# Hypergeometric Functions

Gert Heckman
Radboud University Nijmegen
G.Heckman@math.ru.nl

September 22, 2011

# Contents

## Preface

The Euler–Gauss hypergeometric function

$$F(\alpha,\beta,\gamma;z)=\sum_{k=0}^{\infty}\frac{\alpha(\alpha+1)\cdots(\alpha+k-1)\beta(\beta+1)\cdots(\beta+k-1)}{\gamma(\gamma+1)\cdots(\gamma+k-1)k!}z^k$$

was introduced by Euler in the $18^{th}$ century, and was well studied in the $19^{th}$ century among others by Gauss, Riemann, Schwarz and Klein. The numbers $\alpha,\beta,\gamma$ are called the parameters, and $z$ is called the variable.

On the one hand, for particular values of the parameters this function appears in various problems. For example

$$\begin{aligned}
(1-z)^{-\alpha} &= F(\alpha,1,1;z)\\
\arcsin z &= 2zF(1/2,1,3/2;z^2)\\
K(z) &= \frac{\pi}{2}F(1/2,1/2,1;z^2)\\
P_n^{(\alpha,\beta)}(z) &= \frac{\alpha(\alpha+1)\cdots(\alpha+n)}{n!}F(-n,\alpha+\beta+n+1,\alpha+1;\frac{1-z}{2})
\end{aligned}$$

with $K(z)$ the Jacobi elliptic integral of the first kind given by

$$K(z)=\int_0^1\frac{dx}{\sqrt{(1-x^2)(1-z^2x^2)}}\ ,$$

and $P_n^{(\alpha,\beta)}(z)$ the Jacobi polynomial of degree $n$, normalized by

$$P_n^{(\alpha,\beta)}(1)=\binom{\alpha+n}{n}\ .$$

On the other hand, the hypergeometric differential equation (of which $F(\alpha,\beta,\gamma;z)$ is a solution) served as a guiding example for the general theory of ordinary differential equations in a complex domain. For example, the calculation of the monodromy of the hypergeometric equation led Riemann to formulate the so called Riemann–Hilbert problem, later reformulated by Hilbert in his famous list of 1900 as Problem 21.

A natural generalization of the Euler–Gauss hypergeometric function is the Clausen–Thomae hypergeometric function

$$F(\alpha,\beta;z)=\sum_{k=0}^{\infty}\frac{(\alpha_1)_k(\alpha_2)_k\cdots(\alpha_n)_k}{(\beta_1)_k(\beta_2)_k\cdots(\beta_n)_k}z^k$$

with $(\lambda)_k = \Gamma(\lambda+k)/\Gamma(\lambda) = \lambda(\lambda+1)\cdots(\lambda+k-1)$ the so called Pochhammer symbol. The numbers $\alpha = (\alpha_1, \cdots, \alpha_n)$ are called the numerator parameters and $\beta = (\beta_1, \cdots, \beta_n)$ the denominator parameters. Usually $\beta_n = 1$ so that $(\beta_n)_k = k!$ and therefore the Euler–Gauss hypergeometric function has numerator parameters $(\alpha_1, \alpha_2) = (\alpha, \beta)$ and denominator parameter $\beta_1 = \gamma$. Many (but not all) results of the Euler–Gauss hypergeometric function can be generalized for the Clausen–Thomae hypergeometric function. For very particular values of the parameters the Clausen–Thomae hypergeometric function appeared in modern mathematics in the context of mirror symmetry for Calabi–Yau threefolds.

After a fairly detailed treatment of these two classical hypergeometric functions of the $19^{th}$ century we discuss a multivariable analogue of the Euler–Gauss hypergeometric function: the hypergeometric function

$$F(\lambda, k; t)$$

associated with a root system $R$. These functions generalize the Euler–Gauss hypergeometric function (for the rank one root system) and the elementary spherical functions on a real semisimple Lie group (for particular parameter values). They were introduced and studied in collaboration with Eric Opdam and the lecturer in the eighties and nineties of the $20^{th}$ century. These special functions are intimately connected with the Calogero–Moser system of $n$ points on a circle, under influence of an inverse square potential. The classical integrability of this system was conjectured by Calogero and proved by Moser. The root system hypergeometric functions appear as the simultaneous eigenfunctions of the Schrödinger operator and its conserved operators for the quantum integrable system. In order to make these lecture notes self contained the basic properties of root systems and Weyl groups are included.

These notes are written for a series of lectures at the Tsinghua University of Beijing in the fall of 2011. I would like to thank the Mathematical Sciences Center for their hospitality.

# 1 Linear differential equations

## 1.1 The local existence problem

We shall write $\partial = d/dz$ where $z$ is the standard coordinate in the complex plane. Let us consider the two linear ordinary differential equations

$$(\partial^n + a_1\partial^{n-1} + \cdots + a_{n-1}\partial + a_n)f = 0$$

$$(\partial + A)F = 0$$

with scalar coefficients $a_1, \cdots, a_n$ and matrix coefficient $A = (a_{ij})_{1\le i,j\le n}$ holomorphic on some domain $Z \subset \mathbb{C}$.

The first linear differential equation is a scalar equation of order $n$: the coefficients $a_1(z), \cdots, a_n(z)$ are holomorphic functions on $Z$, and we shall seek holomorphic solutions $f(z)$ on suitable open subsets of $Z$. The second linear differential equation is a first order matrix equation: the entries $a_{ij}(z)$ of the matrix $A(z)$ are holomorphic functions on $Z$, and we shall seek vector valued holomorphic solutions

$$F(z) = (f_1(z), \cdots, f_n(z))^t$$

on suitable open subsets of $Z$. The (local) existence problem of higher order scalar equations can be reduced to the (local) existence problem of first order matrix equations.

**Theorem 1.1.** *Suppose holomorphic functions $a_1(z), \cdots, a_n(z)$ have been given on a domain $Z \subset \mathbb{C}$. Define the matrix valued holomorphic function $A(z)$ on $Z$ by*

$$\begin{pmatrix} 0 & -1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & -1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & -1 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & -1 \\ a_n & a_{n-1} & a_{n-2} & a_{n-3} & \cdots & a_2 & a_1 \end{pmatrix}$$

*If the vector valued function $F(z) = (f_1(z), \cdots, f_n(z))^t$ is a solution of*

$$(\partial + A)F = 0$$

*then $f(z) = f_1(z)$ is a solution of*

$$(\partial^n + a_1\partial^{n-1} + \cdots + a_{n-1}\partial + a_n)f = 0$$

*and $f_{j+1}(z) = \partial f_j(z)$ for $j = 1, \cdots, n-1$.*

*Proof.* With the matrix valued function $A(z)$ as above and the vector valued function $F(z) = (f_1(z), \cdots, f_n(z))^t$ the equation $(\partial + A)F = 0$ amounts to

$$\begin{pmatrix} \partial f_1 \\ \partial f_2 \\ \vdots \\ \partial f_{n-1} \\ \partial f_n \end{pmatrix} + \begin{pmatrix} -f_2 \\ -f_3 \\ \vdots \\ -f_n \\ a_n f_1 + \cdots + a_1 f_n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}$$

which in turn is equivalent to the equations

$$f_1 = f, f_2 = \partial f, \cdots, f_n = \partial^{n-1} f$$

$$(\partial^n + a_1 \partial^{n-1} + \cdots + a_{n-1}\partial + a_n) f = 0$$

which proves the theorem. □

For $r > 0$ let $D_r = \{z \in \mathbb{C}; |z| < r\}$ be the open disc around $z = 0$ with radius $r$. In this section we shall carry out a local analysis for the domain $Z = D_r$. Consider the matrix equation

$$(\partial + A)F = 0$$

with $A = (a_{ij})$ and $a_{ij} = a_{ij}(z)$ holomorphic on $D_r$. Do there exist solutions $F = (f_1, \cdots, f_n)^t$ with $f_j = f_j(z)$ holomorphic on $D_r$, and if yes how many? For this purpose develop $A(z)$ in a power series

$$A(z) = \sum_0^\infty A_k z^k$$

and substitute a formal power series

$$F(z) = \sum_0^\infty F_k z^k$$

with $F_k \in \mathbb{C}^n$ undetermined coefficients.

**Proposition 1.2.** *The formal power series $F(z) = \sum_0^\infty F_k z^k$ is a formal solution of $(\partial + A)F = 0$ with $A(z) = \sum_0^\infty A_k z^k$ if and only if*

$$(k+1)F_{k+1} + \sum_{l=0}^{k} A_{k-l} F_l = 0$$

*for all $k \in \mathbb{N} = \{0, 1, 2, \cdots\}$.*

*Proof.* Substitution in $(\partial + A)F = 0$ gives

$$\sum_{k=1}^{\infty} kF_k z^{k-1} + \sum_{k,l\geq 0} A_k F_l z^{k+l} =$$

$$\sum_{k=0}^{\infty}((k+1)F_{k+1} + \sum_{l=0}^{k} A_{k-l}F_l)z^k = 0$$

and the proposition is clear. □

On the level of formal power series we get:

On the level of formal power series we get:
$F_0 \in \mathbb{C}^n$ is undetermined and can be freely choosen,
$F_1 = -A_0 F_0$,
$F_2 = -(A_1 F_0 + A_0 F_1)/2 = (A_0^2 - A_1)F_0/2$,
$F_3 = -(A_2 F_0 + A_1 F_1 + A_0 F_2)/3 = (-A_0^3 + A_0 A_1 + 2A_1 A_0 - 2A_2)F_0/6$,
$\cdots$

So given $F_0$ the $F_k$ with $k \geq 1$ can be explicitly computed via the recurrence relation. Using Theorem 1.1 we obtain the following result.

**Corollary 1.3.** *The $n^{th}$ order scalar equation*

$$(\partial^n + a_1\partial^{n-1} + \cdots + a_{n-1}\partial + a_n)f = 0$$

*on the disc $D_r$ has a unique formal power series solution $f(z) = \sum f_k z^k$ for freely chosen $f_0, \cdots, f_{n-1} \in \mathbb{C}$.*

Our next goal is to show that these formal power series solutions are in fact convergent power series. The following lemma is familiar from complex function theory.

**Lemma 1.4.** *A formal power series $\sum a_k z^k$ (with coefficients in a Banach space) is convergent on the disc $D_r$ with radius $r > 0$ if and only if for each $\rho \in (0, 1)$ there exists a constant $M_\rho \geq 0$ such that*

$$|a_k| \leq M_\rho(\rho r)^{-k}$$

*for all $k \in \mathbb{N}$.*

**Theorem 1.5.** *If the coefficients of the matrix equation $(\partial + A)F = 0$ are convergent on the disc $D_r$ then the formal power series solution $\sum F_k z^k$ with $F_0$ undetermined and $F_{k+1}$ given by the recurrence relation of Proposition 1.2 also converges on the disc $D_r$.*

*Proof.* The power series $\sum A_k z^k$ converges on $D_r$, and therefore we have an estimate of the form (switch from $M_\rho$ to $M_\rho(\rho r)^{-1}$)

$$\forall \rho \in (0,1)\ \exists M_\rho \geq 0\ ;\ |A_k| \leq M_\rho (\rho r)^{-k-1}\ \forall k \in \mathbb{N}.$$

We claim that this implies an estimate for $F_k$ of the form

$$|F_k| \leq M_\rho(M_\rho+1)\cdots(M_\rho+k-1)(\rho r)^{-k}|F_0|/k!\ \forall k \in \mathbb{N}.$$

We prove this claim by induction on $k \in \mathbb{N}$. The case $k = 0$ is trivial. Using the recurrence relation and the induction hypothesis we get

$$|F_{k+1}| \leq (k+1)^{-1}\{\sum_{l=0}^{k} M_\rho(\rho r)^{-k+l-1} \cdot \frac{\Gamma(M_\rho+l)}{\Gamma(M_\rho)l!}(\rho r)^{-l}|F_0|\}.$$

Using the formula (easily proved by induction on $k$)

$$\sum_{l=0}^{k} \frac{\Gamma(M+l)}{\Gamma(M)l!} = \frac{\Gamma(M+k+1)}{\Gamma(M+1)k!}$$

we arrive at the estimate

$$\begin{aligned}|F_{k+1}| &\leq \frac{1}{(k+1)} \cdot \frac{M_\rho\Gamma(M_\rho+k+1)}{\Gamma(M_\rho+1)k!}(\rho r)^{-k-1}|F_0| \\ &= \frac{\Gamma(M_\rho+k+1)}{\Gamma(M_\rho)(k+1)!}(\rho r)^{-(k+1)}|F_0|\end{aligned}$$

which proves our claim.

