

**Reduced Cancellation in the Evaluation of
Entire Functions and Applications to Certain
Special Functions**

D I S S E R T A T I O N

zur Erlangung des akademischen Grades
eines Doktors der Naturwissenschaften

Dem Fachbereich IV der Universität Trier

vorgelegt von

Martin Reinhard

Trier 2008

Contents

1	Introduction	2
2	Taylor Sections of Entire Functions: Cancellation and How to Reduce It	7
2.1	The problem of cancellation	8
2.1.1	Different kinds of errors	8
2.1.2	Definitons and illustrating examples	9
2.2	A description of the cancellation problem	12
2.2.1	The growth of entire functions	13
2.3	How to quantify cancellation errors	15
2.4	A method to reduce cancellation errors	19
2.4.1	Recurrence relations	20
3	The Computation of Mittag-Leffler and Confluent Hypergeometric Functions by Modified Taylor Series	27
3.1	Mittag-Leffler and confluent hypergeometric functions	28
3.1.1	Order and Type	30
3.1.2	Asymptotic expansions and Indicator Functions	31
3.2	Special cases of Mittag-Leffler and confluent hypergeometric functions	34

3.3	The computation of the error function	35
3.4	The computation of the incomplete gamma function	46
3.5	The Computation of Mittag-Leffler and confluent hypergeometric functions for more general parameters	51
4	Further Methods to Compute Entire Functions	55
4.1	Computation with Asymptotic expansions	55
4.1.1	Asymptotic expansions for the complementary error and the incomplete gamma function	56
4.2	Computation with Continued fractions	60
5	Conclusion	67

Chapter 1

Introduction

It is often necessary to have a mathematical model for a physical process in order to recognize its characteristics. The mathematical description leads in general to differential or integral equations which can have a very complex structure.

In certain cases an explicit form or a series representation of the resulting solutions can be obtained by using special functions. So, considering the numerical simulation of such mathematical models it is necessary to have efficient methods for computing special functions.

We will focus our considerations in particular on the classes of Mittag-Leffler and confluent hypergeometric functions. In the last few years the interest in functions of the Mittag-Leffler type has increased among scientists, engineers and applications-oriented mathematicians. This interest is caused by the close connection of these functions to the solutions of differential equations of fractional order (in the sense of non-integer order) and integral equations of Abel type. Such equations are becoming more and more popular for modelling scientific and technical processes and situations (see [Go1], [Go2], [Ma]).

Although there already exists a rich literature on methods for solving differential equations, in general, numerical solution techniques are required. Lozier [Loz] investigated software available for the computation of special functions and stated the software needs in scientific computing. In particular, software

packages are needed for the computation (in floating point arithmetic with a fixed precision) of Mittag-Leffler type functions and confluent hypergeometric functions with complex parameters and arguments. Since the classes of Mittag-Leffler and confluent hypergeometric functions cover several important special functions (like incomplete gamma or complementary error functions), it is therefore desirable to have an efficient algorithm for computing these functions. A consideration of standard software packages like MATLAB or MATHEMATICA also shows that it is necessary to find such algorithms.

In Chapter 2, entire functions f are considered. If we look at the partial sums of the Taylor series with respect to the origin

$$s_n(f, z) := \sum_{\nu=0}^n a_\nu z^\nu$$

for $n \in \mathbb{N}$, we find that they typically only provide a reasonable approximation of the function f in a small neighborhood of the origin. The main disadvantages of these partial sums are the cancellation errors which occur when computing in fixed precision arithmetic outside this neighborhood. The computation of small function values in modulus causes problems because the summation of the terms of the partial sums $s_n(f, z)$, which are inherently large in modulus, may lead to a loss or cancellation of several decimal digits. Therefore, our first aim is to quantify this cancellation effect using basic growth properties of an entire function f such as order ρ_f , type τ_f and indicator function h_f . We find that the loss of exact decimal digits for $z = re^{i\vartheta}$ under certain conditions can be asymptotically described by

$$(\tau_f - h_f(\vartheta))r_f^\rho / \log(10).$$

We develop and investigate a method for the computation of entire functions which is stable as far as cancellation errors are concerned.

The idea is to modify the function f in such a way that, for certain parts of the complex plane, the order of magnitude of the modified function \tilde{f} and the maximal term of the series $s_n(\tilde{f}, z)$ do not differ as much as the ones of the function f . Then an approximation of $f(z)$ may be obtained

from $s_n(\tilde{f}, z)$ and we can compute the Taylor sections with less cancellation. Such a modification may consist in a multiplication of f by an appropriate function φ which can be numerically evaluated. Under certain conditions, this "multiplier" function φ can be found by solving an adequate optimization problem.

Using this method, we have to consider recurrence formulae as an important tool to find reasonable representations of the coefficients of the modified function \tilde{f} . We give a short review of the basic theory of linear difference equations (of second order) and compute solutions (y_n) satisfying the equation

$$y_{n+1} + a_n y_n + b_n y_{n-1} = 0,$$

with given coefficients a_n and $b_n \neq 0$. Although easy to implement, such recurrence relations are often numerically unstable e.g. due to rounding errors. In particular, forward recurrences may cause problems. In order to circumvent the problems we apply a well-known method for computing recurrence relations in backward direction: the Miller algorithm. Since the application of the Miller algorithm does not require exact starting values for the backward recurrence, a so-called "normalizing series" is usually needed. For our considerations (see Section 2.4.1), we can simply use the known initial values instead. Finally, with some numerical examples we show the efficiency of this method.

In Chapter 3, we first look at the definitions and characteristics of the Mittag-Leffler and confluent hypergeometric functions. We consider the confluent hypergeometric functions of the first kind, denoted by $\Phi(a, c; z)$, with two, in general complex, parameters a , c as well as argument z , whose power series expansions are given by

$$\Phi(a, c; z) = \sum_{\nu=0}^{\infty} \frac{(a)_{\nu} z^{\nu}}{(c)_{\nu} \nu!} \quad (z \in \mathbb{C}).$$

Moreover the Mittag-Leffler functions $M(\alpha, \beta; z)$ for $\alpha > 0$ and β a complex

parameter as well as argument z are defined by

$$M(\alpha, \beta; z) := \sum_{\nu=0}^{\infty} \frac{z^{\nu}}{\Gamma(\alpha\nu + \beta)} \quad (z \in \mathbb{C}).$$

The characteristics, especially the growth properties like order, type and indicator function of the classes of Mittag-Leffler and confluent hypergeometric functions are pointed out. Furthermore, we address the connections with other special functions which are included in the classes of both functions. Then, considering the numerical evaluation of the Mittag-Leffler and confluent hypergeometric functions, we look in particular at the complementary error function and the incomplete gamma function in several parts of the complex plane where the partial sums $s_n(f, z)$ do not provide an appropriate approximation because of cancellation errors. By using the growth properties of the functions, we are able to quantify cancellation problems. Furthermore, using the method we developed in Chapter 2, we can reduce the cancellation problems by "modifying" the function for several parts of the complex plane. With the examples of the complementary error and the incomplete gamma function we illustrate in which way an appropriate multiplier function φ can be appointed and we can see the difficulties to find a reasonable representation for the coefficients of the Taylor series of the modified function \tilde{f} . Numerical results, in which we compare our results of the computation using the method developed in Chapter 2 with reference values, are presented. The reference values for f are generated using exact arithmetic from iRRAM (see [Mue1]). It can be seen that the problem of cancellation errors can be reduced considerably by using the method developed in Chapter 2.

Finally, in Chapter 4 two other approaches to compute Mittag-Leffler type and confluent hypergeometric functions are discussed. If we want to evaluate such functions on unbounded intervals or sectors in the complex plane, we have to consider methods like asymptotic expansions or continued fractions for large arguments z in modulus.

When considering asymptotic expansions, we use an expansion of the form

$$f(z) \sim \sum_{\nu=0}^{\infty} \frac{a_{\nu}}{z^{\nu}} \quad (z \rightarrow \infty)$$

for the computation of a function f in some sectors of the complex plane. After looking at the basic properties of asymptotic expansions (we already did in Chapter 3), we consider the asymptotic expansion of the Mittag-Leffler functions in detail. For numerical considerations, the determination of the number of terms used for the approximation is a crucial point. In [WoZh], a method to determine the number of terms is presented. We will see that from the numerical point of view the monotonicity behaviour of the terms of the series plays a central role for the truncation of the series.

Considering continued fractions for the computation of Mittag-Leffler type and confluent hypergeometric functions, we refer to the works of Gautschi ([Ga1],[Ga3]) and Poppe/Wijers ([PW1]). In their articles the advantages of this method are discussed in detail.

Chapter 2

Taylor Sections of Entire Functions: Cancellation and How to Reduce It

Let f be an entire function given by its Taylor series

$$f(z) = \sum_{\nu=0}^{\infty} a_{\nu} z^{\nu} = \sum_{\nu=0}^{\infty} \frac{f^{(\nu)}(0)}{\nu!} z^{\nu} \quad (z \in \mathbb{C})$$

with respect to the origin. Of course, one way to numerically evaluate $f(z)$ is to truncate the series, that is, to take, for n sufficiently large,

$$s_n(f, z) := \sum_{\nu=0}^n a_{\nu} z^{\nu}$$

as an approximation of $f(z)$.

However, it turns out that, in many interesting cases, serious cancellation problems arise if z is not restricted to a more or less small neighborhood of the origin. So first of all, we have to take a closer look at the problems which occur if we use series expansions as an approximation of $f(z)$. We especially consider the problem of cancellation.

2.1 The problem of cancellation

As mentioned above, our aim is to compute entire functions by series expansions. If we have convergence of the series for all $z \in \mathbb{C}$, it is possible to use some kind of exact arithmetic to compute entire functions by using their Taylor sections, but in this case a variable precision of the decimal digits and in many cases "endless" time is needed.

Because of this, our aim is to find "simple" and "fast" methods to compute entire functions by using floating point arithmetic with fixed precision of p decimal digits.

In the following, our computations will be performed in MATLAB with a fixed precision of 16 decimal digits.

First of all, we give a short overview of the different kinds of errors which occur if numerical computations are performed. After this, we take a closer look at the problem of cancellation.

2.1.1 Different kinds of errors

Different kinds of errors occur when we consider practical methods for mathematical problems.

These include *method errors*, notably *truncating errors*, which occur when an infinite process is replaced by a finite method (for example the partial sum of an infinite series).

Initial errors occur if there are inexact values or statistical variation resulting in errors among the initial data. These must be accepted in subsequent mathematical operations. It is desirable that initial data errors cause only small errors in the result.

Further errors occur when computing practically using real numbers. These depend on the method used and on the computing tool.

Here we have *input data errors* which occur when the input data is rounded into machine-adaptable numbers.

Also, we look at *errors due to rounding*. These occur no matter which arith-

metric operation is used. Of course we have to make sure that we compute using a fixed precision arithmetic.

Primarily, we want to consider the worst possible errors when computing numerically, namely *cancellation errors*.

First, we consider some definitions and examples to describe and demonstrate the phenomenon of cancellation. After this, we will look at the question of how we can quantify and reduce this kind of error.

2.1.2 Definitons and illustrating examples

We give some basic definitions and we take a look at some examples which illustrate the problem of cancellation.

First of all, it is important to say that our considerations regarding the cancellation problem are not only limited to the use of real numbers. As mentioned above, our aim is to compute series expansions of certain entire functions for all $z \in \mathbb{C}$. Actually, we also have to consider the problem of cancellation for complex numbers. But in the following, by using the partial sums of the Taylor sections to compute the entire functions, we will describe the cancellation problem for at least one of the values of the real and the imaginary part of the series. So it makes sense that we focus our considerations on the cancellation problem for real numbers.

Definition 2.1

If \tilde{x} is an approximation of $x \in \mathbb{R}$, then $|\tilde{x} - x|$ is the absolute error and $|(\tilde{x} - x)/x|$ is the relative one, in case of $x \neq 0$.

Let us take a closer look at the behaviour of relative errors using certain operations. It is a common technique to leave out the products of errors.

This enables us to simplify complex terms of errors drastically. We use the symbol \doteq which means that products of errors are ignored.

We refer to [SW] for the following result.

Theorem 2.2

Let $x, y \in \mathbb{R} \setminus \{0\}$ and φ be one of the arithmetic operations $+$, \cdot or $/$. For the relative errors

$$\epsilon_x := (\tilde{x} - x)/x, \quad \epsilon_y := (\tilde{y} - y)/y, \quad \text{and}$$

$$\epsilon_\varphi := (\varphi(\tilde{x}, \tilde{y}) - \varphi(x, y))/\varphi(x, y)$$

regarding two approximations \tilde{x} of x and \tilde{y} of y we get

$$\epsilon_\varphi \doteq \epsilon_x + \epsilon_y, \quad \text{if } \varphi(x, y) = x \cdot y,$$

$$\epsilon_\varphi \doteq \epsilon_x - \epsilon_y, \quad \text{if } \varphi(x, y) = x/y,$$

$$\epsilon_\varphi \doteq \frac{x}{x+y} \epsilon_x + \frac{y}{x+y} \epsilon_y, \quad \text{if } \varphi(x, y) = x + y \neq 0.$$

Conclusion 2.3

Multiplication and division of numbers are not problematic since, in the worst case, they cause an addition of the absolute values of the relative errors found in the input data.

However, **cancellation** occurs in the subtraction of two nearly identical numbers since the factors $x/(x+y)$ and $y/(x+y)$ can become really large.

Example 2.4

Assuming floating point arithmetic, let x and y be two nearly equal numbers. When the difference $x - y$ is formed, some leading digits of the mantissa will be zero, so significant digits have been "cancelled". The mantissa will be normalized by moving the digits to the left and at the tail end of the mantissa there will appear zeros which are meaningless.

For example, let

$$x = \sqrt{9876} = 9.937806599043876 \cdot 10^1, \quad y = \sqrt{9875} = 9.937303457175895 \cdot 10^1.$$

Then we have

$$x - y = 0.000503141867981 \cdot 10^1.$$

Normalization changes this to

$$x - y = 5.031418679810000 \cdot 10^{-3}.$$

The four zeros at the end of the mantissa are meaningless. With the estimate

$$\sqrt{9876} - \sqrt{9875} = \frac{1}{\sqrt{9876} + \sqrt{9875}} = 5.031418679802768 \cdot 10^{-3}$$

we get a result which is accurate to all digits given.

Cancellation does not only occur when two nearly equal numbers are directly subtracted from each other. More generally, this situation occurs in the evaluation of a sum if the partial sums are large compared to the final result.

Let

$$s := \sum_{k=1}^n a_k$$

be the sum to be computed. We assume that the evaluation is done by forming the sequence of partial sums according to

$$s_1 := a_1, \quad s_k := s_{k-1} + a_k \quad (k = 2, 3, \dots, n),$$

so that $s = s_n$.

The phenomenon of cancellation can occur if the sum s is evaluated in floating arithmetic.

Let us assume that one of the intermediate sums s_k is considerably larger than the final sum s in the sense that the exponent of s_k exceeds the exponent of s by several units (of course this can happen only if not all a_k have the same sign).

In [He3] the effect of the loss of significant digits due to large intermediate sums is called "smearing".

Now our aim is to quantify these cancellation errors, especially this smearing-effect by using partial sums of Taylor sections of entire functions.

So we have to consider the relation between the order of magnitude of the function and the largest term of the series that may occur.

2.2 A description of the cancellation problem

Let f be again an entire function given by its Taylor series

$$f(z) = \sum_{\nu=0}^{\infty} a_{\nu} z^{\nu} = \sum_{\nu=0}^{\infty} \frac{f^{(\nu)}(0)}{\nu!} z^{\nu} \quad (z \in \mathbb{C})$$

with respect to the origin. The maximal term of the series is defined by

$$\mu(r) := \mu_f(r) := \max_{\nu \in \mathbb{N}_0} |a_{\nu}| r^{\nu} \quad (r \geq 0).$$

We write

$$s_{n,p}(f, z) := \sum_{\nu=0}^n a_{\nu} z^{\nu}$$

for the n -th partial sum of the above Taylor series of f if the computations are performed in floating point arithmetic with a precision of p decimal digits and with input data a_{ν} and z given with an accuracy of p digits.

