Analytical and numerical aspects of hypergeometric functions

Alfredo Deaño Cabrera and Francisco Marcellán Español

## Contents

1 Hypergeometric functions  ......................................................... 4
   1.1 The Gamma and Beta functions ...................................... 5
   1.2 Second order linear ODEs .......................................... 7
      1.2.1 General theory ................................................. 8
      1.2.2 Normal form and qualitative behaviour of the solutions 10
      1.2.3 Regular and singular points and power series solutions 11
      1.2.4 Regular points ................................................. 11
      1.2.5 Regular singular points .................................... 14
      1.2.6 Irregular singular points .................................. 17
   1.3 The Gauss hypergeometric equation ................................. 18
   1.4 The confluent hypergeometric equation ............................ 22
   1.5 Related topics ......................................................... 24

2 Asymptotic techniques ......................................................... 26
   2.1 Asymptotic analysis of differential equations .................. 28
      2.1.1 The Liouville transformation. Construction .............. 29
      2.1.2 The Liouville transformation. Error bounds .......... 31
      2.1.3 ODEs with parameters ....................................... 33
      2.1.4 Examples ...................................................... 34
      2.1.5 Higher order terms ......................................... 36
   2.2 Asymptotic analysis of integrals .................................... 38
      2.2.1 Laplace transforms .......................................... 38
      2.2.2 Integration by parts and Watson’s lemma .............. 42
      2.2.3 Laplace’s method ............................................ 45
      2.2.4 The method of stationary phase .......................... 49
      2.2.5 The method of steepest descent ......................... 52
   2.3 Related problems ......................................................... 60
3 Real zeros of hypergeometric functions 62
  3.1 Real zeros of solutions of second order ODEs .............. 62
  3.2 Sturm comparison theorem ................................. 65
  3.3 Liouville transformations revisited ......................... 67
    3.3.1 Liouville transformations of Gauss equation ........ 69
    3.3.2 Liouville transformations of Kummer equation ....... 72
  3.4 Systems of differential-difference equations .............. 74
  3.5 Computational aspects .................................. 75
    3.5.1 Gaussian quadrature ................................ 76
    3.5.2 Matrix methods .................................... 76
    3.5.3 Fixed point methods ................................ 80
    3.5.4 Asymptotic methods ................................ 85
  3.6 Related problems ......................................... 87

4 Three term recurrence relations 89
  4.1 General theory ........................................... 91
  4.2 Minimal and dominant solutions and conditioning .......... 92
  4.3 Minimal solutions and continued fractions ............... 99
    4.3.1 Convergence ........................................ 100
  4.4 Computation of minimal solutions .......................... 101
    4.4.1 Initial values ..................................... 102
    4.4.2 Computation of $H_N$ ............................... 102
    4.4.3 Backward and forward methods ....................... 102
  4.5 Miller’s algorithm ....................................... 103
    4.5.1 Miller’s algorithm with a known function value .... 103
    4.5.2 Miller’s algorithm with a normalizing sum .......... 103
  4.6 Related problems ......................................... 106
Chapter 1

Hypergeometric functions

These notes are devoted to the study of several analytical and numerical properties of classical special functions. The concept of “special function” is somewhat fuzzy in the literature, but usually refers to functions that cannot be expressed in terms of elementary functions (polynomials, rational functions, exponential and logarithms), and that have become useful in specific areas of mathematics and/or in applications. That is the case of the Gamma and Beta function in statistics, the Zeta function in number theory, the classical orthogonal polynomials in numerical and real analysis, or the Airy and Bessel functions in Physics and Astronomy.

An incomplete but very important family of classical special functions is that of hypergeometric functions of Gauss and Kummer type. This includes many relevant subfamilies (such as Bessel, Legendre, Weber functions and classical orthogonal polynomials...), and its members are characterized as solutions of second order linear differential equations with three regular singular points. This property allows us to derive a good deal of information about the behaviour of hypergeometric functions using classical tools of analysis. In particular, we can investigate the real zeros of these functions (using Sturm’s comparison theorem and related results), and their asymptotic behaviour (using Liouville-Green transformations).

Additionally, Gauss and Kummer hypergeometric functions depend on several (generally complex) parameters, and they satisfy three term recurrence relations in these parameters. These identities connect functions whose parameters differ by integer quantities, and they can be seen as second order difference equations (in a sense a discrete analogue of the second order ODEs mentioned before). Here we will explore them as an important computational and asymptotic tool.
1.1 The Gamma and Beta functions

The Gamma function does not belong to the big family of hypergeometric functions that we are mainly interested in, but it will appear in many different places and we will need some of its properties.

The Gamma function has its origin in L. Euler’s investigation on how to interpolate the factorial function \(n!\) at noninteger values, see [25]. Since then, it has been widely used in mathematical analysis and statistics, to mention just two examples. We note that

\[
\Gamma(n) = \int_0^{\infty} e^{-t} t^{n-1} dt = (n-1)!
\]

when \(n\) is a nonnegative integer, but that integral does make sense when \(n = x, x\) being a nonnegative real number, and even for \(x\) complex, provided that \(\text{Re} \, x > 0\).

Integration by parts readily shows that

\[
\Gamma(x+1) = x \int_0^{\infty} e^{-t} t^{x-1} dt = x \Gamma(x),
\]

which recovers a well known property of the factorial. Moreover, this result allows us to extend the definition of the Gamma function to negative values of \(x, x\) not a negative integer, using analytic continuation. Indeed, \(\Gamma(x)\) is analytic when \(\text{Re} \, x > 0\), then

\[
f(x) = \frac{\Gamma(x+1)}{x}
\]

is analytic in the strip \(-1 < \text{Re} \, x < 0\), and thus gives the (unique) analytic continuation into that strip in the complex plane, so \(f(x) = \Gamma(x)\) in that region.

The behaviour of the Gamma function at the negative integers can be analysed using Prym’s decomposition of the preceding integral:

\[
\Gamma(x) = \int_0^{1} e^{-t} t^{x-1} dt + \int_1^{\infty} e^{-t} t^{x-1} dt,
\]

and expanding the exponential function, we arrive at

\[
\Gamma(x) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \int_0^{1} t^{x+n-1} dt + \int_1^{\infty} e^{-t} t^{x-1} dt
\]

\[
= \sum_{n=0}^{\infty} \frac{(-1)^n}{n!(x+n)} + \int_1^{\infty} e^{-t} t^{x-1} dt.
\]
We deduce from the previous computation that the Gamma function has poles at the negative integers \( x = -n, \ n = 0, 1, \ldots \), with residue equal to \((-1)^n/n!\).

The multiplicative property of the Gamma function will be very convenient to write the hypergeometric functions in a compact way later on. A particularly beautiful and important special value \( \Gamma(1/2) = \sqrt{\pi} \). This follows from

\[
\Gamma(1/2) = \int_0^\infty e^{-t}t^{-1/2}dt = 2\int_0^\infty e^{-u^2}du,
\]

applying the change of variables \( t = u^2 \). The value of the Gaussian integral can be obtained by squaring and using polar coordinates.

The Beta function, which is also associated with Euler, can be defined by means of an integral as well:

\[
B(p, q) = \int_0^1 t^{p-1}(1-t)^{q-1}dt, \quad p, q > 0. \tag{1.2}
\]

The connection between the Gamma and Beta functions is well known:

**Exercise 1.1.1** Show that

\[
B(p, q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p + q)},
\]

by computing the integral

\[
I(p, q) = \int_0^\infty \int_0^\infty t^{p-1}s^{q-1}e^{-t^2-s^2}dtds
\]

in two different ways, firstly as a product of two integrals and secondly in polar coordinates.

We also note that, iterating the multiplicative property of the Gamma function, we obtain

\[
\frac{\Gamma(x + n)}{\Gamma(x)} = (x + n - 1)(x + n - 2)\ldots(x + 1)x.
\]

A standard notation for this is the so called Pochhammer symbol (or raising factorial):

\[
(x)_n = x(x + 1)\ldots(x + n - 2)(x + n - 1) = \frac{\Gamma(x + n)}{\Gamma(x)}. \tag{1.3}
\]

We take \((x)_0 = 1\) for any \( x \), which is consistent with this definition.
1.2 Second order linear ODEs

Hypergeometric functions of Gauss and Kummer type are solutions of second order linear differential equations with rational coefficients, and many of their properties can be deduced from this fact. In this section, we present first some classical results on existence and properties of functions which are solutions of second order ODEs.

The importance of this type of ODEs in applications comes essentially from the fact that they arise naturally when one applies separation of variables to Helmholtz’s or Laplace’s equation in different coordinate systems. For instance, if we take Helmholtz’s equation

$$\Delta u + k^2 u = 0,$$

where $u = u(x, y, z)$ and $k$ is a constant, and we change to cylindrical coordinates, $x = r \cos \theta$, $y = r \sin \theta$, $z = z$, then the PDE becomes

$$\frac{\partial^2 u}{\partial r^2} + \frac{1}{r} \frac{\partial u}{\partial r} + \frac{1}{r^2} \frac{\partial^2 u}{\partial \theta^2} + \frac{\partial^2 u}{\partial z^2} + k^2 u = 0.$$

Then, if we separate variables $u(r, \theta, z) = f_1(r)f_2(\theta)f_3(z)$, we obtain the following three ODEs:

$$f_1'' + \frac{1}{r} f_1' + \left( k^2 - \alpha^2 - \frac{\mu^2}{r^2} \right) f_1 = 0, \tag{1.4}$$

$$f_2'' + \mu^2 f_2 = 0,$$

$$f_3'' + \alpha^2 f_3 = 0,$$

where $\alpha$ and $\mu$ are constants. The second and third equations are elementary, with solutions

$$f_2(\theta) = e^{\pm i\mu \theta}, \quad f_3(\theta) = e^{\pm i\alpha z}.$$

The first ODE is of the type that we will analyse later on, and it can be solved in terms of cylinder or Bessel functions, $f_1(r) = C_\mu(r\sqrt{k^2 - \alpha^2})$. The change of variables $s = r\sqrt{k^2 - \alpha^2}$ results in

$$s^2 f_1'' + sf_1' + \left( s^2 - \mu^2 \right) f_1 = 0,$$  \tag{1.5}

which is the standard form.

Similarly, in spherical coordinates

$$x = r \sin \phi \cos \theta, \quad y = r \sin \phi \sin \theta, \quad z = r \cos \theta,$$
Helmholtz’s equation is
\[
\frac{1}{r} \frac{\partial}{\partial r} \left( r^2 \frac{\partial u}{\partial r} \right) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial u}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2 u}{\partial \phi^2} + k^2 u = 0.
\]

Then separation of variables \( u(r, \theta, \phi) = f_1(r)f_2(\theta)f_3(\phi) \) yields
\[
\begin{align*}
&f_1'' + \frac{2}{r} f_1' + \left( k^2 - \frac{\nu (\nu + 1)}{r^2} \right) f_1 = 0, \\
&\sin^2 \theta f_2'' + \sin \theta \cos \theta f_2' + (\nu (\nu + 1) \sin^2 \theta - \mu^2) f_2 = 0, \\
&f_3'' + \mu^2 f_3 = 0,
\end{align*}
\]
where \( \mu \) and \( \nu \) are constants. The third equation is elementary again, with solution \( f_3(z) = e^{\pm i \mu z} \), and the first one can be solved using cylinder functions, \( f_1(r) = r^{-1/2} C_{\nu + 1/2}(kr) \). The second equation (after a change of variables \( z = \cos \theta \)) is called Legendre’s equation, and the solutions \( f_2(\theta) = P_{\nu}^\mu(\cos \theta) \) are Legendre functions or spherical harmonics.

More examples can be found in [50, Chapter 10].

### 1.2.1 General theory

Given such an equation
\[
w''(x) + f(x)w'(x) + g(x)w(x) = 0, \tag{1.7}
\]
we recall the basic existence and uniqueness result:

**Theorem 1.2.1** If \( f(x) \) and \( g(x) \) are continuous in a certain interval \( (a, b) \) (finite or infinite), then the differential equation (1.7) has an infinite number of solutions which are twice differentiable in \( (a, b) \). If we prescribe the values of \( w(x_0) \) and \( w'(x_0) \) at a certain point \( x_0 \in (a, b) \), then the solution is unique.

The standard proof of this result is based on Picard iterations, see for instance [44, §5.1]. In these notes we will typically work with ODEs with analytic coefficients (except at certain isolated singular points), and the theorem can be restated as follows:

**Theorem 1.2.2** If \( f(x) \) and \( g(x) \) are analytic in a simply connected domain \( D \) of the complex plane, then the differential equation (1.7) has an infinite number of solutions which are analytic in \( D \). If we prescribe the values of \( w(x_0) \) and \( w'(x_0) \) at a certain point \( x_0 \in D \), then the solution is unique.
The space of solutions of (1.7) has dimension two, which means that any solution can be represented as a linear combination of two solutions $w_1(x)$ and $w_2(x)$ which are independent in the following sense:

**Definition 1.2.1** Two solutions $w_1(x)$ and $w_2(x)$ of (1.7) are independent in a certain domain $D$ if $Aw_1(x) + Bw_2(x) = 0$ for all $x \in D$ implies $A = B = 0$.

The independence of two solutions can be studied by means of the following concept:

**Definition 1.2.2** Given two solutions $w_1(x)$ and $w_2(x)$ of (1.7), the Wronskian is defined as follows:

$$W[y_1, y_2](x) = \begin{vmatrix} w_1(x) & w_2(x) \\ w'_1(x) & w'_2(x) \end{vmatrix}. \quad (1.8)$$

Using the differential equation, it is easy to check that if we write $w(x) = W[w_1, w_2](x)$, then

$$w'(x) = w_1(x)w_2''(x) - w_1''(x)w_2(x)$$

$$= w_1(x)(-f(x)w_2'(x) - g(x)w_2(x)) - w_2(x)(-f(x)w_1'(x) - g(x)w_1(x))$$

$$= -f(x)w(x),$$

so

$$W[w_1, w_2](x) = Ce^{-\int f(t)dt},$$

where $C$ is a constant. A consequence of this result is that in an (open) domain where $f(x)$ is continuous, the Wronskian of two solutions is either identically zero or it never vanishes. Actually, this can be used as a criterion for independence of solutions:

**Theorem 1.2.3** Let $f(x)$ and $g(x)$ be continuous in a certain domain $D$. Then two solutions $w_1(x)$ and $w_2(x)$ of (1.7) are independent in $D$ if and only if their Wronskian never vanishes in $D$.

**Proof 1.2.1** Given $x_0 \in D$, if we impose

$$Aw_1(x_0) + Bw_2(x_0) = 0,$$

$$Aw'_1(x_0) + Bw'_2(x_0) = 0,$$

then it follows that $BW[w_1, w_2](x_0) = 0$ and $-AW[w_1, w_2](x_0) = 0$. If the Wronskian does not vanish, then it follows that $A = B = 0$, so $w_1(x)$ and $w_2(x)$ are independent.
Now, suppose that \( w_1(x) \) and \( w_2(x) \) are independent, and that their Wronskian is identically 0. Then, the solution

\[
z_1(x) = w_2(x_0)w_1(x) - w_1(x_0)w_2(x)
\]
satisfies \( z_1(x_0) = 0 \) and also \( z_1'(x_0) = w_2(x_0)w_1'(x_0) - w_1(x_0)w_2'(x_0) = -W[w_1, w_2](x_0) = 0 \), hence \( z_1 \equiv 0 \), and \( w_1(x_0) = w_2(x_0) = 0 \) because \( w_1 \) and \( w_2 \) are independent. Similarly, if we set

\[
z_2(x) = w_2'(x_0)w_1(x) - w_1'(x_0)w_2(x),
\]
then \( w_1'(x_0) = w_2'(x_0) = 0 \), because \( w_2(x_0) = W[y_1, y_2](x_0) = 0 \) by assumption. We conclude that \( w_1 \equiv 0 \) and \( w_2 \equiv 0 \), which contradicts the fact that the two solutions are independent. Hence, \( W[w_1, w_2](x) \) does not vanish. See also [44].

1.2.2 Normal form and qualitative behaviour of the solutions

It is also well known that it is possible to make a change of variable in order to remove the term multiplying the first derivative. If \( w(x) \) satisfies (1.7), we make the change of variable

\[
w(x) = e^{-\frac{1}{4}\int f(t)dt}y(x),
\]
and then we obtain

\[
y''(x) = q(x)y(x),
\]
where

\[
q(x) = \frac{1}{4}f^2(x) + \frac{1}{2}f'(x) - g(x).
\]

This is known as the normal form of the differential equation. This formulation is useful in several contexts, for instance when analysing real zeros of solutions of the ODE:

**Theorem 1.2.4** Let \( y(x) \) be any non trivial solution of (1.9). If \( q(x) > 0 \) in \((a, b)\) then \( y(x) \) can have at most one zero in \((a, b)\).

The proof of this theorem is a straightforward application of Rolle’s theorem. A sufficient condition for oscillation is the following, see [48, §24]:

**Theorem 1.2.5** Let \( y(x) \) be any non-trivial solution of (1.9), and suppose that \( q(x) < 0 \) for \( x > 0 \), if

\[
\int_1^\infty q(x)dx = -\infty,
\]
then \( y(x) \) has an infinite number of zeros in the negative semiaxis.
For example, this result can be used to prove that any solution of the Airy differential equation \( y''(x) = xy(x) \) has an infinite number of zeros on the negative real axis. On the other hand, when \( x > 0 \) there can be at most one real zero. This transition from oscillatory to monotonic behaviour of the Airy function is a very important property.

Further properties on the real zeros of solutions of this type of ODEs can be obtained via the classical Sturm theorem and related results.

1.2.3 Regular and singular points and power series solutions

The existence and behaviour of solutions of (1.7) around a certain point \( x = x_0 \) (not necessarily finite), depends on the following classification. We say that \( x = x_0 \) is a

- regular point if both \( f(x) \) and \( g(x) \) are analytic at \( x = x_0 \).
- regular singular point if \( f(x) \) or \( g(x) \) are not analytic at \( x = x_0 \) but \( (x - x_0)f(x) \) and \( (x - x_0)^2g(x) \) are analytic.
- irregular singular point in any other case.

The case \( x_0 = \infty \), which is of importance in asymptotic analysis, can be studied by making the change of variables \( x = 1/t \) and investigating the resulting equation at \( t = 0 \). Thus we can show that equation (1.7) can be transformed to

\[
\ddot{w}(t) + F(t)\dot{w}(t) + G(t)w(t) = 0, \quad (1.11)
\]

where dots indicate derivatives with respect to \( t \) and

\[
F(t) = \frac{2}{t} - \frac{1}{t^2} f \left( \frac{1}{t} \right), \quad G(t) = \frac{1}{t^4} g \left( \frac{1}{t} \right). \quad (1.12)
\]

In the sequel we will analyse finite regular or singular points at the origin, that is \( x_0 = 0 \), in order to simplify notation. Clearly there is no loss of generality in doing this.

1.2.4 Regular points

Near a regular point \( x_0 \), it can be shown that any solution of (1.7) can be expanded in power series around \( x = x_0 \), with a radius of convergence that is at least as large as the corresponding one for the coefficients of the ODE.
Theorem 1.2.6 Let \( f(x) \) and \( g(x) \) be analytic in an open neighbourhood \( D \) of a regular point \( x_0 \), then given two initial conditions \( w_0 \) and \( w'_0 \), there exists a unique function \( w(x) \) which is analytic in \( D \) and satisfies (1.7) with \( w(x_0) = w_0 \) and \( w'(x_0) = w'_0 \).

This result is standard, and is usually proved using Picard’s successive approximations or iterates, see for instance [44]. If we suppose that

\[
f(x) = \sum_{k=0}^{\infty} f_k x^k, \quad g(x) = \sum_{k=0}^{\infty} g_k x^k
\]

and we expand

\[
w(x) = \sum_{k=0}^{\infty} a_k x^k, \quad |x| < R,
\]

then we can identify the coefficients \( a_k \) by imposing the differential equation (1.7) term by term. This gives the following conditions on the coefficients:

\[
k(k+1)a_{k+1} + \sum_{j=1}^{k} jf_k a_j + \sum_{j=0}^{k-1} g_k a_{j+1} = 0, \quad k = 1, 2, \ldots
\]

We need to impose two initial values \( a_0 \) and \( a_1 \), which is clearly equivalent to giving the initial values \( w(x_0) \) and \( w'(x_0) \).

Example 1.2.1 We will find the power series expansion around the origin of a solution of the Airy equation

\[
y''(x) = xy(x),
\]

with prescribed \( a_0 = a_1 = 1 \). Since \( x = 0 \) is a regular point of the ODE, we write (1.14), and identifying terms with equal powers of \( x \). We obtain \( a_2 = 0 \) and

\[
a_{k+3} = \frac{a_k}{(k+3)(k+2)}, \quad k \geq 1,
\]

from (1.15), so

\[
a_3 = \frac{a_0}{3 \cdot 2} = \frac{1}{3!}, \quad a_6 = \frac{a_3}{6 \cdot 5} = \frac{4}{6!}, \quad a_9 = \frac{a_6}{9 \cdot 8} = \frac{1 \cdot 4}{9!}, \quad \ldots
\]

and

\[
a_4 = \frac{a_1}{4 \cdot 3} = \frac{2}{4!}, \quad a_7 = \frac{a_4}{7 \cdot 6} = \frac{2 \cdot 5}{7!}, \quad a_{10} = \frac{a_7}{10 \cdot 9} = \frac{2 \cdot 5 \cdot 8}{10!}, \quad \ldots
\]

A particular combination of the terms of these two series is used to construct the classical Airy function \( Ai(x) \), see [2, Eq.10.4.2-3], that we will see again soon.
Example 1.2.2 The Hermite differential equation reads

\[ w''(x) - 2xw'(x) + 2nw(x) = 0, \]

where \( n \) is a constant. We deduce that \( f_1 = -2, g_0 = 2n \) and all the rest of the coefficients are equal to 0. The recursion (1.15) is then

\[ k(k+1)a_{k+1} = (2k - 2 - 2n)a_{k-1}, \quad k \geq 1. \]

If we take \( a_0 = 1, a_1 = 0 \) and set \( k = 2s - 1 \), we get

\[ a_{2s} = \frac{-2(n-2s+2)}{2s(2s-1)}a_{2s-2} = \frac{-2(n-2s+2)}{2s(2s-1)} \frac{-2(n-2s+4)}{(2s-2)(2s-3)}a_{2s-4} \]

\[ = \ldots = \frac{(-1)^s2^n(n-2)\ldots(n-2s+2)}{(2s)!}a_0. \]

If we take \( a_0 = 0, a_1 = 1 \) and set \( k = 2s \), we get

\[ a_{2s+1} = \frac{-2(n-2s+1)}{2s(2s+1)}a_{2s-1} = \frac{-2(n-2s+1)}{2s(2s+1)} \frac{-2(n-2s+3)}{(2s-2)(2s-1)}a_{2s-3} \]

\[ = \ldots = \frac{(-1)^s2^s(n-1)(n-3)\ldots(n-2s+1)}{(2s+1)!}a_1. \]

Observe that if \( n \) is a positive integer one of the series terminates and reduces to a polynomial, whereas the other one continues. Setting \( a_n = 2^n \) we obtain the well known Hermite polynomials as the terminating solution. For example, if \( n = 2 \) we get \( a_0 = 1, a_2 = -n = -2 \), so normalizing

\[ H_2(x) = 4x^2 - 2, \]

if \( n = 4 \) we get \( a_0 = 1, a_2 = -n = -4, a_4 = n(n-2)/6 = 4/3 \), so if we normalize again we obtain

\[ H_4(x) = 16x^4 - 48x^2 + 12. \]

Exercise 1.2.1 We consider the Weber differential equation:

\[ y''(x) = \left(\frac{1}{4}x^2 + a\right)y(x). \]

Show that it has two independent solutions

\[ y_1(x) = \sum_{k=0}^{\infty} a_{2k} \frac{x^{2k}}{(2k)!}, \quad y_2(x) = \sum_{k=0}^{\infty} a_{2k+1} \frac{x^{2k+1}}{(2k+1)!}, \]

where \( a_0 = a_1 = 1, a_2 = a_3 = a \) and the coefficients satisfy the recursion

\[ a_{n+2} = a a_n + \frac{1}{4}n(n-1)a_{n-2}, \quad n \geq 2. \]
1.2.5 Regular singular points

In the neighbourhood of a regular singular point, we expand

\[ xf(x) = \sum_{k=0}^{\infty} f_k x^k, \quad x^2 g(x) = \sum_{k=0}^{\infty} g_k x^k, \quad (1.16) \]

assuming convergence of these series for \( |x| < R \), and we make the ansatz

\[ y(x) = x^\lambda \sum_{k=0}^{\infty} a_k x^k, \quad (1.17) \]

for a certain (not necessarily integer) exponent \( \lambda \). This is motivated by the fact that it seems reasonable to suppose that near the singularity, \( xf(x) \approx f_0 \) and \( x^2 g(x) \approx g_0 \), so one would expect the solutions of the differential equation to behave similarly to those of the equation

\[ x^2 w''(x) + xf_0 w'(x) + g_0 w(x) = 0. \quad (1.18) \]

This equation can be solved explicitly with \( y(x) = x^\lambda \), and this leads to the so called \textit{indicial equation}:

\[ \lambda(\lambda - 1) + f_0 \lambda + g_0 = 0, \quad (1.19) \]

which in general gives two suitable values for the parameter \( \lambda \), say \( \lambda_1 \) and \( \lambda_2 \). If we substitute (1.17) into the ODE, we obtain again a recursion for the coefficients:

\[ [(\lambda + k)(\lambda + k - 1) + (\lambda + k)f_0 + g_0] a_k = -\sum_{j=0}^{k-1} [(\lambda + j)f_{k-j} + g_{k-j}] a_j, \quad (1.20) \]

for \( k \geq 1 \). When \( k = 0 \), this is just the indicial equation, taking the sum equal to 0. Now we observe that the term multiplying \( a_k \) on the left hand side is

\[ (\lambda + k)(\lambda + k - 1) + (\lambda + k)f_0 + g_0 = (\lambda + k - \lambda_1)(\lambda + k - \lambda_2). \]

If \( \lambda_1 \) and \( \lambda_2 \) do not differ by an integer number, then this procedure is fine in order to generate two independent solutions of the differential equation. In this case, it can be proved that there exist two solutions of the ODE (corresponding to the two roots of the indicial equation), that can be expanded in powers of \( x - x_0 \) in a certain domain \( |x - x_0| < R \). This result goes back to Frobenius, whose name is associated with this method.
Theorem 1.2.7 Let the functions \(xf(x)\) and \(x^2g(x)\) admit a power series expansion (1.16), then the series (1.17) converges and defines a solution of the differential equation when \(|x| < R\) for some \(R > 0\), provided that the two exponents given by the indicial equation do not differ by an integer number.