For each $M \geq 0$ (even for $M \in \mathbb{C}$) the binomial series

$$(1-z)^{-M} = \sum_{0}^{\infty} \frac{\Gamma(M+k)}{\Gamma(M)k!} z^k$$

is convergent on the unit disc $D_1$. Hence by Lemma 1.4 we get the estimate

$$\forall \sigma \in (0,1)\ \exists N_{\rho,\sigma} > 0\ ;\ \frac{\Gamma(M_\rho+k)}{\Gamma(M_\rho)k!} \leq N_{\rho,\sigma}\sigma^{-k}\ \forall k \in \mathbb{N}.$$

So finally we arrive at

$$\forall \rho, \sigma \in (0,1)\ \exists N_{\rho,\sigma} > 0\ ;\ |F_k| \leq |F_0| N_{\rho,\sigma}(\rho\sigma r)^{-k}\ \forall k \in \mathbb{N},$$

which in turn implies

$$\forall \rho \in (0,1)\ \exists L_\rho > 0\ ;\ |F_k| \leq L_\rho(\rho r)^{-k}\ \forall k \in \mathbb{N}.$$

Indeed just take $L_\rho = |F_0| N_{\sqrt{\rho},\sqrt{\rho}}$. Now apply Lemma 1.4. □

**Corollary 1.6.** *Let $(\partial + A)F = 0$ be a first order matrix equation with coefficients $A(z) = (a_{ij}(z))_{1\le i,j\le n}$ holomorphic in a domain $Z \subset \mathbb{C}$. For every point $z_0 \in Z$ and $F_0 \in \mathbb{C}^n$ there is a unique local holomorphic solution $F(z)$ around $z_0$ with initial value $F(z_0) = F_0$.*

**Corollary 1.7.** *Let $(\partial^n + a_1\partial^{n-1} + \cdots + a_{n-1}\partial + a_n)f = 0$ be an $n^{th}$ order scalar equation with coefficients $a_1(z), \cdots, a_n(z)$ holomorphic in a domain $Z \subset \mathbb{C}$. For every point $z_0 \in Z$ and complex numbers $f_0, \cdots, f_{n-1}$ there is a unique local holomorhic solution $f(z)$ around $z_0$ with initial conditions*

$$f(z_0) = f_0, \partial f(z_0) = f_1, \cdots, \partial^{n-1} f(z_0) = f_{n-1}.$$

**Example 1.8.** *The second order differential equation $(\partial^2 + (1/z)\partial)f = 0$ on the punctured complex plane $\mathbb{C}^\times$ has*

$$f(z) = \log z = \log(1 + (z-1)) = (z-1) - (z-1)^2/2 + (z-1)^3/3 + \cdots$$

*as unique local holomorphic solution around $z = 1$ with $f(1) = 0, \partial f(1) = 1$. The differential equation provides the analytic continuation*

$$\log z = \int_1^z \frac{d\zeta}{\zeta}$$

*with the line integral taken along a curve in $\mathbb{C}^\times$ from 1 to $z$.*

**Remark 1.9.** *Suppose that the coefficients of the linear differential equation $(\partial + A)F = 0$ in a domain $Z$ also depend in a holomorphic way on a complex parameter $\alpha$, so $A = A(\alpha, z)$ with $\alpha$ a parameter and $z$ the variable of the differential equation, so $\partial = d/dz$. Suppose that the power series*

$$A(\alpha, z) = \sum_0^\infty A_k(\alpha) z^k$$

*converges on $D_r$ in a locally uniform way in $\alpha$, so the constant $M_\rho = M_\rho(\alpha)$ in Lemma 1.4 is locally independent of $\alpha$. The estimates in Theorem 1.5 are also locally uniform in $\alpha$, so the power series $\sum F_k(\alpha)z^k$ on $D_r$ converges in a locally uniform way in $\alpha$. Hence for an initial value $F_0(\alpha) \in \mathbb{C}^n$ that is holomorphic in $\alpha$ the unique solution $F(\alpha, z)$ of $(\partial + A)F = 0$ with initial value $F(\alpha, 0) = F_0(\alpha)$ depends also in a holomorphic way on $\alpha$.*

## 1.2 The fundamental group

Let $Z$ be a connected topological space. The example to have in mind is a domain $Z$ in $\mathbb{C}$.

**Definition 1.10.** *A path in $Z$ is a continuous map $\gamma : [0,1] \to Z, t \mapsto \gamma(t)$. The point $\gamma(0)$ is called the begin point and the point $\gamma(1)$ the end point of $\gamma$. If begin and end point of $\gamma$ coincide then $\gamma$ is called a loop with base point $\gamma(0) = \gamma(1)$.*

**Definition 1.11.** *Let $\gamma_1$ and $\gamma_2$ be two paths in $Z$ with equal begin points $\gamma_1(0) = \gamma_2(0)$ and equal end points $\gamma_1(1) = \gamma_2(1)$. The paths $\gamma_1$ and $\gamma_2$ are called homotopic if there exists a continuous map $h : [0,1] \times [0,1] \to Z, (s,t) \mapsto h(s,t)$ such that*

$$h(0,t) = \gamma_1(t)\ ,\ h(1,t) = \gamma_2(t)\ \forall t \in [0,1]\ ,$$

$$h(s,0) = \gamma_1(0) = \gamma_2(0)\ ,\ h(s,1) = \gamma_1(1) = \gamma_2(2)\ \forall s \in [0,1]\ .$$

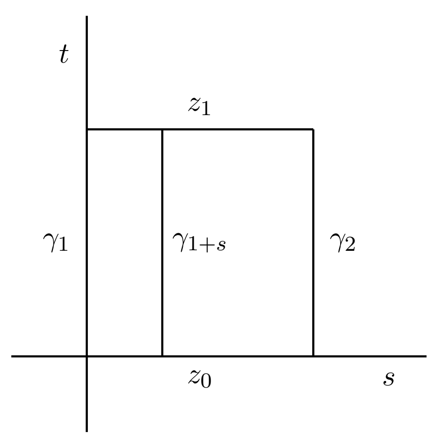
*The map $h$ is called the homotopy between the paths $\gamma_1$ and $\gamma_2$.*

In other words the two paths $\gamma_1$ and $\gamma_2$ are homotopic if there exists a one parameter continuous family (with parameter $s \in [0,1]$) of paths

$$\gamma_{1+s} : [0,1] \to Z$$

$$\gamma_{1+s}(0) = \gamma_1(0) = \gamma_2(0)\ ,\ \gamma_{1+s}(1) = \gamma_1(1) = \gamma_2(1)\ \forall s \in [0,1]\ .$$

The link with our previous notation is $\gamma_{1+s}(t) = h(s,t)$. If $\gamma_1(0) = \gamma_2(0) = z_0$ and $\gamma_1(1) = \gamma_2(1) = z_1$ then we draw the following schematic picture.



We shall write $\gamma_1 \sim \gamma_2$ if the paths $\gamma_1$ and $\gamma_2$ in $Z$ with equal begin points and equal end points are homotopic. It is easy to show that being homotopic is an equivalence relation. The equivalence class of a path $\gamma : [0,1] \to Z$ is denoted by $[\gamma]$.

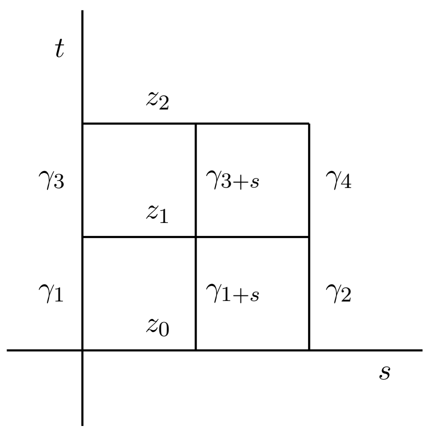
**Definition 1.12.** *Let $z_1, z_1, z_2 \in Z$ be three points and let $\gamma_1, \gamma_2 : [0,1] \to Z$ be two paths in $Z$ with $\gamma_1(0) = z_0, \gamma_1(1) = \gamma_2(0) = z_1, \gamma_2(1) = z_2$. We define a new path $\gamma_2\gamma_1 : [0,1] \to Z$ by*

$$\gamma_2\gamma_1(t) = \gamma_1(2t) \ \forall t \in [0, 1/2] \ ,$$

$$\gamma_2\gamma_1(t) = \gamma_2(2t-1) \ \forall t \in [1/2, 1] \ .$$

*The path $\gamma_2\gamma_1$ is called the product of $\gamma_2$ and $\gamma_1$, and is always taken in the order start with $\gamma_1$ and then followed by $\gamma_2$.*

It is easy to show that if $\gamma_1 \sim \gamma_2$ and $\gamma_3 \sim \gamma_4$ and the end point $z_1$ of $\gamma_1, \gamma_2$ coincides with the begin point $z_1$ of $\gamma_3, \gamma_4$ then $\gamma_3\gamma_1 \sim \gamma_4\gamma_2$. Here is a schematic picture of the homotopy.

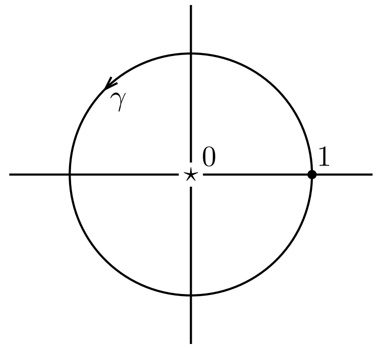



Hence the product $[\gamma_2][\gamma_1]$ of the homotopy classes of paths $\gamma_2$ and $\gamma_1$ as in Definition 1.12 is well defined. We leave it as an exercise to show that the product of paths is associative on homotopy classes of paths.