In this situation, the loss of exact digits in the evaluation of at least one of the values $\operatorname{Re} s_{N(z),p}(f, z)$ and $\operatorname{Im} s_{N(z),p}(f, z)$ may be approximately measured by

$$d_f(z) := \log_{10} \mu_f(|z|) - \log_{10} |f(z)| \quad (2.1)$$

for $N(z)$ sufficiently large (see [He3], p. 27).

We always assume that the number of terms $N(z) = N_f(z)$ is taken so large that errors do not result from truncation of the series (so, increasing the number of terms does not improve the exactness).

This implies that the number of exact digits (here and in what follows always understood in the sense of the minimal number in both the real and the imaginary part) is approximately given by $\max(p - d_f(z), 0)$ in the case of computations with p exact figures. So one can be confronted with serious problems if $d_f(z)$ turns out to be large.

Our aim is to reduce such problems by modifying f in an appropriate way. Before we go into the details, we will first quantify d_f for entire functions of finite order and type and of regular growth, approximately in terms of the indicator function of f .

For this we need some basic knowledge on the growth of entire functions, in

particular we have to explain what it means that an entire function has a finite order and type. Also we define the indicator function of an entire function f and consider some of its properties.

2.2.1 The growth of entire functions

At first we give some basic definitions.

Definition 2.5

Let f be an entire function, defined by

$$f(z) = \sum_{n=0}^{\infty} a_n z^n. \quad (2.2)$$

i) We denote by

$$M_f(r) = \max_{|z| \leq r} |f(z)| \quad (r > 0), \quad (2.3)$$

the maximum modulus of f on $|z| \leq r$.

ii) Then we call $\rho = \rho_f$ the order of f , if

$$\rho_f = \overline{\lim}_{r \rightarrow \infty} \frac{\log \log M_f(r)}{\log r}. \quad (2.4)$$

iii) If $\rho_f \in (0, \infty)$, we say that f is of type $\tau = \tau_f$ of its order, if

$$\tau_f = \overline{\lim}_{r \rightarrow \infty} \frac{\log M_f(r)}{r^{\rho_f}}. \quad (2.5)$$

A function f is said to be of maximal type of its order ρ_f , if $\tau_f = \infty$, of normal type, if $\tau_f \in (0, \infty)$ and of minimal type, if $\tau_f = 0$.

Remark 2.6

So, the growth of an entire function f can be described by its order and type.

A finite order ρ_f means that for each $\epsilon > 0$

$$M_f(r) \leq e^{r^{(\rho_f + \epsilon)}} \quad (r \geq r_0(\epsilon)). \quad (2.6)$$

For the type τ_f we get

$$M_f(r) \leq e^{(\tau_f + \epsilon)r^{\rho_f}} \quad (r \geq r_1(\epsilon)). \quad (2.7)$$

It is possible to characterize the order and type of entire functions by means of the magnitude of the Taylor coefficients.

For the following results, which characterize order and type, we refer to [Bo].

Theorem 2.7

Let $f(z) = \sum_{n=0}^{\infty} a_n z^n$ be an entire function of order ρ_f . Then we have

$$\rho_f = \overline{\lim}_{n \rightarrow \infty} \frac{\log n}{-\log |a_n|^{1/n}}. \quad (2.8)$$

Theorem 2.8

Let $f(z) = \sum_{n=0}^{\infty} a_n z^n$ be an entire function of order $\rho_f \in (0, \infty)$ and type τ_f . Then we have

$$\tau_f e \rho_f = \overline{\lim}_{n \rightarrow \infty} (n |a_n|^{\rho_f/n}). \quad (2.9)$$

Moreover, the growth of f along rays emanating from the origin is asymptotically described by its indicator function $h = h_f$ which can be defined as follows.

Definition 2.9

The (Phragmen-Lindelöf) indicator function $h = h_f : [-\pi, \pi] \rightarrow \mathbb{R}$ is given by

$$h_f(\vartheta) := \overline{\lim}_{r \rightarrow \infty} \frac{\log |f(re^{i\vartheta})|}{r^{\rho_f}} \quad (\vartheta \in [-\pi, \pi]). \quad (2.10)$$

Finally, we give some important remarks about the order, type and indicator function of an entire function f . For this we refer to [Le].

Remark 2.10

i) The indicator function h_f is continuous (with $h_f(-\pi) = h_f(\pi)$) and

$$\max_{[-\pi, \pi]} h_f(\vartheta) = \tau_f.$$

ii) For the function

$$f(z) = e^{az^p} \quad (a \in \mathbb{C})$$

with the finite order $\rho \in \mathbb{N}$, we obtain

$$h_f(\vartheta) = \operatorname{Re} a \cos(\rho\vartheta) + \operatorname{Im} a \sin(\rho\vartheta) \quad (\vartheta \in [-\pi, \pi]).$$

Such indicator functions are called " ρ -trigonometric indicator functions".

- iii) It results directly from the definition that the indicator of the product of two functions does not exceed the sum of the indicators of the factors. That means, if the functions f and φ are functions of the same order, then we have

$$h_{f\varphi}(\vartheta) \leq h_f(\vartheta) + h_\varphi(\vartheta) \quad (\vartheta \in [-\pi, \pi]). \quad (2.11)$$

We can obtain equality in equation (2.11) if one of the functions is of completely regular growth (for a definition, see, e.g., [Le], p.137, for the assertion, see e.g., [Le], p.157).

- iv) If f has completely regular growth, there exists a set E , which is a so-called C^0 -set, that is, E is the union of circles $\{z : |z - z_j| < r_j\}$ with

$$\lim_{R \rightarrow \infty} \frac{1}{R} \sum_{|z_j| < R} r_j = 0,$$

so that for all ϑ

$$h_f(\vartheta) = \lim_{\substack{r \rightarrow \infty \\ re^{i\vartheta} \notin E}} \frac{\log |f(re^{i\vartheta})|}{r^{\rho_f}}.$$

2.3 How to quantify cancellation errors

In this section, we want to quantify the cancellation problem using the properties of the growth of entire functions. Our aim is to find a formula to approximate the difference $d_f(z)$ (see (2.1)) which describes the loss of exact digits. For this we need a relation between the maximal term $\mu_f(r)$ of the series and the growth of the entire function.

This relation is given in the following theorem ([Ru], Theorem 10.1).

Theorem 2.11

Let the function f be of finite order, then

$$\log \mu_f(r) / \log M_f(r) \rightarrow 1 \quad (r \rightarrow \infty).$$

Considering this theorem, we can replace $M_f(r)$ in the above definitions (2.4) and (2.5) by $\mu_f(r)$.

Finally, if f has completely regular growth, we can obtain with the help of Remark 2.10 iv) that

$$\lim_{r \rightarrow \infty, re^{i\vartheta} \notin E} \frac{\log(10) \cdot d_f(re^{i\vartheta})}{r^{\rho_f}} = \tau_f - h_f(\vartheta) =: \delta_f(\vartheta), \quad (2.12)$$

which means that the loss of exact decimal digits for $z = re^{i\vartheta}$ is (up to exceptional values, e.g., near the zeros of f) asymptotically described by $\delta_f(\vartheta)$ in the sense that

$$d_f(re^{i\vartheta}) \sim \delta_f(\vartheta) r^{\rho_f} / \log(10) \quad (r \rightarrow \infty, re^{i\vartheta} \notin E). \quad (2.13)$$

To illustrate the theoretical considerations on the smearing effect by using the partial sums of the Taylor sections of entire functions, we look at the following simple example of computing the exponential function using series expansions on the negative real axis.

Example 2.12

Let f be the exponential function

$$f(x) = \exp(x) = \sum_{k=0}^{\infty} \frac{x^k}{k!} \quad (x \in \mathbb{R}).$$

If e^x is computed numerically with a fixed precision of 16 decimal digits, the series must be truncated and we take the n -th partial sum

$$s_{N(x),16}(f, x) = \sum_{\nu=0}^{N(x)} \frac{x^\nu}{\nu!} \quad (x \in \mathbb{R}). \quad (2.14)$$

as an approximation of $f(x)$.

We again assume that the number of terms $N(x)$ is taken so large that errors do not result from truncation of the series.

Using (2.14) for our computations, we see that this series is not well suited for the evaluation of e^x for $x \ll 0$. The problem of cancellation causes a loss of significant decimal digits on the negative real axis.

Our theoretical considerations above give us the possibility to quantify the cancellation problem.

Using Stirling's formula, we can find an approximation for the largest term of the series (2.14), which is

$$\mu_f(r) \sim \frac{r^r}{r!} \sim \frac{r^r}{\sqrt{2\pi r} \left(\frac{r}{e}\right)^r} = \frac{1}{\sqrt{2\pi r}} e^r.$$

With the help of (2.1) the loss of exact digits in the evaluation of the exponential function can be approximately measured by

$$\begin{aligned} d_f(re^{i\vartheta}) &= \log_{10} \mu_f(r) - \log_{10} e^{r \cos \vartheta} \\ &\sim r(1 - \cos \vartheta)/\log(10) - 1/2 \log_{10}(2\pi r) \\ &\sim r(1 - \cos \vartheta)/\log(10). \end{aligned} \quad (2.15)$$

In contrast to this approximation, we want to quantify the loss of exact digits in the evaluation using the growth properties of the exponential function.

It can be easily seen that the exponential function is an entire function of the order 1 and type 1.

From Remark 2.10 we get the indicator function of the exponential function as follows:

$$h_f(\vartheta) = \cos(\vartheta) \quad (\vartheta \in [-\pi, \pi]).$$

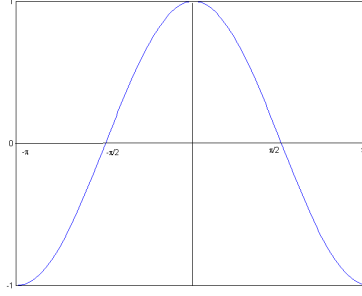


Figure 2.1: Indicator function of the exponential function.

Considering (2.12), we can see that $\delta_f(\vartheta) = 1 - \cos(\vartheta)$, and the loss of exact digits (see (2.13)) can be approximately described by

$$d_f(re^{i\vartheta}) \sim \delta_f(\vartheta)r/\log(10) = r(1 - \cos(\vartheta))/\log(10). \quad (2.16)$$

In fact, this approximation complies with the one in (2.15).

The "worst case" in the evaluation of the exponential function appears in a neighborhood of the negative semiaxis.

It can be seen that $\delta_f(-\pi) = \tau_f - h_f(-\pi) = 2$. Hence, as an approximation for the loss of exact decimal digits on the negative real axis we get

$$d_f(re^{-i\pi}) \sim 2r/\log(10) \quad (r \rightarrow \infty).$$

For the following numerical considerations we evaluate e^x for various $x < 0$ by truncating the Taylor series. The loss of decimal digits, the relation between the largest term of the series and the order of magnitude of the function, described in (2.1), and the approximation for the loss of decimal digits, given in (2.13), are illustrated in the following table. The underlined digits coincide with the correct ones.

x	Value of e^x (Taylor series)	Exact value	$\sim \mu_f(r)$	$[d_f]$
-5	<u>6.737946999086907</u> $\cdot 10^{-3}$	6.737946999085467 $\cdot 10^{-3}$	$2.6 \cdot 10^1$	4
-10	<u>4.539992947267379</u> $\cdot 10^{-5}$	4.539992976248485 $\cdot 10^{-5}$	$2.8 \cdot 10^3$	8
-15	<u>3.062880133869493</u> $\cdot 10^{-7}$	3.059023205018258 $\cdot 10^{-7}$	$3.4 \cdot 10^5$	13

Of course, the determination of accurate values of e^x for $x \ll 0$ does not present a real problem. By exploiting the relation $e^x = (e^{-x})^{-1}$, the computation of e^x may be reduced to the computation of e^{-x} , for which reliable values may be obtained from the truncated Taylor series.

Such simple functional relationships are not always available. So our aim is to find a method to avoid or at least to reduce these cancellation errors in general.

2.4 A method to reduce cancellation errors

We are confronted with the question of how such cancellation problems can be reduced. One idea is to modify the function f in such a way that, at least for certain parts of the complex plane, the order of magnitude of the modified \tilde{f} and $\mu_{\tilde{f}}$ do not differ as much as the ones of the function f . Then the Taylor sections $s_n(\tilde{f}, z)$ can be computed with less cancellation, and an approximation of $f(z)$ may be obtained from $s_n(\tilde{f}, z)$. Such a modification may consist in a multiplication of f by an appropriate function φ which can be numerically evaluated.

More precisely, the general idea is the following:

Suppose that $S \subset \mathbb{C}$ is a given set (where f is to be numerically evaluated).

- i) Choose an appropriate (entire) function $\varphi = \varphi_S$ so that

$$d_{f\varphi}(z) = \log_{10} \mu_{f\varphi}(|z|) - \log_{10} |f(z)\varphi(z)|$$

is “small” for $z \in S$ and $\varphi(z) \neq 0$ in S .

- ii) Take $\frac{1}{\varphi(z)} s_{N(z),p}(f\varphi, z)$ (for $N(z)$ sufficiently large) as an approximation of $f(z)$ for $z \in S$.

Of course, the question arises of how to choose φ appropriately. If f and φ are of the same order ρ and one of them is of completely regular growth, then (see Remark 2.10 iii)) we have

$$h_{f\varphi} = h_f + h_\varphi. \tag{2.17}$$

Remark 2.13

More conveniently, using the definition (2.10) of the indicator function, we also obtain equation (2.17) if there exists a dense subset M of $\partial\mathbb{D}$, so that

$$h_f(\vartheta) = \lim_{r \rightarrow \infty} \frac{\log |f(re^{i\vartheta})|}{r^\rho} \quad \text{or} \quad h_\varphi(\vartheta) = \lim_{r \rightarrow \infty} \frac{\log |\varphi(re^{i\vartheta})|}{r^\rho} \quad \text{for all } \vartheta \in M.$$

With the help of (2.17), we get

$$\delta_{f\varphi} = \tau_{f\varphi} - h_f - h_\varphi.$$

If $S = \{re^{i\vartheta} : r \geq 0, \vartheta \in \Theta\}$ is given, then $\varphi = \varphi_S$ should be chosen so that

$$\max_{\vartheta \in \Theta} \delta_{f\varphi}(\vartheta) = \max_{\vartheta \in \Theta} (\max_{[-\pi, \pi]} (h_f + h_\varphi) - h_f(\vartheta) - h_\varphi(\vartheta))$$

is small (compare also the considerations in [Mue2]).

If φ is taken from a parametrized family $\Phi = \{\varphi(a, \cdot) : a \in A\}$, we can try to solve the problem

$$\max_{\vartheta \in \Theta} (\max_{[-\pi, \pi]} (h_f + h_{\varphi(a, \cdot)}) - h_f(\vartheta) - h_{\varphi(a, \cdot)}(\vartheta)) \rightarrow \min_{a \in A}. \quad (2.18)$$

So we have considered a way to choose φ appropriately.

Now the question we are confronted with is how to numerically evaluate $s_{N(z),16}(f\varphi, z)$ in an efficient way. If we cannot find a closed form for the resulting coefficients of f and φ by convolution, it will often be necessary to find a recurrence relation for the coefficients. For this we need some basic definitions and characteristics of the work with recurrence relations.

2.4.1 Recurrence relations

Recurrence relations are very interesting for computing special functions because in many cases they are easy to implement and the computational input is low. But often these methods are very susceptible to errors in numerical computation, therefore it is important to know whether the recurrence relation is numerically stable.

Linear difference equations

In this section we give some basics about the theory of linear difference equations of second order. For this section we refer to [Te], pp. 335-342, [Ga2] and [Wi].

A three-term recurrence relation

$$y_{n+1} + a_n y_n + b_n y_{n-1} = 0, \quad (2.19)$$

for $n = 1, 2, \dots$ and given coefficients a_n and $b_n \neq 0$, is called a linear homogeneous difference equation of second order.

For the general solution (y_n) of the equation (2.19) we have

$$y_n = p f_n + q g_n, \quad (2.20)$$

and (f_n) , (g_n) are two linearly independent solutions of (2.19), p and q are constants independent of n .

