consider the distances from points to the line in a direction that is parallel to the vertical axis (the vertical offsets). This step would give a different criterion to be minimized:

$$E = \sum_i |y_i - f(x_i, \alpha_1, \alpha_2, \ldots, \alpha_n)|, \tag{9.2}$$

and the minimization problem can be easily solved today using a computer and a suitable numerical routine for minimization.

In the times of Gauss and Legendre, however, it would be natural to find an analytical solution to this problem using the differential calculus. The absolute value is not a good function for this, since its derivative is not continuous. It is possible that thinking in this way, Legendre got a nice idea. The absolute value is a positive-valued function, but does not have a continuous derivative. Is there a positive function close or somewhat similar to absolute value whose derivative is continuous? Of course, there is:



(a) The case of classical least squares fitting from viewpoint of circles

(b) The case of orthogonal distance fitting

Figure 9.4: "Least circles" viewpoint

the square. And this is how could appear the classical least squares fitting, which consists in minimizing the sum of the squares of the vertical deviations of a set of data points

$$E = \sum_i [y_i - f(x_i, \alpha_1, \alpha_2, \ldots, \alpha_n)]^2 \tag{9.3}$$

from a chosen function f. And the minimization problem could be solved by the standard techniques of the differential calculus.

9.5 The Method of Least Circles

To adjust a viewpoint, let us note that the criterion (9.3) can be painlessly replaced with

$$E = \beta \sum_i [y_i - f(x_i, \alpha_1, \alpha_2, \ldots, \alpha_n)]^2. \tag{9.4}$$

Indeed, multiplication by a positive number β does not affect the point of minimum. Only the minimum value of the criterion function (E) will be multiplied by β, which itself is not the subject of interest at this stage, since we look for the values of α_1, α_2, ... α_n.

Taking $\beta = \pi$, we obtain

$$E = \sum_i \pi [y_i - f(x_i, \alpha_1, \alpha_2, \ldots, \alpha_n)]^2. \tag{9.5}$$

Geometrically, the formula (9.5) means the sum of the areas of the circles shown in Fig. 9.4(a). The radii of the circles in Figure 9.4(a) are the vertical offsets of y_i from the fitting line. Figure 9.4(a) is just a reformulation of the standard geometric "illustration" of the least squares method (recall Fig. 9.1). Each of those circles has two points of intersection with the line. It is clear that one cannot consider this picture as elegant. Changing the radii slightly, one can preserve n pairs of intersection of the circles and the line. That is, the circles can be a little bigger or a little smaller and each one will still intersect the line in two places.

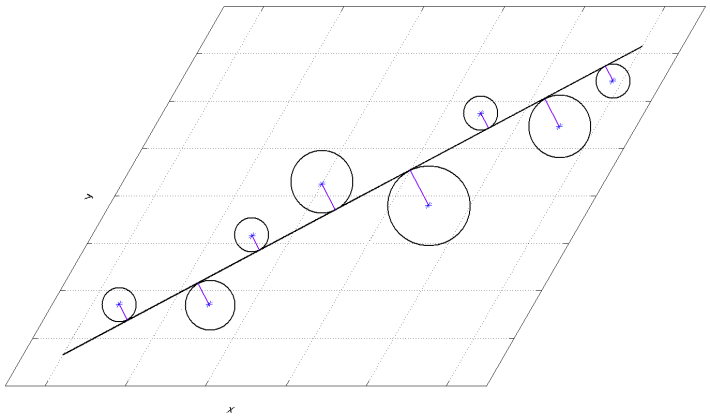


Figure 9.5: Least circles method – non-rectangular coordinates

Instead, suppose we adopt the perspective of the shortest distance to the line in two-dimensional space. The resulting circles are shown in Figure 9.4(b). In this case, the fitting line is a tangent line to all circles. The radii of the circles in Fig. 9.4(b) are equal to distances between the points (x_i, y_i) and the fitting line, and this guarantees the unique picture.