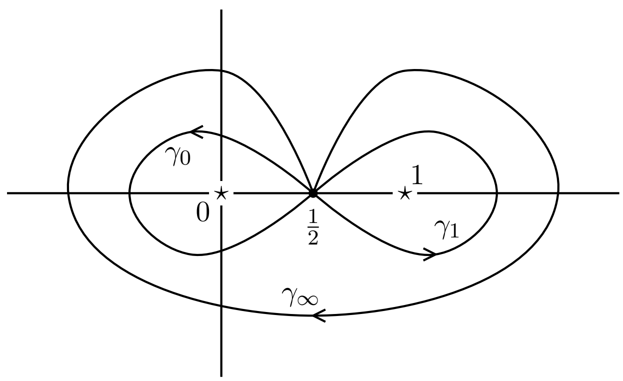
**Theorem 1.13.** *For $z_0 \in Z$ a fixed point let $\Pi_1(Z, z_0)$ denote the collection of homotopy classes of loops in $Z$ with base point (i.e. begin and end point) $z_0$. The product rule on paths in $Z$ according to Definition 1.12 defines a group structure on $\Pi_1(Z, z_0)$. The unit element is represented by the constant path $\epsilon(t) = z_0 \ \forall t \in [0,1]$ at $z_0$. The inverse $[\gamma]^{-1}$ of $[\gamma] \in \Pi_1(Z, z_0)$ is represented by the loop $\gamma^{-1}(t) = \gamma(1-t) \ \forall t \in [0,1]$.*

**Definition 1.14.** *The group $\Pi_1(Z, z_0)$ is called the fundamental group of the connected topological space $Z$ with base point $z_0$.*

**Example 1.15.** *Let $Z = \mathbb{C}^\times = \mathbb{C} - \{0\}$ and $z_0 = 1$. Let $\gamma(t) = exp(2\pi it)$ for $t \in [0, 1]$. Then $\Pi_1(Z, z_0)$ is a cyclic group with generator $[\gamma]$.*




**Example 1.16.** *Let $Z = \mathbb{P} - \{0, 1, \infty\} = \mathbb{C} - \{0, 1\}$ with $\mathbb{P} = \mathbb{C} \cup \{\infty\}$ the projective line and take $z_0 = 1/2$. Choose loops $\gamma_0, \gamma_1, \gamma_\infty$ around the points $0, 1, \infty$ respectively as in the picture.*




*It is easy to see that $[\gamma_\infty][\gamma_1][\gamma_0] = 1$ in $\Pi_1(Z, 1/2)$. It turns out that $\Pi_1(Z, 1/2)$ is isomorphic to the group on three generators $[\gamma_0], [\gamma_1], [\gamma_\infty]$ with the single relation $[\gamma_\infty][\gamma_1][\gamma_0] = 1$.*

## 1.3 The monodromy representation

Suppose $G$ is a group and $V$ is a finite dimensional vector space over the complex numbers $\mathbb{C}$. Let $\mathrm{GL}(V)$ denote the group of invertible linear operators on $V$. A representation $(\pi, V)$ of $G$ on $V$ is a homomorphism $\pi : G \to \mathrm{GL}(V)$. If $(\pi_1, V_1)$ and $(\pi_2, V_2)$ are two representations of a group

$G$, then a linear map $A \in \mathrm{Hom}(V_1, V_2)$ is called an intertwiner from $(\pi_1, V_1)$ to $(\pi_2, V_2)$ if

$$A\pi_1(g) = \pi_2(g)A \;\forall g \in G \;.$$

The intertwiners from $(\pi_1, V_1)$ to $(\pi_2, V_2)$ form a linear subspace of the vector space $\mathrm{Hom}(V_1, V_2)$ of linear maps from $V_1$ to $V_2$, denoted by $\mathrm{Hom}(V_1, V_2)^G$. A bijective intertwiner $A \in \mathrm{Hom}(V_1, V_2)^G$ is called an equivalence between $(\pi_1, V_1)$ and $(\pi_2, V_2)$. It is easy to check that equivalence of representations of a group $G$ is an equivalence relation on the set of representations $(\pi, V)$ of $G$.

Given a representation $(\pi, V)$ of $G$ a linear subspace $U \subset V$ is called invariant if $\pi(g)u \in U \;\forall g \in G \;\forall u \in U$. In this case we denote $\pi_U(g) = \pi(g)|_U$ and call $(\pi_U, U)$ a subrepresentation of $(\pi, V)$. A representation $(\pi, V)$ of $G$ is called irreducible if the only invariant linear subspaces of $V$ are the trivial ones 0 and $V$. Given a representation $(\pi, V)$ of $G$ a Hermitian form $\langle \cdot, \cdot \rangle$ on V (which by definition is linear in the first argument, and antilinear in the second argument) is called invariant if

$$\langle \pi(g)u, \pi(g)v \rangle = \langle u, v \rangle \;\forall u, v \in V \;.$$

The kernel of an invariant Hermitian form is easily seen to be an invariant linear subspace. In particular a nonzero invariant Hermitian form on an irreducible representation space $V$ is nondegenerate. A representation $(\pi, V)$ of $G$ is called unitary if there exists a positive definite invariant Hermitian form on $V$.

Suppose $Z \subset \mathbb{C}$ is a domain. Suppose we are given an $n^{th}$ order linear differential equation

$$(\partial^n + a_1\partial^{n-1} + \cdots + a_{n-1}\partial + a_n)f = 0$$

with coefficients $a_1, \cdots, a_n$ holomorhic in $Z$. Fix a base point $z_0 \in Z$ and let $V_0$ be the linear space of local holomorphic solutions around $z_0$. We know that the dimension of $V_0$ is equal to $n$. Suppose $\gamma$ is a path in $Z$ with begin point $z_0$ and end point $z_1$, and let $V_1$ be the linear space of local holomorphic solutions around $z_1$. Analytic continuation of local solutions along $\gamma$ depends only on the homotopy class $[\gamma]$ of $\gamma$ in $Z$. Therefore we have defined a linear monodromy operator

$$M([\gamma]) : V_0 \to V_1 \;.$$

The monodromy operator corresponding to the product of two paths is clearly equal to the product of the two monodromy operators corresponding to the individual paths. Restricting to loops in $Z$ with base point $z_0$

therefore defines the monodromy representation

$$M : \Pi_1(Z, z_0) \to \mathrm{GL}(V_0) \ .$$

The monodromy representation is a powerful (in general transcendental) invariant of a linear differential equation, and was introduced by Riemann.

For a first order matrix linear differential equation

$$(\partial + A)F = 0$$

of size $n$ by $n$ in a domain $Z$ the monodromy is defined likewise. If $\gamma$ is a path in $Z$ with begin point $z_0$ and end point $z_1$ and $F$ is a local solution around $z_0$ then $M([\gamma])F$ is a local solution around $z_1$. The local solution $F$ around $z_0$ is completely determined by $F(z_0) \in \mathbb{C}^n$. Likewise the local solution $M([\gamma])F$ is completely determined by $(M([\gamma])F)(z_1) \in \mathbb{C}^n$. By abuse of notation there exists an invertible matrix $M([\gamma]) \in \mathrm{GL}(n, \mathbb{C})$ such that

$$(M([\gamma])F)(z_1) = M([\gamma])(F(z_0)) \ .$$

Here $M([\gamma])$ on the left side is the monodromy operator, while $M([\gamma])$ on the right side is the associated matrix. If the scalar and matrix equations

$$(\partial^n + a_1\partial^{n-1} + \cdots + a_{n-1}\partial + a_n)f = 0$$

$$(\partial + A)F = 0$$

are related by Theorem 1.1, so that

$$F(z) = (f_1(z), \cdots, f_n(z))^t \ , \ f(z) = f_1(z) \ , \ f_{j+1}(z) = \partial f_j(z)$$

for $j = 1, \cdots, n-1$, then $M([\gamma])F$ corresponds likewise to $M([\gamma])f$. Indeed the operator $\partial$ commutes with monodromy.

### 1.4 Regular singular points

Suppose $Z \subset \mathbb{C}$ is a domain with base point $z_0$. Consider the linear system of differential equations

$$(\partial + A)F = 0$$

with $A = (a_{ij})_{1 \le i,j \le n}$ and $a_{ij} = a_{ij}(z)$ holomorphic functions on $Z$. Suppose we choose a basis $F_1, \cdots, F_n$ of the local solution space $V_0$ around $z_0$. If we write $F_j = (F_{1j}, \cdots, F_{nj})^t$ then $F = (F_{ij})$ is called a local solution matrix around $z_0$. Let $\gamma \in \Pi_1(Z, z_0)$ be a loop in $Z$ based at $z_0$ and let $M = M([\gamma]) : V_0 \to V_0$ be the corresponding monodromy operator. The monodromy matrix $(m_{jk})_{1\leq j,k\leq n}$ with respect to the basis $F_1, \cdots, F_n$ is defined by the usual relations

$$M(F_k) = \sum m_{jk} F_j \ .$$

Under the monodromy operator $M$ the matrix entry $F_{ik}$ of the local solution matrix $F = (F_1, \cdots, F_n)$ transforms into $M(F_{ik}) = \sum m_{jk} F_{ij}$. In other words we get

$$M(F) = FM$$

so that the monodromy operator $M$ acts on the local solution matrix $(F_{ij})$ by multiplication on the right with the monodromy matrix $(m_{jk})$.

**Example 1.17.** *Let $\theta = z\partial = z\frac{d}{dz}$ and consider the linear system*

$$(\partial + A/z)F = 0 \ \Leftrightarrow (\theta + A)F = 0$$

*with $A = (a_{ij})_{1\leq i,j\leq n} \in \mathrm{Mat}(n, \mathbb{C})$ a scalar matrix. For the domain $Z$ we take $\mathbb{C}^\times$, say with base point $z_0 = 1$. As local solution matrix around $z_0 = 1$ we can take*

$$F(z) = z^{-A} = e^{-A \log z}$$

*which defines a single valued solution matrix on $\mathbb{C} - (-\infty, 0]$ by taking the branch $\log 1 = 0$. If $\gamma(t) = e^{2\pi i t}$ for $t \in [0, 1]$ then the monodromy operator $M = M([\gamma])$ has monodromy matrix*

$$M = e^{-2\pi i A} \ .$$

Let us now take for the domain $Z$ the punctured disc

$$D^\times_{r_0} = \{z \in \mathbb{C}; 0 < |z| < r_0\}$$

for some $r_0 > 0$ with base point $z_0 = r_0/2$, and consider the linear system $(\partial + A)F = 0$ with coefficients holomorphic on $D^\times_{r_0}$. Let $F = (F_1, \cdots, F_n)$ be a local solution matrix around $z_0 = r_0/2$. Let $\gamma$ be the loop $\gamma(t) = r_0 e^{2\pi i t}/2$ and let $M = M([\gamma])$ be the monodromy operator.

**Proposition 1.18.** *The exponential map* $\exp : \mathrm{Mat}(n, \mathbb{C}) \to \mathrm{GL}(n, \mathbb{C})$ *is a surjection.*

*Proof.* This follows from the Jordan decomposition. □

Choose $\Gamma \in \mathrm{Mat}(n, \mathbb{C})$ with $M = e^{2\pi i\Gamma}$. The matrix $\Gamma$ is not unique, but can be fixed uniquely by the requirement $0 \leq \Re\lambda < 1$ for each eigenvalue $\lambda$ of $\Gamma$. Consider the matrix valued holomorphic function

$$G(z) = F(z)z^{-\Gamma} = F(z)e^{-\Gamma \log z}$$

around $z_0 = r_0/2$ with the branch fixed by $\log(r_0/2) > 0$. Both $F(z)$ and $z^{-\Gamma}$ have analytic continuation along loops $\gamma$ in $D^{\times}_{r_0}$ based at $z_0$, and therefore also the product $G(z)$ has this analytic continuation. For the monodromy $M = M([\gamma])$ along $\gamma$ we find

$$M(G(z)) = M(F(z))M(z^{-\Gamma}) = F(z)Mz^{-\Gamma}e^{-2\pi i\Gamma} = G(z)$$

because $z^{-\Gamma}e^{-2\pi i\Gamma} = e^{-2\pi i\Gamma}z^{-\Gamma}$ and $M = e^{2\pi i\Gamma}$. Therefore the function $G(z)$ has trivial monodromy, so is univalued and holomorphic on $D^{\times}_{r_0}$. In particular the function $G(z)$ has a Laurent series expansion

$$G(z) = \sum_{k=-\infty}^{\infty} G_k z^k$$

with coefficients matrices $G_k \in \mathrm{Mat}(n, \mathbb{C})$, which converges on $D^{\times}_{r_0}$. The original local solution matrix $F(z)$ is therefore of the form

$$F(z) = G(z)z^{\Gamma}$$

with $G(z)$ univalued and holomorhic on $D^{\times}_{r_0}$. The multivalued behaviour of $F(z)$ is just a consequence of the factor $z^{\Gamma}$.

**Definition 1.19.** *The solutions of $(\partial + A)F = 0$ have moderate growth near the singular point $z = 0$ if for each sector*

$$\{z \in D^{\times}_{r_0}; \theta_1 < \arg z < \theta_2\}$$

*with $\theta_1 < \theta_2 < \theta_1 + 2\pi$ and each holomorphic solution $F(z)$ on this sector there exist constants $C > 0$ and $D \in \mathbb{R}$ with*

$$|F(re^{i\theta})| \leq Cr^D$$

*on this sector.*

It is clear that the solution matrix $F(z)$ has moderate growth near $z = 0$ if and only if the matrix function $G(z)$ has a pole or a removable singularity at $z = 0$.