In the following, we try to find solutions (f_n) , (g_n) satisfying the condition

$$\lim_{n \rightarrow \infty} \frac{f_n}{g_n} = 0. \quad (2.21)$$

This condition implies

$$\lim_{n \rightarrow \infty} \frac{f_n}{y_n} = 0$$

for any solution (y_n) not being a constant multiple of (f_n) . If (y_n) is not proportional to (f_n) we have $q \neq 0$, and therefore we get

$$\lim_{n \rightarrow \infty} \frac{f_n}{y_n} = \lim_{n \rightarrow \infty} \frac{f_n/g_n}{q + p(f_n/g_n)} = 0.$$

The set of all solutions (f_n) of equation (2.19) having the property (2.21) forms, if existent, a one-dimensional subspace of the space of all solutions. There cannot be two linearly independent solutions (f_n) and (\tilde{f}_n) satisfying condition (2.21). Such solutions are called minimal, any non-minimal solution is called dominant.

The constants p and q can be computed by using the initial values y_0 , y_1 as well as f_0 , f_1 , g_0 , g_1 .

We get (note, that the denominators are different from zero when we have

linearly independent solutions $(f_n), (g_n)$

$$p = \frac{g_1 y_0 - g_0 y_1}{f_0 g_1 - f_1 g_0}, \quad q = \frac{f_1 y_0 - f_0 y_1}{f_1 g_0 - f_0 g_1}.$$

If we calculate such a sequence (f_n) by the relation (2.19) using approximate initial values $y_0 \approx f_0$, $y_1 \approx f_1$, caused by rounding errors for example, the resulting solution (y_n) will be in general linearly independent of the minimal solution (f_n) .

So (2.21) implies

$$\left| \frac{y_n - f_n}{f_n} \right| \rightarrow \infty \quad (n \rightarrow \infty).$$

That means the relative error of the computed approximation (y_n) for (f_n) tends to infinity. Hence, the numerical computation of minimal solutions turns out to be delicate.

So the question arises of how the computation using such recurrence relations can be numerically stable.

The idea (**Miller algorithm**) is to apply (2.19) in backward direction in order to find the values f_1, \dots, f_N for fixed integer N . We usually need starting values f_N and f_{N-1} for this, but for the Miller algorithm these starting values are not required, as we will see in the following.

We assume that a convergent "normalizing" series is known:

$$S := \sum_{k=0}^{\infty} c_k f_k \neq 0. \quad (2.22)$$

Let $\nu > N$ a nonnegative integer, usually large, and compute a solution of (2.19) for $n = \nu - 1, \nu - 2, \dots, 0$ through

$$y_{n-1}^{(\nu)} = -\frac{1}{b_n}(y_{n+1}^{(\nu)} + a_n y_n^{(\nu)})$$

with the chosen initial values

$$y_{\nu+1}^{(\nu)} = 0, \quad y_{\nu}^{(\nu)} = 1.$$

Whereas these initial values are not unique, only one of these must be different from zero.

For $n = 0, 1, \dots, \nu + 1$ the computed solution is a linear combination of the solutions (f_n) and (g_n) of the form

$$y_n^{(\nu)} = p_{\nu} f_n + q_{\nu} g_n,$$

with coefficients

$$p_\nu = \frac{g_{\nu+1}}{g_{\nu+1}f_\nu - g_\nu f_{\nu+1}}, \quad q_\nu = \frac{f_{\nu+1}}{g_{\nu+1}f_\nu - g_\nu f_{\nu+1}}.$$

With (2.21) we get

$$\lim_{\nu \rightarrow \infty} \frac{y_n^{(\nu)}}{p_\nu} = f_n - \lim_{\nu \rightarrow \infty} \frac{f_{\nu+1}}{g_{\nu+1}} g_n = f_n \quad (n = 0, 1, \dots, N). \quad (2.23)$$

So $y_n^{(\nu)}$ and p_ν give us an approximation for f_n if ν is large enough, whereas p_ν is in general unknown. But by using the normalizing series S , we can compute an approximation for f_n if ν is large enough, defining

$$f_n^{(\nu)} := y_n^{(\nu)} \frac{S}{S^{(\nu)}} \quad (2.24)$$

with $S^{(\nu)}$ defined by

$$S^{(\nu)} := \sum_{k=0}^{\nu} c_k y_k^{(\nu)}.$$

Replacing $y_k^{(\nu)}$ by $p_\nu f_k$ in $S^{(\nu)}$ and using (2.23), we get the asymptotic relation $p_\nu \sim S^{(\nu)}/S$.

So we can use $f_n^{(\nu)}$ as an approximation for f_n if the integer ν is large enough.

We call the algorithm convergent if

$$\lim_{\nu \rightarrow \infty} f_n^{(\nu)} = f_n$$

for $n = 0, 1, \dots, N$. Sometimes an initial value $f_0 \neq 0$ is known. If $y_0^{(\nu)} \neq 0$, then (2.24) can be replaced by

$$f_n^{(\nu)} := y_n^{(\nu)} \frac{f_0}{y_0^{(\nu)}}. \quad (2.25)$$

Of course, another question is the determination or estimation of a starting index ν . For this we usually need some information on asymptotic estimates of the dominant and the minimal solution. A very detailed investigation of the Miller algorithm can be found in [Ga2]. Here, Gautschi considered asymptotic estimates of the underlying special functions to obtain determinations of the starting value of the backward recursion. Also, the influence of the choice of the normalizing series is investigated.

Finally, we will look at an example which makes these problems clear and uses a recurrence relation which will play an important role in the following chapter.

Example 2.14

Let us consider the following three-term recurrence relation

$$(2n + 1)(2n + 2)y_{n+1} = (4 + 12n)y_n - 8y_{n-1} \quad (n = 1, 2, \dots) \quad (2.26)$$

with known initial values $y_0 = 1$, $y_1 = \sqrt{\pi}/2$. The results of the forward recursion (\hat{y}_n) have been computed with a precision of 16 decimal digits (MATLAB) using relation (2.26). The exact values (y_n) have been produced using exact arithmetic from iRRAM (see [Mue1]). The results are compared in Table 2.1 where the underlined digits coincide with the correct ones.

n	\hat{y}_n	y_n
0	<u>1.0000000000000000</u> E+000	1.0000000000000000E+000
1	<u>7.267604552648375</u> E-001	7.267604552648373E-0001
2	<u>3.023472736864499</u> E-001	3.023472736864498E-0001
3	<u>8.838800070339661</u> E-002	8.838800070339648E-0002
4	<u>1.994181854721902</u> E-002	1.994181854721895E-0002
5	<u>3.665228431424621</u> E-003	3.665228431424596E-0003
...
10	<u>9.112388225874632</u> E-008	9.112388225872194E-0008
...
20	<u>9.958891087018652</u> E-020	9.958891083304824E-0020
...
30	<u>7.554710592297667</u> E-034	7.554707104224064E-0034
...
40	<u>2.142349065449608</u> E-049	2.141187882624492E-0049
42	<u>1.217332184619682</u> E-052	1.214634895827316E-0052
44	<u>6.334787999685751</u> e-056	6.277762866654770E-0056
45	<u>1.405382835570240</u> e-057	1.380038332000917E-0057
46	<u>3.078657699180013</u> e-059	2.968464205400354E-0059
...

Table 2.1: Approximative solution \hat{y}_n and exact values y_n
using (2.26) (forward computation).

We can see that, if $n = 30$, we already have a loss of 11 exact digits, and if $n = 46$, all digits are incorrect.

As we have mentioned above, one idea (Miller algorithm) is to compute $\hat{y}_n^{(\nu)}$ using relation (2.26) again, but now in backward direction. We may start with the values $\hat{y}_K^{(\nu)} = 1$ and $\hat{y}_{K+1}^{(\nu)} = 0$ (K large enough) and then the exact values are obtained by rescaling with factor $y_0/\hat{y}_0^{(\nu)} = 1/\hat{y}_0^{(\nu)}$.

To illustrate the importance of the starting index, we show two numerical examples; in the first example we have chosen $\nu = 60$ as the starting index with $\hat{y}_{61}^{(60)} = 0$ and $\hat{y}_{60}^{(60)} = 1$, in the second one we have chosen $\nu = 100$ as the starting index with $\hat{y}_{101}^{(100)} = 0$ and $\hat{y}_{100}^{(100)} = 1$ (see Table 2.2).

n	$\hat{y}_n^{(60)}$	$\hat{y}_n^{(100)}$	y_n
101	...	0.000000000000000E+000	1.198710381539549E-0159
90	...	<u>7.922280341562517E-140</u>	7.929633639699585E-0140
80	...	<u>1.743931440570137E-120</u>	1.743932931270254E-0120
70	...	<u>1.112241175786767E-101</u>	1.112241176656552E-0101
60	1.000000000000000E+000	<u>1.726005534410427E-083</u>	1.726005534411647E-0083
50	<u>5.154070593166293E-066</u>	<u>5.158682093073333E-066</u>	5.158682093073330E-0066
40	<u>2.141186203928507E-049</u>	<u>2.141187882624492E-049</u>	2.141187882624492E-0049
30	<u>7.554707099181441E-034</u>	<u>7.554707104224069E-034</u>	7.554707104224064E-0034
20	<u>9.958891083299465E-020</u>	<u>9.958891083304826E-020</u>	9.958891083304824E-0020
10	<u>9.112388225872196E-008</u>	<u>9.112388225872196E-008</u>	9.112388225872194E-0008
...
4	<u>1.994181854721896E-002</u>	<u>1.994181854721895E-002</u>	1.994181854721895E-0002
3	<u>8.838800070339649E-002</u>	<u>8.838800070339649E-002</u>	8.838800070339648E-0002
2	<u>3.023472736864498E-001</u>	<u>3.023472736864498E-001</u>	3.023472736864496E-0001
1	<u>7.267604552648374E-001</u>	<u>7.267604552648374E-001</u>	7.267604552648373E-0001
0	<u>1.000000000000000E+000</u>	<u>1.000000000000000E+000</u>	1.000000000000000E+0000

Table 2.2: Approximative solutions $\hat{y}_n^{(60)}$, $\hat{y}_n^{(100)}$ using (2.26) (in backward direction) and exact values y_n (iRRAM).

Chapter 3

The Computation of Mittag-Leffler and Confluent Hypergeometric Functions by Modified Taylor Series

In the second chapter we developed a general method for computing entire functions in several parts of the complex plane.

Now, in this chapter, we consider special functions, in particular the class of Mittag-Leffler and confluent hypergeometric functions. Both classes include many important special functions. We will consider the error function and the incomplete gamma function in detail. Using the ideas described in Chapter 2, we will see that the partial sums of the resulting series expansions are easily computable and the problem of cancellation errors is considerably reduced compared to the corresponding Taylor sections.

3.1 Mittag-Leffler and confluent hypergeometric functions

First of all, we give some basic definitions and characteristics of the Mittag-Leffler and the confluent hypergeometric functions.

Definition 3.1

For $\alpha > 0$, $\beta \in \mathbb{C}$ the functions

$$M(\alpha, \beta; z) := \sum_{\nu=0}^{\infty} \frac{z^{\nu}}{\Gamma(\alpha\nu + \beta)} \quad (z \in \mathbb{C}) \quad (3.1)$$

are called generalized Mittag-Leffler functions.

For $\beta = 1$ the functions

$$M(\alpha; z) := M(\alpha, 1; z) \quad (z \in \mathbb{C}) \quad (3.2)$$

are briefly called Mittag-Leffler functions.

For certain values of the parameters α and β the functions $M(\alpha, \beta; z)$ reduce to elementary functions. For example we have

$$M(1; z) = e^z, \quad M(2; -z^2) = \cos z \quad (z \in \mathbb{C}), \quad (3.3)$$

$$M(2; z) = \cosh \sqrt{z}, \quad M(2, 2; z) = \frac{\sinh \sqrt{z}}{\sqrt{z}} \quad (z \in \mathbb{C}). \quad (3.4)$$

The functions $M(\alpha; z)$ have been introduced in 1903 by Mittag-Leffler as a generalisation of the exponential function.

We also consider the confluent hypergeometric functions with their basic properties. In order to define the confluent hypergeometric functions, which are also called Kummer functions, we start with the Gaussian hypergeometric equation

$$z(1-z)w'' + (c - (a+b+1)z)w' - abw = 0, \quad (3.5)$$

with its singularities $z = 0$, $z = 1$ and $z = \infty$.

For $\xi \in \mathbb{C}$ and $\nu \in \mathbb{N}$ the Pochhammer symbol is defined by

$$(\xi)_\nu = \xi(\xi + 1)(\xi + 2) \cdots (\xi + \nu - 1), \quad (\xi)_0 = 1.$$

The equation, given in (3.5), is solved by the Gaussian hypergeometric functions $w = F(a, b, c; z)$, defined by

$$F(a, b, c; z) = \sum_{\nu=0}^{\infty} \frac{(a)_\nu (b)_\nu z^\nu}{(c)_\nu \nu!} \quad (|z| < 1) \quad (3.6)$$

with complex parameters a , b and c where $c \notin (-\mathbb{N}_0)$.

We obtain the confluent hypergeometric functions when two of the singularities of (3.5) merge into one singularity ("confluence" of two singularities). In order to describe this formal process, we consider the hypergeometric functions $F(a, b, c; z/b)$ which have a singularity at $z = b$.

We define

$$\Phi(a, c; z) = \lim_{b \rightarrow \infty} F(a, b, c; z/b). \quad (3.7)$$

Since we have

$$\lim_{b \rightarrow \infty} \frac{(b)_n}{b^n} = 1,$$

the computation of the limits of the terms of the power series (3.6) finally yields the power series representation of the confluent hypergeometric functions of the first kind

$$\Phi(a, c; z) = \sum_{\nu=0}^{\infty} \frac{(a)_\nu z^\nu}{(c)_\nu \nu!} \quad (z \in \mathbb{C}) \quad (3.8)$$

with complex parameters a and c , where $c \notin (-\mathbb{N}_0)$.

An important functional relation, namely

$$\Phi(a, c; z) = e^z \Phi(c - a, c; -z), \quad (3.9)$$

is called Kummer transformation.

From their series expansions the following connection between the Mittag-Leffler and the confluent hypergeometric functions can be seen easily:

$$\Phi(1, \alpha + 1; z) = \alpha \Gamma(\alpha) M(1; \alpha + 1; z). \quad (3.10)$$

In the following sections our aim is to characterize the Mittag-Leffler and the confluent hypergeometric functions, especially their growth will be considered.

3.1.1 Order and Type

The order and type of entire functions (see (2.4) and (2.5)) are important quantities to characterize the functions and, as we have seen in the second chapter, they play an essential role if we want to quantify the problem of cancellation.

With the help of the Theorem 2.7 we immediately see that the generalized Mittag-Leffler functions $M(\alpha, \beta; z)$ are entire functions of the order $1/\alpha$ and type 1 for all $\beta \in \mathbb{C}$.

Indeed: By using the Stirling formula

$$\Gamma(\alpha n + \beta) = \sqrt{2\pi} (\alpha n)^{\alpha n + \beta - 1/2} e^{-\alpha n} (1 + o(1)) \quad (n \rightarrow \infty), \quad (3.11)$$

we obtain

$$\rho = \limsup_{n \rightarrow \infty} \frac{\log n}{-\log |1/\Gamma(\alpha n + \beta)|^{1/n}} = 1/\alpha$$

and

$$\tau = \limsup_{n \rightarrow \infty} \frac{n\alpha}{e} \left| \frac{1}{\Gamma(\alpha n + \beta)} \right|^{1/(\alpha n)} = 1.$$

Using Theorem 2.7 again, we obtain that the confluent hypergeometric functions $\Phi(a, c; z)$ are entire functions of the order $\rho = 1$ and type $\tau = 1$.

Indeed: With the use of the asymptotic representation of the Pochhammer symbol

$$(\xi)_{n+1} \sim \frac{n^\xi n!}{\Gamma(\xi)} \quad (n \rightarrow \infty),$$

and a simplified version of the Stirling formula

$$(n!)^{1/n} \sim \frac{n}{e} \quad (n \rightarrow \infty),$$

we obtain

$$\rho = \limsup_{n \rightarrow \infty} \frac{\log n}{-\log |(a)_n / ((c)_n n!)|^{1/n}} = 1$$

and

$$\tau = \limsup_{n \rightarrow \infty} \frac{n}{e} \left| \frac{(a)_n}{(c)_n n!} \right|^{1/n} = 1.$$

3.1.2 Asymptotic expansions and Indicator Functions

Further, we need some information about the functions for large arguments z in modulus. For this we first look at the asymptotic expansions of the Mittag-Leffler and the confluent hypergeometric functions. After that, using the asymptotic expansions, we can find the indicator functions which reflect the asymptotic growth behaviour in a rather coarse manner. As we have seen in the second chapter, the indicator functions are of great importance to quantify and reduce cancellation problems.