Example 1.2.3 The Bessel differential equation is
\[
x^2w''(x) + xw'(x) + (x^2 - \nu^2)w(x) = 0,
\]
where \(\nu\) is a parameter. We deduce that \(f_0 = 1, g_0 = -\nu^2\) and \(g_2 = 1\), the rest of the coefficients being 0. The indicial equation gives solutions \(\lambda_{1,2} = \pm \nu\), so if \(\nu\) is not an integer or a half integer, we can get two independent solutions using the method of Frobenius. If we take \(\lambda = \nu\), then (1.20) gives
\[
a_k = -\frac{a_{k-2}}{(2\nu + k)k}.
\]
Setting \(a_0 = \frac{\Gamma(\nu+1)}{2\nu}, a_1 = 0\) and \(k = 2s\), we obtain
\[
a_{2s} = -\frac{a_{2s-2}}{(2\nu + 2s)2s} = \ldots = \frac{(-1)^s\Gamma(\nu + 1)a_0}{2^{2s}s!\Gamma(\nu + s + 1)} = \frac{1}{2^{2s}s!\Gamma(\nu + s + 1)}\frac{(-1)^s}{\Gamma(\nu + s + 1)}.
\]
This gives the well known series expansion for the Bessel function of the first kind [2, 9.1.10]:
\[
J_\nu(x) = \left(\frac{x}{2}\right)^\nu \sum_{k=0}^{\infty} \frac{(-x^2/4)^k}{k!\Gamma(\nu + k + 1)}.
\]
The second solution \(J_{-\nu}(x)\) can be obtained in a similar way.

If the roots of the indicial equation differ by an integer, then further work is needed, since the term multiplying \(a_k\) in (1.20) can vanish. One of the solutions admits a power series expansion of the form (1.17), but the procedure does not give a second independent solution. This second solution typically involves logarithmic terms. Namely, suppose that the exponents verify \(\lambda_1 - \lambda_2 = m\), \(m\) being an integer, and let \(w_1(x)\) denote a solution of the ODE
\[
x^2w''(x) + xf(x)w'(x) + g(x)w(x) = 0.
\]
Then one sets \(w_2(x) = w_1(x)q(x)\), for a certain function \(q(x)\), that upon substitution satisfies the following ODE:
\[
q''(x) + \left(\frac{2w_1'(x)}{w_1(x)} + \frac{f(x)}{x}\right)q'(x) = 0,
\]
and therefore

\[ \log q'(x) = -2 \log w_1(x) - \int_{x_0}^{x} \frac{f(\zeta)}{\zeta} d\zeta + A, \]

where \( A \) is a constant, that we take equal to 0. Observe now that

\[ q'(x) = \frac{1}{w_1^2(x)} e^{-\int_{x_0}^{x} \frac{f(\zeta)}{\zeta} d\zeta}. \]

The function \( f(x) \) is analytic in a neighbourhood of \( x = 0 \), so

\[ \frac{f(\zeta)}{\zeta} = \frac{f_0}{\zeta} + f_1 + f_2\zeta + \ldots \]

Consequently,

\[ \int_{x_0}^{x} \frac{f(\zeta)}{\zeta} d\zeta = p_0 \log x + h_1(x), \]

where \( h_1(x) \) is analytic near \( x = 0 \), and since \( w_1(x) = x^{\lambda_1} h_2(x) \), where \( h_2(x) \) is analytic again, we have

\[ q'(x) = x^{-f_0 - 2\lambda_1} h_3(x), \]

where \( h_3(x) \) is analytic again near \( x = 0 \). Now, from the indicial equation (1.19), we know that \( \lambda_1 + \lambda_2 = 1 - f_0 \), so \( -f_0 - 2\lambda_1 = -\lambda_1 + \lambda_2 - 1 = -m - 1 \), and therefore

\[ g'(x) = x^{-m - 1} h_3(x). \]

If we expand \( h_3(x) \) around the origin,

\[ h_3(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_m x^m + \ldots, \]

and then integrate term by term, we get

\[ g(x) = a_m \log x + x^{-m} h_4(x), \]

with \( h_4(x) \) analytic once more. Finally,

\[ w_2(x) = w_1(x) g(x) = a_m w_1(x) \log(x) + x^{\lambda_3} \sum_{n \geq 0} b_n x^n, \]

and the second solution includes a logarithmic term, unless the coefficient \( a_m \) happens to be 0. More details can be found for example in [44, §5.5], [50, §4.2.5].
In the example that we considered before, when the order of the Bessel function is an integer $\nu = n$, then the second exponent $-n$ does not give an independent solution, actually $J_{-n}(x) = (-1)^n J_n(x)$. A second independent solution of the ODE is the Bessel function of the second kind, denoted $Y_n(x)$. A power series expansion around $x = 0$ can be computed for this function, and it follows the pattern of this last computation, but the result is complicated, see [2, 9.1.11].

1.2.6 Irregular singular points

The analysis of irregular singular points is more complicated, and in principle there is no general method. In particular, it may happen that the solutions cannot be represented as power series around the singular point in the same way that we have seen before. For example, see [50], the equation

$$x^3 y''(x) + xy'(x) - 2y(x) = 0$$

has an irregular singular point at the origin, and the solution $y(x) = e^{1/x}$ does not admit a power series expansion around the origin.

Another example is given by the Airy equation at infinity. Applying the transformation $t = 1/x$ we get, in accordance with (1.11),

$$\ddot{w}(t) + \frac{2}{t}\dot{w}(t) - \frac{1}{t^3}w(t) = 0,$$

which has an irregular singular point at $t = 0$. We will see that the Airy function has exponential behaviour at $x = \infty$, so further work is needed.

**Definition 1.2.3** A second order differential equation (or more generally a system of ODEs) is called Fuchsian if the only singularities are regular singular points.

It is not difficult to show that there exists no second order Fuchsian differential equation with no singular points (including the point at $\infty$). Indeed, suppose that the coefficients of the ODE $f(x)$ and $g(x)$ are analytic at $x = 0$, so

$$f(x) = f_0 + f_1 x + f_2 x^2 + \ldots, \quad g(x) = g_0 + g_1 x + g_2 x^2 + \ldots,$$

and make the transformation $x = 1/t$. The resulting equation has a singular point at $t = 0$, that is at $x = \infty$, unless $f \equiv 0$ and $g \equiv 0$.

The following exercises are taken from [26], and they motivate the study of classical hypergeometric functions that we present in the next section.
**Exercise 1.2.2** Show that if a second order Fuchsian differential equation has exactly one finite regular singular point at \( x = 0 \), then it is of the form

\[
y''(x) + \frac{2}{x}y'(x) = 0.
\]

Solve this equation in terms of elementary functions. Similarly, show that if a second order Fuchsian differential equation has exactly two finite regular singular points, say at \( x = 0 \) and at \( x = a \), then it is of the form

\[
y''(x) + \frac{2x + c_1}{x(x - a)}y'(x) + \frac{c_2}{x^2(x - a)^2} = 0,
\]

where \( c_1 \) and \( c_2 \) are constants. Make the change of variable \( t = x/(x - a) \) and solve the resulting equation in terms of elementary functions.

### 1.3 The Gauss hypergeometric equation

When we allow three regular singular points, it is no longer possible to solve a second order ODE

\[
w''(x) + f(x)w'(x) + g(x)w(x) = 0 \tag{1.21}
\]

using only elementary functions. The most general formulation of this kind of ODE is the Riemann–Papperitz equation, see for example [44, §5.8.2].

The construction is as follows: let us suppose that \( x = \alpha, \ x = \beta \) and \( x = \gamma \) are the three regular singular points, then we can write

\[
f(x) = \frac{A}{x - \alpha} + \frac{B}{x - \beta} + \frac{C}{x - \gamma} + u_1(x)
\]

\[
(x - \alpha)(x - \beta)(x - \gamma)g(x) = \frac{D}{x - \alpha} + \frac{E}{x - \beta} + \frac{F}{x - \gamma} + u_2(x), \tag{1.22}
\]

where \( u_1(x) \) and \( u_2(x) \) are analytic functions. If \( \infty \) is a regular point, then from (1.11) and (1.12) we know that \( 2x - x^2f(x) \) and \( x^4g(x) \) must be analytic at \( \infty \), and that imposes

\[
A + B + C = 2, \quad u_1(x) = u_2(x) = 0.
\]

We apply Frobenius method to the equation to construct solutions around \( x = \alpha \). Thus

\[
w(x) = x^\lambda \sum_{k=0}^{\infty} a_k x^k,
\]
and the indicial equation is
\[ \lambda(\lambda - 1) + A\lambda + \frac{D}{(\alpha - \beta)(\alpha - \gamma)} = 0, \]
so the exponents \( a_1, a_2 \) corresponding to the singular point \( x = \alpha \) verify
\[ a_1 + a_2 = 1 - A, \quad a_1a_2 = \frac{D}{(\alpha - \beta)(\alpha - \gamma)}. \]

Similarly, the exponents \( b_1, b_2 \) (corresponding to \( x = \beta \)) and \( c_1, c_2 \) (corresponding to \( x = \gamma \)) satisfy
\[ b_1 + b_2 = 1 - B, \quad b_1b_2 = \frac{E}{(\beta - \alpha)(\beta - \gamma)}, \]
\[ c_1 + c_2 = 1 - C, \quad c_1c_2 = \frac{F}{(\gamma - \alpha)(\gamma - \beta)}. \]

We observe that \( A + B + C = 2 \) implies
\[ a_1 + a_2 + b_1 + b_2 + c_1 + c_2 = 1, \]
which is the Fuchs relation for the exponents of the ODE.

The resulting ODE is called the Papperitz (or Riemann, or Riemann–Papperitz) equation:
\[
\begin{align*}
    w''(x) &+ \left( \frac{1 - a_1 - a_2}{x} + \frac{1 - b_1 - b_2}{x - 1} + \frac{1 - c_1 - c_2}{x - \gamma} \right) w'(x) \\
    &- \left( \frac{a_1a_2}{(x - \alpha)(\beta - \gamma)} + \frac{b_1b_2}{(x - \beta)(\gamma - \alpha)} + \frac{c_1c_2}{(x - \gamma)(\alpha - \beta)} \right) \\
    \times \frac{(\alpha - \beta)(\beta - \gamma)(\gamma - \alpha)}{(x - \alpha)(x - \beta)(x - \gamma)} w(x) &= 0.
\end{align*}
\]

It is customary to simplify this as follows: making a change of variables, we can set the singular points of the ODE to be \( \alpha = 0, \beta = 1 \) and \( \gamma = \infty \), and this leads to
\[
\begin{align*}
    w''(x) &+ \left( \frac{1 - a_1 - a_2}{x} + \frac{1 - b_1 - b_2}{x - 1} \right) w'(x) \\
    &- \left( \frac{a_1a_2}{x^2(x - 1)} + \frac{b_1b_2}{x(x - 1)^2} + \frac{c_1c_2}{x(x - 1)} \right) w(x) = 0.
\end{align*}
\]

A very important observation is that if we set
\[ W(x) = x^{-a_1}(1 - x)^{-b_1}, \]
then the exponents at \( x = 0 \) and \( x = 1 \) are decreased by \( a_1 \) and \( b_1 \) (respectively), but because of the Fuchs relation, the exponents at \( \infty \) must increase by \( a_1 + b_1 \). Therefore, the new exponents are 0 and \( a_2 - a_1 \) (corresponding to \( x = 0 \)), 0 and \( b_2 - b_1 \) (corresponding to \( x = 1 \)), and \( c + a_1 + b_1 \) and \( c_2 + a_1 + b_1 \) (corresponding to \( x = \infty \)). Relabeling these new indices as

\[
a = c_1 + a_1 + b_1, \quad b = c_2 + a_1 + b_1, \quad c = 1 + a_1 - a_2,
\]

we arrive at the standard form of the hypergeometric differential equation, already known to Euler but now associated to Gauss and Riemann:

\[
x(1 - x)y''(x) + [c - (a + b + 1)x]y'(x) - aby(x) = 0, \tag{1.23}
\]

where \( a, b \) and \( c \) are parameters. We assume that \( c, a - b \) and \( c - a - b \) are not integers, so that the Frobenius method always gives two independent solutions at each singular point.

The point \( x = 0 \) is a regular singular point, and the indicial equation gives two solutions \( \lambda = 0 \) and \( \lambda = 1 - c \). We note that

\[
xf(x) = \frac{c - (a + b + 1)x}{1 - x}, \quad x^2 g(x) = \frac{-ab}{1 - x},
\]

so \( f_0 = c, f_k = c - a - b - 1 \) for \( k \geq 1 \), \( g_0 = 0 \) and \( g_k = -ab \) for \( k \geq 1 \). If we set \( \lambda = 0 \) in (1.20), we get

\[
k(k + c - 1)a_k = \sum_{j=0}^{k-1} \{ab + (a + b + 1 - c)j\}a_j, \quad k \geq 1.
\]

If we write this formula for \( a_k \) and for \( a_{k-1} \) and subtract, we obtain

\[
k(k + c - 1)a_k = (a + k - 1)(b + k - 1)a_{k-1}, \quad k \geq 1,
\]

so fixing \( a_0 = 1 \), we have

\[
a_k = \frac{(a)_{k(b)_{k}}}{(c)_{k}} \frac{1}{k!}, \quad k \geq 1,
\]

using the standard Pochhammer symbol.

The resulting solution in power series is

\[
y_1(x) = \sum_{k=0}^{\infty} \frac{(a)_{k(b)_{k}} x^k}{(c)_{k}} \frac{1}{k!} = 1 + \frac{ab}{c} x + \frac{a(a + 1)b(b + 1)}{c(c + 1)} \frac{x^2}{2} + \ldots \tag{1.24}
\]
A second solution with a similar form can be computed when \( \alpha = 1 - c \):

\[
y_2(x) = x^{1-c} \sum_{k=0}^{\infty} \frac{(a + 1 - c)_k(b + 1 - c)_k x^k}{(2 - c)_k k!}.
\]

The function \( y_1(x) \) is usually known as the Gauss hypergeometric function. If \( c \notin \mathbb{Z} \) then we write these two independent solutions around \( x = 0 \) using the notation:

\[
y_1(x) = \text{}_2F_1 \left( \begin{array}{c} a, b \\ c \end{array} : x \right), \quad y_2(x) = \text{}_2F_1 \left( \begin{array}{c} a + 1 - c, b + 1 - c \\ 2 - c \end{array} : x \right).
\]

We also observe that the radius of convergence of both series is equal to 1, using the ratio test, so the hypergeometric function is defined only inside the unit disc. By using integral representations and analytic continuation one can extend the domain to the whole complex plane minus a cut that is usually taken on \([1, \infty)\). We refer the reader for example to [2], [38] or [50] for more details.

It is possible to obtain pairs of solutions around the points \( x = 1 \) and \( x = \infty \), with suitable restrictions on the parameters. We note that the exponents are as follows: when \( x = 1 \) we have \( \alpha_1 = 0, \alpha_2 = c - a - b \), and when \( x = \infty \) we obtain \( \alpha_1 = a \) and \( \alpha_2 = b \). It follows from this result that whenever \( c, c - a - b \) and \( a - b \) are not integers, then a pair of independent solutions can be found using Frobenius method. A full description of all possible cases is given in several references, such as [2] or [50], for instance.

**Exercise 1.3.1** Assuming that \( c, c - a - b \) and \( a - b \) are not integers, find pairs of independent solutions of the Gauss hypergeometric differential equation around the regular singular points \( x = 1 \) and \( x = \infty \).

As a particular case of great importance, we mention the Jacobi polynomials as special case of the Gauss function:

\[
P_n^{(\alpha, \beta)}(x) = \binom{n + \alpha}{n} \text{}_2F_1 \left( \begin{array}{c} -n, n + \alpha + \beta + 1 \\ \alpha + 1 \end{array} : \frac{1-x}{2} \right),
\]

and naturally the Gegenbauer (when \( \alpha = \beta \)), Chebyshev (\( \alpha = \beta = \pm 1/2 \)) and Legendre (\( \alpha = \beta = 0 \)) polynomials as well.

21
1.4 The confluent hypergeometric equation

If we consider the change of variables \( x \to x/b \) in the Gauss differential equation, the resulting equation has three regular singular points: \( x = 0 \), \( x = \infty \) and \( x = b \). If we let \( b \to \infty \) then the point \( x = \infty \) becomes an irregular singularity due to the *confluence* of two regular singular points. We thus arrive at the confluent differential equation or Kummer equation:

\[
xy''(x) + (c - x)y'(x) - ay(x) = 0. 
\]

(1.28)

The origin is still a regular singular point, with exponents \( \alpha = 0 \) and \( \alpha = 1 - c \), so we can compute two solutions in power series:

\[
y_1(x) = \sum_{n=0}^{\infty} \frac{(a)_n x^n}{(c)_n n!}, \quad y_2(x) = x^{1-c} \sum_{n=0}^{\infty} \frac{(a + 1 - c)_n x^n}{(2 - c)_n n!},
\]

(1.29)

which are independent if \( c \notin \mathbb{Z} \).

We note that in this case the power series is convergent for all \( x \), unlike what happened with the Gauss function. The usual notation for \( y_1(x) \), which is called *confluent hypergeometric function of the first kind* or *Kummer function*, is the following:

\[
M(a; c; x) := {}_1F_1(a; c; x) = \sum_{n=0}^{\infty} \frac{(a)_n x^n}{(c)_n n!}. 
\]

(1.30)

An alternative second solution is given by a particular linear combination of Kummer functions:

\[
U(a; c, x) := \frac{\Gamma(1-c)}{\Gamma(a+1-c)} M(a; c; x) + \frac{\Gamma(c-1)}{\Gamma(a)} x^{1-c} M(a + 1 - c; 2 - c; x). 
\]

(1.31)

This function is known as the confluent hypergeometric function of the second kind or *Tricomi function*. The functions \( M(a; c; x) \) and \( U(a; c, x) \) are independent if \( a \neq 0, -1, -2, \ldots \).

The confluent hypergeometric function of the second kind \( U(a, c, x) \) has an interesting representation as a Laplace integral:

\[
U(a, c, x) = \frac{1}{\Gamma(a)} \int_0^{\infty} e^{-xt} t^{a-1} (1 + t)^{c-a-1} dt, \quad a > 0, 
\]

(1.32)

which is useful for asymptotic analysis, and it also satisfies the following functional identity:

\[
U(a, c, x) = x^{1-c} U(a + 1 - c, 2 - c, x). 
\]

(1.33)
An alternative to Kummer functions is given by Whittaker functions, which are obtained from the normal form of the confluent hypergeometric differential equation:

$$y''(x) + \left(-\frac{1}{4} + \frac{c/2 - a}{x} + \frac{1-(c-1)^2}{4x^2}\right)y(x) = 0.$$ 

Relabeling the parameters as $\kappa = c/2 - a$ and $\mu = c/2 - 1/2$, we obtain Whittaker’s differential equation:

$$y''(x) + \left(-\frac{1}{4} + \frac{\kappa}{x} + \frac{1/4 - \mu^2}{x^2}\right)y(x) = 0,$$

with solutions

$$M_{\kappa,\mu}(x) = e^{-x/2}x^{1/2+\mu}M\left(\frac{1}{2} + \mu - \kappa, 1 + 2\mu, x\right)$$

$$W_{\kappa,\mu}(x) = e^{-x/2}x^{1/2+\mu}U\left(\frac{1}{2} + \mu - \kappa, 1 + 2\mu, x\right).$$ (1.34)

Again, we have families of classical orthogonal polynomials as examples: namely, Laguerre polynomials [2, 22.5.54]:

$$L_n^{(\alpha)}(x) = \binom{n + \alpha}{n}M(-n; \alpha + 1; x) = \frac{(-1)^n}{n!}U(-n; \alpha + 1; x),$$ (1.35)

and Hermite polynomials:

$$H_n(x) = 2^n x U\left(\frac{1-n}{2}; \frac{3}{2}; x^2\right) = 2^n U\left(-\frac{n}{2}; \frac{1}{2}; x^2\right).$$ (1.36)

Using another process of confluence, we can obtain another hypergeometric function of type $\pFq{1}{0}{-c/2}{c+1;x}$. This function satisfies the ODE:

$$x y''(x) + cy'(x) - y(x) = 0,$$ (1.37)

and it is related to Bessel functions of the first kind by the following formula:

$$\pFq{1}{0}{-c/2}{c+1;-x} = \Gamma(c+1) x^{-c/2} J_c(2\sqrt{x}).$$ (1.38)

The fact that hypergeometric functions of Gauss and Kummer type satisfy second order ODEs allow us to obtain many properties, such as asymptotic behaviour and location of real zeros, using classical tools. We remark that these results are very well understood for some special subfamilies like classical orthogonal polynomials, but they are also general in the sense that they can be applied to any solution of the corresponding differential equation.
1.5 Related topics

In this introduction we have only dealt with classical hypergeometric functions. These are the most important cases in applications, but they also admit several generalizations and extensions:

- We can consider generalized hypergeometric functions:
\[
pFq(a_1, a_2, \ldots, a_p; b_1, b_2, \ldots, b_q; z) = \sum_{n=0}^{\infty} \frac{(a_1)_n(a_2)_n \ldots (a_p)_n z^n}{(b_1)_n(b_2)_n \ldots (b_q)_n n!}.
\]

The expansion in power series is analogous to those in the classical cases, but the convergence properties depend naturally on the number of parameters in the numerator and in the denominator. More explicitly:

- if \( p < q + 1 \), then the series converges for \(|x| < \infty\),
- if \( p = q + 1 \), then the series converges for \(|z| < 1\) (e.g. Gauss).
- if \( p > q + 1 \), then the series diverges.

A function of this type satisfies a higher order differential equation, namely
\[
\vartheta(\vartheta+b_1-1)(\vartheta+b_2-1) \ldots (\vartheta+b_q-1)F = z(\vartheta+a_1)(\vartheta+a_2) \ldots (\vartheta+a_p)F,
\]
where we use the differential operator \( \vartheta = z \frac{d}{dz} \).

In the field of orthogonal polynomials one can find examples of these generalized hypergeometric functions when dealing with non classical families of OP. For example, Hahn polynomials are given by
\[
h_n^{(\alpha, \beta)}(x, N) = \frac{(-1)^n(N-x-n)_n(\beta+x+1)_n}{n!} \times _3F_2\left(\begin{array}{c}
-n, -x, \alpha + N - x \\
N - x - n, -\beta - x - n
\end{array}; 1\right).
\]

- Another area of research deals with \( q \)-hypergeometric functions, and extension that goes back to Heine in the 19th century and has important applications in Physics. A classic reference in this field is [15].
In this chapter we have studied hypergeometric functions as solutions of linear differential equations. Nonlinear second order differential equations
\[ y''(x) = F(x, y(x), y'(x)), \]
have also been studied since the beginning of the 20th century, with several attempts to find a systematic classification.

A particularly important type of equations is found if we impose that the only movable (i.e. dependent on the initial data) singularities are poles. This is known as the Painlevé property, and gives rise (essentially) to six differential equations which are called Painlevé equations. The analytical and numerical properties of these equations and their solutions (the Painlevé transcendents) is a very active area of research nowadays, and it is widely believed that in the 21st century these functions will play a role equivalent to the one played by classical hypergeometric functions in the 19th and 20th centuries. See for example [14] for an extensive treatment.
Chapter 2

Asymptotic techniques

In this chapter we will give a short introduction to the asymptotic analysis of hypergeometric functions. We will focus on two techniques, those derived from integral representations (real or complex) of the solutions of the hypergeometric ODEs, and those given by the asymptotic analysis of the ODEs themselves.

This is a very classical area of mathematical theory with a considerable amount of literature, and we only intend to scratch the surface of it. Some standard references are [3, 54, 50] for asymptotic analysis of integrals and [44, 42] for both approaches.

Given a certain function $F(x)$, asymptotic analysis seeks an approximation of the form

$$F(x) \sim G(x) \sum_{k=0}^{\infty} a_k \phi_k(x),$$

(2.1)

which is valid, in a certain sense that we explain next, when $x$ is close to a limiting value $x_0$, which may be finite or infinite. The function $G(x)$ should be simpler than $F(x)$, in order for the approximation to be useful, and $\phi_k(x)$ should form an asymptotic sequence, which is defined as follows, see [44]:

**Definition 2.0.1** Let $\phi_k(x)$ be a sequence of functions defined in a set $S$ and $x_0$ a limit point (finite or infinite) in $S$, such that for every $k \geq 0$

$$\phi_{k+1}(x) = o(\phi_k(x)), \quad x \to x_0,$$

(2.2)

that is

$$\lim_{x \to x_0} \frac{\phi_{k+1}(x)}{\phi_k(x)} = 0.$$

Then $\{\phi_k(x)\}_{k \geq 0}$ is called an asymptotic sequence.
The statement (2.1) should be understood in the sense that for each nonnegative integer $K$ we have
\[ F(x) = G(x) \sum_{k=0}^{K-1} a_k \phi_k(x) + O(\phi_K(x)). \]

Standard examples of asymptotic sequences are $\phi_k(x) = (x - x_0)^k$ when $x_0$ is finite, or $\phi_k(x) = 1/x^k$ when $x_0$ is infinite. These are called Poincaré asymptotic expansions.

An important observation is that we are not requiring the series in (2.1) to be convergent increasing $n$ and fixed values of $x$. Actually, in most cases the series will be divergent, and it is the property (2.2) which is relevant. That is the reason why we do not write equality in (2.1), but just an asymptotic equivalence.

For example, see [44, §1.1], consider the function
\[ F(x) = \int_0^\infty e^{-xt} \cos t \, dt \]
for positive $x$. If we expand the cosine in power series and integrate term by term, we get
\[ F(x) = \frac{1}{x} - \frac{1}{x^3} + \frac{1}{x^5} - \ldots, \]
which converges to the sum
\[ F(x) = \frac{x}{x^2 + 1} \]
if $x > 1$. On the other hand, if we do the same with the function
\[ G(x) = \int_0^\infty \frac{e^{-xt}}{1 + t} \, dt, \]
then we obtain
\[ G(x) = \frac{1}{x} - \frac{1!}{x^2} + \frac{2!}{x^3} - \frac{3!}{x^4} + \ldots, \]
which diverges for all finite $x$. However, it is still true that defining $\phi_k(x) = k! / x^{k+1}$, so
\[ \lim_{x \to \infty} \frac{\phi_{k+1}(x)}{\phi_k(x)} = \lim_{x \to \infty} \frac{k + 1}{x} = 0, \]
and $\{\phi_k(x)\}_{k \geq 0}$ is indeed an asymptotic sequence.
2.1 Asymptotic analysis of differential equations

A classical method in asymptotic analysis of hypergeometric functions is the study of the second order differential equation itself. Here we will give some basic ideas, and refer the reader to standard references like [44, 42] for more details and examples.

Suppose that we want to analyse the behaviour of \( y(x) \), which is a solution of an ODE

\[
 w''(x) + f(x)w'(x) + g(x)w(x) = 0, \tag{2.3}
\]

for large values of \( x \). Applying the transformation \( x = 1/t \), as we indicated before, we can study the transformed equation

\[
 \ddot{w}(t) + F(t)\dot{w}(t) + G(t)w(t) = 0,
\]

where dots indicate derivatives with respect to \( t \) and

\[
 F(t) = \frac{2}{t} - \frac{1}{t^2} f \left( \frac{1}{t} \right), \quad G(t) = \frac{1}{t^4} g \left( \frac{1}{t} \right)
\]

near the point \( t = 0 \). We essentially find two different situations:

- If \( t = 0 \) is a regular or regular singular point, then the standard method of power series, or the method of Frobenius, provides expansions for the solutions in powers of \( t \) (i.e. inverse powers of \( x \)), which are of asymptotic nature.