**Definition 1.20.** *The linear system* $(\partial + A)F = 0$ *on the punctured disc* $D_{r_0}^{\times}$ *has a regular singular point at* $z = 0$ *if* $z \mapsto zA(z)$ *is holomorphic at* $z = 0$*, or equivalently if the linear system has the form* $(\theta + B)F = 0$ *with* $z \mapsto B(z) = zA(z)$ *holomorphic at* $z = 0$*. Here we always denote* $\theta = z\partial$*.*

**Theorem 1.21.** *If the linear system* $(\partial + A)F = 0$ *has a regular singularity at* $z = 0$ *then all solutions have moderate growth at* $z = 0$*.*

We first prove a lemma.

**Lemma 1.22.** *If* $(a, b) \ni r \mapsto F(r) \in \mathbb{R}^n - \{0\}$ *is smooth then*

$$|\frac{d}{dr}|F(r)|| \leq |\frac{dF}{dr}(r)| \ .$$

*Proof.* Suppose $F(r) = (f_1(r), \cdots, f_n(r))^t$. Then we get

$$\frac{d}{dr}|F(r)| = \frac{d}{dr}(\sum_1^n f_j(r)^2)^{\frac{1}{2}} = (\sum_1^n f_j(r)^2)^{-\frac{1}{2}}(\sum_1^n f_j(r)\frac{df_j}{dr}(r))$$

which in absolute value is $\leq |\frac{dF}{dr}(r)|$ by the Cauchy inequality. □

We now come to the proof of the above theorem.

*Proof.* For $0 < r < r_1 < r_0$ we have

$$\log|F(re^{i\theta})| - \log|F(r_1e^{i\theta})| = \int_{r_1}^{r} \frac{d}{ds}\log|F(se^{i\theta})|ds$$

$$= \int_{r_1}^{r} |F(se^{i\theta})|^{-1}\frac{d}{ds}|F(se^{i\theta})|ds \leq \int_{r}^{r_1} |F(se^{i\theta})|^{-1}|\frac{d}{ds}F(se^{i\theta})|ds$$

$$= \int_{r}^{r_1} |F(se^{i\theta})|^{-1}|\partial F(se^{i\theta})|ds \leq M\int_{r}^{r_1} s^{-1}ds = M\log\frac{r_1}{r}$$

with $M = \max\{|zA(z)|; |z| \leq r_1\} < \infty$ by assumption. Because the natural logarithm is monotonically increasing we find

$$|F(re^{i\theta})| \leq (\frac{r_1}{r})^M|F(r_1e^{i\theta})| \ ,$$

such that for $C = \max\{r_1^M|F(r_1e^{i\theta})|; \theta_1 \leq \theta \leq \theta_2\}\}$ and $N = -M$ the required inequality is obtained. □

The converse of the above theorem is not true, in the sense that for a linear system having a regular singularity at $z = 0$ it is not a necessary condition for the solutions having moderate growth at $z = 0$.

**Example 1.23.** *For $n = 2$ consider the linear system $(\theta + B)F = 0$ on $\mathbb{C}^\times$ with coefficients matrix*

$$B = \begin{pmatrix} 0 & -z^k \\ 0 & k \end{pmatrix}$$

*for some $k \in \mathbb{Z}$ and $F = (f_1, f_2)^t$. Spelled out the linear system becomes*

$$z\partial f_1 - z^k f_2 = 0 \ , \ z\partial f_2 + k f_2 = 0 \ .$$

*The second equation (after multiplication by $z^{k-1}$) becomes $\partial(z^k f_2) = 0$. Hence $f_2 = c_2 z^{-k}$ for some integration constant $c_2$. Substitution in the first equation gives $z\partial f_1 - c_2 = 0$, and therefore $f_1 = c_1 + c_2 \log z$ for a second integration constant $c_1$. Hence the general solution becomes*

$$F(z) = (f_1(z), f_2(z))^t = c_1(1,0)^t + c_2(\log z, z^{-k})^t \ .$$

*These functions have moderate growth for all $k \in \mathbb{Z}$, but clearly for $k \leq -1$ the linear system is not regular singular at $z = 0$. Therefore being regular singular of $(\partial + A)F = 0$ at $z = 0$ is a sufficient, but not a necessary condition for the solutions having moderate growth around $z = 0$.*

Let us now consider a linear system with a regular singularity at $z = 0$, so a linear system of the form

$$(\partial + A)F = 0$$

or equivalently

$$(\theta + B)F = 0$$

with $\theta = z\partial$ and $B = zA$ holomorphic on the disc $D_{r_0} = \{|z| \leq r_0\}$ for some $r_0 > 0$. Hence the power series

$$B(z) = \sum_0^\infty B_k z^k$$

with $B_k \in \mathrm{Mat}(n, \mathbb{C})$ converges on $D_r$. The Frobenius method consists in the substition of a formal series

$$F(z) = z^s \sum_0^\infty F_k z^k = \sum_0^\infty F_k z^{s+k}$$

with $s \in \mathbb{C}$ undetermined and $F_k \in \mathbb{C}^n$ undetermined. We have the following analogue of Proposition 1.2.

**Proposition 1.24.** *The formal series* $F(z) = z^s \sum_0^\infty F_k z^k$ *is a solution of* $(\theta + B)F = 0$ *with* $B(z) = \sum_0^\infty B_k z^k$ *if and only if*

$$(s + B_0)F_0 = 0$$

*and*

$$(s + k + 1 + B_0)F_{k+1} + \sum_{l=0}^{k} B_{k+1-l}F_l = 0$$

*for all* $k \in \mathbb{N}$*.*

*Proof.* This is a direct computation using $\theta(z^{s+k}) = (s + k)z^{s+k}$. □

**Definition 1.25.** *The characteristic equation* $\det(s + B_0) = 0$ *is called the indicial equation and the roots of the indicial equation are called the exponents of the differential equation* $(\theta + B)F = 0$ *at* $z = 0$*.*

**Corollary 1.26.** *Consider the linear system* $(\theta + B)F = 0$ *with a regular singularity at* $z = 0$*. If* $s$ *is an exponent but* $(s+k+1)$ *is not an exponent for all* $k \in \mathbb{N}$ *then there exists for each* $F_0 \in \mathrm{Ker}(s+B_0)$ *a unique formal solution* $F(z) = z^s \sum_0^\infty F_k z^k$ *with* $F_{k+1} \in \mathbb{C}^n$ *given by the recurrence relations in Proposition 1.24.*

*Proof.* This is clear from the recurrence relations in Proposition 1.24 because $(s + k + 1 + B_0)$ is invertible for all $k \in \mathbb{N}$. □

**Theorem 1.27.** *The formal solution* $F(z) = z^s \sum_0^\infty F_k z^k$ *with* $F_k \in \mathbb{C}^n$ *given by Proposition 1.24 converges on* $D_{r_0}$*.*

*Proof.* Because $B(z) = \sum_0^\infty B_k z^k$ converges on $D_{r_0}$ we know by Lemma 1.4 $\forall \rho \in (0, 1)$ the existence of a constant $M_\rho \geq 0$ such that

$$|B_k| \leq M_\rho(\rho r_0)^{-k} \ \forall k \in \mathbb{N} \ .$$

Because $(s + k + 1)$ is not an exponent for $k \in \mathbb{N}$ there exists a constant $K \geq 1$ such that

$$|(s + k + 1 + B_0)^{-1}| \leq K(k + 1)^{-1} \ .$$

Using this inequality one can show by induction on $k$ that

$$|F_k| \leq M_\rho(M_\rho + 1) \cdots (M_\rho + k - 1)K^k(\rho r_0)^{-k}|F_0|/k!$$

for all $k \in \mathbb{N}$. Hence the formal series $F(z) = z^s \sum_0^\infty F_k z^k$ converges on the disc $|z| < r_0/K$. But then the series also converges on $D_{r_0}$ as solution of the differential equation $(\theta + B)F = 0$. □

**Corollary 1.28.** *Consider a linear system* $(\theta + B)F = 0$ *on* $D_{r_0}^\times$ *with a regular singularity at* $z = 0$. *Suppose the exponents* $s_1, \cdots, s_n$ *at* $z = 0$ *are modulo* $\mathbb{Z}$ *distinct:* $s_i - s_j \notin \mathbb{Z}$ *for* $i \neq j$. *Then the* $n$ *solutions*

$$F_j(z) = z^{s_j}(F_{j0} + F_{j1}z + F_{j2}z^2 + \cdots)$$

*with* $0 \neq F_{j0} \in \mathrm{Ker}(s_j + B_0)$ *for* $j = 1, \cdots, n$ *are a basis of the local solution space, say around* $z = r_0/2$ *(with* $z^s = e^{s \log z}$ *and* $\Im(\log(r_0/2)) = 0$*).*

**Proposition 1.29.** *Consider the linear system* $(\theta + B)F = 0$ *with a regular singularity at* $z = 0$. *The matrix* $\mathrm{e}^{-2\pi i B_0}$ *lies in the closure of the conjugation orbit of the monodromy matrix* $M$. *In particular* $\mathrm{e}^{-2\pi i B_0}$ *and* $M$ *have the same characteristic polynomial.*

*Proof.* In polar coordinates $z = re^{i\theta}$ we integrate $(\theta + B)F = 0$ along circles $r =$ constant. Because $\theta = z\partial = -id/d\theta$ we get

$$(\frac{d}{d\theta} + iB_0 + rC(r,\theta))F(r,\theta) = 0$$

for $0 \leq r < r_0, 0 \leq \theta \leq 2\pi$ and $C(r,\theta) = r^{-1}i(B(re^{i\theta}) - B_0)$. Let now $F(r,\theta)$ be the solution matrix with initial value $F(r,0) = I_n$ for all $r \in [0, r_0)$. Then $M(r) = F(r, 2\pi)$ is the monodromy matrix obtained by analytic continuation along paths $\gamma_r(t) = re^{2\pi i t}$ with time $t$ and fixed $r \in (0, r_0)$. Hence $M(r)$ is conjugated with $M$ for all $r \in (0, r_0)$. The function $C(r,\theta)$ is continuous for $(r,\theta) \in [0, r_0) \times [0, 2\pi]$. Hence the solution $F(r,\theta)$ with continuous initial value $F(0,\theta)$ for $\theta \in [0, 2\pi]$ is also continuous for $(r,\theta) \in [0, r_0) \times [0, 2\pi]$. Hence $M(0) = F(0, 2\pi)$ is equal to the limit of $M(r)$ for $r \downarrow 0$. However $M(0) = e^{-2\pi i B_0}$ by direct integration, since $rC(r,\theta)$ vanishes in the limit for $r \downarrow 0$. This proves the proposition. □

**Corollary 1.30.** *A matrix in* $\mathrm{GL}_n(\mathbb{C})$ *is called regular if the conjugacy class has maximal dimension* $n(n-1)$, *or equivalently if the centralizer in* $\mathrm{GL}_n(\mathbb{C})$ *has minimal dimension* $n$. *If the matrix* $\mathrm{e}^{-2\pi i B_0}$ *is regular then* $M$ *and* $\mathrm{e}^{-2\pi i B_0}$ *are conjugated.*

**Example 1.31.** *Consider the linear system* $(\theta + B)F = 0$ *for* $n = 2$ *and with coefficients matrix*

$$B = \begin{pmatrix} 0 & -z^k \\ 0 & k \end{pmatrix}$$

*for some* $k \in \mathbb{Z}$ *discussed in Example 1.23. Relative to the solution matrix*

$$F = \begin{pmatrix} 1 & \log z \\ 0 & z^{-k} \end{pmatrix}$$

*the monodromy matrix around $z = 0$ is given by*

$$M = \begin{pmatrix} 1 & 2\pi i \\ 0 & 1 \end{pmatrix}$$

*while for $k \geq 1$ the matrix $e^{2\pi i B_0}$ is the identity matrix. Hence the matrices $M$ and $e^{-2\pi i B_0}$ need not be conjugated. The conjugacy class of $M$ consists of all regular unipotent matrices, and the identity matrix lies in the closure of this orbit.*

## 1.5 The theorem of Fuchs

Consider the $n^{th}$ order scalar differential equation

$$(\partial^n + a_1\partial^{n-1} + \cdots + a_{n-1}\partial + a_n)f = 0$$

with coefficients holomorphic on the puntured disc $D_{r_0}^\times$ for some $r_0 > 0$.