At first, the definition and fundamental properties of asymptotic expansions are given.

Definition 3.2

Let $\sum_{\nu=0}^{\infty} a_{\nu} z^{\nu}$ be a formal power series and $S \subset \mathbb{C}$ an unbounded region, and let $f : S \rightarrow \mathbb{C}$ be a function such that

$$f(z) = \sum_{\nu=0}^{n-1} \frac{a_{\nu}}{z^{\nu}} + R_n(z) \quad (3.12)$$

where we have for each fixed $n \in \mathbb{N}$

$$R_n(z) = \mathcal{O}(1/|z|^n) \quad (z \rightarrow \infty \text{ in } S). \quad (3.13)$$

Then we call the series $\sum_{\nu=0}^{\infty} a_{\nu} z^{-\nu}$ an asymptotic expansion of $f(z)$ in S and write

$$f(z) \sim \sum_{\nu=0}^{\infty} \frac{a_{\nu}}{z^{\nu}} \quad (z \rightarrow \infty \text{ in } S). \quad (3.14)$$

If the series $\sum_{\nu=0}^{\infty} a_{\nu} z^{-\nu}$ converges for all sufficiently large $|z|$, then it is the asymptotic expansion of its sum without restriction to $\arg(z)$.

We give a necessary and sufficient condition for the existence of asymptotic

expansions.

Remark 3.3

If and only if we have for all $n \in \mathbb{N}_0$

$$z^n \left(f(z) - \sum_{\nu=0}^{n-1} \frac{a_\nu}{z^\nu} \right) \rightarrow a_n \quad (z \rightarrow \infty \text{ in } S), \quad (3.15)$$

the function $f(z)$ has an asymptotic expansion of the form (3.14).

This implies that the sequence (a_n) is uniquely determined. Therefore, every function f can have at most one asymptotic expansion in S .

Asymptotic expansions of Mittag-Leffler and confluent hypergeometric functions

For the asymptotic expansions of Mittag-Leffler functions $M(\alpha, \beta; z)$ we refer to [DZ]. The results are composed in the following two theorems for the cases $\alpha < 2$ and $\alpha \geq 2$.

Theorem 3.4

Let $0 < \alpha < 2$, $z, \beta \in \mathbb{C}$ be given and

$$\frac{\alpha\pi}{2} < \varphi < \min \{ \pi, \alpha\pi \}.$$

For each $N \in \mathbb{N}$ und $|z| \rightarrow \infty$ we get:

a) If $S = \{z : |\arg z| \leq \varphi\}$, we have

$$M(\alpha, \beta; z) = \frac{1}{\alpha} z^{(1-\beta)/\alpha} e^{z^{1/\alpha}} - \sum_{k=1}^N \frac{z^{-k}}{\Gamma(\beta - k\alpha)} + R_N(z) \quad (3.16)$$

with

$$R_N(z) \rightarrow 0 \quad (z \rightarrow \infty \text{ in } S).$$

b) If $S = \{z : \varphi \leq |\arg z| \leq \pi\}$, we have

$$M(\alpha, \beta; z) = - \sum_{k=1}^N \frac{z^{-k}}{\Gamma(\beta - k\alpha)} + R_N(z) \quad (3.17)$$

with

$$R_N(z) \rightarrow 0 \quad (z \rightarrow \infty \text{ in } S).$$

Theorem 3.5

Let $\alpha \geq 2$, $z, \beta \in \mathbb{C}$ and $N \in \mathbb{N}$ be given. Then, we have:

$$M(\alpha, \beta; z) = \frac{1}{\alpha} \sum_{h \in A(z)} \left(z^{1/\alpha} e^{2\pi i h / \alpha} \right)^{(1-\beta)} e^{e^{i2\pi h / \alpha} z^{1/\alpha}} - \sum_{k=1}^N \frac{z^{-k}}{\Gamma(\beta - k\alpha)} + R_N(z),$$

with

$$A(z) = \{h : h \in \mathbb{Z}, |\arg z + 2\pi h| \leq \frac{\alpha\pi}{2}\}$$

and

$$R_N(z) \rightarrow 0 \quad (z \rightarrow \infty).$$

For the asymptotic expansions of the confluent hypergeometric functions $\Phi(a, c; z)$ we refer to [AS].

Theorem 3.6

For fixed parameters a and c we have

$$\begin{aligned} \Phi(a, c; z) &= \frac{\Gamma(c)}{\Gamma(c-a)} \frac{e^{i\pi a}}{z^a} \sum_{k=0}^{N_1} \frac{(a)_k (a-c+1)_k}{k! (-z)^k} + \mathcal{O}(|z|^{-1-N_1}) \\ &+ \frac{\Gamma(c)}{\Gamma(a)} e^z z^{a-c} \sum_{k=0}^{N_2} \frac{(c-a)_k (1-a)_k}{k! z^k} + \mathcal{O}(|z|^{-1-N_2}) \quad (z \rightarrow \infty \text{ in } S) \end{aligned} \quad (3.18)$$

with

$$S = \{z : -\frac{1}{2}\pi < \arg z < \frac{3}{2}\pi\}.$$

It should be mentioned that the term on the right hand side of the formulas (3.16) and (3.18) results in the exponential function for the special cases of $\alpha = \beta = 1$ and $a = c$. In this cases the term on the right hand side of the formula (3.17) results in 0.

Indicator functions of Mittag-Leffler and confluent hypergeometric functions

Using the definition of the indicator function (see (2.10)) and the asymptotic

expansions, given in the theorems above, the indicator functions of the Mittag-Leffler and the confluent hypergeometric functions can be obtained.

The indicator functions of the Mittag-Leffler functions for $\alpha, \beta \neq 1$ are given as

$$h_{M(\alpha, \beta; \cdot)}(\vartheta) = \begin{cases} \cos(\vartheta/\alpha), & |\vartheta| \leq \alpha\pi/2 \\ 0, & \text{else} \end{cases}, \quad (3.19)$$

and for the indicator functions of the confluent hypergeometric functions for $a \neq c$ we find

$$h_{\Phi(a, c; \cdot)}(\vartheta) = \begin{cases} \cos(\vartheta), & |\vartheta| \leq \pi/2 \\ 0, & \text{else} \end{cases}. \quad (3.20)$$

3.2 Special cases of Mittag-Leffler and confluent hypergeometric functions

As we have mentioned above, the classes of Mittag-Leffler and confluent hypergeometric functions include many important special functions.

Thus, at first we give a short summary of certain special functions as well as their connections to the Mittag-Leffler and the confluent hypergeometric functions.

Error functions

The error functions play an important role in statistics and probability theory. The *error function* erf and *complementary error function* erfc are defined by

$$\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt \quad (z \in \mathbb{C}),$$

$$\operatorname{erfc}(z) = 1 - \operatorname{erf}(z) \quad (z \in \mathbb{C}),$$

and the relations to the Mittag-Leffler and the confluent hypergeometric functions are

$$\operatorname{erf}(z) = z \, \Phi\left(\frac{1}{2}, \frac{3}{2}; -z^2\right), \quad (3.21)$$

$$\operatorname{erfc}(-z) = M(1/2; z) e^{-z^2}. \quad (3.22)$$

Here we can also mention DAWSON's function given by

$$\text{daw}(z) = \frac{\sqrt{\pi}}{2i} e^{-z^2} \text{erf}(iz) = e^{-z^2} \int_0^z e^{t^2} dt \quad (z \in \mathbb{C})$$

and the FADEEWA function

$$\omega(z) = M(1/2; iz) \quad (z \in \mathbb{C}).$$

Incomplete gamma functions

The *incomplete gamma function* is defined by

$$\gamma(\alpha, z) = \int_0^z t^{\alpha-1} e^{-t} dt \quad (\text{Re}(\alpha) > 0).$$

For the numerical evaluation of $\gamma(\alpha, x)$ for $x \geq 0$ and $\alpha \in \mathbb{R}$ see Gautschi [Ga3].

The connections with the Mittag-Leffler and the confluent hypergeometric functions are

$$\gamma(\alpha, z) = \frac{z^\alpha}{\alpha} e^{-z} \Phi(1, \alpha + 1; z), \quad (3.23)$$

$$\gamma(\alpha, z) = z^\alpha e^{-z} \Gamma(\alpha) M(1; \alpha + 1; z). \quad (3.24)$$

3.3 The computation of the error function

In the following section we consider the complementary error function in detail and we look at the cancellation errors which occur if we compute this function using series expansions. With the help of the method described in Section 2.4, we will see that the problem of cancellation errors is considerably reduced compared to the corresponding Taylor sections.

Suppose now that

$$f(z) = \text{erfc}(-z) = 1 + \text{erf}(z) = M(1/2, z) e^{-z^2}.$$

Considering this connection with the Mittag-Leffler functions, we obtain that $\rho_f = 2$ and with (3.19) we have the indicator function given as

$$h_f(\vartheta) = \begin{cases} 0, & |\vartheta| \leq \pi/4 \\ -\cos(2\vartheta), & |\vartheta| > \pi/4 \end{cases}.$$

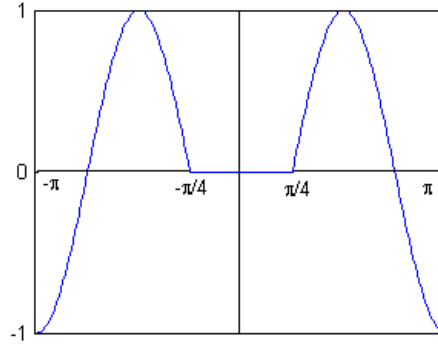


Figure 3.1: Indicator function of $f(z) = \operatorname{erfc}(-z)$.

In Poppe's and Wijers' Algorithm 680 for the computation of the error function (see [PW1], [PW2]), which is based on Gautschi's algorithm [Ga1] and which may be viewed as a benchmark for algorithms concerning the evaluation of the complex error function (see, e.g., [Wei]), truncation of the Taylor series

$$f(z) = 1 + \operatorname{erf}(z) = 1 + \frac{2}{\sqrt{\pi}} \sum_{\nu=0}^{\infty} \frac{(-1)^{\nu} z^{2\nu+1}}{(2\nu+1)\nu!} \quad (3.25)$$

is performed in the second quadrant

$$S = \{z = re^{i\vartheta} : r \geq 0, \pi/2 \leq \vartheta \leq \pi\}$$

and for $|z| = r$ sufficiently small.

Moreover, the calculation in the remaining quadrants is reduced to the second one by elementary operations.

According to the considerations in (2.13), we are faced with a loss of significant decimal digits of about

$$\delta_f(\vartheta)r^2/\log(10) = \begin{cases} r^2/\log(10), & |\vartheta| \leq \pi/4, \\ (1 + \cos(2\vartheta))r^2/\log(10), & |\vartheta| > \pi/4, \end{cases}$$

for $z = re^{i\vartheta}$ and r large.

So, in S , the Taylor sections behave well on the imaginary axis, and the “worst case” appears in a neighborhood of the negative semiaxis (where $\operatorname{erfc}(-z)$ is very small).

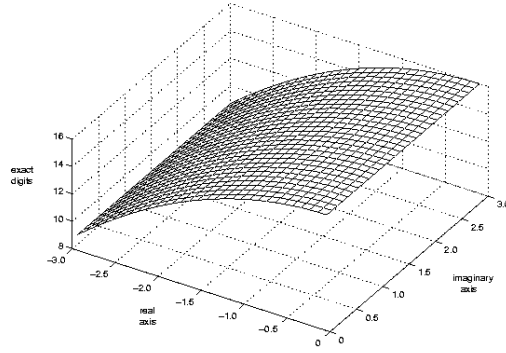


Figure 3.2: $16 - \delta_f(\vartheta)r^2/\log(10)$ in the second quadrant S .

Figure 3.2 shows the values of $16 - \delta_f(\vartheta)r^2/\log(10)$ which, according to (2.13), may be viewed as a theoretical measure for the exact digits if computations are performed in floating point arithmetic with fixed precision of 16 decimal digits.

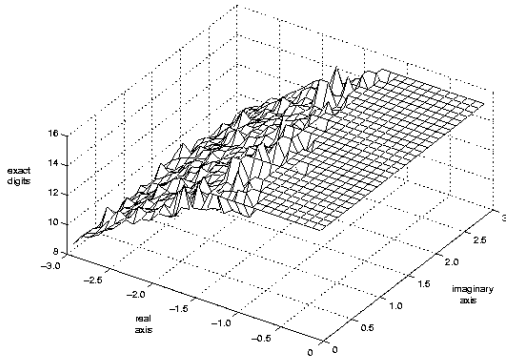


Figure 3.3: $e_{f,s_{N(z),16}(f,\cdot)}(z)$ in S .

Moreover, Figure 3.3 shows $e_{f,s_{N(z),16}(f,\cdot)}(z)$, where

$$e_{f,g}(z) := -\log_{10} \left(\frac{|f(z) - g(z)|}{|f(z)|} \right), \quad (3.26)$$

that is, the decimal logarithm of the relative error when replacing $f(z)$ by $s_{N(z),16}(f, z)$. Once again, the value $e_{f,s_{N(z),16}(f,\cdot)}(z)$ measures approximately the smaller one of the two numbers of exact digits of the real part and the imaginary part of $f(z)$ if $f(z)$ is approximated by $s_{N(z),16}(f, z)$ and the computations are performed in double precision arithmetic providing an accuracy of about 16 decimal digits.

The reference values for $f(z)$ were produced using exact arithmetic from iR-RAM (see [Mue1]) again. Since the usual accuracy requirement of special function routines is 15 digits in double precision, we have cut off the error at a level of 15 digits. Therefore, all values on the 15-digit level represent approximations within the usual tolerance.

The numerical results shown in Figure 3.3 essentially fit the theoretical values from Figure 3.2 and thus support the above considerations on the loss of exact decimal digits.

So we are confronted with the question of how cancellation problems can be reduced. Now, we take a look at the method described in Section 2.4.

Since f is of order 2, an evident choice for Φ is

$$\Phi = \{\varphi(a, \cdot) : a \in A\}, \quad \varphi(a, z) := e^{az^2} \quad (z \in \mathbb{C}),$$

where $A \subset \mathbb{C}$. We (first) restrict our investigations to the case $A = \mathbb{R}$. Then

$$h_{\varphi(a, \cdot)}(\vartheta) = a \cos(2\vartheta) \quad (\vartheta \in [-\pi, \pi]),$$

and thus, according to Remark 2.10 iii),

$$(h_f + h_{\varphi(a, \cdot)})(\vartheta) = \begin{cases} a \cdot \cos(2\vartheta), & |\vartheta| \leq \pi/4, \\ (a - 1) \cos(2\vartheta), & |\vartheta| > \pi/4. \end{cases}$$

In particular, we obtain

$$\max_{[-\pi, \pi]} (h_f + h_{\varphi(a, \cdot)}) = \max(a, |a - 1|) =: \tau(a)$$

and so the optimization problem (2.18) here reads as

$$\max_{\vartheta \in \Theta} \left(\tau(a) - (h_f + h_{\varphi(a, \cdot)})(\vartheta) \right) \rightarrow \min_{a \in \mathbb{R}}. \quad (3.27)$$

If Θ contains one of the points $\pm\pi/4$ or $\pm3\pi/4$ (where $h_f - h_{\varphi(a, \cdot)}$ vanishes), then obviously the maximum in (3.27) is $\geq \tau(a)$, which is minimal exactly for $a = 1/2$. For $a = 1/2$ and $|\vartheta| \leq 3\pi/4$ we find

$$\tau(a) - (h_f + h_{\varphi(a, \cdot)})(\vartheta) = \frac{1}{2}(1 - |\cos(2\vartheta)|) \leq \frac{1}{2},$$

so $a = 1/2$ turns out to be the (unique) solution of (3.27) for all subsets Θ of the interval $[-3\pi/4, 3\pi/4]$, with $\{\pm\pi/4, \pm3\pi/4\} \cap \Theta \neq \emptyset$.