- If \( t = 0 \) is an irregular singularity, then the situation is more complicated. For example, \( x = \infty \) is an irregular singular point of the Airy equation, and we have just seen that the function \( \text{Ai}(x) \) shows an exponentially decaying behaviour for large \( x \). Hence, an ansatz consisting in expanding \( y(x) \) in series that only involves powers of \( 1/x \) is not likely to work, because the behaviour of the function \( y(x) \) can be qualitatively very different.

Before we continue, we define the following concept:

**Definition 2.1.1** Let \( x_0 \) be a singular point of (2.3), and let \( l \) be the least integer such that the functions

\[
 (x - x_0)^lf(x), \quad (x - x_0)^{2l}g(x)
\]

are analytic at \( x = x_0 \), then the point \( x_0 \) is said to be a singularity of rank \( l - 1 \).
We observe that with this definition, a regular singular point is a singularity of rank 0.

For example, if we transform the confluent hypergeometric equation, we obtain

\[ \ddot{w}(t) + \left(\frac{2 - c}{t} + \frac{1}{t^2}\right) \dot{w}(t) - \frac{a}{t^3} w(t) = 0, \]

(2.4)

so the singularity at \( t = 0 \) (and then at \( x = \infty \)) has rank 1. In the case of the Airy equation, the transformed equation reads

\[ \ddot{w}(t) + \frac{2}{t} \dot{w}(t) + \frac{1}{t^5} w(t) = 0, \]

(2.5)

and the singularity at \( t = 0 \) (and then at \( x = \infty \)) has rank 2.

In the next subsection we analyse a general method to deal with the case of irregular singular points in the differential equation. A very simple situation where we encounter this type of singularity at \( x = \infty \) is an equation with constant coefficients:

\[ y''(x) = ky(x), \quad k \in \mathbb{R}, \]

with solution \( y(x) = Ae^{kx} + Be^{-kx} \), \( A \) and \( B \) being constants. This solution has an essential singularity at \( x = \infty \). The idea behind the Liouville transformation is to give an approximation to the solution of a generic ODE near an irregular singular point in terms of the solution of this ODE with constant coefficients.

2.1.1 The Liouville transformation. Construction

Suppose that we have an ODE of the form (2.3), and the coefficients \( f(x) \) and \( g(x) \) in (2.7) admit a power series expansion in inverse powers of \( x \):

\[ f(x) = \sum_{s=0}^{\infty} \frac{f_s}{x^s}, \quad g(x) = \sum_{s=0}^{\infty} \frac{g_s}{x^s}, \]

(2.6)

valid for large enough \( x \). Furthermore, at least one of the coefficients \( f_0, g_0 \) and \( g_1 \) must be different from 0 in order for the singular point to be irregular (this can be seen by applying the standard change of variables \( x = 1/t \)).

Without loss of generality, in this section we will work with equations in normal form, and we will follow the theory exposed in [44, Chapters 6 and 7]. Thus, we consider

\[ y''(x) = q(x)y(x), \]

(2.7)
where
\[ q(x) = \frac{1}{4}f^2(x) + \frac{1}{2}f'(x) - g(x). \]

The idea is to consider a change of independent variables \( z = z(x) \). Then, the function \( y(z) \) satisfies the following ODE in the new variable \( z \):
\[ \ddot{y} - \frac{x}{x}\dot{y} = x^2 q(z)y. \]

Now, assuming that the change of variables is at least three times differentiable and that \( z'(x) \neq 0 \), we make a transformation to normal form:
\[ Y(z) = (\dot{x}(z))^{-1/2} y(z), \quad (2.8) \]
that results in the ODE
\[ \ddot{Y}(z) = \Omega(z) Y(z), \quad (2.9) \]
where
\[ \Omega(z) = \dot{x}^2 q(x) + \sqrt{x} \frac{d^2}{dz^2} \frac{1}{\sqrt{\dot{x}(z)}} = \dot{x}(z)^2 q(x) - \frac{1}{2} \frac{\ddot{x}}{\dot{x}} + \frac{3}{4} \left( \frac{\dddot{x}}{\dot{x}} \right)^2. \]

Here dots denote derivatives with respect to the variable \( z \). Now we use the freedom to choose in \( z = z(x) \) in such a way that the term \( \dot{x}(z)^2 q(x) \) is constant and (for simplicity) equal to 1. Thus
\[ z(x) = \int^x q(t)^{1/2} dt. \]

Observe that we are assuming tacitly that \( q(x) \neq 0 \). If \( q(x) \) vanishes in the interval of study, say \( q(x_0) = 0 \), then equation (2.7) has a turning point at \( x = x_0 \), and the analysis is substantially different, see [44, Chapters 10 and 11].

It is not difficult to check that one can write (2.9) in the form
\[ \ddot{Y}(z) = (1 + \phi) Y(z), \quad (2.10) \]
where
\[ \phi = \frac{4q(x)q''(x) - 5[q'(x)]^2}{16[q(x)]^3} = -\frac{1}{q(x)^{3/4}} \frac{d^2}{dx^2} \left( q(x)^{-1/4} \right). \]

If \( \phi \equiv 0 \), then the exact solution of (2.10) is
\[ Y(z) = Ae^z + Be^{-z}, \]
and since
\[ y(z) = x^{1/2}Y(z) = q(z)^{-1/4}Y(z), \]
we obtain in the original variable \( x \)
\[ y(x) = Aq(x)^{-1/4}e^{\int_x^r q(t)^{1/2}dt} + Bq(x)^{-1/4}e^{-\int_x^r q(t)^{1/2}dt}, \] (2.11)
where \( A \) and \( B \) are constants.

If \( \phi \) is not identically 0, then the quality of (2.11) as an approximation naturally depends on the behaviour of the function \( q(x) \). We will be more precise about this point later on.

The formula (2.11) is usually known as the Liouville–Green (LG) approximation for the general solution of (2.7).

An alternative way to obtain this LG approximation is to take
\[ y(x) = e^{-\int_x^r \phi(t)dt}, \]
and then the new function \( \phi(x) \) satisfies a Riccati equation
\[ \phi'(x) + \phi(x)^2 = q(x). \]

If we neglect the term \( \phi'(x) \), then we obtain \( \phi(x) = \pm q(x)^{1/2} \). As a second approximation, we get
\[ \phi(x) = \pm q(x)^{1/2} - \frac{q'(x)}{4q(x)}, \] (2.12)
provided that \( |q'(x)| \ll 2|q(x)|^{3/2} \). If we integrate this, we get
\[ y(x) = e^{-\frac{1}{4} \log q(x) + \int_x^r q(t)^{1/2}dt} = q(x)^{-1/4}e^{\int_x^r q(t)^{1/2}dt} \]
again.

2.1.2 The Liouville transformation. Error bounds

In some cases (particularly when there are extra parameters involved in the differential equation), it is convenient to split the term multiplying \( w(x) \) in the ODE into two parts:
\[ y''(x) = (f(x) + g(x))y(x), \] (2.13)
and then the Liouville transformation yields
\[ \dot{Y}(z) = \left(1 + \phi(z) + \frac{g(z)}{f(z)} \right), \] (2.14)
where $\phi(z)$ is given by (2.12). In this setting, if $|\phi(z)| \ll 1$ and $|g| \ll |f|$ then we may expect the LG approximation to work well.

A precise result on the validity of the LG approximation is the following, see [44, Chapter 6, §2.2] for the proof:

**Theorem 2.1.1** Let $f(x)$ be a positive function which is $C^2(a_1, a_2)$, and $g(x)$ a continuous real or complex function. Define

$$F(x) = \int \left[ \frac{1}{f^{1/4}} \frac{d^2}{dx^2} \left( \frac{1}{f^{1/4}} \right) - \frac{g}{f^{1/2}} \right] dx.$$  

(2.15)

Then in the interval $(a_1, a_2)$ the ODE

$$y''(x) = (f(x) + g(x))y(x)$$  

(2.16)

has $C^2$ solutions

$$y_1(x) = f^{-1/4}(x) \exp \left( \int f^{1/2}(x)dx \right) (1 + \varepsilon_1(x)),$$

$$y_2(x) = f^{-1/4}(x) \exp \left( - \int f^{1/2}(x)dx \right) (1 + \varepsilon_2(x)),$$

(2.17)

such that

$$|\varepsilon_j| \leq \exp \left( \frac{1}{2} \mathcal{V}_{a_j,x}(F) \right) - 1, \quad j = 1, 2,$$

provided that $\mathcal{V}_{a_j,x}(F) < \infty$.

In this result, $\mathcal{V}_{a_j,x}(F)$ is the total variation of $F$ over $[a_j, x]$, which is defined as the supremum in $n$ of

$$\sum_{s=0}^{n-1} |F(x_{s+1}) - F(x_s)|,$$

taken over all possible subdivisions with $x_0 < x_1 < \ldots < x_n$ in the closure of $(a_j, x)$. In the particular case where $F(x)$ is monotonic in $[a, b]$, we have

$$\mathcal{V}_{a_j,x}(F) = |F(x) - F(a_j)|.$$

In the oscillatory case, we have a similar result:

**Theorem 2.1.2** Let $f(x)$ be a positive function which is $C^2(a_1, a_2)$, $g(x)$ a continuous real or complex function and $a_j$ an arbitrary point in the closure of $(a_1, a_2)$. Then in the interval $(a_1, a_2)$ the ODE

$$y''(x) = (-f(x) + g(x))y(x)$$  

(2.18)
has $C^2$ solutions

\[ y_1(x) = f^{-1/4}(x) \exp \left( i \int f^{1/2}(x) dx \right) (1 + \varepsilon_1(x)), \]
\[ y_2(x) = f^{-1/4}(x) \exp \left( -i \int f^{1/2}(x) dx \right) (1 + \varepsilon_2(x)), \]

such that

\[ |\varepsilon_j| \leq \exp (V_{a,x}(F)) - 1, \quad j = 1, 2, \]

provided that $V_{a,x}(F) < \infty$. Here the function $F(x)$ is defined as in the previous theorem.

### 2.1.3 ODEs with parameters

In many cases that involve classical special functions, there are extra parameters present in the differential equation, and the Liouville transformation is useful to analyse the asymptotic behaviour with respect to both the variable and the parameters.

We consider the equation

\[ y''(x) = (u^2 f(x) + g(x)) y(x), \]

where $u > 0$ is a parameter that can be large, and we suppose that both $f(x)$ and $g(x)$ are independent of $u$. Applying the theorem that we presented before, we have two solutions

\[ y_j(x) = f^{-1/4}(x) u^{-1/2} \exp \left( (-1)^{j-1} u \int f^{1/2}(x) dx \right) (1 + \varepsilon_j(x)), \]

for $j = 1, 2$, where

\[ |\varepsilon_j| \leq \exp \left( \frac{V_{a_j,x}(F)}{2u} \right) - 1, \quad j = 1, 2. \]

We note now that $F(x)$ is independent of $u$, since $f(x)$ and $g(x)$ do not depend on $u$, so

\[ \exp \left( \frac{V_{a_j,x}(F)}{2u} \right) - 1 = O(u^{-1}), \]

for large $u$ and fixed $x$. Moreover, if the total variation is finite for $x$ in some (possibly infinite) interval $(a_1, a_2)$, then this estimation is uniform in $x$. That is,

\[ y_j(x) \sim f^{-1/4}(x) u^{-1/2} \exp \left( (-1)^{j-1} u \int f^{1/2}(x) dx \right), \]
uniformly for \( x \in (a_1, a_2) \).

A similar result holds for solutions of the ODE

\[ y''(x) = (-u^2 f(x) + g(x))y(x), \quad (2.24) \]

namely

\[ y_j(x) \sim f^{-1/4}(x)u^{-1/2}\exp\left((-1)^{j-1}iu \int f^{1/2}(x)dx\right), \quad (2.25) \]

**Remark 2.1.1** In all cases, we suppose that \( V(F) \) converges at noth endpoints. It is possible to derive results when this total variation is divergent, we refer the reader to [44, Chapter 6, §3.4].

### 2.1.4 Examples

As an example, consider again the Airy equation. In that case,

\[ z(x) = \int^x t^{1/2}dt = \frac{2}{3}x^{3/2}. \]

and in the equation (2.9) we have

\[ \Omega(z) = 1 - \frac{5}{36z^2}. \]

The corresponding LG approximation reads

\[ y(x) = Ax^{-1/4}e^{\frac{2}{3}x^{3/2}} + Bx^{-1/4}e^{-\frac{2}{3}x^{3/2}}. \quad (2.26) \]

The standard Airy function \( Ai(x) \) corresponds to the choice \( A = 0, B = \frac{1}{2\sqrt{\pi}} \).

We can also compute a bound for the error. If we consider the interval \((x, \infty)\), for some \( x > 0 \), we obtain

\[ F(x) = \int x^{-1/4} \frac{d^2}{dx^2}(x^{-1/4})dx = \frac{5}{16}x^{-5/2} = -\frac{5}{24}x^{-3/2}, \quad (2.27) \]

and since \( F(x) \) is monotonic, we have

\[ \mathcal{V}_{x, \infty}(F) = \frac{5}{24}x^{-3/2}. \quad (2.28) \]

Hence

\[ |\varepsilon_j| \leq e^{\frac{5}{24}x^{-3/2}} - 1 = \mathcal{O}(x^{-3/2}) \]

34
for large $x$. Therefore,\[Ai(x) = \frac{1}{2\sqrt{\pi x^{1/4}}} e^{-\frac{2}{3}x^{3/2}} \left(1 + O(x^{-3/2})\right), \quad x \to \infty.\tag{2.29}\]

Consider now the Hermite differential equation:
\[w''(x) - 2xw'(x) + 2nw(x) = 0,\]
where $n$ is a parameter. In normal form, setting $y(x) = e^{-x^2/2}w(x)$, we have
\[y''(x) = (x^2 - 2n - 1)y(x).\]

Now we make the change of variable $x = \sqrt{2n + 1}z$, and then
\[\ddot{y}(z) = (2n + 1)^2(z^2 - 1)y(z).\]

We denote $u = 2n + 1$, $f(z) = z^2 - 1$ and $g(z) = 0$. Applying the result that we derived before, we obtain for large $n$ and uniformly for $|z| > 1$:
\[y_j(z) \sim (z^2 - 1)^{-1/4}(2n + 1)^{-1/2}
\times \exp \left((-1)^{j-1}(2n + 1) \left(\frac{z\sqrt{z^2 - 1}}{2} - \frac{1}{2} \log(z + \sqrt{z^2 - 1})\right)\right).\tag{2.30}\]

We pick the minus sign, so that the solution behaves like $x^n$, and writing $z = \cosh \phi$, we can simplify this to
\[y_2(z) \sim (\sinh \phi)^{-1/2}(2n + 1)^{-1/2} e^{(n/2 + 1/4)(2\phi - \sinh 2\phi)}.\tag{2.31}\]

Now we undo all the changes:
\[w(x) = e^{x^2/2}y_2(x) \sim e^{x^2/2}(\sinh \phi)^{-1/2}(2n + 1)^{-1/2} e^{(n/2 + 1/4)(2\phi - \sinh 2\phi)},\tag{2.32}\]
where $x = \sqrt{2n + 1}\cosh \phi$. Normalizing this function so that the leading coefficient is equal to 1, we get the standard Plancherel–Rotach asymptotics for Hermite polynomials, see [49, Eq. 8.22.13]. Furthermore, one can prove that the total variation is finite for $|t| > 1$, so it follows that the error term is $O(u^{-1})$ uniformly in $t$ (and therefore in $|x| > \sqrt{2n + 1}$).

It is possible to derive the asymptotics in the regime $|t| < 1$, using similar techniques (exercise). In this case the approximation is given in terms of trigonometric instead of hyperbolic functions, which is consistent with the fact that the Hermite polynomials have all their zeros in this region. Near the transition point $t \sim \pm 1$, the function $f(x)$ changes sign, which gives a turning point of the differential equation, and the approximation must be constructed in terms of Airy functions.
2.1.5 Higher order terms

Substituting the ansatz (2.6) in the differential equation and grouping equal powers of \( x \), we have

\[
q(x) = \frac{1}{4} f^2(x) + \frac{1}{2} f'(x) - g(x)
\]

\[
= \frac{1}{4} f_0^2 - g_0 + \left( \frac{1}{2} f_0 f_1 - g_1 \right) \frac{1}{x} + \left( \frac{1}{2} f_0 f_2 + \frac{1}{4} f_1^2 - \frac{1}{2} f_1 - g_2 \right) \frac{1}{x^2} + \mathcal{O}(x^{-3}).
\]

Since the Liouville approximation (2.11) gives

\[
y(x) \sim Cq(x)^{-1/4} \exp \left( \pm \frac{1}{2} \int^x q(t)^{1/2} \, dt \right),
\]

then it is possible to check that, to leading order,

\[
w(x) \sim C \exp \left( \pm (\rho x + \sigma \log z) \right),
\]

where

\[
\rho = \left( \frac{1}{4} f_0^2 - g_0 \right)^{1/2}, \quad \sigma = \frac{1}{4} f_0 f_1 - \frac{1}{2} g_1.
\]

This result motivates the general ansatz for the solution in the case of an irregular singular point in the differential equation. We assume that

\[
w(x) = e^{\lambda x} x^\mu \sum_{s=0}^{\infty} \frac{a_s}{x^s},
\]

we substitute this in the differential equation and we try to identify the terms multiplying the same powers of \( x \). This leads to a set of equations to solve for \( \lambda \), \( \mu \) and the coefficients of the expansion \( a_s \). The first two equations are

\[
\lambda^2 + f_0 \lambda + g_0 = 0, \quad (f_0 + 2 \lambda) \mu = -(f_1 \lambda + g_1),
\]

which are used in order to determine possible values of \( \lambda \) and \( \mu \). Higher order terms satisfy a recursion that is presented in [44]:

\[
(f_0 + 2 \lambda) sa_s = (s - \mu)(s - 1 - \mu) a_{s-1}
\]

\[
+ \sum_{j=1}^{s} [\lambda f_{j+1} + g_{j+1} - (s - j - \mu)f_j] a_{s-j}.
\]
In the case of the Airy differential equation, we take the transformed ODE
\[ \ddot{Y}(z) = \left(1 - \frac{5}{36z^2}\right) Y(z), \]
so we have \( f_s = 0 \) for \( s \geq 0 \), \( g_0 = -1 \) and \( g_2 = 5/36 \), all the other coefficients being 0. Therefore, we get \( \lambda_{\pm} = \pm 1 \) and \( \mu_{\pm} = 0 \). If we choose \( \lambda = -1 \), equation (2.35) yields
\[ a_s = -\frac{1}{72s} (6s - 1)(6s - 5)a_{s-1}. \quad (2.36) \]
The standard way to rewrite this is as follows:
\[ a_s = \frac{1}{54s} \frac{(3s - \frac{1}{2}) (3s - \frac{3}{2}) (3s - \frac{5}{2})}{s - \frac{1}{2}} a_{s-1}, \quad (2.37) \]
and iterating we obtain
\[ a_s = \frac{1}{54^s s!} \frac{(-1)^s (3s - \frac{1}{2}) (3s - \frac{3}{2}) \ldots \frac{3}{2} \cdot \frac{1}{2}}{(s - \frac{1}{2}) (s - \frac{3}{2}) \ldots \frac{3}{2} \cdot \frac{1}{2}} a_0 = \frac{1}{54^s s!} \frac{(-1)^s \Gamma (3s + \frac{1}{2})}{\Gamma (s + \frac{1}{2})} a_0. \quad (2.38) \]
Setting \( a_0 = 1 \), we get the large \( x \) expansion for the Airy function \( \text{Ai}(x) \), with the appropriate constants:
\[ \text{Ai}(x) = x^{-1/2} Y(z) = \left(\frac{3}{2} z\right)^{-1/6} Y(z) \sim \frac{1}{2\sqrt{\pi}} x^{-1/4} e^{-\frac{x^{3/2}}{4}} \sum_{s=0}^{\infty} \frac{(-1)^s a_s}{\left(\frac{3}{2} x^{3/2}\right)^s}, \quad (2.39) \]
where \( a_0 \) and the rest of the \( a_s \) are given by (2.38). See [2, Eq. 10.4.59].

**Example 2.1.1** The Bessel differential equation reads
\[ w''(x) + \frac{1}{x} w'(x) + \left(1 - \frac{\nu^2}{x^2}\right) w(x) = 0. \]
In this case \( f_1 = 1, \ g_0 = 1, \ g_2 = -\nu^2/x^2, \) and \( f_s = g_s = 0 \) otherwise. The first equation in (2.33) gives two values \( \lambda_{\pm} = \pm i \), and the second equation yields \( \mu = -1/2 \). Hence we can construct two solutions such that
\[ w_{\pm}(x) \sim C x^{-1/2} e^{\pm ix}, \quad x \to \infty. \]
With the normalizing convention
\[ C_{\nu} = \sqrt{\frac{2}{\pi}} e^{\pi \nu (\frac{1}{2} + \frac{1}{4}) \pi i}, \]
one obtains the classical Hankel functions, see for instance [2].
Exercise 2.1.1 Apply formulas (2.33)-(2.35) to obtain explicit expressions for the coefficients in the asymptotic expansions of the Hankel functions from this last example. Take \( a_0 = 1 \) in both cases.

There is a problem with this strategy when \( f_0^2 - 4g_0 = 0 \), because in that case both values of \( \lambda \) are equal and one does not obtain a pair of independent solutions in general. An alternative transformation is given in [44, §7.1.3] to cover that case.

The asymptotic character of this procedure is justified in [44, §7.2], and in that reference it is proved that the region of validity can be enlarged to cover complex values of \( x \) as well. Namely, we have

**Theorem 2.1.3** Let \( f(z) \) and \( g(z) \) be analytic functions of the complex variable \( z \) that have convergent series expansions

\[
f(z) = \sum_{s=0}^{\infty} \frac{f_s}{z^s}, \quad g(z) = \sum_{s=0}^{\infty} \frac{g_s}{z^s},
\]

in the annulus \( A = \{ |z| > a \} \), with \( f_0^2 \neq 4g_0 \). Then the equation

\[
\ddot{w}(z) + f(z) \dot{w}(z) + g(z) w(z) = 0
\]

has unique solutions \( w_j(z) \), \( j = 1, 2 \), such that in the respective intersections of \( A \) with the sectors

\[
|\arg(\lambda_2 - \lambda_1)z| \leq \pi, \quad j = 1 \\
|\arg(\lambda_1 - \lambda_2)z| \leq \pi, \quad j = 2,
\]

the functions \( w_j(z) \) are holomorphic and

\[
w_j(z) \sim e^{\lambda_j z} z^{\mu_j} \sum_{s=0}^{\infty} \frac{a_{s,j}}{z^s}, \quad z \to \infty.
\]

2.2 Asymptotic analysis of integrals

2.2.1 Laplace transforms

It is possible to derive integral representations for the solutions of second order ODEs by means of (real or complex) integration with respect to a certain kernel.

In this section we will use \( z \) instead of \( x \) as the variable, since in most cases we will have to deal with computations in the complex plane. Let us
suppose that we have a differential equation (or order $n$ in general) of the form:

$$\mathcal{L}_z[y(z)] := \sum_{k=0}^{n} l_k(z)y^{(k)}(z)$$

and we try to find:

$$y(z) = \int_{\alpha}^{\beta} K(z, t)v(t)dt,$$

where $K(z, t)$ is a suitably chosen integral kernel, and $v(t)$, $\alpha$ and $\beta$ are determined by boundary conditions of the following process. We apply the operator $\mathcal{L}_z$ to the integral form (formally):

$$0 = \mathcal{L}_z[y(z)] = \int_{\alpha}^{\beta} \mathcal{L}_z[K(z, t)]v(t)dt$$

If we can find an operator

$$\mathcal{M}_t[u(t)] = \sum_{k=0}^{n} m_k(t)u^{(k)}(t)$$

such that

$$\mathcal{L}_z[K(z, t)] = \mathcal{M}_t[K(z, t)]$$

then (formally):

$$0 = \int_{\alpha}^{\beta} \mathcal{L}_z[K(z, t)]v(t)dt = \int_{\alpha}^{\beta} \mathcal{M}_t[K(z, t)]v(t)dt.$$

Now

$$0 = \int_{\alpha}^{\beta} \mathcal{M}_t[K(z, t)]v(t)dt$$

$$= \int_{\alpha}^{\beta} K(z, t)\mathcal{M}_t^*[v(t)]dt + P[K(z, t), v(t)]\bigg|_{t=\beta}^{t=\alpha}$$

where

$$\mathcal{M}_t^*[v(t)] = \sum_{k=0}^{n} (-1)^k[m_k(t)v(t)]^{(k)}$$

is the adjoint of $\mathcal{M}_t$ and $P(u, v)$ is the so-called bilinear concomitant, such that:

$$v\mathcal{M}_t(u) - u\mathcal{M}_t^*(v) = \frac{d}{dt}P(u, v).$$

The steps are:
- Solve the problem $\mathcal{M}_t[v(t)] = 0$, which is an ODE that gives the function $v(t)$.
- Compute $P[K(z,t), v(t)]$ and find $\alpha$ and $\beta$ by imposing
  \[ P[K(z,\alpha), v(\alpha)] = P[K(z,\beta), v(\beta)], \]
  so that the boundary term vanishes, and we have
  \[ \int_\alpha^\beta \mathcal{M}_t[K(z,t)]v(t)dt = \int_\alpha^\beta K(z,t)\mathcal{M}_t[v(t)]dt. \]

A typical integral kernel is $K(z,t) = e^{\pm zt}$, which corresponds to a Laplace integral (possibly in the complex plane). For more details on this approach, see [50, 28].

**Example 2.2.1** In the Airy case we have
\[ \mathcal{L}_z[y(z)] = y''(z) - zy(z) = 0. \]
We take the kernel $K(z,t) = e^{-zt}$, and then:
\[ \mathcal{M}_t[K(z,t)] = \left(t^2 + \frac{d}{dt}\right)K(z,t), \]
\[ \mathcal{M}_t^*[K(z,t)] = \left(t^2 - \frac{d}{dt}\right)K(z,t) \]
and
\[ P[K(z,t), v(t)] = K(z,t)v(t). \]
Now $v(t)$ solves $\mathcal{M}_t^*[v(t)] = 0$, which gives $v(t) = e^{\frac{t^3}{3}}$, and:
\[ K(z,t)v(t) = e^{\frac{t^3}{3} - zt}. \]
Now we have to find $\alpha$ and $\beta$ such that
\[ e^{\frac{t^3}{3} - zt} \bigg|_{z=\beta} = 0 \]
This exponential is never $0$ if we have finite endpoints, so we take infinite contours such that $\Re t^3 \to -\infty$ at the endpoints. Thus
\[ \arg(t) = \pm \frac{\pi}{3}, \quad \arg(t) = \pi, \]
which gives three contours $C_1, C_2, C_3$ in the complex plane.
The standard Airy function is defined as:

$$Ai(z) = \frac{1}{2\pi i} \int_{C_1} e^{\frac{2}{3}zt} dt,$$

see [2, 1].