**Lemma 1.32.** *If $\theta = z\partial$ then $z^k\partial^k = \theta(\theta - 1)\cdots(\theta - k + 1)$ for $k \in \mathbb{N}$.*

*Proof.* By induction on $k$ we have

$$z^{k+1}\partial^{k+1} = z\theta(\theta - 1)\cdots(\theta - k + 1)\partial = (\theta - 1)(\theta - 2)\cdots(\theta - k)z\partial$$

because $z\theta = (\theta - 1)z \Leftrightarrow \theta z = z(\theta + 1)$ by the Leibniz product rule. □

Multiplying the above differential equation by $z^n$ we can rewrite this equation in the form

$$(\theta^n + b_1\theta^{n-1} + \cdots + b_{n-1}\theta + b_n)f = 0$$

with the transition from the functions $\{1, za_1, z^2a_2, \cdots, z^na_n\}$ to the new coefficients $\{1, b_1, b_2, \cdots, b_n\}$ given by an integral unitriangular matrix (with unitriangular meaning upper triangular with 1 on the diagonal). Note that the collection of unitriangular matrices in $\mathrm{GL}_n(\mathbb{Z})$ is a group.

**Definition 1.33.** *The point $z = 0$ is a regular singular point of the above $n^{th}$ order scalar differential equation if $z^ja_j$ is holomorphic at $z = 0$ $\forall j$ or equivalently if $b_j$ is holomorhic at $z = 0$ $\forall j$.*

The next result is called the theorem of Fuchs. It marks an important difference between first order matrix systems and $n^{th}$ order scalar differential equations.

**Theorem 1.34.** *The point $z = 0$ is a regular singularity of the above $n^{th}$ order scalar differential equation if and only if all solutions around $z = 0$ have moderate growth.*

*Proof.* Suppose $z = 0$ is a regular singular point of the $n^{th}$ order scalar differential equation, so $b_1, \cdots, b_n$ are holomorphic around $z = 0 \ \forall j$. We associate to the $n^{th}$ order scalar differential equation a first order matrix system

$$(\theta + B)F = 0$$

with the coefficient matrix $B$ given by

$$\begin{pmatrix} 0 & -1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & -1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & -1 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & -1 \\ b_n & b_{n-1} & b_{n-2} & b_{n-3} & \cdots & b_2 & b_1 \end{pmatrix}$$

and

$$F = (f, \theta f, \theta^2 f, \cdots, \theta^{n-1} f)^t \ .$$

This $F$ is a solution of this first order matrix system if and only if the first coordinate $f$ of $F$ is a solution of the $n^{th}$ order scalar differential equation. Hence $f$ has moderate growth around $z = 0$ by Theorem 1.21.

Conversely, suppose that all solutions of the $n^{th}$ order scalar differential equation have moderate growth around $z = 0$. We prove the statement by induction on the order $n$ of the scalar equation. There always exists a solution of the form

$$f_0(z) = z^s(1 + O(z)) \ , \ z \to 0$$

with exponent $s \in \mathbb{C}$. Indeed, just take a suitably normalized eigenvector of the monodromy operator $M(\gamma(t) = r_0 e^{2\pi i t}/2)$ in the local solution space around $z_0 = r_0/2$. Consider the linear differential operators

$$D = \theta^n + b_1\theta^{n-1} + \cdots + b_{n-1}\theta + b_n$$

and

$$E = f_0^{-1} \circ D \circ f_0 = \theta^n + c_1\theta^{n-1} + \cdots + c_{n-1}\theta + c_n \ .$$

Here $f_0$ stands for the $0^{th}$ order linear differential operator of multiplication by $f_0$. Because $f_0^{-1} \circ \theta \circ f_0 = \theta + \theta(f_0)/f_0$ with $\theta(f_0)/f_0$ (univalued) holomorphic around $z = 0$, we conclude that $b_1, \cdots, b_n$ are holomorphic around $z = 0$ if and only if $c_1, \cdots, c_n$ are holomorphic around $z = 0$. Moreover $E(1) = 0$ hence $c_n = 0$. In other words $E$ factorizes as

$$E = F\theta$$

with $F = \theta^{n-1} + c_1\theta^{n-2} + \cdots c_{n-1}$. The solutions $g$ of $E(g) = 0$ and $h$ of $F(h) = 0$ are related by $h = \theta g$. The solutions $g$ of $E(g) = 0$ are of the form $g = f/f_0$ with $f$ a solution of $Df = 0$. The solutions $f$ of $D(f) = 0$ have moderate growth around $z = 0$ by assumption. Hence the solutions $g$ of $E(g) = 0$ have moderate growth around $z = 0$, but then also the solutions $h$ of $F(h) = 0$ have moderate growth around $z = 0$. By induction on the the order $n$ of the scalar equation we can assume that $c_1, \cdots, c_{n-1}$ are holomorphic around $z = 0$. Because $c_n = 0$ is holomorphic as well we conclude that $b_1, \cdots, b_n$ are holomorphic around $z = 0$. This completes the proof of the theorem of Fuchs. □

**Definition 1.35.** *If the $n^{th}$ order scalar linear differential equation*

$$Df = 0\ ,\ D = \theta^n + b_1\theta^{n-1} + \cdots + b_{n-1}\theta + b_n$$

*on $D_{r_0}^{\times}$ has a regular singularity at $z = 0$ then the degree $n$ polynomial equation*

$$s^n + b_1(0)s^{n-1} + \cdots + b_{n-1}(0)s + b_n = 0$$

*is called the indicial equation and its roots are called the exponents of $Df = 0$ at $z = 0$.*

A solution of $Df = 0, D = \theta^n + b_1\theta^{n-1} + \cdots + b_{n-1}\theta + b_n$ around the regular singular point $z = 0$ of the form

$$f(z) = z^s \sum_0^{\infty} f_k z^k$$

with $f_k \in \mathbb{C}, f_0 \neq 0$ is called a formal solution with exponent $s$. Such a formal solution is only possible if $s$ is a root of the indicial equation. The coefficients $f_k \in \mathbb{C}, k \in \mathbb{Z}$ of such a solution are again given by recurrence relations, and these have a unique solution (for given $f_0$) if $s + k$ is not an exponent $\forall k \in \mathbb{Z}, k \geq 1$. If the $n$ exponents at $z = 0$ are all distinct modulo $\mathbb{Z}$ then there exists a basis of formal solutions with these exponents. Using

Theorem 1.27 it follows that these formal solutions have a positive radius of convergence.

The eigenvalues of the monodromy operator around $z = 0$ are of the form $e^{2\pi is}$ with $s$ an exponent at $z = 0$. However, just like in Proposition 1.29 the Jordan normal form of the monodromy operator around $z = 0$ can not in general be deduced from the indicial equation, namely in case some exponents coincide modulo $\mathbb{Z}$.

**Definition 1.36.** *An $n^{th}$ order scalar linear differential equation on the projective line $\mathbb{P}$ minus the singular points of the form*

$$Df = 0\ ,\ D = \partial^n + a_1\partial^{n-1} + \cdots + a_{n-1}\partial + a_n$$

*with rational coefficients $a_1, \cdots, a_n \in \mathbb{C}(z)$ is called a Fuchsian equation if all singular points (including $z = \infty$) are regular singular.*

In order to analyze the behaviour of an $n^{th}$ order scalar linear differential equation at $z = \infty$ one makes the substitution $w = z^{-1}$ and considers the behaviour of the transformed equation at $w = 0$. The same strategy works for first order matrix systems. Remark that $\theta = zd/dz = -wd/dw$.

## 1.6 Exercises

**Exercise 1.1.** *Show the formula*

$$\sum_{l=0}^{k} \frac{\Gamma(M+l)}{\Gamma(M)l!} = \frac{\Gamma(M+k+1)}{\Gamma(M+1)k!}$$

*by induction on $k$.*

**Exercise 1.2.** *Suppose we have given a second order differential equation of the form $(\partial^2 + a_1\partial + a_2)f = 0$ with coefficients $a_1, a_2$ holomorphic on $\mathbb{C}^\times$, and suppose that $f(z) = \log z = \log r + i\theta$ is a local solution around $z = 1$. Show that $f(z) \equiv 1$ is also a local solution around $z = 1$. Conclude that $a_1 = 1/z, a_2 \equiv 0$. Is the same conclusion valid if we only know that the coefficients $a_1, a_2$ are holomorphic on $\Re z > 0$, while still assuming that $f(z) = \log z$ is a local solution around $z = 1$?*

**Exercise 1.3.** *Show that the Euler-Gauss hypergeometric equation*

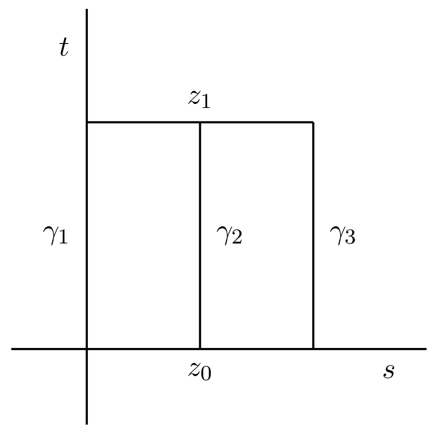
$$[z(1-z)\partial^2 + (\gamma - (\alpha+\beta+1)z)\partial - \alpha\beta)]f = 0$$

*can be transformed to the form*

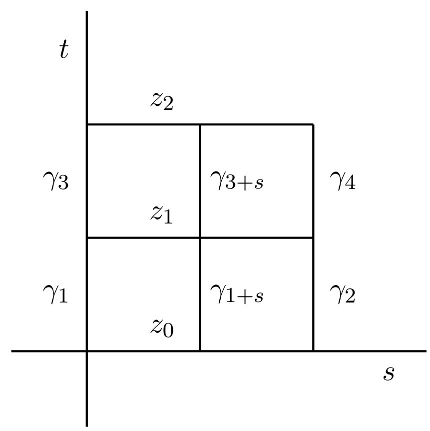
$$[(\theta+\gamma-1)\theta - z(\theta+\alpha)(\theta+\beta)]f = 0$$

*with $\theta = z\partial$. Hint: Use that $z^2\partial^2 = \theta(\theta-1)$.*

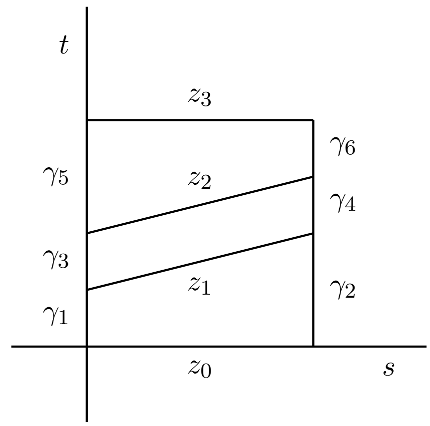
**Exercise 1.4.** *Show that homotopy equivalence for paths in $Z$ is an equivalence relation. Here is a schematic picture of the argument.*