As already mentioned above, in the algorithm of Poppe and Wijers, truncation of the Taylor series

$$f(z) = 1 + \frac{2}{\sqrt{\pi}} \sum_{\nu=0}^{\infty} \frac{(-1)^{\nu} z^{2\nu+1}}{(2\nu+1)\nu!} \quad (3.28)$$

is performed in the second quadrant $S = \{z = re^{i\vartheta} : r \geq 0, \pi/2 \leq \vartheta \leq \pi\}$ and for $|z| = r$ sufficiently small.

The advice is to use Taylor sections of

$$(f\varphi)(z) := f(z)\varphi\left(\frac{1}{2}, z\right) = \operatorname{erfc}(-z)e^{z^2/2} \quad (3.29)$$

instead. The following figures show the indicator function of $f\varphi$ and the corresponding approximation for the number of exact digits.

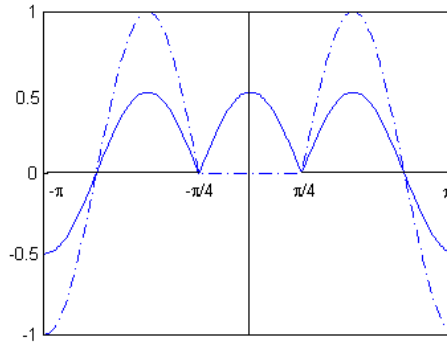


Figure 3.4: Indicator functions of $f\varphi$ (solid curve) and f (dotted curve).

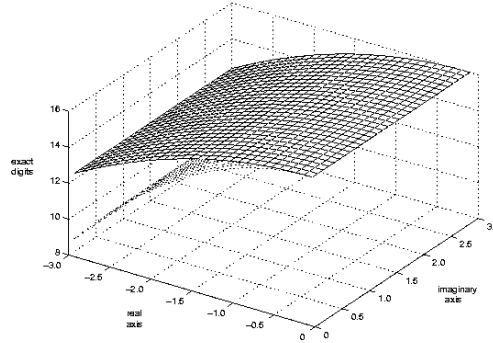


Figure 3.5: $16 - \delta_{f\varphi}(\vartheta)r^2/\log(10)$ (solid line) and $16 - \delta_f(\vartheta)r^2/\log(10)$ (dotted line) in the second quadrant S .

For the computation of the Taylor sections of (3.29) it is important to have a reasonable representation for the coefficients.

Theorem 3.7

For $z \in \mathbb{C}$ we have

$$(f\varphi)(z) = f(z)\varphi\left(\frac{1}{2}, z\right) = \sum_{\nu=0}^{\infty} b_{\nu}z^{\nu}, \quad (3.30)$$

with

$$b_{\nu} = \begin{cases} \frac{1}{l!2^l} & \text{if } \nu = 2l, \\ \frac{1}{\Gamma(l+3/2)} \sum_{k=0}^l \binom{k-1/2}{k} (-1)^k & \text{if } \nu = 2l+1. \end{cases} \quad (3.31)$$

Proof: From (3.22) we have the relation

$$\operatorname{erfc}(-z) = M(1/2; z)e^{-z^2}.$$

Hence, in terms of the Cauchy product we get

$$f(z)\varphi\left(\frac{1}{2}, z\right) = M(1/2; z)e^{-\frac{1}{2}z^2} = \sum_{\nu=0}^{\infty} b_{\nu}z^{\nu},$$

with

$$b_{\nu} = \begin{cases} \frac{1}{l!2^l} & \text{if } \nu = 2l, \\ \frac{1}{\Gamma(l+3/2)} \sum_{k=0}^l \binom{l+1/2}{k} \left(\frac{-1}{2}\right)^k & \text{if } \nu = 2l+1. \end{cases} \quad (3.32)$$

So it remains to consider the case $\nu = 2l+1$.

The binomial theorem and an index shift show that for $x \in \mathbb{R}$, $z \in \mathbb{C}$, and $n \in \mathbb{N}_0$, we have

$$\sum_{k=0}^n \binom{x-n+k}{k} z^k (1+z)^{n-k} = \sum_{\mu=0}^n z^{\mu} \sum_{k=0}^{\mu} \binom{x-n+k}{k} \binom{n-k}{\mu-k}.$$

Furthermore, we obtain for $a, b \in \mathbb{C}$, $\mu \in \mathbb{N}_0$

$$\sum_{k=0}^{\mu} \binom{a+k}{k} \binom{b+\mu-k}{\mu-k} = \binom{a+b+\mu+1}{\mu}.$$

Setting $a = x - n$, $b = n - \mu$, we have

$$\sum_{k=0}^n \binom{x+1}{k} z^k = \sum_{k=0}^n \binom{x-n+k}{k} z^k (1+z)^{n-k}. \quad (3.33)$$

Then applying (3.33) and setting $x = l - 1/2$, $n = l$, $z = -1/2$ in (3.32), we finally get the assertion. \diamond

For

$$m_k := \binom{k - 1/2}{k}$$

we have $m_{k+1} = \frac{k + 1/2}{k + 1} m_k$, $k \in \mathbb{N}_0$. Thus, the sum $\sum_{k=0}^l \binom{k - 1/2}{k} (-1)^k$ in (3.30) and therefore also the Taylor coefficients b_ν of (3.29) can be evaluated recursively (note that $\Gamma(l + \frac{3}{2}) = (l + \frac{1}{2})(l - \frac{1}{2}) \dots \frac{1}{2}\sqrt{\pi}$).

Even more suitable for numerical purposes is the two-term recursion

$$(\nu + 1)(\nu + 2)b_{\nu+2} = b_\nu + b_{\nu-2}, \quad b_{-2} := b_{-1} := 0, \quad b_0 = 1, \quad b_1 = \frac{2}{\sqrt{\pi}}$$

for the coefficients b_ν , which may be found by applying the above relations twice. More directly, this recursion follows from the fact that $F := f\varphi$ satisfies the differential equation $F'' = (1 + z^2)F$.

We take $\frac{1}{\varphi(z)} s_{N(z),16}(f\varphi, z)$ (for $N(z)$ sufficiently large) as an approximation of $f(z)$ for $z \in S$ of small modulus and compare the results with the exact values of $f(z)$, which again were produced using exact arithmetic (iRRAM).

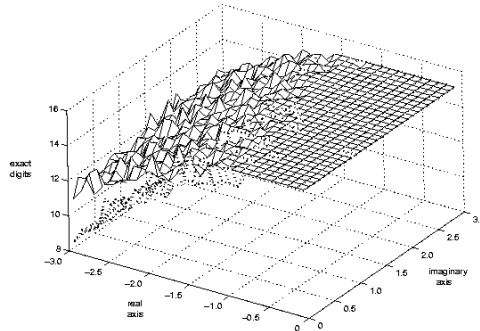


Figure 3.6: $e_{f,s_{N(z),16}(f\varphi,\cdot)}/\varphi(z)$ (solid line), $e_{f,s_{N(z),16}(f,\cdot)}(z)$ (dotted line) in S .

The numerical results shown in Figure 3.6 support our theoretical considerations which are illustrated in Figure 3.5 and demonstrate the advantages of

this method compared to the computation of the Taylor sections of f (see Figures 3.2 and 3.3).

Although there is an improvement, the Taylor sections of $f\varphi$ also turn out to be numerically unstable with respect to cancellation near the negative axis where $\delta_{f\varphi}$ is maximal (cf. Figure 3.4). So the question arises, if there is an appropriate entire function ψ such that $\delta_{f\psi}(\vartheta)$ is smaller than $\delta_{f\varphi}(\vartheta)$ for $\vartheta \approx \pi$. If we take

$$\psi(z) = \operatorname{erfc}(z) e^{2z^2},$$

then we obtain for the indicator function of $f\psi$, using Remark 2.13 and (3.19),

$$h_{f\psi}(\vartheta) = \begin{cases} 0, & \pi/4 \leq |\vartheta| \leq 3\pi/4 \\ \cos(2\vartheta), & \text{else} \end{cases},$$

and therefore $\delta_{f\psi}(\pi) = 0$.

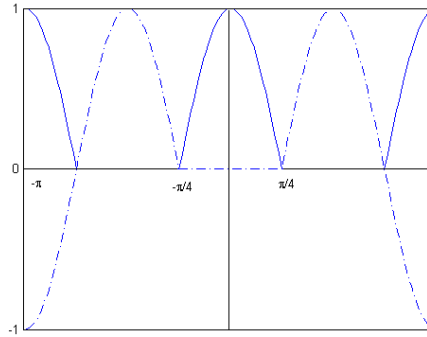


Figure 3.7: Indicator functions of $f\psi$ (solid curve) and f (dotted curve).

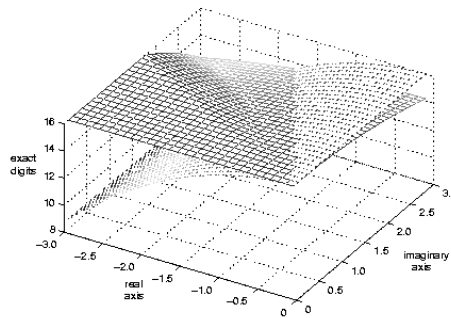


Figure 3.8: $16 - \delta_{f\psi}(\vartheta)r^2/\log(10)$ (solid), $16 - \delta_f(\vartheta)r^2/\log(10)$ (dotted) in S .

Thus it is reasonable to take $\frac{1}{\psi(z)} s_{N(z),16}(f\psi, z)$ as an approximation of $f(z)$, in particular, close to the negative real axis. The multiplication by $1/\psi(z)$ in the second quadrant S requires the evaluation of $\operatorname{erfc}(w)$ in the fourth (or the first) quadrant. In this part of the plane the Taylor sections of $f\varphi$ from (3.29) turn out to be sufficiently well behaved. So we actually replace $\psi(z)$ by $\tilde{\psi}(z) := e^{2z^2} s_{N(-z),16}(f\varphi, -z)/\varphi(-z)$.

In order to find the Taylor coefficients of $f\psi$ with respect to the origin, we just have to apply the Cauchy product again. This leads to

$$(f\psi)(z) = \operatorname{erfc}(-z)\psi(z) = M(1/2; z) M(1/2; z) = \sum_{\nu=0}^{\infty} c_{\nu} z^{\nu}, \quad (3.34)$$

with

$$c_{\nu} = \begin{cases} 0 & \text{if } \nu \text{ is odd,} \\ \sum_{\mu=0}^{\nu} \frac{(-1)^{\mu}}{\Gamma(\frac{1}{2}\mu + 1)\Gamma(\frac{1}{2}(\nu - \mu) + 1)} & \text{if } \nu \text{ is even.} \end{cases} \quad (3.35)$$

As in the case of $f\varphi$ above, the coefficients c_{ν} , ν even, can be evaluated by recursion.

Setting $r := \nu/2$, we write

$$d_r := \sum_{\mu=0}^{2r} \frac{(-1)^{\mu}}{\Gamma(\frac{1}{2}\mu + 1)\Gamma(r + 1 - \frac{1}{2}\mu)} \quad (r \in \mathbb{N}_0).$$

We get

$$d_r = \frac{2^r}{r!} - s_{r-1}, \quad (3.36)$$

with

$$s_n := \sum_{\nu=0}^n \frac{1}{\Gamma(\nu + \frac{3}{2})\Gamma(n - \nu + \frac{3}{2})} \quad (n \in \mathbb{N}_0), \text{ and } s_0 = \frac{4}{\pi}.$$

Then we can obtain that

$$s_{n+1} = \frac{2}{n+2} s_n + \frac{1}{(n+2)\Gamma(\frac{3}{2})\Gamma(n + \frac{5}{2})} \quad (n \in \mathbb{N}_0), \text{ and } s_0 = \frac{4}{\pi}. \quad (3.37)$$

If we understand (3.37) as a difference equation, we get

$$s_n = \frac{2^n}{(n+1)!} \sum_{\nu=0}^n \frac{\nu!}{2^{\nu}\Gamma(\frac{3}{2})\Gamma(\nu + \frac{3}{2})}. \quad (3.38)$$

For the recursion to evaluate the coefficients d_r we obtain

$$d_{r+1} - \frac{2}{r+1} d_r = \frac{2^{r+1}}{(r+1)!} - s_r - \frac{2^{r+1}}{(r+1)!} + \frac{2}{r+1} s_{r-1} = -\frac{1}{(r+1)\Gamma(\frac{3}{2})\Gamma(r + \frac{3}{2})}$$

and therefore

$$d_{r+1} = \frac{2}{r+1}d_r - \frac{2}{(r+1)\pi} \prod_{\nu=0}^r \frac{2}{2\nu+1} \quad (r \in \mathbb{N}_0), \text{ and } d_0 = 1. \quad (3.39)$$

It turns out, however, that this recursion (3.39) tends to be unstable if performed upwards. Fortunately, this problem no longer occurs if we apply it in the *backward* direction, that is, we compute with an appropriate starting value $d_{R(z)}$ (or an approximation $\tilde{d}_{R(z)}$)

$$d_r = \frac{r+1}{2}d_{r+1} + \frac{1}{\pi} \prod_{\nu=0}^r \frac{2}{2\nu+1}$$

for $r = R(z) - 1, \dots, 1, 0$.

Of course, in this case the question arises of how to get the starting value. Since d_r tends to 0 very rapidly as r tends to ∞ , it is possible to simply take $\tilde{d}_{R(z)} = 0$ for $R(z)$ sufficiently large.

In our case suitable values of $R(z)$ could be computed explicitly by using the representation

$$d_r = \frac{4}{\pi r!} \int_0^1 \frac{(1-\xi^2)^r}{1+\xi^2} d\xi \quad (r \in \mathbb{N}_0), \quad (3.40)$$

which results from (3.36) and (3.38) after some routine calculations employing Euler's Beta integral.

From (3.40) we also have the estimate

$$\sqrt{r} \, r! \, d_r \leq \frac{2}{\sqrt{\pi}} \quad (3.41)$$

being asymptotically sharp as $r \rightarrow \infty$.

Similarly as in the case of $f\varphi$, a two-term recursion for the coefficients c_ν can be obtained from the fact that the function $G := f\psi$ satisfies the differential equation

$$G'' = 4(1-2z^2)G + 6zG' - \frac{8}{\pi},$$

namely,

$$(\nu+1)(\nu+2)c_{\nu+2} = (4+6\nu)c_\nu - 8c_{\nu-2}.$$

This recursion equals the recursion (2.26) we considered in the first chapter in detail. We know this recursion is again stable (only) in the backward direction. The exact value $c_0 = 1$ is known, and so the backward recurrence may be started with the values $\hat{c}_K = 1$ and $\hat{c}_{K+2} = 0$ (for K large enough). Then the exact values are obtained by rescaling with factor $c_0/\hat{c}_0 = 1/\hat{c}_0$ as we have seen in Example 2.14.

The numerical results shown in the Figure 3.9 demonstrate the efficiency of the proposed method for the evaluation of $f(z)$ in particular near the negative real axis.

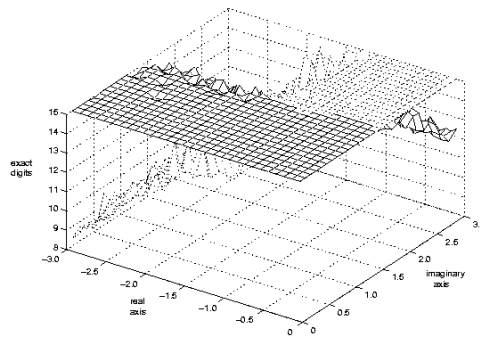


Figure 3.9: $e_{f, s_{N(z), 16}}(f\psi, \cdot)/\tilde{\psi}(z)$ (solid line), $e_{f, s_{N(z), 16}}(f, \cdot)(z)$ (dotted line) in S .

It turns out that a further improvement concerning the accuracy is possible. Using Remark 2.13 and (3.19), multiplication of $f\psi$ with $\varphi(-1/2, z) = e^{-z^2/2}$ results in the indicator function

$$h_{f\psi\varphi(-1/2, \cdot)}(\vartheta) = \frac{1}{2} |\cos(2\vartheta)|.$$

The above theory shows that $f\psi\varphi(-1/2, \cdot)$ shares the advantages of $f\varphi$ and $f\psi$ concerning the reduction of cancellation.