Another example is given by the Weber parabolic cylinder function $U(a, x)$, which is a solution of the equation:

$$L_x[y] = y''(x) - \left(\frac{x^2}{4} + a\right) y(x) = 0.$$

We take the kernel $K(x, t) = e^{-xt}$, and then:

$$\mathcal{M}_t[K(x, t)] = \left(t^2 + t \frac{d}{dt} - a + \frac{1}{2}\right) K(x, t).$$

$$\mathcal{M}_t^*[K(x, t)] = \left(t^2 - t \frac{d}{dt} - a - \frac{1}{2}\right) K(x, t).$$

Now $v(t)$ solves $\mathcal{M}_t^*[v(t)] = 0$, which gives $v(t) = e^{\frac{t^2}{2} t^{a - \frac{1}{2}}}$, and

$$P[K(x, t), v(t)] = K(x, t)v(t) = e^{-xt + \frac{t^2}{2} t^{a - \frac{1}{2}}}.$$
The standard definition reads:

\[ U(a, x) = \frac{e^{x^2}}{i\sqrt{2\pi}} \int_{\mathcal{C}} e^{-xt + \frac{i}{2} t^{-a - \frac{1}{2}}} dt, \]

where \( \mathcal{C} \) is a vertical line in the complex plane such that \( \text{Re}(t) > 0 \). See also [50, Chapters 4,7].

As we have seen before, Laplace transforms are a useful tool for deriving integral representations of the solutions of second order ODEs. The standard form of these integrals is

\[ F(x) = \int_0^\infty e^{-xt} f(t) dt, \quad (2.40) \]

where \( \text{Re} x > 0 \) and we assume that \( f(t) \) does not grow exponentially, in order for the integral to be defined.

In the next subsections we will analyse some methods to find asymptotic approximations for this kind of integrals (and more general instances of it) when \( x \) is large.

### 2.2.2 Integration by parts and Watson’s lemma

In order to obtain asymptotic approximations for \( F(x) \) when \( x \) is large, a naive approach would be to integrate by parts, assuming that \( f(t) \) is smooth enough and that the derivatives are small enough at \( \infty \). That would yield

\[ \int_0^\infty e^{-xt} f(t) dt \sim \sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{x^{k+1}}. \]

We do not use equality here because it is not clear in general that this process gives a convergent series on the right. What is true, however, is that if

\[ S_{K-1}(x) := \sum_{k=0}^{K-1} \frac{f^{(k)}(0)}{x^{k+1}}, \]

then

\[ F(x) - S_{K-1}(x) = \mathcal{O}(x^{-K}), \quad x \to \infty, \]

which is satisfactory in the sense of asymptotic analysis that we have given before.
Example 2.2.2 The expansion

\[ F(x) = \int_0^\infty e^{-xt} \cos t \, dt = \sum_{k=0}^\infty \frac{(-1)^k}{x^{2k+1}}, \]

is convergent for \( x > 1 \), because the Taylor series of \( \cos t \) converges everywhere (and uniformly on compact intervals). But

\[ F(x) = \int_0^\infty \frac{e^{-xt}}{1 + t} \, dt \sim \sum_{k=0}^\infty \frac{(-1)^k k!}{x^{k+1}} \]

is divergent for all value of \( x \).

More in general, if we can write

\[ f(t) = \sum_{k=0}^\infty \lambda_k t^{\mu_k}, \quad (2.41) \]

for an increasing sequence of exponents \( \mu_k \), with \( \mu_0 > -1 \), then interchanging summation and integration gives

\[ \int_0^\infty e^{-xt} f(t) \, dt \sim \sum_{k=0}^\infty \frac{\lambda_k \Gamma(\lambda_k + 1)}{k! x^{\lambda_k+1}}, \quad x \to \infty, \]

in terms of the Euler Gamma function. Of course, the validity of this procedure is subject to the (uniform) convergence of the power series \( (2.41) \). If the radius of convergence of this power series is not large enough, then the expansion should be understood in an asymptotic sense.

This idea is formalized in the classical Watson’s lemma:

Lemma 2.2.1 (Watson) Suppose that a function \( f(t) \):

- is analytic at \( t = 0 \) and within a sector \( |\arg t| < \alpha \).
- admits a power series expansion:

\[ f(t) = \sum_{n=0}^\infty a_n t^n, \quad |t| < R, \]

- does not grow exponentially fast when \( |\arg t| < \alpha \).
Then the Laplace-type integral:

\[ F(x) = \int_0^\infty e^{-xt}t^{\lambda-1}f(t)dt, \quad \text{Re} \lambda > 0 \] (2.42)

satisfies:

\[ F(x) \sim \sum_{k=0}^{\infty} \frac{a_k}{x^{\lambda+k}}, \quad x \to \infty, \quad |\arg x| < \alpha + \pi/2. \]

Informally, Watson’s lemma states that under certain conditions, integration term by term of the power series of \( f(t) \) yields an asymptotic expansion in inverse powers of \( x \) for the Laplace integral (2.42).

**Exercise 2.2.1** Apply Watson’s lemma to obtain asymptotic expansions of the modified Bessel function of the second kind

\[ K_\nu(x) = \frac{\sqrt{\pi}}{\Gamma(\nu + 1/2)} \left( \frac{x}{2} \right)^\nu \int_1^\infty e^{-xt}(t^2 - 1)^{\nu-1/2}dt \]

and the confluent hypergeometric function of the second kind

\[ U(a, c, x) = \frac{1}{\Gamma(a)} \int_0^\infty e^{-xt}t^{a-1}(1 + t)^{c-a-1}dt, \quad a > 0, \]

when \( x \to \infty \).

**Remark 2.2.1** It is possible to consider similar integrals on finite intervals of \((0, \infty)\):

\[ \int_a^b e^{-xt}f(t)dt. \]

However, observe that if \( f(t) \) is integrable on the real line, then

\[ \int_a^b e^{-xt}f(t)dt = \left( \int_a^\infty - \int_b^\infty \right) e^{-xt}f(t)dt, \]

and

\[ \left| \int_b^\infty e^{-xt}f(t)dt \right| \leq e^{-xb} \int_b^\infty |f(t)|dt. \]

If \( b > 0 \) then this integral is exponentially small, and negligible with respect to any expansion in inverse powers of \( x \). Therefore, the integral from \( a \) to \( b \) and the integral from \( a \) to \( \infty \) are equivalent in this asymptotic sense, and that is the reason why the upper limit of integration will not play any significant role in the following sections.
2.2.3 Laplace’s method

More generally, we can study integrals of the form

\[ F(x) = \int_a^b e^{-x \phi(t)} f(t) dt, \]  

(2.43)

where \( f(t) \) and \( \phi(t) \) are regular enough. In order to obtain asymptotic approximations for \( F(x) \) when \( x \) is large, the intuitive idea is that the main contribution to an integral like (2.43) is given by the neighbourhood of those points where the phase function \( \phi(t) \) attains a local minimum. In the simplified case (2.40), this corresponds to \( t = 0 \), so we recover the analysis presented before.

Let us suppose that \( R(t) \) attains a maximum value at \( t = t^\ast \) inside \((a, b)\), then we write

\[ F(x) = \int_a^b e^{-x \phi(t)} f(t) dt = e^{-x \phi(t^\ast)} \int_a^b e^{-x(\phi(t) - \phi(t^\ast))} f(t) dt. \]

Note that if \( U_\delta(t^\ast) \) is a small neighbourhood of the point \( t^\ast \), then for \( t \in (a, b) \setminus U_\delta(t^\ast) \) we have

\[ \phi(t) - \phi(t^\ast) \geq K_\delta > 0. \]

Hence, for \( t \in [a, b] \setminus U_\delta(t^\ast) \)

\[ \left| \int_{[a, b] \setminus U_\delta(t^\ast)} e^{-x(\phi(t) - \phi(t^\ast))} f(t) dt \right| \leq e^{-xK_\delta} \int_a^b |f(t)| dt, \]

and this contribution is exponentially small (since \( K_\delta > 0 \)), and therefore negligible with respect to a series in inverse powers of \( x \), like the one obtained through Watson’s lemma. An alternative derivation of this result can be found in [3], with the use of cut-off functions.

If the minimum of \( \phi(t) \) is attained at one of the endpoints (for instance at \( t = a \)), then we can make the change of variables

\[ s = \phi(t) - \phi(a), \]  

(2.44)

so that now the minimum is located at \( s = 0 \). Thus

\[ \int_a^b e^{-x \phi(t)} f(t) dt = e^{-x \phi(a)} \int_0^{\phi(b) - \phi(a)} e^{-xs} f(\phi^{-1}(s + \phi(a))) \frac{dt}{ds} ds, \]
where we suppose that $\phi'(a) > 0$, so that the function $\phi(t)$ is locally invertible near the endpoint. The idea now is to expand the function

$$g(s) = f(\phi^{-1}(s + \phi(a))) \frac{dt}{ds}$$

in powers of $s$ and to extend the integral to $\infty$ using the remark that we presented before. This may not be immediate, since the change of variables (2.44) may not be easily (or explicitly) invertible, but it can usually be accomplished using Lagrange’s inversion theorem or similar results. Note that if $g(s)$ is analytic at the origin, then

$$g(s) = \sum_{k=0}^{\infty} a_k s^k,$$

with

$$a_0 = g(0) = f(\phi^{-1}(s + \phi(a))) \frac{dt}{ds} \bigg|_{s=0} = \frac{f(a)}{\phi'(a)},$$

since

$$\frac{dt}{ds} = \frac{1}{\phi'(t)},$$

using (2.44). Similarly, $a_1 = g'(0)$, which can be computed using the chain rule. The first two terms are therefore

$$\int_0^{\phi(b) - \phi(a)} e^{-xs} f(s) \frac{dt}{ds} ds = \frac{f(a)}{\phi'(a)x} + \frac{f'(a)\phi'(a) - f(a)\phi''(a)}{[\phi'(a)]^2 x^2} + \mathcal{O}(x^{-3}),$$

see also [42].

Observe that the endpoint $\phi(b) - \phi(a)$ does not play any role in the expansion. This is again a consequence of the fact that, away from the local minima of $\phi(t)$, the contributions to the integral are exponentially small. Therefore, we can extend the integral from 0 to $\infty$ (provided that the integral is well defined). Hence

$$\int_0^{\infty} e^{-xs} f(s) \frac{dt}{ds} ds = \frac{f(a)}{\phi'(a)x} + \mathcal{O}(x^{-2}).$$

If there is one local minimum inside, $t^* > 0$, then one can split the integral:

$$F(x) = \int_a^b f(t)e^{-x\phi(t)} dt = \left( \int_a^{t^*} + \int_{t^*}^b \right) f(t)e^{-x\phi(t)} dt,$$

46
use the change of variable $\phi(t) - \phi(t^*) = s$ and apply Watson’s lemma. Alternatively, we can write

$$\phi(t) - \phi(t^*) = \frac{1}{2} s^2,$$

with the condition $\text{sgn} \ s = \text{sgn}(t - t^*)$. This maps $t = t^*$ to $s = 0$, and thus

$$F(x) = e^{-x\phi(t^*)} \int_{-\sqrt{2(\phi(b) - \phi(t^*))}}^{\sqrt{2(\phi(a) - \phi(t^*))}} f(s) \frac{dt}{ds} e^{-\frac{1}{2}xs^2} ds.$$

Now we extend the interval of integration to the whole real line, with an exponentially small error, and apply a similar method of integration term by term.

Now it is not difficult to show that an integral of the form

$$\int_{-\infty}^{\infty} s^{2n} e^{-\frac{1}{2}xs^2} ds = 2 \int_{0}^{\infty} s^{2n} e^{-\frac{1}{2}xs^2} ds, \quad n \geq 0$$

can be written in terms of the Gamma function. Indeed, making the change of variables $s = \sqrt{2t/x}$, we have

$$\int_{0}^{\infty} s^{2n} e^{-\frac{1}{2}xs^2} ds = \frac{1}{2} \left( \frac{2}{x} \right)^{n+\frac{1}{2}} \Gamma \left( n + \frac{1}{2} \right),$$

so

$$\int_{-\infty}^{\infty} s^{2n} e^{-\frac{1}{2}xs^2} ds = \left( \frac{2}{x} \right)^{n+\frac{1}{2}} \Gamma \left( n + \frac{1}{2} \right).$$

As an example, we consider the Euler Gamma function:

$$\Gamma(x + 1) = \int_{0}^{\infty} t^x e^{-t} dt, \quad \text{Re} \ x > -1,$$

and suppose that we want an asymptotic estimate for large $x$. If we write $t = xs$, we obtain the integral

$$\Gamma(x + 1) = x^{x+1} \int_{0}^{\infty} e^{-x\phi(s)} ds, \quad \phi(s) = s - \log s.$$

The function $\phi(s)$ has a minimum at $s^* = 1$, and we make the change of variable

$$\phi(s) - \phi(1) = s - \log s - 1 = \frac{1}{2} u^2,$$
with the condition \( \text{sgn} \, u = \text{sgn}(s - 1) \). Then

\[
\Gamma(x + 1) = x^{x+1} \left( e^{-x} \int_{-\infty}^{\infty} e^{-\frac{1}{2} xu^2} \frac{ds}{du} du \right).
\]

The problem is that this change of variable is not easily invertible and we need an expansion of the form

\[
\frac{ds}{du} = \sum_{n=0}^{\infty} a_n u^n.
\]

Note that if we differentiate the change of variable, we get

\[
\frac{du}{ds} = 1 - \frac{1}{s} \Rightarrow \frac{ds}{du} = \frac{s}{su} = \frac{s - 1}{su}.
\]

Taking limit when \( u \to 0 \) (or \( s \to 1 \)) we obtain that

\[
\left. \frac{ds}{du} \right|_{u=0} = 1.
\]

Higher order terms can be computed, although things become a bit cumbersome. Another possibility is the following:

\[
\frac{1}{2} u^2 = s - \log s - 1 = \frac{(s - 1)^2}{2} - \frac{(s - 1)^3}{3} + O(s - 1)^4.
\]

so for small \( u \) (and hence small \( s - 1 \)):

\[
u = (s - 1) - \frac{1}{3}(s - 1)^2 + \frac{7}{36}(s - 1)^3 + O(s - 1)^4.
\]

Applying Lagrange’s inversion theorem, we obtain

\[
s = 1 + u + \frac{1}{3} u^2 + \frac{1}{36} u^3 + O(u^4),
\]

so

\[
\frac{ds}{du} = 1 + \frac{2}{3} u + \frac{1}{12} u^2 + O(u^3),
\]

and therefore

\[
\Gamma(x + 1) = \sqrt{2\pi} x^{x+1/2} e^{-x} \left( 1 + \frac{1}{12x} + O(x^{-2}) \right)
\]

when \( x \to \infty \), which is Stirling’s approximation of the Gamma function, see [2].
A very important observation is that the contribution from an interior critical point and from an endpoint (in the absence of critical points) are of different order, namely $O(x^{-1})$ and $O(x^{-1/2})$ respectively. Similarly, one can prove that in the presence of a critical point of higher order, namely $t^*$ such that

$$\phi'(t^*) = \phi''(t^*) = \ldots = \phi^{(2k-1)}(t^*) = 0, \quad \phi^{(2k)}(t^*) \neq 0,$$

with $k \geq 1$, then the leading term of the asymptotic expansion is $O(x^{-\frac{1}{2}k})$. This mismatch is one of the motivations for the development of uniform asymptotic expansions, see the remark at the end of the chapter.

2.2.4 The method of stationary phase

The method of stationary phase is usually applied to integrals of the form

$$F(x) = \int_a^b e^{i\lambda\phi(t)} f(t) dt,$$

(2.45)

where $\phi(t)$ is real and $\lambda$ is a large parameter. Observe that from a numerical point of view, such an integral represents a considerable challenge, because it is highly oscillatory and it can be very expensive to evaluate by standard quadrature.

The main idea behind the method of stationary phase is that the integrand in (2.45) will be oscillatory in $[a, b]$ except near those points $t^*$ where $\phi'(t^*) = 0$, where the function does not oscillate (locally). The intuitive idea is that away from $t^*$ and the endpoints, the positive and negative parts of the integrand will almost cancel each other, and that the main contributions to the integral will come from neighbourhoods of these distinguished points. If $\phi'(t^*) = 0$ we say that $t^*$ is a stationary point, and hence the name of the method.

First suppose that $\phi'(t)$ does not vanish and $f(t)$ is regular enough, then integration by parts gives

$$F(\lambda) = \int_a^b e^{i\lambda\phi(t)} f(t) dt = \int_a^b e^{i\lambda\phi(t)} \frac{f(t)}{i\lambda\phi'(t)} d(e^{i\lambda\phi(t)})$$

$$= \frac{1}{i\lambda} \left[ e^{i\lambda\phi(b)} \frac{f(b)}{\phi'(b)} - e^{i\lambda\phi(a)} \frac{f(a)}{\phi'(a)} \right] + O(\lambda^{-2}).$$

It is clear that one can compute more terms in this expansion under suitable assumptions on $f(t)$. See for instance [31]. We observe that
the endpoints will play an important role in this analysis and also that
$I(x) = \mathcal{O}(\lambda^{-1})$, something which is known (in a more general setting) as
the Riemann–Lebesgue lemma.

We mention in passing that this approach has proved useful in recent
years in the development of efficient numerical methods for oscillatory inte-
grals. Namely, if we interpolate $f(x)$ (and its derivatives) by a polynomial
$p(t)$ in a Hermite sense, then
\[
\int_{a}^{b} e^{i\lambda\phi(t)} f(t) \, dt = \int_{a}^{b} e^{i\lambda\phi(t)} p(t) \, dt + \int_{a}^{b} e^{i\lambda\phi(t)} [f(t) - p(t)] \, dt. \tag{2.46}
\]

Provided that the first of the last two integrals can be computed explicitly,
then the second one has a higher order behavior in terms of $\omega$ because of
the matching between $f(t)$ and $p(t)$ at the endpoints. This is the philosophy
of Filon–type methods.

Suppose that we have a stationary point of order 1, that is, a point
t* ∈ [a, b] such that $\phi'(t^*) = 0$ and $\phi''(t^*) > 0$ (local minimum), then we can
make the change of variables
\[
\phi(t) - \phi(t^*) = -\frac{1}{2} \phi''(t^*) s^2,
\]
again with the condition $\text{sgn } s = \text{sgn } (t - t^*)$. Similarly, if $\phi''(t^*) < 0$ (local
maximum), then we can make the change of variables

\[ \phi(t) - \phi(t^*) = -\frac{1}{2} \phi''(t^*) s^2, \]

so in general we have

\[ \phi(t) - \phi(t^*) = \frac{1}{2} \sigma \phi''(t^*) s^2, \]

where \( \sigma = \text{sgn} \phi''(t^*) \). In this way, the local contribution from a neighbourhood of the stationary point is equal to

\[ F(\lambda) = e^{i\lambda \phi(t^*)} \int_{-\infty}^{\infty} e^{\frac{\pm i \alpha s^2}{2}} f(s) ds, \]

Now the integral

\[ \int_{-\infty}^{\infty} e^{\pm i \alpha s^2} ds = e^{\pm \pi i/4} \sqrt{\frac{\pi}{\alpha}} \]

exists as an improper Riemann integral (and it can be computed in terms of the error function). Therefore, extending the limits of integration again, we obtain

\[ \int_{a}^{b} e^{i\lambda \phi(t)} f(t) dt = f(t^*) e^{i \lambda \phi(t^*)} \left( \frac{2\pi}{\lambda \phi''(t^*)} \right)^{1/2} + O(\lambda^{-1}), \]

We note that in this case the contribution of the stationary point is of a larger order of magnitude than the one from an endpoint, compare \( O(x^{-1/2}) \) and \( O(x^{-1}) \). Even more important, unlike what happened in the analysis of Laplace integrals, in the case of Fourier integrals the contributions from the regions away from the stationary points are only algebraically small (and not exponentially small) with respect to the one coming from the stationary point. See a more detailed analysis in [42].

Similar results can be obtained when we have a stationary point of higher order (that is, when \( \phi'(t^*) = \phi''(t^*) = \ldots = \phi^{(k)}(t^*) = 0 \) but \( \phi^{(k+1)}(t^*) \neq 0 \)). In that case, we write

\[ \phi(t) - \phi(t^*) = \frac{1}{(k+1)!} \sigma |\phi^{(k+1)}(t^*)| s^{k+1}, \]

where now \( \sigma = \text{sgn} \phi^{(k+1)}(t^*) \). Now

\[ \int_{-\infty}^{\infty} e^{\pm i \alpha s^{k+1}} ds = e^{\pm \pi i/4} \left( \frac{\pi}{\alpha} \right)^{\frac{1}{k+1}}, \]

so the local contribution from the stationary point is of order \( O(x^{-\frac{1}{k+1}}) \).
2.2.5 The method of steepest descent

The method of steepest descent can be seen as a generalization of Laplace’s method, devised to analyse the asymptotic behaviour of integrals of the form

\[ \int_{C} e^{\lambda \phi(z)} f(z) dz, \quad \lambda \to \infty \tag{2.47} \]

where \( C \) is a suitable contour in the complex plane. We will suppose that \( \lambda \) is real and that \( f(z) \) and \( \phi(z) \) are analytic in suitable domains in order to justify the deformations of the contour that are presented below.

We first observe that if we write \( \phi(z) = R(z) + iI(z) \), we have

\[ \int_{C} e^{\lambda [R(z)+iI(z)]} f(z) dz, \quad \lambda \to \infty \tag{2.48} \]

The imaginary part of the phase function \( I(z) \) introduces fast oscillations in the integrand when \( \lambda \) is large. This constitutes a problem both from an asymptotic and from a numerical point of view, since the integral is difficult and expensive to evaluate, and the dominant asymptotic behaviour is not clear.

The first idea of the method (subject to analyticity of the integrand) is to deform the contour of integration in such a way that the imaginary part of the phase function is constant along the new contour. That is:

\[ \int_{C} e^{\lambda [R(z)+iI(z)]} f(z) dz = e^{i\lambda I_0} \int_{\tilde{C}} e^{\lambda R(z)} f(z) dz, \tag{2.49} \]

where \( I_0 = I(z_0) \) is the value of the imaginary part at a suitable chosen point. This is determined by finding the points in the complex plane that will give the main contributions to (2.49). These are the endpoints of \( \tilde{C} \) (if the contour is finite), the singularities of \( f(z) \) and those points where the real part of \( \phi(z) \) attains a maximum along the level curve \( I(z) = I_0 \).

Let us write \( z = x + iy \). Using the method of Lagrange multipliers for this problem of constrained extrema, we know that if \( R(x, y) \) has an extremum at a point \( z_0 = x + iy_0 \), subject to \( I(x, y) = I_0 \), then there exists a real number \( \mu \) such that

\[ \nabla R(x_0, y_0) = \mu \nabla I(x_0, y_0) \tag{2.50} \]

However, we know that the gradient \( \nabla I(x, y) \) is orthogonal to the level curve \( I(x, y) = I_0 \). Using Cauchy-Riemann equations, we get:

\[ \nabla I(x, y) = \left( \frac{\partial I}{\partial x}(x, y), \frac{\partial I}{\partial y}(x, y) \right) = \left( -\frac{\partial R}{\partial y}(x, y), \frac{\partial R}{\partial x}(x, y) \right), \]

where \( R(x, y) \) and \( I(x, y) \) are real functions of complex variables.
which is clearly orthogonal to $\nabla R(x, y)$ at each point of the level curve. This has two crucial consequences:

- The only possibility to satisfy (2.50) is that $\nabla R(x_0, y_0) = \nabla I(x_0, y_0) = (0, 0)$. This is seen to be equivalent to the condition $\phi'(z) = 0$ (taking this derivative with respect to $z$). This gives a very convenient method to compute the relevant critical points.

- At any point on the level curve $I(x, y) = I_0$, the tangent vector to that curve has the direction of $\pm \nabla R(x, y)$, and therefore it follows that $R(x, y)$ changes as fast as possible along that curve. This explains the name of method of steepest descent. A direction where $R(x, y)$ decreases as fast as possible is very advantageous for numerical quadrature. Naturally, there will be directions of steepest ascent as well, along which the real part of the phase function grows, and it is important to distinguish both cases.

Moreover, since $R(x, y)$ and $I(x, y)$ are harmonic functions (because they are the real and imaginary parts of an analytic function), then a critical point $z_0 = x_0 + iy_0$ cannot be a local maximum or a minimum, because in this case the real and the imaginary part of $\phi(z)$ would be constant. Instead, $(x_0, y_0)$ must be a saddle point, and this is the reason why this method is sometimes referred to as the saddle point method.

In the neighbourhood of a saddle point, it is possible to describe the directions of steepest descent, steepest ascent and constant real part. More explicitly, if $\phi'(z_0) = \phi''(z_0) = \ldots = \phi^{(n-1)}(z_0) = 0$ but $\phi^{(n)}(z_0) = ae^{i\alpha}$, with $a > 0$, then the directions of steepest descent are

$$\theta = -\frac{\alpha}{n} + (2p + 1)\frac{\pi}{n}, \quad p = 0, 1, \ldots n - 1,$$

the directions of steepest ascent are

$$\theta = -\frac{\alpha}{n} + 2p\frac{\pi}{n}, \quad p = 0, 1, \ldots n - 1,$$

and the directions of constant real part are

$$\theta = -\frac{\alpha}{n} + \left(p + \frac{1}{2}\right)\frac{\pi}{n}, \quad p = 0, 1, \ldots 2n - 1.$$

For a proof see [3, Chapter 7] and also [42].
Once these transformations have been carried out, the path of steepest
descent can be parametrized, say by $z = z(t)$, and then we have
\[
e^{i\lambda t_0} \int_a^b e^{\lambda R(z(t))} f(z(t))z'(t)dt.
\] (2.51)

As an example to illustrate the method and also the connection with
the method of stationary phase, we consider the following integral:
\[
I[f] = \int_a^b f(t)e^{i\lambda t^2} dt, \quad \lambda \to \infty
\] (2.52)
where $a < 0 < b$ and we suppose that $f(t)$ can be extended to an analytic
function $f(z)$ in $\mathbb{C}$. In this case $\phi(t) = it^2$, so the critical point is given by
$\phi'(t^*) = 0$, which clearly yields $t^* = 0$. Moreover $\phi''(0) = 2i$, so $a = 2$ and
$\alpha = \pi/2$ in the notation that we have used before. Hence, near this critical
point we have the following direction of steepest descent:
\[
\theta = \frac{\pi}{4}, \frac{5\pi}{4}.
\]
The directions of steepest ascent are
\[
\theta = -\frac{\pi}{4}, \frac{3\pi}{4},
\] and the directions of constant real part are
\[
\theta = 0, \pm \frac{\pi}{2}, \pi.
\]
See Figure 2.3 for a graphical illustration.