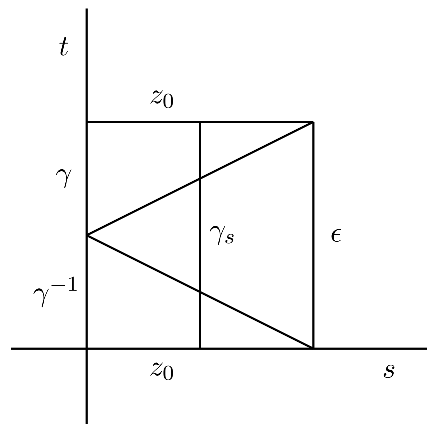

**Exercise 1.5.** *Show that if the end points of $\gamma_1 \sim \gamma_2$ coincide with the begin points of $\gamma_3 \sim \gamma_4$ then $\gamma_3\gamma_1 \sim \gamma_4\gamma_2$.*




**Exercise 1.6.** *Suppose that $\gamma_1, \cdots, \gamma_6$ are paths in $Z$ such that $\gamma_1 \sim \gamma_2$, $\gamma_3 \sim \gamma_4$, $\gamma_5 \sim \gamma_6$ and the products $\gamma_5(\gamma_3\gamma_1)$ and $(\gamma_6\gamma_4)\gamma_2$ are well defined. In other words we assume that the begin points of $\gamma_1, \gamma_2$ equal $z_0$, the end points of $\gamma_1, \gamma_2$ and the begin points of $\gamma_3, \gamma_4$ equal $z_1$, the end points of $\gamma_3, \gamma_4$ and the begin points of $\gamma_5, \gamma_6$ equal $z_2$, and finally the end points of $\gamma_5, \gamma_6$ equal $z_3$. Show that $\gamma_5(\gamma_3\gamma_1) \sim (\gamma_6\gamma_4)\gamma_2$. In turn this implies that the group law on the fundamental group is associative. A picture of the homotopy is given by the picture below.*




**Exercise 1.7.** *Show that in the notation of Theorem 1.13 we have* $\epsilon\gamma \sim \gamma\epsilon \sim \gamma$ *and* $\gamma\gamma^{-1} \sim \gamma^{-1}\gamma \sim \epsilon$.




*Here* $\gamma_s(t)$ *is equal to* $\gamma^{-1}(2t)$ *for* $t \in [0, (1-s)/2]$, *is constant equal to* $\gamma^{-1}(1-s) = \gamma(s)$ *for* $t \in [(1-s)/2, (1+s)/2]$, *and is equal to* $\gamma(2t-1)$ *for* $t \in [(1+s)/2, 1]$

**Exercise 1.8.** *Let* $\gamma : [0,1] \to Z$ *be a path in* $Z$ *with begin point* $z_0$ *and end point* $z_1$. *Show that conjugation by* $[\gamma]$ *induces an isomorphism* $\Pi_1(Z, z_0) \to \Pi_1(Z, z_1)$.

**Exercise 1.9.** *Compute the monodromy representation for the second order linear equation*

$$(z\partial^2 + \partial)f = 0$$

*on the domain* $Z = \mathbb{C}^\times$ *relative to the basis of solution* $f_1(z) = 1, f_2(z) = \log z$ *around* $z = 1$.

**Exercise 1.10.** *Prove Proposition 1.18. Give a counterexample for the failure of the proposition if we replace the field* $\mathbb{C}$ *by* $\mathbb{R}$.

**Exercise 1.11.** *Consider the linear system* $(\partial + A)F = 0$ *with coefficient matrix* $A$ *holomorphic on the domain* $Z$. *Suppose* $F = (F_1, \cdots, F_n)$ *is a local solution matrix. Show that* $\det(F)$ *is a local solution of the first order scalar equation*

$$(\partial + \mathrm{tr}(A)) \det(F) = 0 \ .$$

*In particular if* $\mathrm{tr}(A) = 0$ *then the monodromy group is contained in the special linear group* $\mathrm{SL}_n(\mathbb{C})$.

**Exercise 1.12.** *Show that the Euler–Gauss hypergeometric equation (introduced in Exercise 1.3)*

$$[z(1-z)\partial^2 + (\gamma - (\alpha + \beta + 1)z)\partial - \alpha\beta)]f(z) = 0$$

*or equivalently*

$$[(\theta + \gamma - 1)\theta - z(\theta + \alpha)(\theta + \beta)]f(z) = 0$$

*is a Fuchsian equation on* $\mathbb{P}$ *with regular singular points at* $z = 0, 1, \infty$. *Show that the exponents are* $0, 1-\gamma$ *at* $z = 0$, *and* $0, \gamma - (\alpha + \beta)$ *at* $z = 1$, *and* $\alpha, \beta$ *at* $z = \infty$.

**Exercise 1.13.** *Consider the Clausen-Thomae hypergeometric equation*

$$[(\theta + \beta_1 - 1) \cdots (\theta + \beta_n - 1) - z(\theta + \alpha_1) \cdots (\theta + \alpha_n)]f(z) = 0$$

*with so called numerator parameters* $\alpha = (\alpha_1, \cdots, \alpha_n)$ *and denominator parameters* $\beta = (\beta_1, \cdots, \beta_n)$. *Show that* $z = 0, 1, \infty$ *are the only singular points. Show that they are regular singular with exponents* $1-\beta_1, \cdots, 1-\beta_n$ *at* $z = 0$, *exponents* $\alpha_1, \cdots, \alpha_n$ *at* $z = \infty$ *and exponents* $0, 1, \cdots, (n-2)$ *and* $\gamma = \sum_1^n (\beta_j - \alpha_j) - 1$ *at* $z = 1$.

**Exercise 1.14.** *Show that the point* $z = \infty$ *is a regular point of the second order linear differential equation*

$$[\partial^2 + a_1(z)\partial + a_2(z)]f(z) = 0$$

*if and only if* $a_1(z) = 2z^{-1} + O(z^{-2}), a_2(z) = O(z^{-4})$ *for* $z \to \infty$. *Conclude that the most general second order Fuchsian equation on* $\mathbb{P}$ *with* $n$ *distinct regular singular points at* $z_1, \cdots, z_n \in \mathbb{C}$ *has the form*

$$[\partial^2 + \frac{G_{n-1}(z)}{F(z)}\partial + \frac{G_{2n-4}(z)}{F(z)^2}]f(z) = 0$$

*with* $F(z) = (z - z_1)\cdots(z - z_n)$ *and* $G_{n-1}, G_{2n-4} \in \mathbb{C}[z]$ *polynomials in* $z$ *of degrees* $= (n-1)$ *and* $\leq (2n-4)$ *respectively with leading coefficient of* $G_{n-1}$ *equal to* $2$.

**Exercise 1.15.** *Show that the most general form of a second order Fuchsian equation on* $\mathbb{P}$ *with* $n$ *distinct regular singular points* $z_1, \cdots, z_n \in \mathbb{C}$ *and exponents* $\alpha_j, \beta_j$ *at* $z = z_j$, *which satisfy* $\sum_1^n(\alpha_j + \beta_j) = (n-2)$, *is of the form*

$$[\partial^2 + \{\sum_1^n \frac{1 - \alpha_j - \beta_j}{(z - z_j)}\}\partial + \frac{1}{F(z)}\{\sum_1^n \frac{F_j(z_j)\alpha_j\beta_j}{(z - z_j)} + G_{n-4}(z)\}]f(z) = 0$$

*with* $F(z) = (z - z_1)\cdots(z - z_n)$ *and* $F_j(z) = F(z)/(z - z_j)$. *Finally* $G_{n-4} \in \mathbb{C}[z]$ *is a polynomial of degree* $\leq (n-4)$. *The* $(n-3)$ *coefficients of* $G_{n-4}$ *are called the accesory parameters.*

*Note that in case* $n = 3$ *the differential equation is completely determined by the three singular points* $z_1, z_2, z_3$ *and the exponents at* $z_1, z_2, z_3$ *(which are restricted to sum up to 1). A differential equation with no accesory parameters is called a rigid equation. The Euler–Gauss hypergeometric equation is a standard example of a rigid equation.*

# 2 The Euler–Gauss hypergeometric function

## 2.1 The hypergeometric function of Euler–Gauss

The Euler–Gauss hypergeometric equation, introduced by Euler in the $18^{th}$ century, is the second order linear differential equation on the projective line $\mathbb{P} = \mathbb{C} \sqcup \{\infty\}$ of the form

$$[z(1-z)\partial^2 + (\gamma - (\alpha+\beta+1)z)\partial - \alpha\beta)]f = 0 \; ,$$

or equivalently

$$[(\theta+\gamma-1)\theta - z(\theta+\alpha)(\theta+\beta)]f = 0 \; .$$

Here as before $\partial = \mathrm{d}/\mathrm{d}\,z$, $\theta = z\partial$ and $z$ is a complex variable. The numbers $\alpha, \beta, \gamma$ are called the parameters of the hypergeometric equation. It is a Fuchsian equation with regular singular points at $z = 0, 1, \infty$. The local exponents of the hypergeometric equation are given by the so called Riemann scheme

| | | |
|---|---|---|
| $0$ | $1$ | $\infty$ |
| $0$ | $0$ | $\alpha$ |
| $1-\gamma$ | $\gamma-(\alpha+\beta)$ | $\beta$ |

The first line gives the three singular points and the next two lines the exponents at the three singular points.

The Euler–Gauss hypergeometric function with parameters $\alpha, \beta, \gamma \in \mathbb{C}$ (but $\gamma \notin -\mathbb{N}$) is defined as the power series

$$F(\alpha, \beta, \gamma; z) = \sum_{k=0}^{\infty} \frac{(\alpha)_k (\beta)_k}{(\gamma)_k \; k!} z^k$$

with

$$(\alpha)_k = \alpha(\alpha+1)\cdots(\alpha+k-1) = \Gamma(\alpha+k)/\Gamma(\alpha)$$

the Pochhammer symbol. Its domain of convergence is equal to the unit disc $\mathbb{D}$, unless $\alpha$ or $\beta$ is a negative integer, in which case the series terminates and converges on all of $\mathbb{C}$. It is the unique holomorphic solution of the hypergeometric differential equation around $z = 0$ (an easy verification), normalized to be 1 at $z = 0$. In other words, the hypergeometric function is the normalized solution of the hypergeometric equation around $z = 0$ with exponent 0.

Besides the differential equation and the power series there is yet a third way of defining the hypergeometric function by means of a contour integral, obtained by Euler in 1748.

**Theorem 2.1.** *For $0 < \Re(\beta) < \Re(\gamma)$ the hypergeometric function is given by the Euler integral*

$$F(\alpha,\beta,\gamma;z) = \frac{\Gamma(\gamma)}{\Gamma(\beta)\Gamma(\gamma-\beta)} \int_0^1 t^{\beta-1}(1-t)^{\gamma-\beta-1}(1-zt)^{-\alpha}dt$$

*with $z \in \mathbb{D}$.*

*Proof.* Note that the condition $0 < \Re(\beta) < \Re(\gamma)$ ensures the convergence of the integral. Moreover the integral defines an analytic continuation from $\mathbb{D}$ to $\mathbb{C} - [1,\infty)$. The theorem is an immediate consequence of the binomial series

$$(1-w)^{-\alpha} = \sum_{k=0}^{\infty} (\alpha)_k w^k/k!$$

for $w \in \mathbb{D}$ and the Euler Beta integral

$$\mathrm{B}(\alpha,\beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha+\beta)$$

for $\Re(\alpha), \Re(\beta) > 0$. Details are left to the reader. □

A direct corollary of the Euler integral is the exact evaluation of the Gauss hypergeometric series at $z = 1$, a result of Gauss from 1812.