We were, however, not able to find a reasonable recursion relation for the coefficients, and, unfortunately, the computation from the coefficients of $f\psi$ and $\varphi(-1/2, \cdot)$ by convolution leads again to cancellation. So it seems necessary to evaluate the coefficients using exact arithmetic and then to implement them as data.

If we agree with the same disadvantage, then we can do even better in the area in which we are interested, namely, near the line $\arg(z) = 3\pi/4$.

With

$$a := e^{3\pi i/4} - 1/2$$

and $\varphi(a, z) = e^{az^2}$, we obtain for $f\psi\varphi(a, \cdot)$ the indicator function

$$h_{f\psi\varphi(a, \cdot)}(\vartheta) = h_{f\psi\varphi(-1/2, \cdot)}(\vartheta) + \cos(2\vartheta - 3\pi/2)$$

(see Figure 3.10), which is comparably "large" for $\vartheta \approx 3\pi/4$.

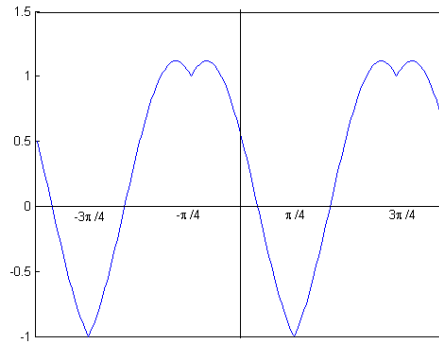


Figure 3.10: Indicator function of $f\psi\varphi(a, \cdot)$.

3.4 The computation of the incomplete gamma function

In the following section we consider the incomplete gamma function and we look at the relation to the Mittag-Leffler functions (see (3.24))

$$\gamma(\alpha, z) = z^\alpha e^{-z} \Gamma(\alpha) M(1, \alpha + 1; z).$$

Considering this equation, we will analyse the cancellation errors of the function $M(1, \beta; z)$ especially in the area near the negative axis.

Using the method described in Section 2.4, we will see that the cancellation problem is again strictly reduced compared to the corresponding Taylor

sections and that the partial sums are easily computable.

We are not interested in the special case of $M(1, \beta; z)$ being the exponential function. So for the following considerations we always assume that $\beta \neq 1$.

Suppose that

$$f(z) := M(1; \beta; z) = \sum_{k=0}^{\infty} \frac{z^k}{\Gamma(k + \beta)}.$$

Then we have already seen that $\rho_f = 1$ and for the indicator function (see (3.19)) we have

$$h_f(\vartheta) = \begin{cases} \cos(\vartheta), & |\vartheta| \leq \pi/2 \\ 0, & |\vartheta| > \pi/2 \end{cases}.$$

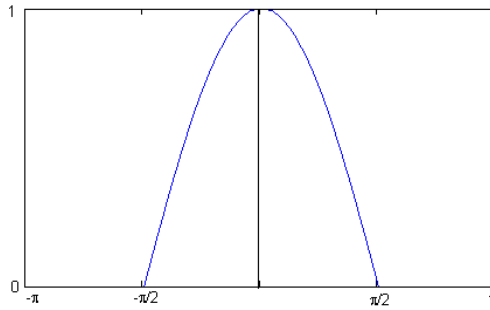


Figure 3.11: Indicator function of $f(z) = M(1; \beta; z)$.

Again we consider the second quadrant

$$S = \{z = re^{i\vartheta} : r \geq 0, \pi/2 \leq \vartheta \leq \pi\}$$

for $|z| = r$ sufficiently small, especially the area near the negative axis.

According to the considerations in (2.13), we are faced with a loss of significant decimal digits of about

$$\delta_f(\vartheta)r/\log(10) = \begin{cases} (1 - \cos(\vartheta))r/\log(10), & |\vartheta| \leq \pi/2 \\ r/\log(10), & |\vartheta| > \pi/2 \end{cases}$$

for $z = re^{i\vartheta}$ and r large. So the Taylor sections do not behave well in the whole quadrant S , except for a small neighborhood of the origin.

Using the ideas which are described in Section 2.4, we have to solve the question

if there is an appropriate entire function ψ such that $\delta_{f\psi}(\vartheta)$ is smaller than $\delta_f(\vartheta)$ for $\vartheta \approx \pi$.

If we take

$$\psi(z) = M(1, \beta, -z),$$

then we obtain for the indicator function of $f\psi$ with the help of Remark 2.13 and (3.19)

$$h_{f\psi}(\vartheta) = |\cos(\vartheta)|,$$

and therefore $\delta_{f\psi}(\pi) = 0$.

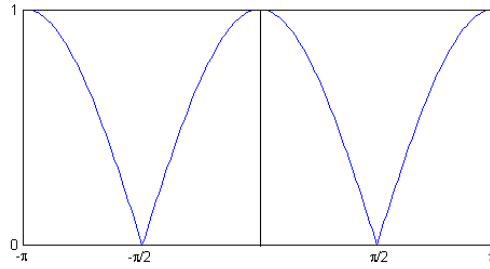


Figure 3.12: Indicator function of $f\psi$.

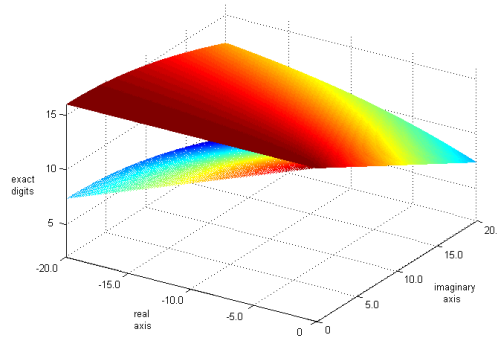


Figure 3.13: $16 - \delta_{f\psi}(\vartheta)r / \log(10)$ (solid line), $16 - \delta_f(\vartheta)r / \log(10)$ (dotted line) in S .

Thus, it is reasonable to take $\frac{1}{\psi(z)}s_{N(z),16}(f\psi, z)$ as an approximation of $f(z)$, in particular, close to the negative real axis. The multiplication by $1/\psi(z)$ in the second quadrant S requires the evaluation of $M(1, \beta, z)$ in the fourth (or the first) quadrant. In this part of the plane the Taylor sections of f turn out to be sufficiently well behaved.

It has to be noticed that all these considerations are independent from the parameter β .

In order to find the Taylor coefficients of $f\psi$ with respect to the origin, we just have to apply Cauchy product again. This leads to

$$(f\psi)(z) = M(1, \beta; z) M(1, \beta; -z) = \sum_{\nu=0}^{\infty} c_{\nu} z^{\nu}, \quad (3.42)$$

with

$$c_{\nu} = \begin{cases} 0 & , \text{ if } \nu \text{ is odd,} \\ \frac{1}{(\nu/2 + \beta - 1)\Gamma(\beta - 1)\Gamma(\nu + \beta)}, & \text{ if } \nu \text{ is even.} \end{cases} \quad (3.43)$$

Indeed, setting $r := \nu/2$ and applying Cauchy product, we have to prove that for $r \in \mathbb{N}_0$ and $\beta \in \mathbb{C}$

$$\sum_{k=0}^{2r} \frac{(-1)^k}{\Gamma(k + \beta)\Gamma(2r - k + \beta)} = \frac{1}{(r + \beta - 1)\Gamma(\beta - 1)\Gamma(2r + \beta)}. \quad (3.44)$$

Therefore we need the following equality which can be proved by induction

$$\sum_{\nu=0}^n (-1)^{\nu} \binom{\alpha}{\nu} = (-1)^n \binom{\alpha - 1}{n}. \quad (3.45)$$

At first we want to prove (3.44) for $r \in \mathbb{N}_0$ and $\beta \in \mathbb{N}$.

If $\beta = 1$, the identity follows easily.

For $\beta = l + 1 \geq 2$ an index shift show that

$$\sum_{k=0}^{2r} \frac{(-1)^k}{(k + l)!(2r - k + l)!} = \frac{(-1)^l}{(2r + 2l)!} \sum_{\nu=l}^{2r+l} (-1)^{\nu} \binom{2r + 2l}{\nu}.$$

This sum can be split up to

$$\frac{(-1)^l}{(2r + 2l)!} \left(\sum_{\nu=0}^{2r+l} (-1)^{\nu} \binom{2r + 2l}{\nu} - \sum_{\nu=0}^{l-1} (-1)^{\nu} \binom{2r + 2l}{\nu} \right). \quad (3.46)$$

Applying (3.45), we can write (3.46) as

$$\frac{(-1)^l}{(2r + 2l)!} \left((-1)^{2r+l} \binom{2r + 2l - 1}{2r + l} - (-1)^{l-1} \binom{2r + 2l - 1}{l - 1} \right).$$

Finally we can reduce this to

$$\frac{1}{(2r+2l)!} \left(\frac{(2r+2l-1)!}{(2r+l)!(l-1)!} + \frac{(2r+2l-1)!}{(2r+l)!(l-1)!} \right) = \frac{1}{(r+l)\Gamma(2r+l+1)\Gamma(l)}.$$

Equation (3.44) gives us

$$\sum_{k=0}^{2r} (-1)^k (2r+\beta-1) \dots (2r-k+\beta) (r+\beta-1) (k+\beta) \dots (2r+\beta-1) = \prod_{\nu=0}^{2r} (\nu+\beta-1).$$

Now the assertion follows with the identity theorem for polynomials.

The following numerical results shown in Figure (3.14) support our theoretical considerations which are illustrated in Figure (3.13) and demonstrate the advantages of this method compared to the computation of the Taylor sections of f .

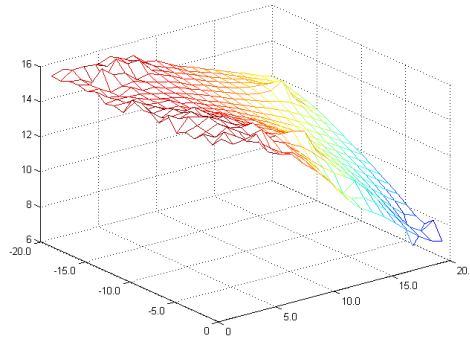


Figure 3.14: $e_{f,s_{N(z),16}(f\psi,\cdot)/\tilde{\psi}}(z)$ in S (rem. $\tilde{\psi}(z) = s_{N(-z),16}(\psi, -z)$).

Finally, we give some ideas of how the evaluation of the function $f(z) = M(1, \beta; z)$ in the second quadrant S can be improved.

Remark 3.8

- i) Similar to the considerations in the case of the complementary error function it turns out that a further improvement concerning the accuracy of the computation of f is possible.

The multiplication of f by $\varphi(z) = e^{-z/2}$ results with the help of Remark

2.10 iii) in the indicator function

$$h_{f\varphi} = \frac{1}{2}|\cos(\vartheta)|.$$

Obviously, $f\varphi$ intensifies the advantages of $f\psi$ which are illustrated in Figure 3.13.

Unfortunately, the computation of the coefficients of f and φ by convolution leads to cancellation again, and we were not able to find a reasonable recursion relation for the coefficients.

So it seems necessary to evaluate the coefficients with the help of exact arithmetic and implement them as data.

- ii) Another way to evaluate the incomplete gamma function in several parts of the second quadrant S is the use of the Kummer transformation, which has been mentioned in (3.9).

3.5 The Computation of Mittag-Leffler and confluent hypergeometric functions for more general parameters

We consider similar ideas which we have applied to the special cases of f being the complementary error function or the incomplete gamma function for Mittag-Leffler and hypergeometric functions for more general parameters.

Let us first take a look at the Mittag-Leffler functions $M(\alpha; z)$, defined in (3.2). We already have seen that the functions $M(\alpha; z)$ are entire functions of the order $\rho_\alpha = 1/\alpha$.

The indicator function for $\alpha \neq 1$ is given explicitly in (3.19) as

$$h_{M(\alpha; \cdot)}(\vartheta) = \begin{cases} \cos(\vartheta/\alpha), & |\vartheta| \leq \alpha\pi/2 \\ 0, & \text{else.} \end{cases}.$$

Let us consider the case of $\alpha \in \mathbb{N}$, $\alpha \neq 1$.

Resulting from the considerations in Chapter 2, it may be of advantage to use

the Taylor sections

$$(f\varphi)(z) := M(1/\alpha; z)e^{-\frac{1}{\alpha}z^\alpha} \quad (\alpha \in \mathbb{N}, \alpha > 2) \quad (3.47)$$

in several parts of the complex plane. Analogous considerations as in Theorem 3.7 give us for $z \in \mathbb{C}$

$$(f\varphi)(z) = M(1/\alpha; z)e^{-\frac{1}{\alpha}z^\alpha} = \sum_{\nu=0}^{\infty} b_\nu z^\nu, \quad (3.48)$$

with

$$b_\nu = \begin{cases} \frac{(\alpha-1)^l}{\alpha^l l!}, & \text{if } \nu = \alpha l, \ l \in \mathbb{N}, \\ \frac{(\alpha-1)^l}{\alpha^l \Gamma(l+1+r\alpha)} \sum_{k=0}^l \binom{k-1+r/\alpha}{k} \frac{(-1)^k}{(\alpha-1)^k}, & \text{if } \nu = \alpha l + r, \ 0 \leq r < \alpha. \end{cases} \quad (3.49)$$

For

$$m_k := \binom{k-1+r/\alpha}{k} \frac{1}{(\alpha-1)^k}$$

we have

$$m_{k+1} = \frac{k+r/\alpha}{(k+1)(\alpha-1)} m_k \quad (k \in \mathbb{N}_0).$$

Thus, the sum

$$\sum_{k=0}^l \binom{k-1+r/\alpha}{k} \frac{(-1)^k}{(\alpha-1)^k}$$

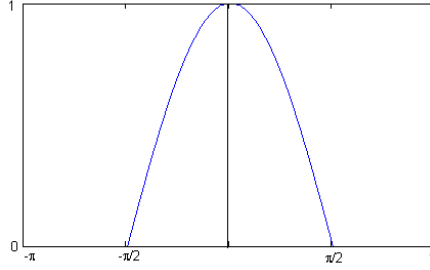
and also the Taylor coefficients b_ν of (3.49) can be evaluated recursively.

Finally we tried to apply this method for the general case $\alpha \in \mathbb{R}$. However, there was the problem of how an adequate function φ can be found, satisfying the conditions described in Section 2.4. So, this is still an open problem.

Also we want to look at the confluent hypergeometric functions $\Phi(a, c; z)$ (see (3.8)) for more general parameters.

The indicator function for $a \neq c$ is given in (3.20) as

$$h_{\Phi(a, c; \cdot)}(\vartheta) = \begin{cases} \cos(\vartheta), & |\vartheta| \leq \pi/2 \\ 0, & \text{else} \end{cases}. \quad (3.50)$$

Figure 3.15: Indicator function of $f(z) = \Phi(a, c; z)$.

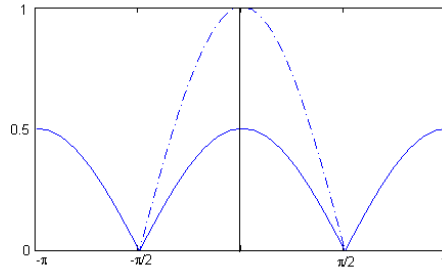
The Taylor sections only behave well on the positive real axis. In the whole second quadrant

$$S = \{z = re^{i\vartheta} : r \geq 0, \pi/2 \leq \vartheta \leq \pi\}$$

we have $\delta_{\Phi(a, c; \cdot)}(\vartheta) = 1$ (see (2.12)) and so in this area the cancellation problem causes a great loss of significant decimal digits.

According to the considerations in Chapter 2, the advice is to use Taylor sections of

$$(f\varphi)(z) := \Phi(a, c; z)e^{-z/2}. \quad (3.51)$$

Figure 3.16: Indicator functions of $f\varphi$ (solid curve) and f (dotted curve).

Using MATHEMATICA, we find the following representation for the coefficients to compute the Taylor sections:

$$\sum_{k=0}^n \frac{(-1/2)^k (a)_{n-k}}{(c)_{n-k} k! (n-k)!} = \frac{2^{(1-c-n)} \Gamma(c) \Gamma(a+n) F(1-a, 1-c-n, 1-a-n; -1)}{\Gamma(a) \Gamma(1+n) \Gamma(c+n)},$$

where F is the hypergeometric function defined in (3.6).