The criterion to minimize in this case is

$$E = \sum_i \pi \left[d\Big((x_i, y_i), f(x, \alpha_1, \alpha_2, \ldots, \alpha_n) \Big) \right]^2, \tag{9.6}$$

which is up to a constant multiplier π the formula known under the name of orthogonal regression. It is also known as total least squares or as the errors in variables method. Here $d\Big((x_i, y_i), f \Big)$ denotes the distance between the point (x_i, y_i) and the fitting line f.

There are several obvious advantages in using least circles (squared orthogonal distance) fitting.

1. *The shortest (orthogonal) distance is the most natural viewpoint on any fitting.*
2. *The sum of orthogonal distances is invariant with respect to the choice of the system of coordinates (see Fig. 9.5).*
3. *There are no conjugate regression lines, which appear after swapping x and y, because in the case of orthogonal regression the fitting $y = f(x)$ gives exactly the same line as the fitting $x = f^{-1}(y)$.* (So, the student from the story in the beginning of this article could be absolutely right, if he used the orthogonal "least circles" method to produce the first coefficients a and b instead of the classical least squares method!)
4. *There are no problems with causality (normally, determination of what is an independent variable and what is a dependent variable is simply unclear or even impossible; this is always postulated).*
5. *Implementation of the orthogonal fitting does not depend on the number of spatial dimensions.*

The 4^{th} point above could be a sticking point for some. Often the goal is to predict an outcome. In that case, one dimension (y) is of primary interest and one often considers distance in that dimension of primary importance. Still, orthogonal regression can be less sensitive to outlying observations and useful as a form of robust regression.

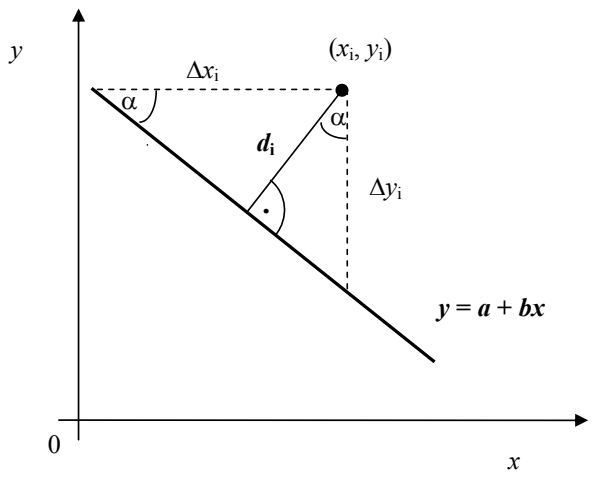


Figure 9.6: Orthogonal distance from a point to a straight line.

9.6 Orthogonal Distance Linear Regression

In general, the use of orthogonal distance fitting requires the use of numerical routines for minimization of the criteria. Fortunately for the student in the econometrics course, in the case of orthogonal distance fitting it is possible to obtain simple formulas for evaluating the parameters of a straight line that fits a given set of points in a plane (orthogonal linear regression problem). Indeed, the orthogonal distance between a point $P_i(x_i, y_i)$ and a straight line $y = a + bx$ is illustrated in Figure 9.6 with values

$$d_i = \frac{\Delta y_i}{\sqrt{1 + \tan^2\alpha}} = \frac{|y_i - (a + bx_i)|}{\sqrt{1 + b^2}} \tag{9.7}$$

Following Legendre, instead of minimizing the sum of orthogonal distances minimize the sum of their squares:

$$E_\perp^2 = \sum_{i=1}^{n} \frac{[y_i - (a + bx_i)]^2}{1 + b^2} = \frac{1}{1 + b^2} \sum_{i=1}^{n} \Big[y_i - (a + bx_i) \Big]^2 \tag{9.8}$$

As usual, take partial derivatives with respect to the parameters a and b equal zero:

$$\frac{\partial E_\perp^2}{\partial a} = 0, \qquad \frac{\partial E_\perp^2}{\partial b} = 0.$$

One obtains a system of two equations for determining the values of a and b.