**Theorem 2.2.** *If $\Re(\gamma - \alpha - \beta) > 0$ then*

$$F(\alpha,\beta,\gamma;1) = \frac{\Gamma(\gamma)\Gamma(\gamma-\alpha-\beta)}{\Gamma(\gamma-\alpha)\Gamma(\gamma-\beta)}$$

*which is called the Gauss summation formula.*

*Proof.* Using the Euler integral formula we obtain

$$F(\alpha,\beta,\gamma;1) = \frac{\Gamma(\gamma)}{\Gamma(\beta)\Gamma(\gamma-\beta)} \int_0^1 t^{\beta-1}(1-t)^{\gamma-\alpha-\beta-1}dt$$

which is valid for $\Re\beta > 0$ and $\Re(\gamma-\alpha-\beta) > 0$. So the Gauss summation formula is clear from the Euler Beta integral formula. □

The hypergeometric equation

$$[z(1-z)\partial^2 + (\gamma - (\alpha+\beta+1)z)\partial - \alpha\beta)]f = 0$$

is the unique second order Fuchsian equation with regular singular points at $\{0, 1, \infty\}$ and with the given Riemann scheme. Hence the hypergeometric series $F(\alpha-\gamma+1, \beta-\gamma+1, 2-\gamma; z)$ is a solution of the hypergeometric equation with Riemann scheme

| | | |
|---|---|---|
| 0 | 1 | $\infty$ |
| 0 | 0 | $\alpha-\gamma+1$ |
| $\gamma-1$ | $\gamma-(\alpha+\beta)$ | $\beta-\gamma+1$ |

Therefore the function $z^{1-\gamma}F(\alpha-\gamma+1,\beta-\gamma+1,2-\gamma;z)$ is a solution of a hypergeometric equation with Riemann scheme

| | | |
|---|---|---|
| 0 | 1 | $\infty$ |
| 0 | 0 | $\alpha$ |
| $1-\gamma$ | $\gamma-(\alpha+\beta)$ | $\beta$ |

which is the Riemann scheme of our original hypergeometric equation. This shows that the formal series

$$z^{1-\gamma}F(\alpha-\gamma+1,\beta-\gamma+1,2-\gamma;z)$$

is the upto scalar unique solution of the original hypergeometric equation around $z=0$ with exponent $(1-\gamma)$. By a similar reasoning (going back to Riemann in 1857) we obtain the following result of Kummer from 1836.

**Proposition 2.3.** *The solution space of the hypergeometric equation*

$$[z(1-z)\partial^2+(\gamma-(\alpha+\beta+1)z)\partial-\alpha\beta)]f=0$$

*has a basis of the form*

$$F(\alpha,\beta,\gamma;z)$$
$$z^{1-\gamma}F(\alpha-\gamma+1,\beta-\gamma+1,2-\gamma;z)$$

*around the point $z=0$,*

$$F(\alpha,\beta,\alpha+\beta-\gamma+1;1-z)$$
$$(1-z)^{\gamma-\alpha-\beta}F(\gamma-\alpha,\gamma-\beta,\gamma-\alpha-\beta+1;1-z)$$

*around the point $z=1$,*

$$(-z)^{-\alpha}F(\alpha,\alpha-\gamma+1,\alpha-\beta+1;1/z)$$
$$(-z)^{-\beta}F(\beta,\beta-\gamma+1,\beta-\alpha+1;1/z)$$

*around the point $z=\infty$. Here the parameters $\alpha,\beta,\gamma$ are restricted such that the various hypergeometric series are well defined. For example, the first solution round $z=0$ is defined for $\gamma\notin-\mathbb{N}$, the second solution around $z=0$ is defined for $(2-\gamma)\notin-\mathbb{N}\Leftrightarrow\gamma\notin\mathbb{N}+2$, while they are linearly independent if $\gamma\notin\mathbb{Z}$. These solutions of the hypergeometric equation are called Kummer solutions.*

**Proposition 2.4.** *Analytic continuation from* $0$ *to* $1$ *along the interval* $[0,1]$ *yields the identity*

$$F(\alpha,\beta,\gamma;z)=\frac{\Gamma(\gamma)\Gamma(\gamma-\alpha-\beta)}{\Gamma(\gamma-\alpha)\Gamma(\gamma-\beta)}F(\alpha,\beta,\alpha+\beta-\gamma+1;1-z)$$

$$+\frac{\Gamma(\gamma)\Gamma(\alpha+\beta-\gamma)}{\Gamma(\alpha)\Gamma(\beta)}(1-z)^{\gamma-\alpha-\beta}F(\gamma-\alpha,\gamma-\beta,\gamma-\alpha-\beta+1;1-z)$$

*under the parameter restrictions* $\gamma\notin-\mathbb{N}$ *and* $(\gamma-\alpha-\beta)\notin\mathbb{Z}$.

*Proof.* Using the basis of Kummer solutions around $z=1$ we obtain by analytic continuation along the interval $[0,1]$ an identity of the form

$$F(\alpha,\beta,\gamma;z)=\kappa F(\alpha,\beta,\alpha+\beta-\gamma+1;1-z)$$

$$+\lambda(1-z)^{\gamma-\alpha-\beta}F(\gamma-\alpha,\gamma-\beta,\gamma-\alpha-\beta+1;1-z)$$

for certain complex numbers $\kappa,\lambda$ depending on $\alpha,\beta,\gamma$. If $\Re(\gamma-\alpha-\beta)>0$ we can take the limit $z\uparrow 1$ and apply the Gauss summation formula to find for $\kappa$ the quotient of $\Gamma$-factors as given in the proposition. The validity of this expression for $\kappa$ holds for all $\alpha,\beta,\gamma$ by analytic continuation in the parameters whenever it is defined. The substitution $z\mapsto z/(z-1)$ gives for $z\in(-\infty,0]$ the relation

$$F(\alpha,\beta,\gamma;z/(z-1))=\kappa F(\alpha,\beta,\alpha+\beta-\gamma+1;1/(1-z))$$

$$+\lambda(1-z)^{\alpha+\beta-\gamma}F(\gamma-\alpha,\gamma-\beta,\gamma-\alpha-\beta+1;1/(1-z)).$$

The substitution $\beta\mapsto(\gamma-\beta)$ and the multiplication by $(1-z)^{-\alpha}$ give for $z\in(-\infty,0]$ the relation

$$(1-z)^{-\alpha}F(\alpha,\gamma-\beta,\gamma;z/(z-1))=$$

$$\mu(1-z)^{-\alpha}F(\alpha,\gamma-\beta,\alpha-\beta+1;1/(1-z))$$

$$+\nu(1-z)^{-\beta}F(\beta,\gamma-\alpha,\beta-\alpha+1;1/(1-z))$$

with

$$\mu=\frac{\Gamma(\gamma)\Gamma(\beta-\alpha)}{\Gamma(\beta)\Gamma(\gamma-\alpha)}.$$

The Kummer relation

$$F(\alpha,\beta,\gamma;z)=(1-z)^{-\alpha}F(\alpha,\gamma-\beta,\gamma;z/(z-1))$$

around $z = 0$ makes the symmetry $\alpha \leftrightarrow \beta$ visible, and so

$$\nu = \frac{\Gamma(\gamma)\Gamma(\alpha - \beta)}{\Gamma(\alpha)\Gamma(\gamma - \beta)},$$

which in turn implies (via the substitution $\beta \mapsto (\gamma - \beta)$) that

$$\lambda = \frac{\Gamma(\gamma)\Gamma(\alpha + \beta - \gamma)}{\Gamma(\alpha)\Gamma(\beta)}.$$

This proves the desired formula for the analytic continuation of the hypergeometric function $F(\alpha, \beta, \gamma; z)$ from 0 to 1 along the interval $[0, 1]$. □

Using the above proof and the Kummer relations

$$z^{-\alpha}F(\alpha, \alpha-\gamma+1, \alpha-\beta+1; 1/z) = (z-1)^{-\alpha}F(\alpha, \gamma-\beta, \alpha-\beta+1; 1/(1-z))$$

$$z^{-\beta}F(\beta, \beta-\gamma+1, \beta-\alpha+1; 1/z) = (z-1)^{-\beta}F(\beta, \gamma-\alpha, \beta-\alpha+1; 1/(1-z))$$

around $z = \infty$ also gives the analytic continuation of the hypergeometric function $F(\alpha, \beta, \gamma; z)$ along the negative real axis.

**Theorem 2.5.** *Analytic continuation from* 0 *to* $-\infty$ *along the negative real axis yields the identity*

$$F(\alpha, \beta, \gamma; z) = \frac{\Gamma(\gamma)\Gamma(\beta - \alpha)}{\Gamma(\beta)\Gamma(\gamma - \alpha)}(-z)^{-\alpha}F(\alpha, \alpha - \gamma + 1, \alpha - \beta + 1; 1/z)$$
$$+\frac{\Gamma(\gamma)\Gamma(\alpha - \beta)}{\Gamma(\alpha)\Gamma(\gamma - \beta)}(-z)^{-\beta}F(\beta, \beta - \gamma + 1, \beta - \alpha + 1; 1/z)$$

*under the parameter restrictions* $\gamma \notin -\mathbb{N}$ *and* $(\alpha - \beta) \notin \mathbb{Z}$*.*

This formula is an important ingredient for the solution of the singular Sturm-Liouville problem (using Weyl-Titchmarsh theory [24]) of the hypergeometric operator on the unbounded interval $(-\infty, 0]$.

## 2.2 The monodromy according to Schwarz

The question posed and solved by Schwarz in 1873 was: For which of the parameter values $\alpha, \beta, \gamma \in \mathbb{Q}$ are the solutions of the hypergeometric equation algebraic functions of its variable $z$?

The essential ingredient for the proof is the concept of monodromy, that was introduced by Riemann in his fundamental paper from 1857 on the hypergeometric equation [15]. It turns out that the solutions of the hypergeometric equation are algebraic if and only if the monodromy group of this equation is finite. Schwarz gave a beautiful alternative description of the (projective) monodromy group of the hypergeometric equation using the reflection principle, that he invented exactly for this purpose [17]. Subsequently Klein extended the work of Schwarz to deal not only with finite monodromy groups acting on the (elliptic) Riemann sphere, but also with infinite monodromy groups acting on the (parabolic) Euclidean plane and the (hyperbolic) Poincaré disc [10]. In turn this gave a boost to the theory of automorphic forms and functions.

The local exponents of the hypergeometric equation are given by the Riemann scheme

| | | |
|---|---|---|
| $0$ | $1$ | $\infty$ |
| $0$ | $0$ | $\alpha$ |
| $1-\gamma$ | $\gamma-(\alpha+\beta)$ | $\beta$ |

The exponent differences at the three singular points $0, 1, \infty$ are given by

$$\kappa = \pm(\gamma - 1) \ , \ \lambda = \pm(\alpha + \beta - \gamma) \ , \ \mu = \pm(\alpha - \beta)$$

respectively. Let us assume that the parameters $\alpha, \beta, \gamma$ are real numbers. In addition we shall assume that $0 \leq \kappa, \lambda, \mu$ and $\kappa + \lambda, \kappa + \mu, \lambda + \mu \leq 1$, which can always be arranged after shifting $\alpha, \beta, \gamma$ by integers (equivalently shifting $\kappa, \lambda, \mu$ by integers with even sum and performing sign changes on $\kappa, \lambda, \mu$). Indeed, we can arrange $0 \leq \kappa, \lambda, \mu \leq 1$ by sign changes and shifts by even integers. Say $0 \leq \kappa \leq \lambda \leq \mu \leq 1$ and so $\kappa + \lambda \leq \kappa + \mu \leq \lambda + \mu$. If $\lambda + \mu \leq 1$ we are done. If $\lambda + \mu > 1$ replace $(\kappa, \lambda, \mu)$ by $(\kappa, 1 - \lambda, 1 - \mu)$ which satisfies the requirements.