For known values of the hypergeometric function for certain parameters, we

can compute the coefficients in this way and it is possible to compute the Kummer function near the negative real axis without cancellation.

Although there is an improvement, the Taylor sections of $f\varphi$ also turn out to be numerically unstable with respect to cancellation near the imaginary axis where $\delta_{f\varphi}$ is maximal (cf. Figure 3.5). So the question arises again if there is an appropriate entire function ψ such that $\delta_{f\psi}(\vartheta)$ is smaller than $\delta_{f\varphi}(\vartheta)$ for $\vartheta \approx \pi/2$. For this we can use similar methods we used to compute the complementary error function (cf. Figure 3.10). We should use a function $\psi(a, z) = e^{az}$, $a \in \mathbb{C}$ and find the parameter a in such a way that $\delta_{f\psi}(\vartheta)$ is small for $\vartheta \approx \pi/2$.

However, in this general case it is difficult to find a reasonable relation for the coefficients, or the computation of the coefficients of $f\varphi$ by convolution leads to cancellation and we have to implement them as data using exact arithmetic again.

In [Schw] different methods to compute the Kummer functions on bounded intervals can be found, especially with good results near the imaginary axis using modified kinds of Chebyshev polynomials.

Chapter 4

Further Methods to Compute Entire Functions

In this chapter we will consider two other methods, which can be used to numerically evaluate entire functions. In particular we will look again at the computation of some special cases of the Mittag-Leffler and the confluent hypergeometric functions. The modified series expansions, described in the second chapter, are only applicable in several parts of the complex plane. So, in this chapter, we consider asymptotic expansions and continued fractions, and we use them to compute the functions on unbounded sectors S .

4.1 Computation with Asymptotic expansions

In Section 3.1.2 definitions and basic properties of asymptotic expansions have been considered. There we have seen that it is possible to find the indicator function of certain entire functions with the help of asymptotic expansions. This fact is of great importance to quantify and reduce cancellation errors. Now, in this section, we compute certain entire functions using their asymptotic expansions. We consider a sector $S = \{z \in \mathbb{C} : |\arg(z) - \alpha| \leq \delta\}$ and large arguments z in modulus.

4.1.1 Asymptotic expansions for the complementary error and the incomplete gamma function

In the following section we will consider in particular the asymptotic expansions of the complementary error function and the incomplete gamma function. We already have considered the connections of these functions with the class of Mittag-Leffler functions (see (3.22) and (3.24)). So we have to study the asymptotic expansion of the Mittag-Leffler functions for $\alpha < 2$ in detail.

In Theorem 3.4 the asymptotic expansion for the case $\alpha < 2$ is given explicitly. From the expansions (3.16) and (3.17) we can see that the result depends on a sector. As $\arg z$ varies, the exponential term is dominant or recessive. Of course, if we use the expansions (3.16) and (3.17) to numerically evaluate the Mittag-Leffler functions, the question arises of how the remainder $R_N(z)$ can be estimated and of how the number of terms N in the sum

$$\sum_{n=1}^{N-1} \frac{z^{-n}}{\Gamma(\beta - \alpha n)} \quad (4.1)$$

can be chosen adequately.

In [WoZh], where the asymptotic expansion of the functions $M(\alpha, \beta; z)$ are investigated in detail, the questions are answered. The authors point out an estimation for the remainder $R_N(z)$ in the following form:

$$|R_N(z)| \leq C(\alpha, \beta)(\alpha N)^{-\operatorname{Re}\beta+1} e^{N(-\alpha+\alpha\log(\alpha N)-\log z)} \quad (4.2)$$

where $\arg z \in [-\pi, -\alpha\pi]$ and $C(\alpha, \beta)$ is a positive constant whose values depend only on α and β .

Considering the estimation (4.2) and minimizing the remainder, an optimal choice for the index $N = N(z)$ is given by

$$N(z) = \frac{1}{\alpha} |z|^{\frac{1}{\alpha}}. \quad (4.3)$$

Remark 4.1

We numerically consider the non-zero terms of

$$\sum_{n=1}^{N-1} \frac{z^{-n}}{\Gamma(\beta - \alpha n)} \quad (4.4)$$

for a real parameter β , as well as $\alpha < 2$ and $z \in \mathbb{C}$ with $\text{Im } z > 0$. Then, for increasing index n , we can see that the absolute values of the terms first decrease and subsequently increase monotonically. Numerical evaluations have shown that the index of the smallest term, according to amount, essentially equals the index $N(z)$ (see (4.3)), deduced in [WoZh].

Hence, using the sum

$$\sum_{n=1}^{N(z)-1} \frac{z^{-n}}{\Gamma(\beta - \alpha n)} \quad (4.5)$$

to approximate a function of Mittag-Leffler type (β real, $\alpha < 2$, $z \in \mathbb{C}$, $\text{Im } z > 0$), we just have to add terms which decrease according to amount. From numerical evaluations it can be seen that the order of magnitude of the result of the summation equals the order of magnitude of the first term of the sum if the modulus of the argument z is large enough. Because of this, we do not expect any cancellation problems if we use the sum (4.5) to approximate a function of Mittag-Leffler type for a sufficiently large argument z in modulus.

Now we want to apply the asymptotic expansions in particular to the computation of the complementary error function and the incomplete gamma function for large arguments z in modulus.

The complementary error function

To compute the complementary error function for large arguments z in modulus, we will use the asymptotic expansion of the Mittag-Leffler function $M(1/2; z)$, given in Theorem 3.4. With the help of (4.2) we can obtain that the remainder behaves like $\mathcal{O}(e^{-|z|^2})$, and considering (4.3), we should use $\sum_{k=1}^{[2|z|^2]} \frac{z^{-k}}{\Gamma(1 - k/2)}$ as an approximation of the function $M(1/2; z)$. (In the first quadrant the leading term $2e^{z^2}$ has to be taken into account.)

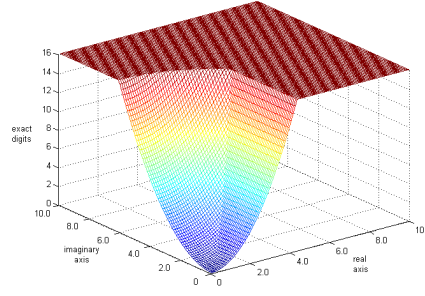


Figure 4.1: $-\log_{10}((\exp(-|z|^2)/\exp(|z|^2) h_{M(1/2;\cdot)}(\arg z)))$ in the first quadrant.

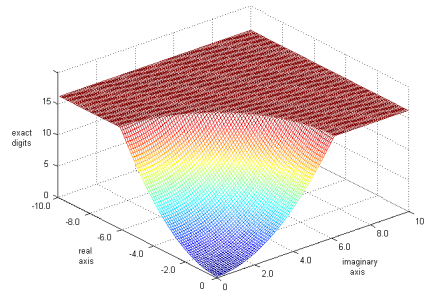


Figure 4.2: $-\log_{10}((\exp(-|z|^2)/\exp(|z|^2) h_{M(1/2;\cdot)}(\arg z)))$ in the 2nd quadrant.

The theoretical loss of significant digits using this kind of approximation to compute the Mittag-Leffler function $M(1/2; z)$ is illustrated in the Figures 4.1 and 4.2. The values of $-\log_{10}((\exp(-|z|^2)/\exp(|z|^2) h_{M(1/2;\cdot)}(\arg z)))$ in the first and in the second quadrant are shown and can be viewed as a theoretical measure for the exact digits if computations are again performed in floating point arithmetic with a fixed precision of 16 decimal digits.

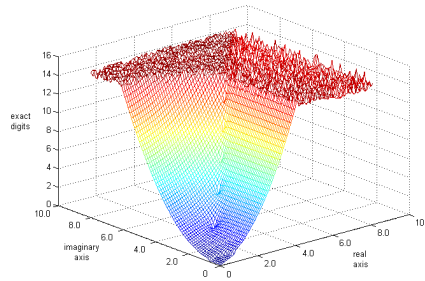


Figure 4.3: Numerical results of the asymptotic expansion in the first quadrant.

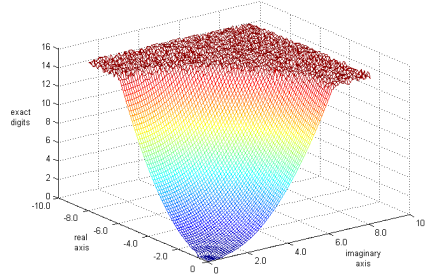


Figure 4.4: Numerical results of the asymptotic expansion in the 2nd quadrant.

The numerical results shown in the Figures 4.3 and 4.4 support our theoretical considerations again. Furthermore, we have found a method to compute the complementary error function for large arguments z in modulus.

The incomplete gamma function

Similar to the case of computing the complementary error function, we now look at the incomplete gamma function for large arguments z in modulus. So it makes sense to consider the asymptotic expansion of the Mittag-Leffler functions again for $\alpha = 1$ and β arbitrary ($\beta \neq 1$). According to (4.2) and (4.3), we can use

$$z^{1-\beta}e^z - \sum_{k=1}^{\lfloor |z| \rfloor} \frac{z^{-k}}{\Gamma(\beta - k)} \quad (4.6)$$

as an approximation of the function $M(1, \beta; z)$ and for the remainder we obtain the estimation $\mathcal{O}(e^{-|z|}|z|^{-\beta+1})$.

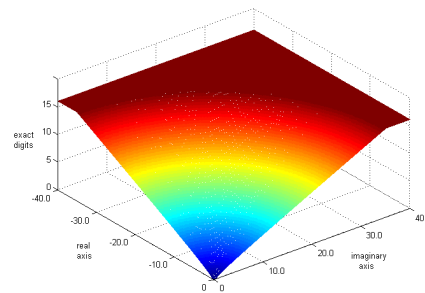


Figure 4.5: $-\log_{10}((\exp(-|z|)|z|^{-1/2}/\exp(|z| h_{M(1,3/2;)}(\arg z))))$ in the second quadrant.

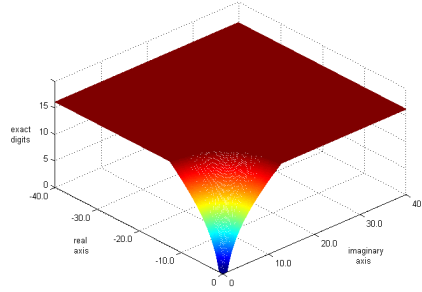


Figure 4.6: $-\log_{10}((\exp(-|z|)|z|^{-19/2}/\exp(|z|) h_{M(1,3/2,\cdot)}(\arg z)))$ in the second quadrant.

Again we illustrate (see Figures 4.5 and 4.6) the theoretical loss of exact digits using this approximation to compute the function $M(1, \beta; z)$ for arguments z of large modulus if the evaluations are performed in floating point arithmetic with a fixed precision of 16 decimal digits. We look at the special cases of $\beta = 3/2$ and $\beta = 21/2$. From the Figures 4.5 and 4.6 we can also see that the area in which we have a loss of exact digits decreases if β increases.

For arguments z of large modulus we can use the approximation (4.6) to compute the incomplete gamma function.

An alternative method to compute special functions for arguments z of large modulus will be described in the following section.

4.2 Computation with Continued fractions

At first the definition and some basic properties of continued fractions (see [JoTh]) are given. A relation to three-term recurrence formulas will be described (Theorem of Pincherle) and the continued fractions of the complementary error function and the incomplete gamma function will be computed.

Definition 4.2

A continued fraction is an ordered pair $[[(a_n), (b_n)], (f_n)]$ where a_1, a_2, \dots and

b_0, b_1, b_2, \dots are complex numbers with all $a_n \neq 0$ and where (f_n) is a sequence in the extended complex plane defined as follows:

$$f_n = S_n(0) \quad (n = 0, 1, 2, \dots), \quad (4.7)$$

where

$$S_0(w) = s_0(w); \quad S_n(w) = S_{n-1}(s_n(w)) \quad (n = 1, 2, 3, \dots), \quad (4.8)$$

and

$$s_0(w) = b_0 + w; \quad s_n(w) = \frac{a_n}{b_n + w} \quad (n = 1, 2, 3, \dots). \quad (4.9)$$

The numbers a_n and b_n are called the n -th partial numerator and denominator of the continued fraction.

Sometimes they are simply called the elements; (f_n) is called the n -th approximant.

If (a_n) and (b_n) are infinite sequences, then $[(a_n), (b_n)], (f_n)]$ is called an infinite (or non terminating) continued fraction.

It is called a finite (or terminating) continued fraction if (a_n) and (b_n) only have a finite number of terms a_1, a_2, \dots, a_m and b_0, b_1, \dots, b_m . Hereafter a continued fraction will be assumed to be infinite unless otherwise stated.

It can be seen that the n -th approximant is given by

$$f_n = b_0 + \frac{a_1}{b_1 + \frac{a_2}{b_2 + \frac{a_3}{\ddots + \frac{a_n}{b_n}}}}.$$

For convenience we will now denote a continued fraction $[(a_n), (b_n)], (f_n)]$ by the symbol

$$b_0 + \frac{a_1}{b_1} + \frac{a_2}{b_2} + \frac{a_3}{b_3} + \dots \quad (4.10)$$

A continued fraction is said to converge if its sequence of approximants (f_n) converges to a point in the extended complex plane.

When convergent, the value of the continued fraction is $\lim f_n$.

Now our aim is to get representations of analytic functions by continued fractions.

In [JoTh] and [Wa] different approaches are given. In one of these approaches three-term recurrence relations (see Section 2.4.1) play an important role for representing analytic functions by continued fractions.

The method is based on a theorem of Pincherle concerning minimal solutions of three-term recurrence relations (see Section 2.4.1). In the paper [Ga2] about computational aspects of three-term recurrence relations Gautschi recognized the importance of Pincherle's result and applied it extensively.

A generalization of Pincherle's theorem for continued fractions is given in [JoTh].

Theorem 4.3

For each $n = 1, 2, 3, \dots$ let a_n and b_n be elements of \mathbb{C} , with

$$a_n \neq 0 \quad (n = 1, 2, 3, \dots).$$

1. The three-term recurrence relation

$$y_{n+1} = b_n y_n + a_n y_{n-1} \quad (n = 1, 2, 3, \dots), \quad (4.11)$$

has a minimal solution (h_n) , $h_n \in \mathbb{C}$ if and only if the continued fraction

$$\frac{a_1}{b_1} + \frac{a_2}{b_2} + \frac{a_3}{b_3} + \dots \quad (4.12)$$

converges (to a finite value in \mathbb{C} or to infinity).

2. Suppose that (4.12) has a minimal solution (h_n) , $h_n \in \mathbb{C}$. Then for each $m = 1, 2, 3, \dots$

$$\frac{h_m}{h_{m-1}} = -\frac{a_m}{b_m} + \frac{a_{m+1}}{b_{m+1}} + \frac{a_{m+2}}{b_{m+2}} + \dots \quad (4.13)$$

By (4.13) we mean the following: If $h_{m-1} = 0$, then $h_m \neq 0$ and the continued fraction (4.13) converges to $\infty = h_m/h_{m-1}$. If $h_{m-1} \neq 0$ then the continued fraction (4.13) converges to the finite value $h_m/h_{m-1} \in \mathbb{C}$.