Hence, a reasonable deformation of the contour, invoking Cauchy’s the-
orem, is the following
\[
I[f] = \left( \int_{\Gamma_a} + \int_{\Gamma_0^+} + \int_{\Gamma_0^-} + \int_{\Gamma_b} \right) f(z)e^{i\lambda z^2} dz,
\] (2.53)
where the contours are given approximately in Figure 2.4.

What are exactly the paths of steepest descent? If we write $z = x + iy$, then
\[
\text{Re} \phi(z) = -2xy, \quad \text{Im} \phi(z) = x^2 - y^2,
\] (2.54)
so if we set $\text{Im} \phi(z) = \text{Im} \phi(a) = a^2$, we have
\[
x^2 - y^2 = a^2 \Rightarrow y = \pm \sqrt{x^2 - a^2}.
\] (2.55)
Figure 2.3: Level curves of the real part of $\phi(z) = iz^2$.

Figure 2.4: Path deformation for the integral (2.52).
We take the minus sign of the root, bearing in mind the structure of 
Re \( \phi(z) \). Similarly, we have \( y = \sqrt{x^2 - b^2} \) from \( z = b \). From the critical 
point \( z = 0 \) we take \( y = -x \) for the path \( \Gamma_0^- \) and \( y = x \) for the path \( \Gamma_0^+ \).

For instance, at the right endpoint \( z = b \), we have \( z = b \) \( \implies \)
\( x = \sqrt{x^2 - b^2} \), therefore
\[
\int_{\Gamma_b} f(z)e^{i\lambda z^2} \, dz = e^{i\lambda b^2} \int_b^\infty (x+i\sqrt{x^2 - b^2})e^{-2\lambda x\sqrt{x^2 - b^2}} \left( 1 + \frac{i x}{\sqrt{x^2 - b^2}} \right) \, dx.
\]  

(2.56)

Now the function \( x\sqrt{x^2 - b^2} \) is monotonic for \( x \geq b \), so we can make the 
change of variables
\[
2x\sqrt{x^2 - b^2} = \tau,
\]
and then
\[
\tau = -(2b)^{3/2}\sqrt{x - b} + O((x - b)^{3/2}).
\]  

(2.58)

From here, using perturbation theory, we conclude that
\[
x = b + \frac{1}{8b^3} \tau^2 - \frac{5}{128b^7} \tau^4 + O(\tau^6).
\]  

(2.59)

Finally,
\[
\left( 1 + \frac{i x}{\sqrt{x^2 - b^2}} \right) \frac{dx}{d\tau} = \frac{i}{2b} + O(\tau),
\]

(2.60)

hence an application of Watson’s lemma gives that the contribution of the 
right endpoint is
\[
\int_{\Gamma_b} f(z)e^{i\lambda z^2} \, dz = ie^{i\lambda b^2} f(b) \frac{1}{2b} + O(\lambda^{-2}).
\]  

(2.61)

Similarly, one can work out the contribution from the other endpoint and 
from the stationary point \( z = 0 \). Namely, we have \( z = x + iy(x) = x + ix \), and we can combine the integrals over \( \Gamma_0^- \) and \( \Gamma_0^+ \) and integrate in \( x \) on the 
whole real line:
\[
\left( \int_{\Gamma_0^-} + \int_{\Gamma_0^+} \right) f(z)e^{i\lambda z^2} \, dz
\]
\[
= (1 + i) \int_{-\infty}^\infty f(x + ix)e^{-2\lambda x^2} \, dx = e^{\pi i} f(0) \frac{\sqrt{\pi}}{\sqrt{\lambda}} + O(\lambda^{-1}),
\]

(2.62)

which is consistent with the approximation given by the method of stationary phase.
Figure 2.5: Level curves of the real part of $\phi(z) = iz^3$.

A similar analysis is possible for higher stationary points: for example, for $\phi(z) = iz^3$, the structure of $\text{Re}\,\phi(z)$ is given in Figure 2.5. In this case, the directions of steepest descent from the origin would be $\theta = \pi/6$, $\theta = 5\pi/6$ and $\theta = -\pi/2$.

As another important example, we take the integral representation for the Airy function given before:

$$\text{Ai}(z) = \frac{1}{2\pi i} \int_{C_1} e^{\frac{2}{3}z^3 - zt} dt,$$

where $C_1$ is a contour that starts with $\text{arg}\,t = -\pi/3$ and ends with $\text{arg}\,t = \pi/3$. We suppose that $z$ is real, and we are interested in an asymptotic expansion for large values of $z$. First, we make a change of variables $t = \sqrt{z}w$, so that

$$\text{Ai}(z) = \frac{1}{2\pi i} \int_{C_1} e^{\frac{3}{2}z - z^3} dt = \frac{\lambda^{1/3}}{2\pi i} \int_{C_1} e^{\lambda(w^3 - w)} dw,$$

where $\lambda = z^{3/2}$, hence large $z$ implies large $\lambda$. Now, in this case

$$f(w) = 1, \quad \phi(w) = \frac{w^3}{3} - w,$$

and solving for the critical points of $\phi(w)$, we find $w_\pm = \pm 1$. The new path of integration should run through the critical point $w = 1$. 

57
Figure 2.6: Level curves of $\text{Re} \phi(w)$ for the Airy integral (left) and saddle points at $w = 1$ and at $w = -1$ (right).

The directions of steepest descent from $w = 1$ are $\theta = \pm \pi/2$, and those of steepest ascent are given by $\theta = 0, \pi$. For the other saddle point, $w = -1$, the directions of steepest descent are $\theta = 0, \pi$, and those of steepest ascent are given by $\theta = \pm \pi/2$.

For real $z$, the original path can be deformed to a different one running through $w = 1$, using Cauchy’s theorem, and such that the imaginary part of the phase function is constant.

**Exercise 2.2.2** Complete the asymptotic analysis of the Airy function $A_i(z)$ from the previous integral. A possible way to proceed is to follow these steps:

- Write the phase function $\phi(w)$ as $\phi(u + iv)$, and derive explicitly $\text{Re} \phi(u + iv)$ and $\text{Im} \phi(u + iv)$.
- Compute the equation of the path of steepest descent $u = u(v)$ through the point $w = 1$.
- Write the integral on the real axis in terms of the variable $v$, that is
  \[
  \frac{\lambda^{1/3}}{2\pi} \int_{-\infty}^{\infty} e^{\lambda \phi(v)} dv.
  \]
- Check that
  \[
  \phi(v) = \sqrt{\frac{v^2}{3} + 1} \left( -\frac{8}{9} v^2 - \frac{2}{3} \right),
  \]
verify that \( \phi(v) \) has a maximum at \( v = 0 \) and make the change of variables

\[
\phi(v) - \phi(0) = -\frac{1}{2} s^2.
\]

to write it as a standard Laplace integral.

• Apply Watson’s lemma and deduce that

\[
Ai(z) \sim \frac{1}{2\sqrt{\pi}} z^{-1/4} e^{-\frac{2}{3} z^{3/2}}, \quad z \to \infty.
\]

The analysis for complex values of \( z \) is more involved, but very instructive, see for example [20, 42]. It is particularly interesting that for some values of \( \arg z \), one needs to take into account both saddle points. This is reasonable if the imaginary part of the phase function is equal at both points.

If \( z = re^{i\alpha} \), then

\[
\phi(w) = \frac{w^3}{3} - we^{i\alpha},
\]

and we denote the saddle points by \( w_\pm = \pm e^{i\alpha/2} \). The directions of steepest descent from \( w_+ \) are given by \( \theta = \pm \pi/2 - \alpha/4 \), and those of steepest ascent are \( \theta = -\alpha/4, \pi - \alpha/4 \). The directions of steepest descent from \( w_- \) are given by \( \theta = \pm \pi/2 + \alpha/4 \), and those of steepest ascent are \( \theta = \alpha/4, \pi + \alpha/4 \).

Moreover, one can check that

\[
\text{Re} \, \phi(w_+) = -\frac{2}{3} \cos \frac{3\alpha}{2}, \quad \text{Im} \, \phi(w_+) = -\frac{2}{3} \sin \frac{3\alpha}{2},
\]

and

\[
\text{Re} \, \phi(w_-) = \frac{2}{3} \cos \frac{3\alpha}{2}, \quad \text{Im} \, \phi(w_-) = \frac{2}{3} \sin \frac{3\alpha}{2}.
\]

The imaginary part of the phase function at both saddle points is equal when \( \alpha = 0 \) and when \( \alpha = \pm 2\pi/3 \). When \( 0 < \alpha < 2\pi/3 \), only one of the saddle points is needed in order to give the asymptotic approximation of the Airy function. However, when \( 2\pi/3 \leq \alpha < \pi \), one needs to take into account both \( w_+ \) and \( w_- \). See the details in [42, §4.7].

Another remarkable feature of this analysis is that, whereas the Airy function \( Ai(x) \) is entire (analytic in the whole complex plane), the asymptotic estimate that we have derived before has a branch cut because of the factor \( z^{-1/4} \). It follows that the asymptotic behaviour of \( Ai(z) \) is different when \( |\arg z| < \pi \) and when \( |\arg z| = \pi \), and this is a simple example of
the so-called *Stokes phenomenon*, which is very important in asymptotic analysis, see for instance [44, 42].

The precise formulation of the method of steepest descent in the complex plane is quite delicate, see for example [44]. See also [3, 50, 54] for many more examples.

### 2.3 Related problems

The asymptotic analysis of special functions is a huge area of research, and here we have just presented the basic idea of two different methods. Other topics of study are the following:

- **Uniform asymptotic expansions.** When several parameters are involved in the analysis of a certain function, it may be relevant to allow not only one but several to become large. The fact that certain asymptotic expansions are sensitive to external parameters taking large or small values is a sign of non-uniformity. For example, the incomplete gamma function has the integral representation

\[
\Gamma(a + 1, z) = \int_z^\infty t^a e^{-t}dt, \quad |\arg z| < \pi,
\]

or if \( t = zs \):

\[
\Gamma(a + 1, z) = z^{a+1} \int_1^\infty s^a e^{-zs}ds = z^{a+1} e^{-z} \int_0^\infty (s+1)^a e^{-zs}ds,
\]

and standard application of Watson’s lemma yields a well known asymptotic expansion for large values of \( z \):

\[
\Gamma(a + 1, z) = z^{a} e^{-z} \left[ 1 + \frac{a}{z} + \frac{a(a-1)}{z^2} + \frac{a(a-1)(a-2)}{z^3} + \ldots \right],
\]

when \( |\arg z| < \pi/2 \), see [2, Eq.6.5.32]. However, if \( a \) is large as well (for instance if \( a \approx z \), then this asymptotic expansion is no longer useful. An alternative is to write

\[
\Gamma(a + 1, z) = z^{a+1} e^{-z} \int_0^\infty (s+1)^a e^{-zs}ds = z^{a+1} e^{-z} \int_0^\infty e^{-z\phi(s)}ds,
\]

where

\[
\phi(s) = s - \lambda \log(1 + s), \quad \lambda = \frac{a}{z}.
\]

This function \( \phi(s) \) has a minimum at \( s^* = \lambda - 1 \).
– When $\lambda \ll 1$ (i.e. when $z \gg a$) then $s^* < 0$, i.e. the critical point is outside $[0, \infty)$ and we can apply Watson’s lemma as before.

– When $\lambda \gg 1$ (i.e. when $a \gg z$) then $s^* > 0$, i.e. the critical point is in $[0, \infty)$ and therefore should be taken into account.

– When $\lambda \sim 1$ (i.e. when $a \sim z$) then the critical point coalesces with the endpoint $s = 0$.

In order to derive asymptotic expansions that hold uniformly with respect to external parameters, such as $a$ in this last example, more complicated special functions are needed as approximants. In this example, the coalescence of a critical point and an endpoint of the interval of integration requires the use of the complementary error function. Similarly, when two saddle points of the integrand coalesce, then Airy functions are used, see for example [3, Chapter 9].

• Other approaches based on recurrence relations or generating functions can be also used. We will see an example in the chapter on recurrence relations.

• The method of stationary phase and the method of steepest descent have proved to be extremely useful in recent approaches to the asymptotic and numerical analysis of oscillatory integrals, see for instance [29, 30] for recent accounts. In the field of oscillatory integrals, this has represented a change of paradigm in the last few years, which is leading to highly efficient numerical methods for this kind of problems.

• Riemann–Hilbert techniques provide a very powerful method to derive the asymptotic behaviour of special functions in the complex plane. An analogue of the classical method of steepest descent was introduced in the seminal paper [11], and since then this method has been successfully applied to classical and multiple orthogonal polynomials, see for example [36, 33, 34, 32], and also to Painlevé transcendents, where integral representations are no longer available for the solutions, see a recent and comprehensive account in [14].
Chapter 3

Real zeros of hypergeometric functions

In this chapter we will present several results on real zeros of classical orthogonal polynomials and special functions. The analysis in the complex case is considerably more involved, and much less information is available in general, see for instance [27].

We will be interested in the following properties:

- Bounds on the distances between consecutive (real) zeros of a function.
- Monotonicity of distances between consecutive (real) zeros of a function.
- Interlacing of zeros of functions which are contiguous.

The basic tools that we will use are the Sturm comparison theorem and related results, as well as the fact that classical orthogonal polynomials (and more in general, Gauss and Kummer hypergeometric functions) are solutions of second order linear ordinary differential equations in the variable \( x \).

3.1 Real zeros of solutions of second order ODEs

As explained before, most classical special functions are solutions of second order linear ODEs of the form:

\[
w''(x) + f(x)w'(x) + g(x)w(x) = 0, \quad (3.1)
\]

where \( A(x) \) and \( B(x) \) are analytic except for the singular points of the equation.
The classical theory of zeros of solutions of second order differential equations is usually stated in terms of equations in normal form:

\[ y''(x) + p(x)y(x) = 0, \quad (3.2) \]

where

\[ p(x) = g(x) - \frac{1}{2}f'(x) - \frac{1}{4}f^2(x). \]

Throughout this chapter, we will suppose that \( p(x) \) is continuous, except at the singular points of the differential equation. Observe that we have changed notation from the previous chapter, our function \( p(x) \) equals \( -q(x) \) before. This is just a matter of clarity when stating the Sturm theorem later on.

The properties of the real zeros of any solution \( y(x) \) of this type of equation will naturally depend on the function \( p(x) \). We first present some classical results:

**Theorem 3.1.1** Let \( y(x) \) be any non trivial solution of (3.2). If \( p(x) < 0 \) in \((a, b)\) then \( y(x) \) can have at most one zero in \((a, b)\).

**Proof 3.1.1** It is left as an exercise. Hint: suppose that \( y(x) \) has two different zeros \( x_0 < x_1 \) in \((a, b)\) and use Rolle’s theorem.

It is clear that this case is not very interesting in our setting, and for this reason in the sequel we will always suppose that we are in a situation where any solution of the differential equation has at least two zeros. It is worth noting that when there are parameters involved (as happens in the hypergeometric case), we will need to impose conditions on them to ensure that we have at least two real zeros in a certain interval. We will refer to these as oscillatory conditions.

It is more complicated to state conditions under which the solutions of the differential equation do oscillate. Of course, this is clear in some special cases, such as classical orthogonal polynomials, where the number and location of the zeros is very well known, see [49, 4]. In general, observe that if \( p(x) > 0 \) when \( x > x_0 \), say, then \( y''(x) = -p(x)y(x) \), so \( y''(x) \) and \( y(x) \) have opposite signs. Suppose, without loss of generality, that \( y(x) > 0 \) when \( x > x_0 \), then \( y''(x) < 0 \) and \( y'(x) \) is decreasing. If \( y'(x) \) becomes negative, then \( y(x) \) will have at least one zero on the right of \( x_0 \). However, it may happen that \( y'(x) \) decreases steadily but remains positive, so that \( y(x) \) is always increasing and it will not have any zeros for \( x > x_0 \).

A sufficient condition for oscillation is the following, see [48, §24]:
Theorem 3.1.2 Let \( y(x) \) be any non-trivial solution of (3.2), and suppose that \( p(x) > 0 \) for \( x > 0 \), if
\[
\int_1^\infty p(x)dx = \infty,
\]
then \( y(x) \) has an infinite number of zeros in the positive semiaxis.

Proof 3.1.2 We proceed by contradiction. Suppose that there exists a point \( x_0 > 1 \) such that \( y(x) \) has constant sign for \( x > x_0 \). Without loss of generality, we suppose that \( y(x) > 0 \) for \( x > x_0 \). If we define the function
\[
v(x) = -\frac{y'(x)}{y(x)},
\]
then one can verify that \( v(x) \) is a solution of the Riccati equation
\[
v'(x) = p(x) + v(x)^2.
\]
Then, integrating this expression from \( x = x_0 \) to \( x \),
\[
v(x) - v(x_0) = \int_{x_0}^x p(t)dt + \int_{x_0}^x v(t)^2dt,
\]
hence for large enough \( x \) we have that \( v(x) > 0 \), since the first integral diverges. If \( v(x) > 0 \) then \( y(x) \) and \( y'(x) \) have opposite signs, hence if \( y(x) > 0 \) then \( y'(x) < 0 \) and \( y(x) \) will have a zero for \( x > x_0 \). This contradiction proves that \( y(x) \) has an infinite number of zeros when \( x > x_0 \).

Exercise 3.1.1 Apply the previous results to the case where \( y(x) \) is any non-trivial solution of
- the Airy differential equation:
  \[
y''(x) - xy(x) = 0,
  \]
- the Weber differential equation:
  \[
y''(x) - \left(\frac{1}{4}x^2 + a\right)y(x) = 0.
  \]

The next result concerns separation and interlacing of zeros of independent solutions of the same ODE. This type of result is well known in the theory of orthogonal polynomials, but it also holds for any two independent solutions of a second order ODE.
Theorem 3.1.3 Let $y_1(x)$ and $y_2(x)$ be two independent solutions of (3.2). Then $y_1(x)$ and $y_2(x)$ do not have common zeros, and their zeros alternate, in the sense that between two zeros of $y_1(x)$ there is exactly one zero of $y_2(x)$ and vice versa.

Proof 3.1.3 We follow the lines in [48]. The fact that $y_1(x)$ and $y_2(x)$ have no common zeros follows immediately from the fact that the two solutions are independent and therefore their Wronskian is different from 0.

Let $x_1$ and $x_2$ be two consecutive zeros of $y_2(x)$, then the Wronskian at $x = x_1$ is $W(x_1) = y_1(x_1)y'_2(x_1)$, and similarly for $x = x_2$. Since the Wronskian does not vanish, it follows that $y'_2(x_1)$ and $y'_2(x_2)$ are different from 0 and have opposite signs. Now, since the Wronskian has constant sign, $y_1(x_1)$ and $y_1(x_2)$ have opposite signs too, and using Bolzano’s theorem we deduce that there at least one zero of $y_1(x)$ in $(x_1, x_2)$. Furthermore, this zero is unique, since otherwise reversing the argument $y_2(x)$ would have a zero between $x_1$ and $x_2$, which is a contradiction.

Another result on interlacing of zeros, which will be useful to compare functions depending on external parameters, is a consequence of Sturm theorem, that we present next.

3.2 Sturm comparison theorem

The main result that we will use in this chapter is the following classical theorem:

Theorem 3.2.1 (Sturm) Let $y''(x) + p(x)y(x) = 0$ be a second order differential equation in normal form, with $q(x)$ continuous in $(a, b)$. Let $x_k < x_{k+1} < \ldots$ denote consecutive zeros of $y(x)$ in $(a, b)$ in increasing order. Then

1. If there exists $p_M > 0$ such that $p(x) < p_M$ in $(a, b)$ then

$$\Delta x_k = x_{k+1} - x_k > \frac{\pi}{\sqrt{p_M}}.$$ 

2. If there exists $p_m > 0$ such that $p(x) > p_m$ in $(a, b)$ then

$$\Delta x_k = x_{k+1} - x_k < \frac{\pi}{\sqrt{p_m}}.$$
3. If \( p(x) \) is strictly increasing in \((a, b)\) then \( \Delta^2 x_k = x_{k+2} - 2x_{k+1} + x_k < 0 \).

4. If \( p(x) \) is strictly decreasing in \((a, b)\) then \( \Delta^2 x_k = x_{k+2} - 2x_{k+1} + x_k > 0 \).

Exercise 3.2.1 Give results about the real zeros of \( y(x) \), if \( y(x) \) is any non-trivial solution of

- the Airy differential equation:
  \[ y''(x) - xy(x) = 0, \]

- the Bessel differential equation:
  \[ x^2 y''(x) + xy'(x) + (x^2 - \nu^2)y(x) = 0. \]

There are several different formulations of this classical result in the literature. Note that in this version we are comparing the function \( p(x) \) with a constant in order to bound the distances between consecutive zeros. It is possible to compare two equations in normal form with two different functions \( p_1(x) \) and \( p_2(x) \). Namely, we have

Theorem 3.2.2 Suppose that \( y(x) \) and \( z(x) \) are nontrivial solutions of the equations

\[ y''(x) + p_1(x)y(x) = 0, \quad z''(x) + p_2(x)z(x) = 0, \]

where \( p_1(x) \) and \( p_2(x) \) are positive functions such that \( p_1(x) > p_2(x) \), then the function \( y(x) \) vanishes at least once between two consecutive zeros of \( z(x) \).

Proof 3.2.1 Let \( x = x_1 \) and \( x = x_2 \) be two consecutive zeros of \( z(x) \), and let us suppose that \( y(x) \) does not vanish in \( I = (x_1, x_2) \). Without loss of generality, we assume that both \( z(x) > 0 \) and \( y(x) > 0 \) in \( I \). If we consider the Wronskian of \( y(x) \) and \( z(x) \),

\[ W[y, z](x) = y(x)z'(x) - y'(x)z(x), \]

then differentiating with respect to \( x \) we obtain

\[ W[y, z]'(x) = y(x)z''(x) - y''(x)z(x) = (p_2(x) - p_1(x))y(x)z(x) > 0. \]
Integrating this function between \( x = x_1 \) and \( x = x_2 \), we get

\[
W(x_2) - W(x_1) > 0 \Rightarrow W(x_2) > W(x_1).
\]

However, \( W(x_2) = y(x_2)z'(x_2) \leq 0 \) and \( W(x_1) = y(x_1)z'(x_1) \geq 0 \) because of the hypothesis on \( z(x) \) and \( y(x) \), so we have a contradiction and \( y(x) \) must vanish in \( I \).

The idea behind this theorem is that the larger the function \( p(x) \) is, the faster the oscillations of the corresponding solution are. Observe that the theorem does not say that there is exactly one zero of \( y(x) \) between two consecutive zeros of \( z(x) \), in fact this is clearly false if we set for example \( p_1(x) = k^2 \) and \( p_2(x) = 4k^2 \). In order to deduce strict interlacing of zeros extra information is needed, for instance the total number of zeros of \( y(x) \) and \( z(x) \) in a given interval. This can be deduced in some important cases, such as classical orthogonal polynomials.

**Exercise 3.2.2** Consider the differential equations satisfied by Laguerre polynomials

\[
xy''(x) + (\alpha + 1 - x)y'(x) + ny(x) = 0,
\]

and prove that for \( x > 0 \) the zeros of \( L_n^{(\alpha)}(x) \) and \( L_{n+1}^{(\alpha)}(x) \) interlace strictly. Prove an analogous result for Hermite polynomials, that satisfy

\[
y''(x) - 2xy'(x) + 2ny(x) = 0.
\]

### 3.3 Liouville transformations revisited

The Liouville transformation that we saw before in the context of asymptotic analysis of ODEs is quite useful here as well. We recall that we make a change of independent variables \( z = z(x) \) and a change of dependent variables

\[
Y(z(x)) = \sqrt{z(x)}y(x) = \sqrt{z(x)} \exp \left( \frac{1}{2} \int x f(t)dt \right) w(x), \quad (3.3)
\]

where \( y(x) \) satisfies (3.2) and \( w(x) \) is a solution of (3.1). Then we have

\[
Y(z) + \Omega(z)Y(z) = 0, \quad (3.4)
\]

where

\[
\Omega(z) = \dot{x}(z)^2 q(z) + \sqrt{x(z)} \frac{d^2}{dz^2} \frac{1}{\sqrt{x(z)}}.
\]
Here dots denote derivatives with respect to the variable $z$. It is possible, and quite convenient, to write $\Omega(z)$ as a function of $x$, which results in

$$\Omega(z(x)) = \frac{1}{z'(x)^2} \left( q(x) + \frac{3d'(x)^2}{4d(x)^2} - \frac{d''(x)}{2d(x)} \right),$$

(3.5)

where $d(x) = z'(x)$.

It is clear that we have quite a lot of freedom to choose the Liouville transformation, provided that the manipulations are feasible. We also note two important points:

- The function $Y(x)$ and the function $w(x)$ have the same zeros if $f(x)$ is continuous, see (3.3), so an analysis based on (3.4) and (3.5) will give results on the real zeros of the original function.
- Since $\Omega'(x) = \dot{\Omega}(z)z'(x)$, the monotonicity properties will be the same in both variables provided that $z'(x) > 0$, so we can analyse the function $\Omega$ in the variable $x$, which turns out to be simpler.

Following the theory exposed in [7], a reasonable criterion that we will apply to determine suitable Liouville transformations is that the problem of solving $\Omega'(x) = 0$ is equivalent to solving a quadratic equation on the interval of interest for any values of the parameters involved. As we will see, this covers a wide range of different transformations (including several cases already studied in the literature), and at the same time it leads to reasonably simple computations.

In the sequel, we will use the following notation for the difference between consecutive (functions of) zeros:

$$\Delta z_k = z(x_{k+1}) - z(x_k), \quad k = 0, 1, \ldots, \quad (3.6)$$

where $x_k$ are the zeros of the function we are interested in. It is also convenient to use the parameters related to classical orthogonal polynomials instead of $a$, $b$ and $c$. The connection between both are as follows:

$$a = -n, \quad b = n + \alpha + \beta + 1, \quad c = \alpha + 1$$

for Gauss hypergeometric functions, and

$$a = -n, \quad c = \alpha + 1$$

for Kummer functions.
We will also need some oscillatory conditions in order to make sure that the ranges of parameters that we consider are consistent with the oscillatory behaviour of the solutions. In the case of Gauss functions, we will use the following:
\[
a < 0, \quad b > 1, \quad c - a > 1, \quad c - b < 0,
\]
(3.7)
or in Jacobi notation,
\[
n > 0, \quad n + \alpha > 0, \quad n + \beta > 0, \quad n + \alpha + \beta > 0,
\]
(3.8)
which ensure existence of real zeros in the interval \((0, 1)\). For Kummer functions, the conditions are
\[
a < 0, \quad c - a > 1,
\]
(3.9)
or with Laguerre parameters
\[
n > 0, \quad n + \alpha > 0.
\]
(3.10)

These are not necessary and sufficient conditions, and a more detailed analysis can be found in [55], see also [18], but they will be enough for our purposes.