Partial derivative with respect to the parameter a equal zero is

$$\frac{\partial E_\perp^2}{\partial a} = \frac{-2}{(1 + b^2)} \sum_{i=1}^{n} (y_i - (a + bx_i)) = 0$$

which implies

$$\sum_{i=1}^{n} (y_i - (a + bx_i)) = 0$$

$$\sum_{i=1}^{n} y_i = na + b\sum_{i=1}^{n} x_i,$$

where

$$\bar{y} = a + b\bar{x} \quad \text{and} \quad a = \bar{y} - b\bar{x}$$

Partial derivative with respect to the parameter b equal zero is

$$\begin{aligned} \frac{\partial E_\perp^2}{\partial b} &= \frac{(1+b^2)[-2\sum_{i=1}^{n}(y_i-(a+bx_i)) - 2b\sum_{i=1}^{n}(y_i-(a+bx_i))]^2}{(1+b^2)^2} \\ &= \frac{-2b\sum_{i=1}^{n}(y_i-(a+bx_i))^2}{(1+b^2)} = 0 \end{aligned}$$

Since

$$\sum_{i=1}^{n}(y_i-(a+bx_i)) = 0$$

so

$$\begin{aligned} 0 = \sum_{i=1}^{n}(y_i-(a+bx_i))^2 &= \sum_{i=1}^{n}(y_i-(\bar{y}-b\bar{x}+bx_i))^2 \\ &= \sum_{i=1}^{n}[(y_i-\bar{y})-b(x_i-\bar{x})]^2 \end{aligned}$$

Quadratic equation for finding b is:

$$b^2 S_{xx} - 2bS_x S_y + S_{yy} = 0$$

where

$$b_{1,2} = \frac{-S_x S_y \pm \sqrt{(S_x S_y)^2 - 4S_x S_y}}{S_{xx}^2}$$

So, there are two possible fitting lines, $y = a_1 + b_1 x$ and $y = a_2 + b_2 x$, which both run though the centroid $(\bar{x}, \bar{y})$ and are mutually orthogonal, since $b_1 b_2 = -1$. The proper fitting line (the proper pair of the values of a and b) can be determined by a smaller value of the criterion (9.8).

This has a very simple geometric interpretation. Indeed, the set of point that we try to fit with a straight line is a "cloud" of points in 2D. One of the possible fitting lines coincides with the direction of the main "axis" of that "cloud", and the second line corresponds to the direction of the "width" of that "cloud". It is worth mentioning that this is an elementary illustration of the relationship between the orthogonal distance fitting, on one side, and the principal components analysis (PCA), on the other. There are a few mathematical tools which can be used for orthogonal distance fitting. For solving the linear problems in n-dimensional space the PCA method is appropriate. Generally, for solving the linear and non-linear problem the singular value decomposition (SVD) method and QR decomposition method are suitable. SVD is widely used in statistics where it is related to the PCA method. Now PCA is mostly used as a tool in exploratory data analysis and for making predictive models but the applicability of the PCA is limited by several assumptions (linearity, statistical importance of mean and covariance, etc.).

9.7 Least Circles, Least Spheres, Least Hyperspheres!

Now let us consider the 3D case. Suppose we have a set of data points, which look to be close to a straight or curved line in 3D, and we want to obtain the equation of the optimal fitting line. First, moving from 2D to 3D makes the idea of "least squares" absolutely useless. The only natural criterion is the minimum sum of distances from the data points to the fitting line, and this criterion can be in

Legendre's manner replaced with the volumes of spheres with the radii equal to the distances from the data points to the fitting line,

$$E_{\perp}^{3} = \sum_{i} \frac{4\pi}{3} \left[d\Big((x_i, y_i, z_i), F_{x,y,z,\alpha_1,\alpha_2,\ldots,\alpha_n} \Big) \right]^3, \tag{9.9}$$

where $F_{x,y,z,\alpha_1,\alpha_2,\ldots,\alpha_n}$ denotes a line in 3D described by implicit or explicit equations containing n parameters $\alpha_1, \alpha_2, \ldots, \alpha_n$.