Continued fraction of the complementary error function

We again look at the complementary error function

$$f(z) = \operatorname{erfc}(z) = 1 - \operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \int_z^\infty e^{-t^2} dt \quad (z \in \mathbb{C}).$$

Here the path of integration is subject to the restriction $\arg t \rightarrow \alpha$ with $|\alpha| < \pi/4$ as $t \rightarrow \infty$ along the path (see, [AS], p. 297). Repeated integrals of $\operatorname{erfc}(z)$ are defined by

$$I^{-1}\operatorname{erfc}(z) = \frac{2}{\sqrt{\pi}} e^{-z^2}, \quad I^0\operatorname{erfc}(z) = \operatorname{erfc}(z),$$

$$I^n\operatorname{erfc}(z) = \int_z^\infty I^{n-1}\operatorname{erfc}(t) dt \quad (n = 1, 2, 3, \dots).$$

Now let (h_n) and (g_n) defined by

$$h_n = e^{z^2} I^n \operatorname{erfc}(z) \quad (n = -1, 0, 1, 2, \dots),$$

$$g_n = (-1)^n e^{z^2} I^n \operatorname{erfc}(-z) \quad (n = -1, 0, 1, 2, \dots).$$

Gautschi has shown in [Ga2] that (h_n) is a minimal solution of the system of three-term recurrence relations

$$y_{n+1} = -\frac{z}{n+1} y_n + \frac{1}{2(n+1)} y_{n-1} \quad (n = 0, 1, 2, \dots),$$

if $\operatorname{Re} z > 0$, and (g_n) is a minimal solution, if $\operatorname{Re} z < 0$. It follows from Theorem 4.3 that for each $m = 0, 1, 2, \dots$,

$$\frac{I^m \operatorname{erfc}(z)}{I^{m-1} \operatorname{erfc}(z)} = \frac{1/2}{z} + \frac{(m+1)/2}{z} + \frac{(m+2)/2}{z} + \dots \quad (\operatorname{Re} z > 0).$$

In the special case $m = 0$ we get

$$2e^{z^2} \int_z^\infty e^{-t^2} dt = \frac{1}{z} + \frac{1/2}{z} + \frac{1}{z} + \frac{3/2}{z} + \frac{2}{z} + \frac{5/2}{z} + \frac{3}{z} + \dots \quad (\operatorname{Re} z > 0),$$

hence,

$$\operatorname{erfc}(z) = \frac{e^{-z^2}}{\sqrt{\pi}} \left(\frac{1}{z} + \frac{1/2}{z} + \frac{1}{z} + \frac{3/2}{z} + \frac{2}{z} + \frac{5/2}{z} + \frac{(\nu-1)/2}{z} + \dots \right) \quad (\operatorname{Re} z > 0). \quad (4.14)$$

With the help of the continued fraction (4.14) Gautschi [Ga1] and Poppe/Wijers [PW1] have computed the complementary error function for

large arguments z in modulus. While Gautschi used this continued fraction outside a rectangular region with a fixed chosen index ν , Poppe/Wijers optimized this region in which it is reasonable to use the continued fraction of the complementary error function. They pointed out that it makes sense to define an elliptic contour Γ in the second quadrant S by the condition that points z on Γ have to obey the relation

$$\sigma(z) = \sqrt{\left(\frac{\operatorname{Re} z}{x_0}\right)^2 + \left(\frac{\operatorname{Im} z}{y_0}\right)^2} = 1, \quad \text{with } x_0 = -4.4, y_0 = 6.3. \quad (4.15)$$

Outside of this contour Γ the continued fraction attains a required accuracy of 14 decimal digits.

Also Poppe/Wijers have chosen an optimal index ν which depends on z .

An approximation of the index ν for each z outside the contour Γ is defined by

$$\nu(z) = \left\lceil 3 + \frac{1442}{26\sigma(z) + 77} \right\rceil.$$

The numerical results using the continued fraction (4.14) to compute the complementary error function can be seen in Figure 4.7.

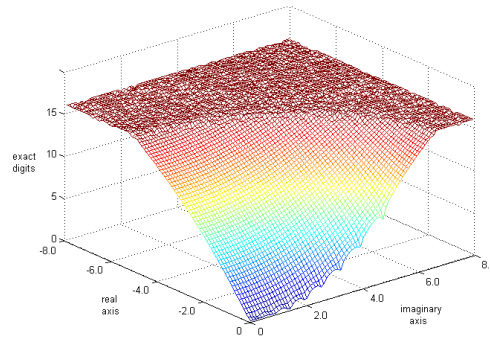


Figure 4.7: Numerical results using continued fraction of $f(z) = \operatorname{erfc}(-z)$ in the second quadrant S .

The use of the elliptic contour described in (4.15) is illustrated in the following Figure 4.8.

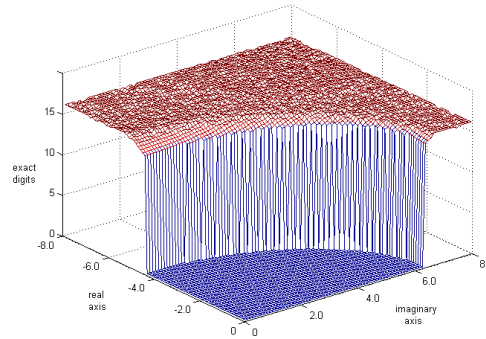


Figure 4.8: Elliptic contour of the continued fraction of $f(z) = \operatorname{erfc}(-z)$.

Finally, we compare the continued fraction and the asymptotic expansion for the numerical evaluation of the complementary error function in the second quadrant for large arguments z in modulus.

From the Figures 4.4 and 4.7 we can see that using the continued fraction with the elliptic contour illustrated in Figure 4.8 is more advantageous for the computation than using the asymptotic expansion.

Continued fraction of the incomplete gamma function

Similar considerations as in the case of the complementary error function give us a continued fraction of the incomplete gamma function.

For all z with $|\arg z| < \pi$ we have (see [Ga3])

$$\gamma(a, z) = \frac{1}{2} + \frac{1-a}{1} + \frac{1}{z} + \frac{2-a}{1} + \frac{2}{z} + \frac{3-a}{1} + \frac{3}{2} + \dots \quad (4.16)$$

The numerical results using the continued fraction (4.16) to compute the incomplete gamma function for the special case $a = 1/2$ can be seen in the following Figure 4.9.

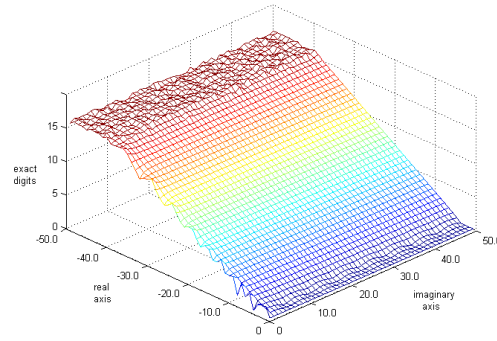


Figure 4.9: Numerical results using continued fraction of $f(z) = \gamma(1/2, z)$ in the second quadrant S .

Finally, we again compare the continued fraction and the asymptotic expansion for the numerical evaluation of the incomplete gamma function in the second quadrant for large arguments z in modulus. In this case it can be seen from the Figures 4.5 and 4.9 that the continued fraction should be used in a neighborhood of the negative real axis (see [Ga3]). In the remaining part of the second quadrant we should use the asymptotic expansion.

Chapter 5

Conclusion

In the previous chapters we have given some concluding remarks in order to make suggestions to combine the presented methods for computing Mittag-Leffler type and confluent hypergeometric functions.

In particular, the complementary error function has been considered in detail. At first, the computation method investigated in Section 2.4 leads to reliable results for several parts of the complex plane if the modulus of the arguments z is not too large. In the Figures 3.6 and 3.9, numerical results are illustrated. For large arguments z in modulus we have seen in Chapter 4 that the use of continued fractions in the second quadrant is more advantageous for the computation than the use of the asymptotic expansion. In the Figure 4.7, corresponding numerical results are illustrated. Overall, numerical experiments confirm that for the computation of the complementary error function a combination of the four types of approximants based on the continued fractions and the Taylor series of $\operatorname{erfc}(-z)e^{z^2/2}$, $\operatorname{erfc}(-z)\operatorname{erfc}(z)e^{2z^2}$ and $\operatorname{erfc}(-z)\operatorname{erfc}(z)e^{az^2}$ ($a = e^{3\pi i/4} - 1/2$), provide a reasonable method in the sense that at least 14 exact figures are reached in the second quadrant (and thus in all of the complex plane; cf. [PW1]).

In the case of the incomplete gamma function, the computation method described in Section 2.4 has been applied for values of z near the negative real axis and small modulus of the arguments z . In Figure 3.14, numerical results

are illustrated. For methods which are efficient near the imaginary axis, we refer to [Schw]. In Chapter 4, continued fractions and the asymptotic expansion of the incomplete gamma function have been compared for large arguments z in modulus. The continued fractions should be used in a neighborhood of the negative real axis and in the remaining part of the second quadrant we should use the asymptotic expansion. For the computation of the incomplete gamma function, similarly to the case of the complementary error function, a combination of different types of approximants based on the asymptotic expansion, continued fractions and modified Taylor series is possible.

The important question in which way the method investigated in Section 2.4 is available for Mittag-Leffler type and confluent hypergeometric functions in general, has been discussed in Chapter 3.5. For the Mittag-Leffler functions $M(\alpha; z)$, with parameter $\alpha \in \mathbb{N}$, $\alpha \neq 1$, some ideas concerning the numerical evaluation have been presented. Applying this method to the general case $\alpha \in \mathbb{R}$, we have the problem in which way a multiplier function φ can be chosen satisfying the conditions which are described in Section 2.4.

In the case of the confluent hypergeometric function $\Phi(a, c; z)$, computational methods for arguments z near the negative real axis have been obtained. However, in several parts of the complex plane it is difficult to find a reasonable relation for the coefficients of $f\varphi$ by convolution, and the computation of the coefficients by convolution leads to cancellation again. The only way is to implement them as data using exact arithmetic.

Finally, as already mentioned in Section 2.1, for the numerical evaluation of entire functions it is in principle possible to use some kind of exact arithmetic instead of floating point arithmetic with a fixed precision. In particular, the reference values which we used for our numerical considerations have been generated using such exact arithmetic from iRRAM. The computational effort was comparatively high. Therefore, our suggestion is to use the method investigated in Section 2.4 also in exact arithmetic software with the aim to reduce the computational effort and to accelerate the evaluations.

Bibliography

- [AS] Abramovitz, M., Stegun, I. A., Handbook of mathematical functions with formulas, graphs, and mathematical tables. National Bureau of Standards, Washington D.C., 1964.
- [Bo] Boas, R. P., Entire Functions, Academic Press, New York, 1954.
- [DZ] Dzrbasjan, M. M., Integral transforms and representations of functions in the complex plane, Nauka, Moskau, 1966.
- [EMO] Erdelyi, A., Magnus, F., Oberhettinger, F., Higher Transcendental Functions, vol. II, McGraw-Hill, New York, 1953.
- [Ga1] Gautschi, W., Efficient computation of the complex error function, SIAM J. Numer. Anal., 7 (1970) 197-198.
- [Ga2] Gautschi, W., Computational aspects of three-term recurrence relations, SIAM Review, 9 (1967) 24-82.
- [Ga3] Gautschi, W., A Computational Procedure for Incomplete Gamma Functions, ACM Trans. Math. Software, 5 (1979) 466-481.
- [Go] Gorenflo R., Luchko Y., Loutchko I., Numerische Berechnungen der Mittag-Leffler Funktionen $E_{\alpha,\beta}(z)$ und ihrer Ableitungen, Preprint No. A-06/99, Freie Universität Berlin, Berlin, 1999.
- [Go1] Gorenflo R., Mainardi F., Fractional oscillations and Mittag-Leffler functions, Preprint No. A-14/96, Freie Universität Berlin, Berlin, 1994.

- [Go2] Gorenflo R., Mainardi F., Fractional Calculus: Introduction to Integral and Differential Equations of Fractional Order, in: Fractals and Fractional Calculus in Continuum Mechanics (Ed. A. Carpinteri and F. Mainardi), Springer Verlag, Wien, 1997.
- [He1] Henrici, P., Applied and computational complex analysis, Vol. I, John Wiley & Sons, New York, 1974.
- [He2] Henrici, P., Applied and computational complex analysis, Vol. II, John Wiley & Sons, New York, 1977.
- [He3] Henrici, P., Essentials of numerical analysis with pocket calculator demonstrations, Wiley, New York, 1982.
- [Hi] Hille E., Analytic function theory, Vol. II, 2nd ed., New York, 1987.
- [JoTh] Jones, W. B., Thron, W. J., Continued Fractions, Analytic Theory and Applications, Addison-Wesley Publishing Company, London, 1980.
- [Le] Levin, B. Y., Distribution of zeros of entire functions, Transl. Math. Monographs, vol. 5, Amer. Math. Soc., Providence, R. I., 1980.
- [Le1] Levin, B. Y., Lectures on Entire Functions, Transl. Math. Monographs, vol. 150, Amer. Math. Soc., Providence, R. I., 1996.
- [LoWa] Lorentzen, L., Waadeland, H., Continued Fractions with Applications, North-Holland, Amsterdam, 1992.
- [Loz] Lozier, D. W., Software needs in special functions, J. Comp. and Appl. Math., 66 (1996) 345-358.
- [Lu1] Luke, Y. L., The special functions and their approximations, Vol. I, Academic Press, New York, 1969.
- [Lu2] Luke, Y. L., The special functions and their approximations, Vol. II, Academic Press, New York, 1969.

- [Ma] Mainardi, F., Fractional Calculus: Some Basic Problems in Continuum and Statistical Mechanics, in: Fractals and Fractional Calculus in Continuum Mechanics (Ed. A. Carpinteri and F. Mainardi), Springer Verlag, Wien, 1997.
- [Mue3] Müller, J., Convergence Acceleration of Polynomial Expansions for Finite Order Entire Functions, Habilitationsschrift, 1995.
- [Mue2] Müller, J., Accelerated polynomial approximation of finite order entire functions by growth reduction, Math. Comp., 66 (1997) 743-761.
- [Mue1] Müller, N. T., The iRRAM: Exact arithmetic in C++, Lecture Notes in Computer Science, 2064, pp. 222-252, Springer, Berlin, 2001.
- [Ol1] Olver, F. W. J., Error analysis of Miller's recurrence algorithm, Math. Comput., 18 (1964) 65-74.
- [Ol2] Olver, F. W. J., Introduction to asymptotics and special functions, Academic Press, New York, 1974.
- [Pe] Perron, O., Die Lehre von den Kettenbrüchen, vol. II, 3rd ed., Teubner, Stuttgart, 1957.
- [PW1] Poppe, G. P. M., Wijers, C. M. J., More efficient evaluation of the complex error function, ACM Trans. Math. Software, 16 (1990) 38-46.
- [PW2] Poppe, G. P. M., Wijers, C. M. J., Algorithm 680: Evaluation of the complex error function, ACM Trans. Math. Software, 16 (1990) 47.
- [Ru] Rubel, L. A., Entire and meromorphic functions, Springer, New York, 1996.
- [Sa] Sansone, G., Gerretsen, J., Lectures on the theory of functions of a complex variable, Noordhoff, Groningen, 1960.
- [Schw] Schwarz, C., Computation of Confluent Hypergeometric Functions on Compact Intervals., In Proceedings of the Alexits Memorial Conference: Functions, Series, Operators, Aug. 9-14, 1999, Budapest.

-
- [SW] Schaback, R., Wendland, H., Numerische Mathematik, Springer, Berlin, Heidelberg, New York, 2005.
- [Te] Temme, N. M., Special functions: An introduction to the classical functions of mathematical physics, Wiley, New York, 1996.
- [Wa] Wall, H. S., Analytic Theory of Continued Fractions, Chelsea Publishing Company, New York, 1948.
- [Wat] Watanabe, T., Natori, M., Oguni, T., Mathematical Software for the P.C. and Work Stations, North-Holland, Amsterdam, 1994.
- [Wei] Weideman, J. A. C., Computation of the complex error function, SIAM J. Numer. Anal., 31 (1994) 1497-1518.
- [Wi] Wimp, J., Computation with recurrence relations, Pitman, Boston, London, Melbourne, 1984.
- [WoZh] Wong, R., Zhao, Y. Q., Exponential Asymptotics of the Mittag-Leffler Function, Constr. Approx., 18 (2002) 355-385.

Tabellarische Zusammenfassung des Bildungsweges

Name	Martin Reinhard
Geburtstag	2. Juni 1977
Geburtsort	Bitburg
08/1983 - 07/1987	Grundschule Kyllburg
08/1984 - 06/1996	Gymnasium Bitburg
03/1998 - 04/2003	Studium der Wirtschaftsmathematik an der Universität Trier mit Abschluss Diplom
06/2003 - 03/2004	Promotionsstipendiat der Landesgraduiertenförderung des Landes Rheinland-Pfalz
04/2004 - 03/2008	Wissenschaftlicher Mitarbeiter am Lehrstuhl Angewandte Analysis im Fachbereich IV der Universität Trier