### 3.3.1 Liouville transformations of Gauss equation

In view of (3.5), a general family of transformations is given by \(z = z(x)\) such that
\[
z'(x) = x^{p-1}(1-x)^{q-1}.
\]
(3.11)

In this case, one can check that
\[
\Omega'(x) = x^{-2p-1}(1-x)^{-2q-1}P(x),
\]
where \(P(x) = a_3x^3 + a_2x^2 + a_1x + a_0\) is a cubic polynomial. Furthermore,
\[
a_3 = \frac{1}{2}(1-p-q)[L^2 - (1-p-q)^2],
\]
\[
P(0) = -\frac{1}{2}p(p^2 - \alpha^2), \quad P(1) = \frac{1}{2}q(q^2 - \beta^2),
\]
where \(L = 2n + \alpha + \beta + 1\), see [7]. Therefore, if we want that \(\Omega'(x) = 0\) is equivalent to solving a quadratic equation we need to impose one of the following conditions:
\[
p + q = 1, \quad p = 0, \quad q = 0.
\]
(3.12)
Observe that because of (3.11), we have

\[
z(x) = \int_0^x t^{p-1} (1-t)^{q-1} \, dt = B_x(p,q), \quad p > 0,
\]

where \( B_x(p,q) \) is an incomplete beta function. When \( p = 0 \), we take

\[
z(x) = -\int_x^1 t^{p-1} (1-t)^{q-1} \, dt = -B_{1-x}(q,p), \quad q > 0.
\]

Note that the information on the zeros provided by the Sturm theorem will be given in terms of \( z(x_k) \), where \( x_k \) are the zeros of \( w(x) \). For this reason, we will be especially interested in those cases where \( z(x) \) reduces to an elementary function.

We also remark that we can interchange \( x \leftrightarrow 1-x \) together with \( \alpha \leftrightarrow \beta \), which reduces the number of transformations that we need to analyse. Furthermore, since we will be interested in results about zeros in either \( (-\infty, 0) \), \( (0, 1) \) or \( (1, \infty) \) (these intervals are determined by the singular points of the differential equation), this means that information on the zeros in one interval can be used to give results about the zeros in a different one.

**Example 3.3.1** If we consider \( p = q = 1/2 \) we obtain

\[
z'(x) = \frac{1}{\sqrt{x(1-x)}}
\]

and we can take \( z(x) = \arccos(1-2x) \). We will use the notation \( \theta(x) \) instead of \( z(x) \) for this trigonometric change of variables.

**Example 3.3.2** When \( p = 0 \) and \( q = 1 \) we obtain

\[
z'(x) = \frac{1}{x},
\]

so \( z(x) = \log x \).

In the case \( p = q = 1/2 \), the functions are:

\[
\Omega(x) = \frac{1}{4} \left[ L^2 - \frac{\alpha^2 - 1/4}{x} - \frac{\beta^2 - 1/4}{1-x} \right], \quad (3.13)
\]

\[
\Omega'(x) = \frac{1}{4} \left[ \frac{\alpha^2 - 1/4}{x^2} - \frac{\beta^2 - 1/4}{(1-x)^2} \right]. \quad (3.14)
\]

As a consequence, we obtain:
**Theorem 3.3.1** Let us suppose that the parameters $n$, $\alpha$, $\gamma$, $\beta$ satisfy the oscillatory conditions (3.8). Except in the case $|\alpha| = |\beta| = 1/2$, the differences $\Delta \theta_k$, defined as in (3.6), satisfy:

1. If $|\alpha| \geq 1/2$ and $|\beta| \leq 1/2$, then $\Delta \theta_k$ is decreasing as a function of $k$, that is, $\Delta \theta_{k+1} < \Delta \theta_k$.

2. If $|\alpha| \leq 1/2$ and $|\beta| \geq 1/2$, then $\Delta \theta_k$ is increasing as a function of $k$, that is, $\Delta \theta_{k+1} > \Delta \theta_k$.

3. If $|\alpha| < 1/2$ and $|\beta| < 1/2$, then $\Delta \theta_k < \pi/\sqrt{\Omega_m}$.

4. If $|\alpha| > 1/2$ and $|\beta| > 1/2$, then $\Delta \theta_k > \pi/\sqrt{\Omega_M}$.

If $|\alpha| = |\beta| = 1/2$ then $\Delta \theta_k$ is constant. The third and fourth cases can be extended to the cases where equality holds, bearing in mind that in those cases the maximum or minimum of $\Omega(x)$ is attained at $x = 0$ (when $|\alpha| = 1/2$) or at $x = 1$ (when $|\beta| = 1/2$).

Observe that the case $|\alpha| = |\beta| = 1/2$ corresponds to the Chebyshev polynomials, in which case the spacing of the zeros is well known. The values of the constants in the general case can be computed explicitly:

$$x_e = \frac{\sqrt{|1/4 - \alpha^2|}}{|\sqrt{|1/4 - \alpha^2|} + \sqrt{|1/4 - \beta^2|}}$$

$$\Omega(x_e) = \frac{1}{4} \left[ L^2 \pm \left( \sqrt{|1/4 - \alpha^2|} + \sqrt{|1/4 - \beta^2|} \right)^2 \right] > 0,$$

the plus sign corresponding to the case where the function has a maximum and the minus sign when there is a minimum. The bounds on the difference between consecutive zeros represent an improvement on the results in [49] for the zeros of Jacobi polynomials.

**Exercise 3.3.1** Apply the change of variables $z(x) = \log x$, and check that in this case

$$\Omega(x) = \frac{1}{4} \left[ -L^2 + \frac{L^2 - \alpha^2 + \beta^2 - 1}{1 - x} + \frac{1 - \beta^2}{(1 - x)^2} \right].$$

Show that if $|\beta| \leq 1$ then $\Delta^2 z_k < 0$, and therefore that

$$x_k^2 > x_{k+1} x_{k-1}, \quad k \geq 1,$$

71
where \( x_k \) are consecutive zeros of any Gauss hypergeometric function in \((0, 1)\). Conclude that if \( |\beta| \leq 1 \), then the zeros of Jacobi polynomials \( \tilde{x}_k \) satisfy
\[
(1 - \tilde{x}_k)^2 > (1 - \tilde{x}_{k+1})(1 - \tilde{x}_{k-1}), \quad k \geq 1.
\]

Using the property that we can interchange \( \alpha \leftrightarrow \beta \) and \( x \leftrightarrow 1 - x \), show that if \( |\alpha| \leq 1 \), then the zeros of Jacobi polynomials \( \tilde{x}_k \) satisfy
\[
(1 + \tilde{x}_k)^2 > (1 + \tilde{x}_{k+1})(1 + \tilde{x}_{k-1}), \quad k \geq 1.
\]

### 3.3.2 Liouville transformations of Kummer equation

In the case of the confluent hypergeometric equation, a family of admissible Liouville transformations is given by
\[
z(x) = \begin{cases} 
  x^m/m & m \neq 0 \\
  \log x & m = 0
\end{cases} \tag{3.17}
\]

Thus we obtain
\[
\Omega(x) = -\frac{1}{4}x^{-2m}(x^2 - 2Lx + \alpha^2 - m^2),
\]
where in this case \( L = 2n + \alpha + 1 \). A complete analysis of the different possibilities can be found in [7], and here we will only present two cases.

If we take \( m = 1 \) the change of variables is trivial, \( z(x) = x \), and
\[
\Omega(x) = -\frac{1}{4} \left( 1 - \frac{2L}{x} - \frac{1 - \alpha^2}{x^2} \right),
\]
which is decreasing for \( x \in (0, \infty) \) if \( |\alpha| \leq 1 \). If \( |\alpha| > 1 \) we have a maximum at
\[
x_e = \frac{\alpha^2 - 1}{L},
\]
and
\[
\Omega(x_e) = \frac{L^2 - (\alpha^2 - 1)}{\alpha^2 - 1} > 0.
\]

Consequently, we have

**Theorem 3.3.2** Let us suppose that the parameters \( n \) and \( \alpha \) satisfy the oscillatory conditions (3.10). Then the zeros \( x_k \) of the confluent hypergeometric functions in \((0, \infty)\), and in particular the zeros of Laguerre polynomials, satisfy:
1. If $|\alpha| \leq 1$ then $\Delta^2 x_k > 0$, so
   \[ x_k < \frac{x_{k-1} + x_{k+1}}{2}, \]
   \[ x_k^2 < \frac{x_{k-1}^2 + x_{k+1}^2}{2}, \]

2. If $|\alpha| > 1$ then
   \[ \Delta x_k = x_{k+1} - x_k > \frac{\pi \sqrt{\alpha^2 - 1}}{L^2 - (\alpha^2 - 1)}. \]

The zeros of Hermite polynomials satisfy
   \[ x_k^2 < \frac{x_{k-1}^2 + x_{k+1}^2}{2}, \]
as a direct consequence of the theorem with $\alpha = -1/2$, recall (1.36).

Exercise 3.3.2 Use the change of variables $z(x) = x^{1/2}$ to prove that the zeros of Laguerre polynomials $x_k$ satisfy

1. If $|\alpha| \leq 1/2$ then $\Delta^2 z_k > 0$, so
   \[ \sqrt{x_k} < \frac{\sqrt{x_{k-1}^2 + x_{k+1}^2}}{2}, \]

2. If $|\alpha| > 1/2$ then
   \[ \sqrt{x_{k+1}^2 - \sqrt{x_k^2}} > \frac{\pi}{\sqrt{2(L - \sqrt{\alpha^2 - 1/4})}}. \]

Show that the zeros of Hermite polynomials $x_k$ verify
   \[ x_k < \frac{x_{k-1} + x_{k+1}}{2}, \quad x_{k+1} - x_k > \frac{2}{\sqrt{2n + 1}}. \]

These bounds are classical, see [49, Eq. 6.31.21].

The same family of changes can be applied to the solutions of the other confluent hypergeometric equation:
   \[ xy''(x) + cy'(x) + y(x) = 0, \]
recall (1.37) One solution is the confluent hypergeometric function $\,_{0}\!F_1(-; c; -x)$, which it is related to Bessel functions of the first kind by the following formula:
   \[ \,_{0}\!F_1(-; c + 1; -x) = \Gamma(c + 1) x^{-c/2} J_c(2\sqrt{x}), \]
or
\[ J_\nu(x) = \frac{1}{\Gamma(\nu)} \left( \frac{t}{2} \right)^{\nu-1} 0F_1(-; c + 1; -\frac{t^2}{4}), \]
see (1.38). In this case, with the same family of transformations as in (3.17), we obtain
\[ \Omega(x) = \frac{4x + m^2 - \nu^2}{4x^{2m}}. \]

The Bessel function \( J_\nu(x) \) has an infinite number of zeros on the positive real axis. See the detailed analysis in [7].

**Exercise 3.3.3** Using the appropriate change of variables from (3.17), prove the following results on the zeros \( c_{\nu,k}, k \geq 0 \), of the Bessel function \( J_\nu(x) \):

- If \( |\nu| > 1/2 \) then \( c_{\nu,k+1} - c_{\nu,k} > \pi \);
- If \( |\nu| = 1/2 \) then \( c_{\nu,k+1} - c_{\nu,k} = \pi \);
- If \( |\nu| < 1/2 \) then \( c_{\nu,k+1} - c_{\nu,k} < \pi \);
- \( c_{\nu,k} > \sqrt{c_{\nu,k+1}c_{\nu,k-1}} \).

### 3.4 Systems of differential-difference equations

An alternative approach to the analysis of real zeros of hypergeometric functions, both analytically and numerically, is provided by systems of differential-difference equations (DDEs). More precisely, suppose that we have two ODEs of the form
\[ y''(x) + p_y(x)y(x) = 0, \quad w''(x) + p_w(x)w(x) = 0. \quad (3.18) \]

We say that \( w(x) \) is a contrast function with respect to the function \( y(x) \). As explained in [47] and [45], from a general result proved in [40, 41], given two sets of independent solutions of the previous equations, \( \{ y^{(1)}(x), y^{(2)}(x) \} \) and \( \{ w^{(1)}(x), w^{(2)}(x) \} \) respectively, there exists a unique system of DDEs of the form:
\[ \begin{align*}
y'(x) &= a(x)y(x) + d(x)w(x) \\
w'(x) &= b(x)w(x) + e(x)y(x) \end{align*} \quad (3.19) \]
which is satisfied by both \( \{ y^{(1)}(x), w^{(1)}(x) \} \) and \( \{ y^{(2)}(x), w^{(2)}(x) \} \).

What is relevant from the point of view of hypergeometric functions is that we can choose \( w(x) \) to be a contiguous function of the same family as
y(x). For example, if \( y(x) = M(a, c; x) \) and \( w(x) = M(a - 1; c; x) \), then the corresponding system of DDEs is

\[
\begin{align*}
y'(x) &= \frac{a - c + x}{x} y(x) - \frac{a - c}{x} w(x) \\
w'(x) &= \frac{1 - a}{x} w(x) + \frac{a - 1}{x} y(x)
\end{align*}
\] (3.20)

In the case of classical orthogonal polynomials, we can set \( y(x) = P_n(x) \) and \( w(x) = P_{n+1}(x) \). The equations of the system of DDEs are usually known as structure relations in this context.

Under certain conditions on the coefficients of the system (3.19), it is possible to obtain results on interlacing of the real zeros of the functions \( y(x) \) and \( w(x) \). More precisely, we quote [45, Lemma 2.4]:

**Lemma 3.4.1** Let \( y(x) \) and \( w(x) \) be two nontrivial solutions of (3.19) in a given interval \( I \), and let \( d(x) \) be continuous in \( I \). If \( y(x) \) or \( w(x) \) have at least two zeros in \( I \), then

- The zeros of \( y(x) \) and \( w(x) \) are simple.
- The functions \( y(x) \) and \( w(x) \) have no common zeros.
- The zeros of \( y(x) \) and \( w(x) \) are interlaced.
- It is true that \( d(x) e(x) < 0 \) in \( I \).

This lemma has important consequences both in order to get analytic information on the zeros of \( y(x) \) and \( w(x) \) and to construct fixed point methods for the numerical computation of those zeros, as we will see later. Moreover, the fact that \( d(x) e(x) < 0 \) whenever the solutions oscillate gives an easy way to check conditions on external parameters such that \( y(x) \), for example, has at least two zeros in a given interval. We refer the reader to [18], where the authors carry out a systematic analysis of several DDEs for Gauss, Kummer and Bessel functions. This is the way how the general conditions (3.8) and (3.10) can be obtained.

### 3.5 Computational aspects

The accurate and efficient computation of zeros of special functions is highly relevant in several problems in applied mathematics. For example, the zeros of classical orthogonal polynomials are crucial in the computation of Gauss quadrature rules, as we will see later on. In the literature, several numerical
methods have been proposed, see [45] and the references therein. In the case of orthogonal polynomials, an excellent source of information both on theoretical and numerical aspects, together with MATLAB routines, is [17].

Since the family of hypergeometric functions is quite large and comprises functions that depend on several parameters and exhibit very different behaviours, there is normally a trade-off between generality and efficiency. Below we give a few details on some of the different approaches proposed.

### 3.5.1 Gaussian quadrature

The problem of numerical integration is one of the cornerstones of numerical analysis, see for instance the classical reference [6]. The basic problem is how to construct a quadrature rule that approximates the value of a definite integral on a (possibly infinite) interval of the real line:

\[
I[f] = \int_a^b f(x)w(x)dx \approx \sum_{k=1}^n w_k f(x_k).
\]

Here \(x_k\) are the nodes and \(w_k\) the weights of the quadrature rule. The function \(w(x)\) is supposed to be integrable and nonnegative in \([a, b]\), and is called weight function.

If one places the nodes equally spaced in \([a, b]\), the result is known as a Newton-Cotes method, and this includes the popular trapezoidal and Simpson rules. The big advantage of these methods is their adaptivity, since adding extra nodes to increase accuracy is simple. One substantial problem of this approach is that if the number of nodes is increased, then the rule becomes unstable and leads to bad numerical results. Some alternatives are the composite rules or extrapolation methods such as Romberg’s method.

A different approach is considered in Gaussian quadrature: one seeks nodes and weights in such a way that the quadrature rule is exact for polynomials of degree as high as possible. This maximum degree turns out to be \(2n-1\) when we use \(n\) nodes and weights, and the key result is that the nodes are precisely the zeros of \(p_n(x)\), where the family \(\{p_n(x)\}_{n=0}^{\infty}\) is orthogonal with respect to the weight function \(w(x)\), see for example [4, §1.6] or [17, §1.4].

### 3.5.2 Matrix methods

This approach is especially successful for the zeros of orthogonal polynomials. Suppose that the family \(\{P_n(x)\}_{n=0}^{\infty}\) are monic orthogonal polynomials
with respect to a weight function \( w(x) \). We write the three term recurrence relation in the form

\[
x P_n(x) = P_{n+1}(x) + c_n P_n(x) + \lambda_n P_{n-1}(x),
\]

(3.21)

with \( P_{-1}(x) = 0 \), \( P_1(x) = 1 \). Alternatively, if we work with orthonormal polynomials (which is irrelevant as far as zeros are concerned), then we have

\[
x p_n(x) = a_n p_{n+1}(x) + b_n p_n(x) + a_{n-1} p_{n-1}(x),
\]

(3.22)

where \( b_n = c_n \) and \( a_n = \sqrt{\lambda_n} \), and the initial values are now \( p_{-1}(x) = 0 \), \( p_0(x) = 1/\sqrt{a_0} \), where \( a_0 = \int w(x) dx \).

We can express the recurrence relation in matrix form \( x \mathbf{p} = J \mathbf{p} \), where \( J \) is the Jacobi matrix and \( \mathbf{p} = [p_0(x), p_1(x), \ldots]^T \). If we truncate this expression, we obtain

\[
\begin{pmatrix}
p_0(x) \\
p_1(x) \\
\vdots \\
p_{n-1}(x)
\end{pmatrix}
= \begin{pmatrix}
b_0 & a_1 & 0 & & \\
a_1 & b_1 & a_2 & & \\
& \ddots & \ddots & \ddots & \ddots \\
& & a_{n-2} & b_{n-2} & a_{n-1} \ & \\
& & & 0 & a_{n-1} & b_{n-1}
\end{pmatrix}
\begin{pmatrix}
p_0(x) \\
p_1(x) \\
\vdots \\
p_{n-1}(x)
\end{pmatrix} + \begin{pmatrix}
0 \\
0 \\
\vdots \\
0
\end{pmatrix}
\]

(3.23)

Observe that \( J \) is tridiagonal and symmetric. We denote by \( J_n \) the truncated Jacobi matrix that appears in this last expression, and also \( \mathbf{p}_n = [p_0(x), p_1(x), \ldots, p_{n-1}(x)]^T \). That is, we write for short

\[
x \mathbf{p}_n(x) = J_n \mathbf{p}_n(x) + a_n p_n(x) \mathbf{e}_{n+1},
\]

where \( \mathbf{e}_{n+1} \) is the \( n + 1 \)-th vector of the canonical basis.

This formulation leads to a key result, that is the basis for the most important and widely used method for the computation of zeros of orthogonal polynomials:

**Theorem 3.5.1** Let \( w(x) \) be a weight function in an interval \([a, b] \subset \mathbb{R}\), possibly infinite. Given a Gauss quadrature formula

\[
I[f] = \int_a^b f(x) w(x) dx \approx \sum_{k=1}^n w_k f(x_k),
\]

(3.24)

which is exact for polynomials of degree \( \leq 2n - 1 \), the nodes \( x_k \) are the eigenvalues of the truncated Jacobi matrix \( J_n \) associated with the sequence
of polynomials \( P_n(x) \) which are orthogonal with respect to the weight \( w(x) \). Moreover, the quadrature weights are

\[
w_k = a_0 v_{k,1}^2, \quad k = 1, 2, \ldots, n,
\]

where again \( a_0 = \int w(x) \, dx \), and \( v_{k,1} \) is the first component of the (normalized) eigenvector of the Jacobi matrix corresponding to the eigenvalue \( x_k \).

**Proof 3.5.1** This result goes back to Golub and Welsch, [24], see also [17]. Suppose that \( x_k \) is a node of Gaussian quadrature, then it is a zero of \( p_n(x) \).

We deduce from (3.23) that \( x_k p_n(x_k) = J_n p_n(x_k) \), and therefore \( x_k \) is an eigenvalue of the matrix \( J_n \) for \( k = 1, 2, \ldots, n \).

It follows from (3.23) that \( p_n(x_k) \) is the eigenvector of \( J_n \) associated with the eigenvalue \( x_k \). Now we need to normalize it, and from the scalar product

\[
||p_n(x_k)||^2 = p_n^T(x_k) p_n(x_k) = \sum_{j=0}^{n-1} p_j^2(x_k),
\]

we obtain

\[
v_k = \frac{p_n(x_k)}{\left[ \sum_{j=0}^{n-1} p_j^2(x_k) \right]^{1/2}},
\]

so comparing the first component on both sides and squaring (recall that \( p_0(x) = 1/\sqrt{a_0} \)), it follows that

\[
a_0 v_{k,1}^2 = \frac{1}{\sum_{j=0}^{n-1} p_j^2(x_k)}.
\]

We only need to show that this last sum is equal to the Gaussian weight \( w_k \). Now, applying the Gauss quadrature formula (3.24) to the polynomial \( p_m(x) \), for \( 0 \leq m \leq n - 1 \), we get

\[
a_0^{1/2} \delta_{m,0} = \sum_{j=1}^{n} w_j p_m(x_j), \quad 0 \leq m \leq n - 1,
\]

because of orthonormality. In matrix form,

\[
Q w = a_0^{1/2} e_1,
\]

where \( Q \) is the matrix of eigenvectors

\[
Q = [p(x_1), p(x_2), \ldots, p(x_n)]
\]

78
and \( w = [w_1, w_2, \ldots, w_n]^T \). Since the columns of \( Q \) are mutually orthogonal, we have \( Q^T Q = D \), where \( D \) is a diagonal matrix with diagonal entries

\[
d_{k-1} = \sum_{j=0}^{n-1} p_j^2(x_k), \quad k = 1, 2, \ldots, n.
\]

Now

\[
Dw = Q^T Qw = a_0^{1/2} Q^T e_1 = a_0^{1/2} a_0^{-1/2} e = e,
\]

where \( e = [1, 1, \ldots, 1]^T \). Consequently, \( w = D^{-1} e \), which implies that

\[
w_k = \frac{1}{\sum_{j=0}^{n-1} p_j^2(x_k)}, \quad 0 \leq k \leq n - 1.
\]

This result makes it possible to use well known and efficient eigenvalue methods in Numerical Linear Algebra to compute the zeros of orthogonal polynomials as eigenvalues of tridiagonal symmetric matrices. Furthermore, as explained in [37], the computation of Gauss nodes and weights from recursion coefficients is well conditioned, therefore when these coefficients are available this provides a very attractive method.

This method is also applicable for minimal solutions of recurrence relations which are not polynomials, see [22, Section 7.4.2]. The formulation is more delicate, since one needs to deal with finite approximations \( J_n \) (truncations of the Jacobi matrix) of an operator \( J \), and there are convergence issues depending on the entries of the matrix.

It is also possible to modify the Jacobi matrix in order to deal with quadrature rules with preassigned nodes. For instance, suppose that \( x = a \) must be a node in our approximation, then we seek nodes and weights for the Gauss–Radau formula

\[
I[f] = \int_a^b f(x) w(x) dx \approx w_0^a f(a) + \sum_{k=1}^n w_k^a f(x_k). \quad (3.25)
\]

We construct the matrix

\[
J_{n+1}^R = \begin{pmatrix} J_n & a_n e_n^T \\ a_n e_n & \alpha_n^R \end{pmatrix},
\]

where

\[
\alpha_n^R = a - a_n \frac{p_{n-1}(a)}{p_n(a)}
\]

and \( e_n \) as before. Recall that with this formulation, the \( p_n(x) \) must be orthonormal polynomials. Then it can be proved, see [17, §3.1.1.2], that the
zeros of the matrix $J_{n+1}^R$ are the nodes for (3.25), and the weights can be computed as before:

$$w^a_k = a_0 v_{k,1}^2, \quad k = 0, 1, 2, \ldots, n,$$

where $v_{k,1}$ is the first component of the $k$-th normalized eigenvector of the matrix $J_{n+1}^R$.

In a similar way, if we are interested in a Gauss–Lobatto quadrature:

$$I[f] = \int_a^b f(x)w(x)dx \approx w_0^L f(a) + w_{n+1}^L + \sum_{k=1}^n w_k^L f(x_k),$$

(3.26)

then we need the matrix

$$J_{n+2}^L = \begin{pmatrix} J_{n+1} & \sqrt{\beta_{n+1}} e_{n+1} \\ \sqrt{\beta_{n+1}} e_{n+1}^T & \alpha_{n+1}^L \end{pmatrix},$$

where now the coefficients $\alpha_{n+1}^L$ and $\beta_{n+1}^L$ solve the system

$$\begin{pmatrix} P_{n+1}(a) & P_n(a) \\ P_{n+1}(b) & P_n(b) \end{pmatrix} \begin{pmatrix} \alpha_{n+1}^L \\ \beta_{n+1}^L \end{pmatrix} = \begin{pmatrix} aP_{n+1}(a) \\ bP_{n+1}(b) \end{pmatrix},$$

this time in terms of the monic orthogonal polynomials, see [17, §3.1.1.3].

Gauss–Radau and Gauss–Lobatto quadrature rules are of interest if a particular point gives a relevant contribution to the total integral. An example of this kind of situation was encountered in the last chapter, when discussing the method of stationary phase for oscillatory integrals on the real axis.

### 3.5.3 Fixed point methods

The systems of DDEs that we have presented before can be used to construct fixed point methods to compute the zeros of hypergeometric functions. We will present here the basic ideas contained in [47, 19, 45]. Suppose that we have a system of DDEs

$$\begin{align*}
y'(x) &= a(x)y(x) + d(x)w(x) \\
w'(x) &= b(x)w(x) + e(x)y(x)
\end{align*}$$

(3.27)

with continuous coefficients, satisfied both by the pair $\{y^{(1)}(x), w^{(1)}(x)\}$ and $\{y^{(2)}(x), w^{(2)}(x)\}$, which are solutions of the ODEs

$$\begin{align*}
y''(x) + p_y(x)y(x) &= 0, \\
w''(x) + p_w(x)w(x) &= 0.
\end{align*}$$
We have seen before that under quite general conditions, if the functions $y(x)$ and $w(x)$ have at least two real zeros in an interval $[a, b]$, then these zeros are distinct and they interlace. If this holds, then $d(x)e(x) < 0$ for $x \in [a, b]$, and without loss of generality we will assume that $d(x) > 0$.