The idea is illustrated in Figure 9.7, where F is a straight line.

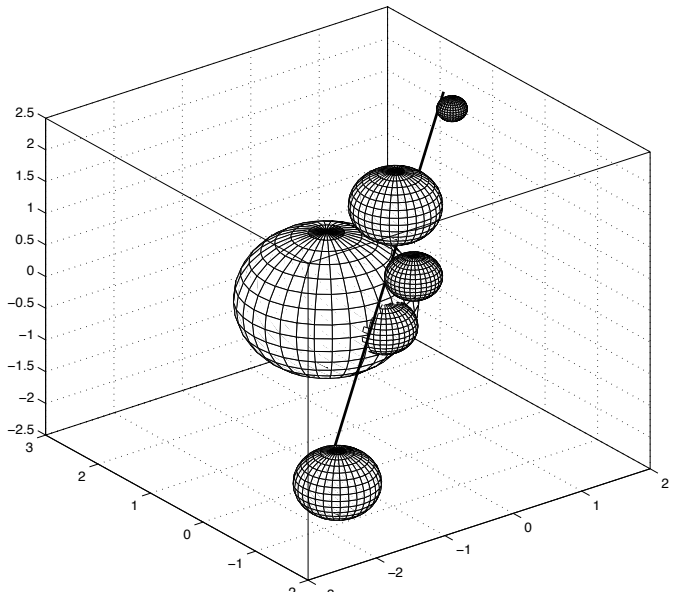


Figure 9.7: Least spheres fitting of data by a line in 3D

Obviously, this approach can be used in n-dimensional space too, where we have to minimize the sum of hypervolumes of the hyperspheres with radii that are equal to the distances from the data points to the fitting line. Besides fitting data by lines, we can consider fitting data by geometric shapes in n-dimensional space (recall the example of fitting data points by an ellipse in previous section, or imagine fitting a set of 3D data by the surface of an ellipsoid), which is a part of image processing theory. Such an approach can have many unexpected applications, like, for example, the description of national economies in state space, where the 3D data describing the behavior of national economies have been fitted by planes. In other words, there is a uniform approach to fitting lines (either straight or curved) and shapes in n-dimensional space by minimizing the volumes of n-dimensional spheres with radii equal to the orthogonal distances from the data points to the fitting line or shape. Maybe Gauss's method, that has not been successfully reproduced until today, was close to such a viewpoint?

9.8 References

[1] Ahn, S. J. (2005). *Least Squares Orthogonal Distance Fitting of Curves and Surfaces in Space*. Berlin: Springer.

[2] Celmins, A. (1998). "The method of Gauss in 1799." *Statistical Science*, 13(2):123–135.

[3] van Huffel, S. and Vandewalle, J. (1991). *The Total Least Squares Problem: Computational Aspects and Analysis*. Philadelphia: SIAM.

[4] Nievergelt, Y. (1994). "Total least squares: State-of-the-art regression in numerical analysis." *SIAM Review*, 36(2):258–264.

[5] Pernkopf, F. and O'Leary, P. (2003). "Image acquisition techniques for automatic visual inspection of metallic surfaces." *NDT & E International*, 36(8):609–617.

[6] Petras, I. and Podlubny, I. (2007). "State space description of national economies: the V4 countries." *Computational Statistics & Data Analysis*, 52(2):1223–1233.

[7] Sardelis, D. and Valahas, T. (2004). "Least Squares Fitting – Perpendicular Offsets.": *library.wolfram.com/infocenter/MathSource/5292/*. Last accessed June 12, 2008.

[8] Sprott, D. A. (1978). "Gauss's contributions to statistics." *Historia Mathematica*, 5:183–203.

[9] Stigler, S. M. (1981). "Gauss and the invention of least squares." *Annals of Statistics*, 9(3):465–474.