If we differentiate the first equation in (3.27) and substitute
\[ w(x) = \frac{y'(x) - a(x)y(x)}{d(x)}, \]
we will obtain a second order ODE for the function $y(x)$. The coefficient multiplying $y'(x)$ in this equation is
\[ f(x) = -(a(x) + b(x) + d'(x)/d(x)), \]
and it follows that the function
\[ \tilde{y}(z(x)) = \frac{1}{\sqrt{|d(x)|}} \exp \left( -\frac{1}{2} \int^x (a + b) \right) y(x) \] (3.28)
satisfies an equation in normal form. Analogously, we can transform
\[ \tilde{w}(z(x)) = \frac{1}{\sqrt{|e(x)|}} \exp \left( -\frac{1}{2} \int^x (a + b) \right) w(x). \] (3.29)

It is straightforward to check that these new functions satisfy the system of DDE:
\[ \begin{align*}
\tilde{y}'(x) &= \tilde{a}(x)\tilde{y}(x) + \tilde{d}(x)\tilde{w}(x) \\
\tilde{w}'(x) &= \tilde{c}(x)\tilde{y}(x) + \tilde{b}(x)\tilde{w}(x),
\end{align*} \]

where additionally we have $\tilde{d}(x) = -\tilde{c}(x)$. We note that the functions $y(x)$ and $\tilde{y}(x)$ have the same zeros in $[a, b]$.

Now we apply a change of independent variable:
\[ z(x) = \int^x \sqrt{-d(t)} dt, \]
and the transformed system reads
\[ \begin{align*}
\hat{y}(z) &= \alpha(z)\hat{y}(z) + \hat{w}(z) \\
\hat{w}(z) &= -\hat{y}(z) + \beta(z)\hat{w}(z),
\end{align*} \]

where $\beta(z) = -\alpha(z) =: \eta(z)$ and dots indicate differentiation with respect to $z$. We can express both the change of variable $z(x)$ and the function $\eta(z)$ in terms of the original system:
\[ z(x) = \int^x \sqrt{-d(t)e(t)} dt. \]
\[ \eta(z) = \frac{b(z) - a(z)}{2} \dot{x}(z) + \frac{d}{4} \frac{d}{dz} \log \left| \frac{d(z)}{e(z)} \right|. \]

The idea now is that the system

\[
\begin{pmatrix}
\dot{\tilde{y}}(z) \\
\dot{\tilde{w}}(z)
\end{pmatrix}
= \begin{pmatrix}
-\eta(z) & 1 \\
-1 & \eta(z)
\end{pmatrix}
\begin{pmatrix}
\tilde{y}(z) \\
\tilde{w}(z)
\end{pmatrix},
\]

(3.30)

resembles that satisfied by the functions \( \tilde{y}(z) = \sin z \) and \( \tilde{w}(z) = \cos z \), which is exactly the solution if \( \eta(z) \equiv 0 \). Naturally, \( \eta(z) \neq 0 \) in general, but it can be proved that if \( |\eta(z)| > 1 \) in a certain interval \( I \) then \( H(z) \) can have at most one zero in \( I \), see [47, Th. 2.1].

If we construct the ratio of functions

\[ H(z) := \frac{\tilde{y}(z)}{\tilde{w}(z)}, \]

(3.31)

then it can be shown that \( H(z) \) has interlaced zeros and poles, as a consequence of the (strict) interlacing of the zeros of \( \tilde{y}(z) \) and \( \tilde{w}(z) \). Moreover \( H(z) \) is related to \( \eta(z) \) through a Riccati equation:

\[ \dot{H}(z) = 1 + H^2(z) - 2\eta(z)H(z). \]

(3.32)

Let us consider two consecutive zeros of \( \tilde{w}(z) \), say \( z(x_{w,k}) \) and \( z(x_{w,k+1}) \), and let \( J = (z(x_{w,k}), z(x_{w,k+1})) \). Thanks to interlacing, we know that there is exactly one zero of \( \tilde{y}(z) \) in \( J \), say \( z = z_{y,k} \). If \( \eta(z) = 0 \), then given an initial value \( z_0 \) in the interval \( J \), the zero of \( H(z) \) would be directly

\[ z(x_{y,k}) = z_0 - \arctan(H(z_0)). \]

In the general case \( \eta(z) \neq 0 \), we consider the following fixed point iteration: given \( z_0 \),

\[ z_{n+1} = z_n - \arctan(H(z_n)), \quad n = 0, 1, 2, \ldots, \]

and this can be seen to converge globally to the zeros of \( H(z) \) (and therefore to the zeros of \( \tilde{y}(z) \), and of \( y(x) \) undoing the change of variable), in those intervals where \( \eta(z) \) does not change sign, see [22, §7.5].

The properties of \( \eta(z) \) determine the detailed form of the function \( H(z) \) and the distribution of its zeros. More precisely, observe that from the Riccati equation (3.32), we have

\[ \frac{\dot{H}(z)}{1 + H(z)^2} - 1 = -\frac{2\eta(z)H(z)}{1 + H(z)^2}. \]
so
\[
\text{sign(}\eta H,\frac{\dot{H}(z)}{1+H(z)^2}-1) = \frac{-2|\eta(z)H(z)|}{1+H(z)^2} \leq 0,
\]

If we now integrate between \( z = z_{w,k} \) and \( z \in J = (z(x_w,k), z(x_w,k+1)) \), assuming that \( \eta \neq 0 \) in \( J \), then
\[
\text{sign(}\eta)\text{sign}(z - z_{w,k})\left(\frac{\pi}{2} - |z - z_{w,k}|\right) < 0,
\]
because \( \text{sign}(H) = \text{sign}(z - z_{w,k}) \), since \( \dot{H}(z_{y,k}) = 1 \) and the function \( H(z) \) is therefore increasing in \( J \). As a consequence:

- If \( \eta(z) < 0 \) in \( J \), then
  \[
  |z(x_{w,2}) - z(x_{y,j})| < \frac{\pi}{2}, \quad |z(x_{y,j}) - z(x_{w,1})| > \frac{\pi}{2}.
  \]

  Therefore \( z(x_{y,j}) > \frac{z(x_{w,2}) - z(x_{w,1})}{2} \).

- If \( \eta(z) = 0 \) en \( J \), then \( z(x_{y,j}) = \frac{z(x_{w,2}) - z(x_{w,1})}{2} \).

- If \( \eta(z) > 0 \) in \( J \), then
  \[
  |z(x_{w,2}) - z(x_{y,j})| > \frac{\pi}{2}, \quad |z(x_{y,j}) - z(x_{w,1})| < \frac{\pi}{2}.
  \]

  Therefore \( z(x_{y,j}) < \frac{z(x_{w,2}) - z(x_{w,1})}{2} \).

An important consequence of this result is that the distance between consecutive zeros of \( \tilde{y}(z) \) is larger than \( \pi/2 \), and this is convenient for numerical purposes if the zeros in the original variable \( x \) tend to cluster.
The idea now is to construct an iterative algorithm that computes all the zeros of \( \tilde{y}(z) \) in a given interval by backward or forward sweep, depending on the sign of \( \eta(z) \). If \( \eta(z) \) changes sign in \([a, b]\), then it is possible to adapt the scheme appropriately, see the details in [45, 19, 22].

**Example 3.5.1** Bessel functions \( J_\nu(x) \) and \( J_{\nu-1}(x) \) satisfy the system

\[
J'_\nu(x) = -\frac{\nu}{x} J_\nu(x) + J_\nu(x) \tag{3.33}
\]
\[
J'_{\nu-1}(x) = \frac{\nu + 1}{x} J_\nu(x) - J_\nu(x), \tag{3.34}
\]

so

\[
a(x) = -\frac{\nu}{x}, \quad b(x) = \frac{\nu - 1}{x}, \quad d(x) = -e(x) = 1,
\]

hence

\[
z(x) = x, \quad \eta(z) = \eta(x) = \frac{2\nu - 1}{2x}.
\]

In this case, the sign of \( \eta \) only depends on \( \nu \), for positive \( x \).

**Example 3.5.2** We consider the Laguerre polynomials

\[
y_n(x) = L_n^{(\alpha)}(x), \quad y_{n-1}(x) = L_{n+1}^{(\alpha)}(x),
\]
and the corresponding system of DDEs reads

\[ y_n'(x) = \frac{-n - \alpha - 1 + x}{x} y_n(x) + \frac{n + 1}{x} y_{n-1}(x) \]
\[ y_{n-1}'(x) = -\frac{n + \alpha + 1}{x} y_n(x) + \frac{n + 1}{x} y_{n-1}(x) \]

In this case

\[ z(x) = \sqrt{(n + 1)(n + \alpha + 1)} \log(x), \quad \eta(z(x)) = \frac{2n + \alpha + 2 - x}{2\sqrt{(n + 1)(n + \alpha + 1)}} \]

Let \( x^* = 2n + \alpha + 2 \), let \( x_{n+1,1} \) and \( x_{n+1,2} \) be two consecutive zeros of \( L_{n+1}^{(\alpha)}(x) \), and \( x_{n,j} \) the zero of \( L_{n}^{(\alpha)}(x) \) between them.

- If \( x_{n+1,2} < x^* \), then \( x_{n,j}^2 < x_{n+1,1} x_{n+1,2} \).
- If \( x_{n+1,1} > x^* \), then \( x_{n,j}^2 > x_{n+1,1} x_{n+1,2} \).

We highlight two important points: firstly, the freedom in choosing the contrast function can be used to our advantage, since some iterative methods will be more efficient than others. See [18] for a detailed analysis of Gauss, Kummer and Bessel functions. Secondly, it is important to bear in mind that when applying this method (and other root-finding schemes), one needs to evaluate the ratio \( H(z) \). In some cases this is a minor problem, but sometimes further work is needed. In the next chapter we will see a possible approach to this question using continued fractions.

### 3.5.4 Asymptotic methods

If we have asymptotic information about a certain function, under some circumstances it is possible to derive approximations to its zeros, via inversion of the asymptotic expansion. This procedure can be used either to compute the zeros or to obtain initial values for a different method such as a root-finding scheme. The general result is as follows: suppose that we have a function \( f(z) \) that admits an asymptotic expansion

\[
f(z) \sim z + f_0 + \frac{f_1}{z} + \frac{f_2}{z^2} + \ldots,
\]

when \( z \to \infty \) in a certain sector of the complex plane. Then, under mild conditions on \( f(z) \), see for instance [13, Theorem 2.3] and also [44], the equation \( f(z) = w \) has a solution, and furthermore

\[
z \sim w - F_0 - \frac{F_1}{w} - \frac{F_2}{w^2} + \ldots
\]
The coefficients in this expansion can be computed in terms of the coefficients of the original one \( f_j \), and the first one are

\[
F_0 = f_0, \quad F_1 = f_1, \quad F_2 = f_0 f_1 + f_2, \quad F_3 = f_0^2 f_1 + f_1^2 + 2 f_0 f_2 + f_3.
\]

A general formula can be obtained for the \( F_j \), see [13]. As an example, taken from this reference, we present the real zeros of the Airy function \( \text{Ai}(-x) \), for \( x > 0 \). The asymptotic expansion is

\[
\text{Ai}(-x) = \frac{1}{\sqrt{\pi x^{1/4}}} \left\{ \sin \left( \xi + \frac{\pi}{4} \right) P(\xi) - \cos \left( \xi + \frac{\pi}{4} \right) Q(\xi) \right\},
\]

where \( \xi = \frac{2}{3} x^{3/2} \) and the functions \( P(\xi) \) and \( Q(\xi) \) can be expanded in inverse powers of \( \xi \):

\[
P(\xi) = \sum_{j=0}^{\infty} (-1)^j \frac{u_{2j}}{\xi^{2j}}, \quad Q(\xi) = \sum_{j=0}^{\infty} (-1)^j \frac{u_{2j+1}}{\xi^{2j+1}},
\]

where

\[
u_0 = 1, \quad \nu_j = \frac{1}{54 j!} \frac{\Gamma(3j + 1/2)}{\Gamma(j + 1/2)}, \quad j \geq 1.
\]

see for instance [2, Eq. 10.4.60]. Therefore, at a zero of the Airy function we have

\[
\tan \left( \xi + \frac{\pi}{4} \right) = \frac{Q(\xi)}{P(\xi)} = \frac{5}{72\xi} - \frac{39655}{1119744\xi^3} + \ldots,
\]

and therefore

\[
\xi + \frac{\pi}{4} - s\pi \sim \arctan \frac{Q(\xi)}{P(\xi)} = \frac{5}{72\xi} - \frac{1105}{31104\xi^3} + \ldots,
\]

where \( s \) is an arbitrary integer. Now we have a relation between the variables \( \xi \) and \( s \):

\[
w = \left( s(\xi) - \frac{1}{4} \right) \pi \sim \xi - \frac{5}{72\xi} + \frac{1105}{31104\xi^3} + \ldots.
\]

Since

\[
f_0 = 0, \quad f_1 = -\frac{5}{72}, \quad f_2 = 0, \quad f_2 = \frac{1105}{31104},
\]

we obtain

\[
F_0 = 0, \quad F_1 = -\frac{5}{72}, \quad F_2 = 0, \quad F_3 = \frac{1255}{31104}.
\]
and therefore

\begin{equation}
\xi \sim w + \frac{5}{72w} - \frac{1055}{31104w^3} + \ldots \\
\sim \left( s - \frac{1}{4} \right) \pi + \frac{5}{72 \left( s - \frac{1}{4} \right) \pi} - \frac{1055}{31104 \left( s - \frac{1}{4} \right) \pi^3} + \ldots
\end{equation}

Finally, since \( \xi = \frac{2}{3} x^{3/2} \), we find

\[ x = t^{2/3} \left( 1 + \frac{5}{48t^2} \right), \quad t \to \infty, \quad (3.35) \]

where

\[ t = \frac{3\pi}{8 (4s - 1)}. \]

Extra work is needed in order to prove that each value of \( s \) indeed corresponds to the \( s \)-th zero of the Airy function, see [13] for more details. Table 3.1 shows the accuracy of the approximation for some of the first few real zeros, taking the first terms shown in (3.35). Observe that the method is of asymptotic nature and still the approximations for the first (small) zeros can be quite good.

Other examples can be found for instance in [22, §7.3.1 and §7.6].

### 3.6 Related problems

- One problem of Gaussian quadrature is that it is not easily adaptive, in the sense that if one needs extra nodes in order to increase the accuracy, then the computation normally has to be redone from scratch (there are exceptions such as Gauss–Kronrod quadrature, see [17]). The trapezoidal rule, with modifications, is an interesting alternative
in some situations, since the computation of nodes and weights is much simpler, see for instance [22, §5]. Yet another important alternative to Newton-Cotes and Gaussian rules is given by Clenshaw-Curtis methods, see [51].

- Gaussian quadrature imposes high degree of accuracy on the space of polynomials, but this is not the only possible choice. If the integrand shows poles near the interval of integration, for instance, then rational functions may be more suitable. This is an area of considerable research activity, closely related to problems in rational approximation.

- Gaussian quadrature rules with complex nodes and weights seem to be quite effective to approximate integrals that come from the application of the method of steepest descent to oscillatory integrals on the real axis, see [8]. One of the main problems in this case is how to compute or approximate the nodes and weights, because there are no results on existence and distribution of zeros of the corresponding sequence of orthogonal polynomials, as happens in the real case.

- Interlacing properties of the zeros of hypergeometric functions can be further investigated using systems of DDEs. For instance, one can try to find out if interlacing is maintained when one chooses as contrast function a contiguous one whose parameters differ in more than one unit, see for example [46].

- Obtaining information about complex zeros of hypergeometric functions is usually much harder, and in a general setting only zero-free regions or estimations of the number of zeros in a given region can be computed, using different methods. However, for some families of functions, the distribution of zeros in the complex plane is an active area of research. In this sense, there has been a lot of interest recently in the zeros of orthogonal polynomials with non-standard parameters, using Riemann–Hilbert techniques. See for example [35].
Chapter 4

Three term recurrence relations

We have seen that hypergeometric functions depend on one or several parameters. One of the most important type of identity, both theoretically and computationally, is the three term recurrence relations satisfied by functions of the same families with different values of the parameters. More precisely, these relations connect three functions which are contiguous, which means that their parameters differ by integer numbers. Given a Gauss function

\[ 2F_1 \left( \frac{a, b}{c} ; x \right), \]

its contiguous functions are

\[ 2F_1 \left( \frac{a \pm 1, b}{c} ; x \right), \quad 2F_1 \left( \frac{a, b \pm 1}{c} ; x \right), \quad 2F_1 \left( \frac{a, b}{c \pm 1} ; x \right). \]

It is clear that we can combine the recursions to connect any three Gauss functions whose parameters differ by integer numbers. It is known from the times of Gauss that any three contiguous Gauss functions satisfy a three term recurrence relation with rational coefficients in \( a, b, c \) and \( x \), which is naturally similar to the structure of the differential equations that we have studied before.

Similarly, confluent hypergeometric functions \( \phi(a; c; x) \), where \( \phi(a; c; x) \) is either \( M(a; c; x) \) or \( U(a; c; x) \), satisfy three term recurrence relations with rational coefficients in \( a, c \) and \( x \), and furthermore, \( M(a; c; x) \) and \( U(a; c; x) \) satisfy the same recurrence relation with a suitable normalization.
In these notes we will consider TTRR with the following notation:

\[ y_{n+1}(x) + b_n(x)y_n(x) + a_n(x)y_{n-1}(x) = 0, \quad (4.1) \]

where \( n \) is an integer parameter. In order to determine a specific solution of (4.1), we need two initial values \( y_0(x) \) and \( y_1(x) \). In the sequel we will assume that \( a_n(x) \neq 0 \) for all \( n \) and \( x \) in the interval of interest, so we indeed have a recursion involving three consecutive terms.

It is clear that we can have different directions of recursion, for instance if we choose

\[
y_n(x) = \, _2F_1\left( \begin{array}{c} a + n, b \\ c \end{array} ; x \right), \quad y_n(x) = \, _2F_1\left( \begin{array}{c} a + n, b + n \\ c \end{array} ; x \right)
\]

or

\[
y_n(x) = \, _2F_1\left( \begin{array}{c} a + n, b + n \\ c + n \end{array} ; x \right), \quad y_n(x) = \, _2F_1\left( \begin{array}{c} a, b \\ c + n \end{array} ; x \right).
\]

We will speak of recurrence directions to indicate which parameters we are varying. For example, the previous examples will be denoted by (0+0), (0+0) (+++) and (00+) respectively. The number of independent recursions is smaller than one would think due to the functional identities satisfied by Gauss functions, see [21] for a detailed analysis of the different cases.

**Remark 4.0.1** In the sequel we will speak of forward direction of recursion (or forward recursion) when we increase \( n \) in (4.1), that is, when we go from \( y_{n-1}(x) \) and \( y_n(x) \) to \( y_{n+1}(x) \). When we decrease \( n \) we will speak of backward direction.

Three term recurrence relations are a very important tool to analyse the behaviour of special functions, and they are also important from a computational point of view. In the case of some subfamilies, like classical orthogonal polynomials, TTRR are an essential part of the theory, but as happened before, these relations can be analysed in general, and the numerical part of the theory is especially rich if we allow a general setting and do not concentrate only on polynomial solutions.
4.1 General theory

The space of solutions of (4.1) has dimension two, in other words, it can be generated by two independent solutions \( f_n(x) \) and \( g_n(x) \). It follows that any generic solution \( y_n(x) \) of the recursion can be written as

\[
y_n(x) = Af_n(x) +Bg_n(x),
\]

where \( A \) and \( B \) are constants. The notion of independent solution is analogous to the idea in the theory of differential equations:

**Definition 4.1.1** Two solutions \( f_n(x) \) and \( g_n(x) \) of (4.1) are independent if the Casorati determinant

\[
D_n = \begin{vmatrix} f_n & g_n \\ f_{n-1} & g_{n-1} \end{vmatrix}
\]

is different from zero for all positive integer \( n \) and \( x \) in a given domain.

**Example 4.1.1** The Bessel function of the first kind \( J_n(x) \) satisfies the recursion

\[
J_{n+1} - \frac{2n}{x}J_n + J_{n-1} = 0, \quad n \geq 1.
\]

Another solution is given by the Bessel function of the second kind \( Y_n(x) \).

It can be easily shown that \( D_{n+1} = b_n D_n \) for \( n \geq 0 \). Using this fact one can prove that the functions \( J_n(x) \) and \( Y_n(x) \) are independent solutions of (4.3) for \( x \neq 0 \). Additionally, one needs to use the fact that

\[
D_1 = J_1(x)Y_0(x) - J_0(x)Y_1(x) = -J'_0(x)Y_0(x) + J_1(x)Y'_1(x),
\]

together with the Wronskian for the Bessel functions

\[
W[J_n(x), Y_n(x)] = \frac{2}{\pi x}, \quad x \neq 0,
\]

see [2, Eq. 9.1.28].

From a theoretical point of view, any two independent solutions are enough to describe the set of solutions of (4.1), but from a numerical point of view, it may be necessary to include a specific solution in our basis in order to have a so called *numerically satisfactory solution*. Informally, this roughly means that from the pair of solutions that we choose as a basis it is possible to compute any other solution without heavy errors due to cancellation. This is quite naturally related to the behaviour of the different solutions when \( n \) is large. We present a more detailed analysis below.
4.2 Minimal and dominant solutions and conditioning

From a computational perspective, given two initial values $y_0$ and $y_1$ it is easy to compute any desired solution $y_n$ of the recurrence relation by iterating the recursion (4.1). However, great care is needed in order to generate accurate numerical results in the process, since the conditioning of the recursion can be quite bad. This depends on the relative size of the different solutions of (4.1) for large $n$. More precisely, we present the following classical definition:

**Definition 4.2.1** A solution $f_n(x)$ is said to be minimal (for increasing $n$), if

$$\lim_{n \to \infty} \frac{f_n(x)}{g_n(x)} = 0,$$

(4.4)

for any other solution $g_n(x)$ independent of $f_n(x)$. In this case the solution $g_n(x)$ is said to be dominant.

**Remark 4.2.1** It is very important to specify the direction of recursion when discussing minimal and dominant solutions, because these roles typically change when we consider a different direction of recursion.

**Remark 4.2.2** Note that the minimal solution $f_n(x)$ is unique up to a multiplicative constant, since any other independent solution can be written as a combination of $f_n(x)$ and $g_n(x)$ and thus will be dominant.

Observe the example shown in Figure (4.1). The computation of the function $Y_n(x)$ is satisfactory (assuming round-off error), but the error for $J_n(x)$ becomes very large as soon as $n > x$.

Suppose that we want to compute the minimal solution $f_n(x)$ of (4.1), with initial values $f_0$ and $f_1$. Since these initial values are contaminated by round-off errors, we do not start with the exact initial values but with $y_0$ and $y_1$, and we generate a solution $y_n(x)$ that we write again as

$$y_n(x) = Af_n(x) + Bg_n(x),$$

where $A$ and $B$ are constants. Now, the relative error between $f(x)$ and $g(x)$ is

$$e_n = \left| \frac{y_n(x)}{f_n(x)} \right| = \left| \frac{Af_n(x) + Bg_n(x)}{Af_n(x)} \right| = \left| 1 + \frac{B}{Ar_n(x)} \right|,$$

where $r_n(x) = f_n(x)/g_n(x)$. Observe that this ratio can be arbitrarily large when $f_n(x)$ is minimal, since in that case $r_n \to 0$ when $n \to \infty$, and therefore
Figure 4.1: Absolute errors (in $\log_{10}$ scale) in the computation of the Bessel functions $g_n(x) = Y_n(x)$ (left) and $f_n(x) = J_n(x)$ (right) using the recurrence relation. Here $x = 31.1$.

$r_n$ will be large as well. On the other hand, the computation of a dominant solution $B g_n$ is well conditioned.

Conversely, when $n$ decreases then $r_n$ will decrease, so this suggests that backward recursion can be effective for computing solutions which are minimal when $n$ increases. We will use this fact later, and a more detailed and rigorous discussion can be found in [22, Section 4.2].

As a consequence, forward recursion is acceptable for dominant solutions (for increasing $n$) and backward recursion is the right choice for minimal solutions. Thus, the numerical analysis of a TTRR normally involves two steps: firstly, we try to determine the existence of minimal solutions, and secondly we try to identify the minimal solution. The first step can be analysed using classical results like Perron’s theorem, that we present below. The second part typically requires more information on the asymptotic behaviour of the solutions of (4.1).

**Theorem 4.2.1 (Perron)** Let us consider the recursion (4.1), and suppose that

$$a_n \sim an^\alpha, \quad b_n \sim bn^\beta, \quad ab \neq 0.$$

Let $t_1$ and $t_2$ be the roots of the characteristic polynomial $\phi(t) = t^2 + bt + a$.

(1) If $\beta > \alpha/2$, then (4.1) has two independent solutions $y_{n,1}(x)$ and
\( y_{n,2}(x) \) such that

\[
\frac{y_{n+1,1}}{y_{n,1}} \sim -bn^\beta, \quad \frac{y_{n+1,2}}{y_{n,2}} \sim -\frac{a}{b}n^{\alpha-\beta}, \quad n \to \infty.
\]

(2) If \( \beta = \alpha/2 \) and \( |t_1| > |t_2| \), then (4.1) has two independent solutions \( y_{n,1}(x) \) and \( y_{n,2}(x) \) such that

\[
\frac{y_{n+1,1}}{y_{n,1}} \sim t_1n^\beta, \quad \frac{y_{n+1,2}}{y_{n,2}} \sim t_2n^\beta, \quad n \to \infty.
\]

(3) If \( \beta = \alpha/2 \) and \( |t_1| = |t_2| \), then any non-trivial solution of (4.1) satisfies

\[
\limsup_{n \to \infty} \left[ \frac{|y_n|}{(n!)^{\alpha/2}} \right]^{1/n} = |t_1|.
\]

(4) If \( \beta < \alpha/2 \), then any non-trivial solution of (4.1) satisfies

\[
\limsup_{n \to \infty} \left[ \frac{|y_n|}{(n!)^{\alpha/2}} \right]^{1/n} = \sqrt{|a|}.
\]

See for instance [16], [50] and [22] for a proof of this result.

In cases (1) and (2) the solution \( y_{n,2} \) is minimal, whereas in cases (3) and (4) the theorem is inconclusive and further analysis is needed. Furthermore, in cases (1) and (2) it is possible to deduce a similar result for backward recursion, that is, when \( n \to -\infty \). If we rewrite (4.1) as follows:

\[
y_{n-1} + \frac{b_n}{a_n} y_n + \frac{1}{a_n} y_{n+1} = 0,
\]

(recall that we are assuming that \( a_n \neq 0 \)), then we can determine the existence of minimal solutions by analysing the behaviour of the coefficients when \( n \to -\infty \) and applying Perron’s theorem. More precisely:

**Corollary 4.2.1** Consider the recursion (4.1) and let us suppose that

\[
a_n \sim an^\alpha, \quad b_n \sim bn^\beta, \quad ab \neq 0.
\]

Let \( t_1 \) and \( t_2 \) be the roots of the characteristic polynomial \( \phi(t) = t^2 + bt + a \).

(1) If \( \beta > \alpha/2 \), then (4.1) admits two independent solutions \( y_{n,1}(x) \) and \( y_{n,2}(x) \) such that

\[
\frac{y_{n-1,1}}{y_{n,1}} \sim -\frac{1}{b}n^{-\beta}, \quad \frac{y_{n-1,2}}{y_{n,2}} \sim -\frac{b}{a}n^{\beta-\alpha}, \quad n \to -\infty.
\]
(2) If \( \beta = \alpha/2 \) and \( |t_1| > |t_2| \), then (4.1) has two independent solutions \( y_{n,1}(x) \) and \( y_{n,2}(x) \) such that

\[
\frac{y_{n-1,1}}{y_{n,1}} \sim \frac{1}{t_2} n^{-\beta}, \quad \frac{y_{n-1,2}}{y_{n,2}} \sim \frac{1}{t_1} n^{-\beta}, \quad n \to -\infty.
\]

In both cases the solution \( y_{n,2} \) is minimal when \( n \to -\infty \).

We present some examples and refer the reader to [16] and [50] for more cases. See also [21], [23] for a detailed study of recursions for Gauss hypergeometric functions, and [10] for the corresponding analysis for Kummer functions.

Example 4.2.1 The recursion for the Bessel functions \( J_n(x) \) and \( Y_n(x) \) satisfies

\[
a_n = 1, \quad b_n = \frac{-2n}{x},
\]

so \( a = 1, \alpha = 0, b = -2/x \) and \( \beta = 1 \), which corresponds to case (1) of Perron’s theorem. As a consequence, we have two solutions \( y_{n,1}(x) \) and \( y_{n,2}(x) \) such that

\[
\frac{y_{n+1,1}}{y_{n,1}} \sim \frac{2n}{x}, \quad \frac{y_{n+1,2}}{y_{n,2}} \sim \frac{x}{2n}, \quad n \to \infty,
\]

the solution \( y_{n,2} \) being minimal for increasing \( n \). From asymptotic information of the Bessel functions when \( n \) is large, see [2, Eq. 9.3.1], we can identify the solutions \( y_{n,1}(x) = Y_n(x) \) and \( y_{n,2}(x) = J_n(x) \). Therefore \( J_n(x) \) is minimal for increasing \( n \), and that is consistent with the numerical results shown before.

If we are working with classical orthogonal polynomials, it is customary to take a function of the second kind as a second independent solution of the recursion. If we have a sequence of polynomials \( P_n(x) \) orthogonal on an interval \( \gamma \) with respect to a weight function \( w(x) \) and satisfying (4.1), we take the associated polynomials (also called polynomial of the second kind, see [43]):

\[
Q_n(x) = \int_{\gamma} \frac{P_n(x) - P_n(t)}{x - t} w(t) dt. \tag{4.5}
\]

The function \( Q_n(x) \) is a polynomial of degree \( n - 1 \), and satisfies the recursion (4.1), with initial values

\[
Q_0(x) = 0, \quad Q_1(x) = \mu_0 = \int_{\gamma} w(t) dt.
\]
Furthermore, \( P_n(x) \) and \( Q_n(x) \) are independent solutions of the recurrence relation, see for instance [12, pg. 422]. Note also that if the limit
\[
\lim_{n \to \infty} \frac{Q_n(x)}{P_n(x)} = F(x)
\]
exists, then the solution
\[
S_n(x) = F(x)P_n(x) - Q_n(x)
\]
is minimal, see [17]. This is just a reformulation of Pincherle’s theorem that we shall see later on, since the polynomials \( P_n(x) \) and \( Q_n(x) \) are the denominators and the numerators (respectively) of the Jacobi continued fraction associated with the recurrence relation:
\[
a_0 \frac{x}{x - b_0} - \frac{a_1}{x - b_1} - \frac{a_2}{x - b_2} - \frac{a_3}{x - b_3} - \cdots
\]
(4.6)

If we truncate this continued fraction, then we have
\[
\frac{a_0}{x - b_0} - \frac{a_1}{x - b_1} - \cdots - \frac{a_{n-1}}{x - b_{n-1}} = \frac{Q_n(x)}{P_n(x)}.
\]

If it exists, it is possible to derive an integral representation of the function \( S_n(x) \) in the form of a Stieltjes transform, namely
\[
S_n(x) = \int_{\gamma} \frac{P_n(t)}{x-t} w(t) dt, \quad x \notin \gamma.
\]

**Example 4.2.2** In the recursion for Jacobi polynomials \( P^{(\alpha,\beta)}_n(x) \) we have
\[
a_n = \frac{(n + \alpha)(n + \beta)(2n + \alpha + \beta + 2)}{(n + 1)(n + \alpha + \beta + 1)(2n + \alpha + \beta)} \sim 1,
\]
\[
b_n = -\frac{(2n + \alpha + \beta + 1) [(2n + \alpha + \beta + 2)(2n + \alpha + \beta)x + \alpha^2 - \beta^2]}{(2n + 2)(n + \alpha + \beta + 1)(2n + \alpha + \beta)} \sim -2x.
\]

A second solution is given by \( Q^{(\alpha,\beta)}_n(x) \), a Jacobi function of the second kind. In this case \( a = 1 \), \( \alpha = 0 \), \( b = -2x \) and \( \beta = 0 \), so \( \beta = 2\alpha \), and we need to analyse the characteristic polynomial, which is
\[
\phi(t) = t^2 - 2xt + 1,
\]
with solutions
\[
t_1 = x + \sqrt{x^2 - 1}, \quad t_2 = x - \sqrt{x^2 - 1}.
\]
• If \(|x| > 1\) then both roots are real and \(|t_1| > |t_2|\), so we are in case (2) of Perron’s theorem and we have two solutions \(y_{n,1}(x)\) and \(y_{n,2}(x)\) such that
\[
\frac{y_{n+1,1}}{y_{n,1}} \sim t_1, \quad \frac{y_{n+1,2}}{y_{n,2}} \sim t_2. \quad n \to \infty,
\]
the solution \(y_{n,2}\) being minimal.

• If \(|x| < 1\) then we have two complex conjugate solutions and we are in case (3). Therefore, for every non-trivial solution of the recursion \(y_n(x)\) we have
\[
\limsup_{n \to \infty} \left[ \frac{|y_n|}{(n!)^2} \right]^{1/n} = |t_1| = 1.
\]

This Jacobi function of the second kind is given by a Stieltjes transform of the Jacobi polynomials:
\[
Q_n^{(\alpha, \beta)}(x) = \frac{1}{2} (x-1)^{-\alpha} (x+1)^{-\beta} \int_{-1}^{1} \frac{P_n^{(\alpha, \beta)}(t)}{x-t} (1-t)^{\alpha} (1+t)^{\beta} dt, \quad (4.7)
\]
see [49, 4.61.4]. It is understood that \(\alpha, \beta > -1\) and \(x \notin [-1,1]\), otherwise the integral is a Cauchy principal value.

**Exercise 4.2.1** Show that the Jacobi function of the second kind can be identified as a Gauss hypergeometric function:
\[
Q_n^{(\alpha, \beta)}(x) = 2^{n+\alpha+\beta} \frac{\Gamma(n+\alpha+1)\Gamma(n+\beta+1)}{\Gamma(2n+\alpha+\beta+2)} (x-1)^{-n-\alpha-1}(x+1)^{-\beta} \times \ _2F_1 \left( \begin{array}{c} n + \alpha + 1, n + 1 \\ 2n + \alpha + \beta + 2 \end{array} ; \frac{2}{1-x} \right). \quad (4.8)
\]

In order to show this, use the representation of Jacobi polynomials in terms of the Rodrigues formula:
\[
(1+t)^{\alpha}(1-t)^{\beta} P_n^{(\alpha, \beta)}(t) = \frac{(-1)^n}{2^n n!} \left( \frac{d}{dt} \right)^n \left[ (1-t)^{n+\alpha}(1+t)^{n+\beta} \right],
\]
see [49, 4.3.1], and integrate by parts. Use then the integral representation
\[
_2F_1 \left( \begin{array}{c} a, b \\ c \end{array} ; x \right) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)} \int_0^1 t^{b-1}(1-t)^{c-b-1}(1-xt)^{-a} dt, \quad (4.9)
\]
when \(c > b > 0\), see [2, 15.3.1].
The asymptotic behaviour of the Jacobi polynomial $P^{(\alpha, \beta)}(x)$ and the Jacobi function of the second kind is well known, see for example [49, §8.21]. From that, it can be seen that there are no minimal solutions when $x \in [-1, 1]$.

Outside the interval of orthogonality, the behaviour is also known. For example, from [49, Eq. 8.21.9, 8.71.19], we deduce that

$$\frac{Q^{(\alpha, \beta)}_n(x)}{P^{(\alpha, \beta)}_n(x)} \to 0, \quad n \to \infty, \quad x \in \mathbb{C} \setminus [-1, 1],$$

so the Jacobi function of the second kind is minimal outside of the support of the orthogonality measure. The Jacobi polynomials therefore are not minimal for any $x$, so they can be computed (in principle) using the standard recurrence relation.

**Example 4.2.3** In the recursion for Laguerre polynomials $L^{(\alpha)}_n(x)$, we have

$$a_n = \frac{-2n - \alpha - 1 + x}{n + 1} \sim -2, \quad b_n = \frac{n + \alpha}{n + 1} \sim 1.$$ 

Therefore $a = -2$, $\alpha = 0$, $b = 1$ and $\beta = 0$. The characteristic polynomial has a double root $t_{1,2} = 1$, hence Perron’s theorem is inconclusive about the existence of minimal solutions.

It is possible to deduce the behaviour of the solutions of this recursion using confluent hypergeometric functions.

**Exercise 4.2.2** Show that the Stieltjes transform of the Laguerre polynomials is

$$\int_0^\infty \frac{L^{(\alpha)}_n(t)}{x - t^\alpha} e^{-t} dt = -\Gamma(n + \alpha + 1)U(n + 1, 1 - \alpha, xe^{\pm \pi i}),$$

for $x \in \mathbb{C} \setminus [0, \infty)$, in terms of the confluent hypergeometric function of the second kind. Use the corresponding Rodrigues formula

$$t^\alpha e^{-t} L^{(\alpha)}_n(t) = \frac{1}{n!} \left( \frac{d}{dt} \right)^n [t^{\alpha+n} e^{-t}]$$

and (1.32).

It is known that in this case the function $U(a, c, x)$ is minimal with respect to the $M(a, c, x)$ function, see the analysis in [10]. From those results, it turns out that the Laguerre polynomials are never minimal, and therefore their computation by means of the recurrence relation is safe.
Exercise 4.2.3 Consider the following three term recursion:

\[ y_{\nu+1}(x) - \frac{2\nu}{x} y_{\nu}(x) - y_{\nu-1}(x) = 0, \]

where \( x \) and \( \nu \) are real and positive. Use Perron’s theorem to determine the asymptotic behavior of the solutions of this recursion when \( \nu \to \infty \).

Two independent solutions of this recursion are given by the standard modified Bessel functions \( K_\nu(x) \) and \( e^{i\nu\pi} I_\nu(x) \). Decide which one of them is the minimal solution, using the following information:

\[ I_\nu(x) = \left( \frac{x}{2} \right)^\nu \sum_{n=0}^{\infty} \frac{\left( \frac{x}{2} \right)^n}{n! \Gamma(n + \nu + 1)} \]

and

\[ K_\nu(x) = \frac{\pi}{2} \frac{I_{-\nu}(x) - I_\nu(x)}{\sin \nu \pi}. \]

The last formula should be understood as a limit when \( \nu \) assumes integer values.

4.3 Minimal solutions and continued fractions

We presented before the idea that backward recursion can be potentially useful to compute the minimal solution of a three term recurrence relation. That is, if we want to compute \( f_n \), we fix an integer \( N > n \), we compute \( f_N \) and \( f_{N-1} \) and then we apply (4.1) backwards.

Instead of working with function values \( f_N \) and \( f_{N-1} \), it is common to rephrase the problem in terms of the ratios \( H_N = f_N/f_{N-1} \), if there is no danger of dividing by 0. There are several reasons for this: one can prevent possible underflow/overflow problems, asymptotic estimates are usually easier to obtain (via Perron’s theorem, for instance), and it is possible to use the theory of associated continued fractions. For many more details on this last point, we refer the reader to [39] or [5].

Dividing by \( y_n(x) \) in (4.1) and rearranging (supposing that \( y_n(x) \neq 0 \)) gives

\[ \frac{y_n(x)}{y_{n-1}(x)} = \frac{-a_n}{b_n + \frac{y_{n+1}(x)}{y_n(x)}}. \]

Iterating this scheme we get

\[ \frac{y_n(x)}{y_{n-1}(x)} = \frac{-a_n}{b_n + \frac{-a_{n+1}}{b_{n+1} + \frac{-a_{n+2}}{b_{n+2} + \ldots}}}, \quad (4.10) \]
where we have used a standard notation for continued fractions in order to save space.

The backward algorithm consists of computing an initial index $N$ and ratio $H_N = f_N/f_{N-1}$ and then use the continued fraction expansion:

$$H_k = \frac{-a_k}{b_k + H_{k+1}}, \quad k = N - 1, N - 2, \ldots, 1. \quad (4.11)$$

So far this process is formal, and it is not clear if this continued fraction converges at all, and if it does to which ratio of solutions. This question is answered by Pincherle’s theorem below.

### 4.3.1 Convergence

The convergence of this continued fraction is given by the following theorem:

**Theorem 4.3.1 (Pincherle)** The continued fraction (4.10) converges if and only if the recursion (4.1) possesses a minimal solution $f_n(x)$. In this case the continued fraction converges to the ratio $f_n(x)/f_{n-1}(x)$.

**Proof 4.3.1** See for instance [22, Theorem 4.7]. Given a positive integer $m$, we set the initial values $y_{n+m} = 0, y_{n+m-1} = 1$ (any other value different from 0 is valid for $y_{n+m-1}$), and iterate the continued fraction:

$$\frac{y_n}{y_{n-1}} = \frac{-a_n}{b_n} = \frac{-a_n - a_{n+1} - a_{n+2} \ldots - a_{n+m-1}}{b_n + b_{n+1} + b_{n+2} + \ldots + b_{n+m-1}}.$$

Now we write $y_n(x) = Af_n(x) + Bg_n(x)$, $f_n(x)$ and $g_n(x)$ being two independent solutions of (4.1). If we impose the initial values and solve for $A$ and $B$ we obtain:

$$A = -\frac{g_{n+m}}{D_{n+m}}, \quad B = \frac{f_{n+m}}{D_{n+m}},$$

where

$$D_{n+m} = \begin{vmatrix} f_{n+m} & g_{n+m} \\ f_{n+m-1} & g_{n+m-1} \end{vmatrix}$$

is the Casorati determinant again, see (4.2), and is different from 0 because of $f_n$ and $g_n$ are independent solutions. Now

$$\frac{y_n}{y_{n-1}} = \frac{Af_n + Bg_n}{Af_{n-1} + Bg_{n-1}} = \frac{f_n - r_{n+m}g_n}{f_{n-1} - r_{n+m}g_{n-1}}.$$
where $r_{m+n} = f_{n+m}/g_{n+m}$. Since $g_n$ and $g_{n-1}$ cannot be 0 simultaneously, and the same holds for $f_n$ and $f_{n-1}$, we deduce that the limit

$$\lim_{m \to \infty} \frac{f_n - r_{n+m}g_n}{f_{n-1} - r_{n+m}g_{n-1}}$$

exists if and only if either $\lim_{m \to \infty} r_{n+m}$ or $\lim_{m \to \infty} 1/r_{n+m}$ exist. Therefore, the continued fraction (4.10) converges if and only if there is a minimal solution.

Finally, if $f_n$ is minimal then $r_{n+m} \to 0$ when $m \to \infty$, and

$$\lim_{m \to \infty} \frac{f_n - r_{n+m}g_n}{f_{n-1} - r_{n+m}g_{n-1}} = \frac{f_n}{f_{n-1}}.$$

**Example 4.3.1** The continued fraction

$$\frac{x}{2n} - \frac{x^2}{2(n+1)} - \frac{x^2}{2(n+2)} - \frac{x^2}{2(n+3)} - \cdots$$

converges to the ratio of Bessel functions $J_n(x)/J_{n-1}(x)$.

### 4.4 Computation of minimal solutions

Pincherle’s theorem provides an attractive method for evaluating minimal solutions of three term recurrence relations by means of the associated continued fractions. However, in order to use this approach effectively we need to pay attention to three computational issues:

1. How to estimate a suitable initial value $N$;
2. How to compute or approximate the initial value $H_N$;
3. How to recover the value $f_n$ from the ratio $H_n$ given by the continued fraction.

The first two problems will be analysed in this section, and an alternative is given in the form of forward methods for the continued fraction. The last one leads to Miller’s algorithm, that will be explained in Section 4.5.
4.4.1 Initial values

In some cases the initial value $N$ can be worked out from the proof of Pincherle’s theorem, see also [22, Section 4.4]. The relative error between the computed ratio $y_n/y_{n-1}$ and the minimal one $f_n/f_{n-1}$ can be written as follows:

$$e_{n,N} = \frac{y_n}{y_{n-1}} - 1 = \frac{r_N}{r_{n-1}} \frac{1 - r_{n-1}/r_n}{1 - r_N/r_{n-1}},$$

(4.12)

where $N = n + m$. Then we only need to compute this value for increasing $N$ until the error is smaller than a given tolerance.

Observe, however, that if we want to estimate the ratios in the previous formula using asymptotic information, then $n$ should be large. If this is not the case then we may need extra work to make sure that the estimation is correct. Alternatively, we can repeat the process with increasing values of $N$ and compare the results obtained for $H_1$ until they coincide within a given tolerance. This is the approach used in the example below.

4.4.2 Computation of $H_N$

The initial ratio $H_N$ may be available through asymptotic expansions (or Perron’s theorem), but it is not really needed, because one can start the backward recursion with false values, say $y_N = 0$ and $y_{N-1} = 1$ (at least one of them must be different from 0). Since backward recursion will single out a constant multiple of the minimal solution, then $y_n = C f_n$, and the constant can be determined if we have some sort of normalizing relation involving the minimal solution. This is the basis of Miller’s algorithm, that we explore in more detail in Section 4.5.

4.4.3 Backward and forward methods

Yet another possibility is given by the so called forward algorithms for the continued fraction given by $H_n = f_n/f_{n-1}$. These methods are based on the approximants of the continued fraction, which are obtained by truncation of the expansion. Namely, for $k \geq 1$ we define:

$$b_0 + \frac{a_1}{b_1 + b_2} + \frac{a_2}{b_3} + \ldots + \frac{a_k}{b_k} = \frac{A_k}{B_k}. \quad (4.13)$$

The $k$-th numerator $A_k$ and denominator $B_k$ satisfy the same three term recurrence relation:

$$\begin{pmatrix} A_k \\ B_k \end{pmatrix} = b_k \begin{pmatrix} A_{k-1} \\ B_{k-1} \end{pmatrix} + a_k \begin{pmatrix} A_{k-2} \\ B_{k-2} \end{pmatrix}, \quad (4.14)$$

102
with initial values \( A_{-1} = 1, A_0 = b_0, B_{-1} = 0, B_0 = 1 \). In some cases this recursion can be used directly in the forward direction. It is advisable, however, to use the ratios \( C_k = A_k/B_k \), because the numerators and denominators may overflow while the value of the ratio remains moderate.

Steed’s algorithm, see [22, Section 6.6.1], uses the ratio \( C_k \), which satisfies

\[
\nabla C_k = C_k - C_{k-1} = (b_k D_k - 1) \nabla C_{k-1}, \tag{4.15}
\]

where \( D_k = B_{k-1}/B_k \). This quantity can be computed recursively:

\[
D_k = \frac{1}{b_k + a_k D_{k-1}}. \tag{4.16}
\]

Hence, we initialize with the values \( C_0 = b_0, D_1 = 1/b_1, \nabla C_1 = C_1 - C_0 = a_1/b_1 \) and iterate (4.15) together with (4.16) until \( \nabla C_k/C_k \) is small enough.

Steed’s algorithm can be used in the computation of continued fractions provided that the denominators of the convergents \( C_k \) are different from 0. If the convergents may vanish, then a more sophisticated and robust alternative is given by the modified Lentz algorithm, see [22, Section 6.6.2].

### 4.5 Miller’s algorithm

#### 4.5.1 Miller’s algorithm with a known function value

If we know \( f_0 \) (or some value \( f_k \) for \( k < n \)), then we start with the value \( H_N = f_N/f_{N-1} = 0 \) and then iterate (4.11). The last computed value is \( H_1 = f_1/f_0 \), if \( f_0 \) is known then we can recover \( f_1 = H_1 f_0 \), and then move up to \( f_n \) by computing

\[
f_k = H_k f_{k-1}, \quad k = 2, 3, \ldots, n.
\]

Figure 4.2 illustrates the computation of the minimal solution \( J_n(x) \) of the Bessel recursion using Miller’s algorithm with one known value, in this case \( J_0(x) \).

#### 4.5.2 Miller’s algorithm with a normalizing sum

In a more general case, we can have a normalizing relation involving the minimal solution:

\[
\sum_{n=0}^{\infty} \lambda_n f_n = s. \tag{4.17}
\]

103
Figure 4.2: Absolute errors (in $\log_{10}$ scale) in the computation of the Bessel function $f_n(x) = J_n(x)$ using Miller’s algorithm with known value $J_0(x)$. Here $x = 31.1$.

Clearly this includes the previous case, if all but one of the coefficients $\lambda_n$ vanish. If we start the backward recursion with values $y_N = 0$ and $y_{N-1} = 1$, with $N > n$, then we obtain as before

$$y_n^{(N)} = A_N f_n + B_N g_n = \frac{1}{D_N} [-g_N f_n + f_N g_n] = \frac{g_N}{D_N} [-f_n + r_N g_n],$$

where $D_N$ is the Casorati determinant of the two solutions $f_N$ and $g_N$ and $r_N = f_N / g_N$. Since $r_N \to 0$ when $N \to \infty$, because $f_n$ is minimal, then

$$\lim_{N \to \infty} \frac{y_n^{(N)}}{A_N} = f_n.$$

In order to estimate $A_N$, we define the sum

$$s^{(N)} = \sum_{n=0}^{N} \lambda_n y_n^{(N)} \approx A_N \sum_{n=0}^{N} \lambda_n f_n,$$

and

$$f_n^{(N)} = \frac{s}{s^{(N)}} y_n^{(N)}.$$

The idea now is that imposing certain extra conditions we can have $f_n^{(N)} \to f_n$ when $N \to \infty$. More precisely, we have:
Proposition 4.5.1 It is true that \( f^N_n \to f_n \) when \( N \to \infty \) if and only if 
\[
\tau_N = r_N \sum_{n=0}^N \lambda_n g_n. \tag{4.18}
\]

Proof 4.5.1 We follow the proofs presented in [22, Section 4.6.2] and [50]. See also [16]. If we write the relative error between these two functions, we obtain:
\[
\varepsilon^N_n = \frac{f^N_n - f_n}{f_n} = \frac{sA_N}{s^{(N)}} \left( 1 - \frac{r_N}{r_n} \right) - 1.
\]
Since \( r_N \to 0 \), we only need to impose that \( A_N s/s^{(N)} \to 1 \) as \( N \to \infty \) in order for this relative error to be small for large \( N \). But
\[
A_N = \sum_{n=0}^N \lambda_n f_n - r_N \sum_{n=0}^N \lambda_n g_n = \sum_{n=0}^N \lambda_n f_n - \tau_N,
\]
and the first sum converges to \( s \). Hence, we have that \( \varepsilon^N_n \to 0 \), and hence that \( f^N_n \to f_n \) if and only if \( \tau_N \to 0 \).

Remark 4.5.1 The sum \( s^{(N)} \) can be evaluated in nested fashion:
\[
s^{(N)} = \sum_{n=0}^N \lambda_n y_n^{(N)}
\]
\[
= y_0^{(N)}(\lambda_0 + H_1^{(N)}(\lambda_1 + H_2^{(N)}(\lambda_2 + H_3^{(N)}(\ldots + H_{N-1}^{(N)}(\lambda_{N-1} + H_N^{(N)})\ldots))))),
\]
where as usual \( H_k^{(N)} = y_k^{(N)}/y_{k-1}^{(N)} \), assuming that \( y_k^{(N)} \) do not vanish.

In an algorithm, we take as initial value \( s_N = H_N^{(N)} = 0 \), and then
\[
s_k^{(N)} = H_k^{(N)}(\lambda_k + s_{k+1}^{(N)}), \quad k = N - 1, N - 2, \ldots, 1.
\]
Finally, we have \( s_0^{(N)} = y_0^{(N)}(\lambda_0 + s_1^{(N)}) \). Then
\[
f_0^{(N)} = \frac{s}{s^{(N)} y_0^{(N)}} = \frac{s}{\lambda_0 + s_1^{(N)}},
\]
and then
\[
f_k^{(N)} = H_k^{(N)} f_{k-1}^{(N)}, \quad k = 1, 2, \ldots, n.
\]
In most cases, normalizing relations can be found in the literature, but care must be paid to stability issues, since some of them might not be suitable numerically. For a more detailed discussion see [22].
Example 4.5.1 For the Bessel functions $J_n(x)$ we have the following normalization, see \cite[Eq. 9.1.46]{2}:

$$J_0(x) + 2 \sum_{n=1}^{\infty} J_{2n}(x) = 1,$$

so $\lambda_0 = 1$, $\lambda_{2n} = 2$ and $\lambda_{2n-1} = 0$ for $n \geq 1$. Other possibilities are explained in \cite{16}.

4.6 Related problems

We conclude this chapter with related questions and more general problems connected with recurrence relations. Standard references are \cite{16} and \cite{53}.

- Anomalous behaviour of recurrence relations. The definitions of minimal and dominant solutions that we have given here are asymptotic (when $n \to \infty$). In practice, when computing for finite values of $n$ it may happen that solutions have temporarily a different and even opposite behaviour to the one predicted by the theory. In \cite{9} several examples are given in which solutions that are eventually dominant behave as minimal for moderate $n$. This can have important numerical consequences, like convergence of the continued fraction to the wrong ratio of solutions.

- Inhomogeneous and higher order recurrence relations. Inhomogeneous and higher order recurrence relations are an important part of the general theory, see \cite{12}, and they also feature in the study of non-classical orthogonal polynomials (including matrix orthogonal polynomials). In this case, the ideas of minimal and dominant solutions remain the same, but the asymptotic analysis may be more complicated because we can have superminimal or subdominant solutions. For instance, if $f_n$ is minimal, $g_n$ is dominant and we have $h_n$ such that

$$\lim_{n \to \infty} \frac{h_n}{g_n} = 0, \quad \lim_{n \to \infty} \frac{h_n}{f_n} = \infty.$$  

In this case neither forward nor backward recursion is advisable for computing $h_n$. An alternative method for these situations was provided by F. Olver, see \cite{22}. The idea of the method is to formulate the problem of computing a subdominant solution as a boundary value problem, instead of an initial value one.
Nonlinear recursions. The theory that we have exposed deals only with linear recurrence relations, but there is an increasing interest in nonlinear difference equations, in connection with the relations satisfied by the recursion coefficients themselves $a_n$ and $b_n$, and also in the context of discrete Painlevé equations. See [52] for a recent account.
Bibliography