Now let us pick two linearly independent solutions $f_1, f_2$ on the upper half plane $\mathbb{H} = \{\Im(z) > 0\}$, and consider the projective evaluation map (also called the Schwarz map)

$$\mathrm{Pev} : \mathbb{H} \to \mathbb{P} \ , \ \mathrm{Pev}(z) = f_1(z)/f_2(z)$$

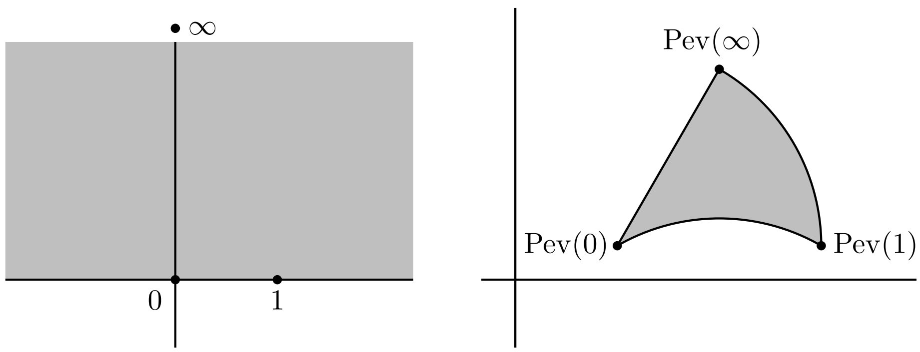
with $\mathbb{P}$ the complex projective line. Because of the ambiguity of the base choice $f_1, f_2$ the Schwarz map is only canonical upto action of $\mathrm{Aut}(\mathbb{P})$. We claim that the the Schwarz map Pev maps the upper half plane $\mathbb{H}$ conformally onto the interior of a triangle with sides circular arcs, and with angles $\kappa\pi$, $\lambda\pi$ and $\mu\pi$ at the vertices $\mathrm{Pev}(0)$, $\mathrm{Pev}(1)$ and $\mathrm{Pev}(\infty)$ respectively. This circular triangle is called the Schwarz triangle of the hypergeometric equation.

The Schwarz map is conformal because its derivative

$$\partial(\mathrm{Pev}) = \frac{\partial(f_1)f_2 - f_1\partial(f_2)}{f_2^2}$$

vanishes nowhere. Indeed the numerator is the Wronskian, which does not vanish, because $f_1, f_2$ are linearly independent solutions on $\mathbb{H}$. In order to understand the image of the Schwarz map we look at its behaviour on the real axis as boundary of $\mathbb{H}$.

For example, for the boundary interval $(0,1)$ we can choose the solutions $f_1, f_2$ to be real on $(0,1)$. This is possible because the hypergeometric equation is a real differential equation, since the parameters $\alpha, \beta, \gamma$ were assumed to be real numbers. In that case the image of the interval $(0,1)$ under the Schwarz map is a real interval. For a general choice of $f_1, f_2$ the image of $(0,1)$ is the transform under an element of $\mathrm{Aut}(\mathbb{P})$, so a fractional linear transformation, of a real interval, and therefore equal to a real interval or a circular arc.




The angles of the Schwarz triangle at the vertices Pev(0), Pev(1) and Pev(∞) are equal to $\kappa\pi$,$\lambda\pi$ and $\mu\pi$ respectively. For example, near the origin 0 let us choose the solutions $f_1, f_2$ of the form

$$f_1(z) = (1 + \cdots)\ ,\ f_2(z) = z^{1-\gamma}(1 + \cdots)$$

which in turn implies that

$$f_1(z)/f_2(z) = z^{\kappa}(1 + \cdots)$$

which indeed gives an angle $\kappa\pi$ at the vertex Pev(0) of the Schwarz triangle. For a general choice of $f_1, f_2$ this angle $\kappa\pi$ is conserved by some fractional linear transformation.

By continuity we can extend the Schwarz map

$$\mathrm{Pev} : \mathbb{H} \sqcup (-\infty, 0) \sqcup (0, 1) \sqcup (1, \infty) \to \mathbb{P}$$

with image the Schwarz triangle minus its vertices. The key step in the argument of Schwarz is the beautiful insight that the analytic continuation of Pev is given by the reflection principle. Indeed, there are three possibilities for analytic continuation from the upper half plane $\mathbb{H}$ to the lower half plane $-\mathbb{H}$, namely through the intervals $(-\infty, 0)$, $(0, 1)$ and $(1, \infty)$. The analytic continuation of the Schwarz map is obtained by reflecting the Schwarz triangle in the corresponding sides. Now we can iterate the above construction with the new triangle, which allows one to understand the full analytic continuation of the Schwarz map, step by step reflecting in sides of circular triangles. The domain of this full analytic continuation is the universal covering space $\widetilde{Z}$ of $Z = \mathbb{P} - \{0, 1, \infty\}$, say relative to the base point $z_0 = 1/2$, and we write

$$\widetilde{\mathrm{Pev}} : \widetilde{Z} \to \mathbb{P}$$

for the analytic continuation of the Schwarz map as a univalued map.

The range of this map can get messy, as the triangles start overlapping. However in case the Schwarz triangle is dihedral, which means that

$$\kappa = 1/k,\ \lambda = 1/l,\ \mu = 1/m$$

for some integers $k, l, m \geq 2$, we do get a regular tesselation by congruent images of the Schwarz triangle. These conditions on the parameters are called the Schwarz integrality conditions. The range $\mathbb{G}$ of this tesselation is equal to

$$\mathbb{G} = \mathbb{P},\ \mathbb{C},\ \mathbb{D}$$

upto an action of $\mathrm{Aut}(\mathbb{P})$, depending on whether the angle sum of the Schwarz triangle

$$(\kappa + \lambda + \mu)\pi = (1/k + 1/l + 1/m)\pi$$

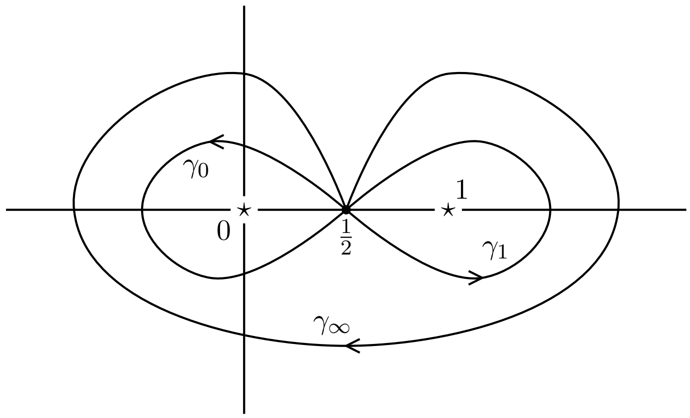
is greater than $\pi$, equal to $\pi$, or smaller than $\pi$ respectively. Here

$$\mathbb{D} = \{w \in \mathbb{C}; |w| < 1\}$$

denotes the unit disc. In this last case the disc $\mathbb{D}$ is bounded by a circle (Klein's Orthogonalkreis) which is orthogonal to the three circles bounding the Schwarz triangle. Hence the range $\mathbb{G}$ of the Schwarz map equals the

Riemann sphere $\mathbb{P}$, the Euclidean plane $\mathbb{C}$ or the Poincaré disc $\mathbb{D}$ respectively. Note that in all three cases $\mathbb{G}$ is simply connected.

The image of the analytically continued Schwarz map $\widetilde{\text{Pev}} : \widetilde{Z} \to \mathbb{G}$ is equal to $\mathbb{G}$ minus all vertices of the triangular tesselation. These vertices can be filled in by the following construction.




The fundamental group $\Pi$ of $Z = \mathbb{P} - \{0, 1, \infty\}$ with base point $z_0 = 1/2$ has three generators $g_0 = [\gamma_0], g_1 = [\gamma_1], g_\infty = [\gamma_\infty]$ with a single relation $g_\infty g_1 g_0 = 1$ as indicated in the above picture. Under the above Schwarz integrality conditions the Schwarz map factors through the intermediate covering

$$\widehat{\text{Pev}} : \widehat{Z} = \Pi(k,l,m)\backslash\widetilde{Z} \to \mathbb{G}$$

with $\Pi(k, l, m)$ be the normal subgroup of $\Pi$ generated by $g_0^k$, $g_1^l$ and $g_\infty^m$. The group

$$\Gamma = \Gamma(k, l, m) \simeq \Pi/\Pi(k, l, m)$$

is the projective monodromy group of the hypergeometric equation. At this level we can lift the compactification $Z \hookrightarrow Z^+ = \mathbb{P}$ to a partial compactification $\widehat{Z} \hookrightarrow \widehat{Z}^+$, resulting in a commutative diagram

$$\begin{array}{ccccc}
\widehat{Z} & \longrightarrow & \widehat{Z}^+ & \xrightarrow{\widehat{\text{Pev}}^+} & \mathbb{G} \\
\downarrow & & \downarrow & & \downarrow \\
Z = \mathbb{P} - \{0, 1, \infty\} & \longrightarrow & Z^+ = \mathbb{P} & \xrightarrow{\text{Pev}^+} & \Gamma\backslash\mathbb{G}
\end{array}$$

The left vertical arrow is an unramified $\Gamma$-covering, while the middle vertical arrow is a ramified covering, branched of orders $k$, $l$ and $m$ above the points 0, 1 and $\infty$ respectively. The extended Schwarz map

$$\widehat{\mathrm{Pev}}^{+} : \widehat{Z}^{+} \to \mathbb{G}$$

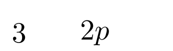
becomes an unramified covering. Since $\widehat{Z}^{+}$ is connected and $\mathbb{G}$ simply connected we conclude that the Schwarz map yields a conformal isomorphism between $\widehat{Z}^{+}$ and $\mathbb{G}$. In other words the projective monodromy group $\Gamma(k,l,m) \cong \Pi/\Pi(k,l,m)$ acts on $\mathbb{G}$ with quotient $\Gamma(k,l,m)\backslash\mathbb{G} \cong \mathbb{P}$. This quotient map is given by the inverse of the Schwarz map, and is ramified above $0,1,\infty$ of orders $k,l,m$ respectively. The projective monodromy group $\Gamma(k,l,m)$ is a subgroup of $\mathrm{Aut}(\mathbb{P})$, and is called the Schwarz triangle group. The group $W(k,l,m)$ generated by the reflections in the sides of the Schwarz triangle is called the Coxeter triangle group. It consists of holomorphic and antiholomorphic transformations of $\mathbb{P}$. The Schwarz triangle group is the index two subgroup of the Coxeter triangle group, consisting of even products of reflections in the sides of the Schwartz triangle.

Algebraic hypergeometric functions appear in case the monodromy group is finite, and for rational parameters $\alpha,\beta,\gamma$ this is equivalent with the projective monodromy group $\Gamma(k,l,m) \simeq \Pi/\Pi(k,l,m)$ being finite. For the integers $k,l,m \geq 2$ this amounts to $1/k + 1/l + 1/m > 1$. In that case the order $n$ of $\Gamma(k,l,m)$ is given by $1/k + 1/l + 1/m - 1 = 2/n$ as there are $2n$ triangles of area $(1/k + 1/l + 1/m - 1)\pi$ needed to tesselate the unit sphere of area $4\pi$. The results of this section are essentially due to Riemann [15], Schwarz [17] and Klein [10].

**Example 2.6.** *For $p = 3,4,5$ the hypergeometric function*

$$F((p+6)/(12p),(p-6)/(12p);2/3;z)$$

*is an algebraic function with projective monodromy group $\Gamma$ equal to the three Platonic rotation groups $A_4$,$S_4$,$A_5$ of tetrahedron, octahedron, icosahedron respectively. Indeed the exponent differences are $1/3, 1/2, 1/p$ at the points $0, 1, \infty$ respectively. The monodromy around $0$ and $1$ is a complex reflection of order $3$ and $2$ respectively. The linear monodromy group is the finite complex reflection group with Coxeter diagram*




*in the notation of Coxeter [5],[18]. The order of the group $\Gamma(3,2,p)$ is equal to $12p/(6-p)$ for $p = 3,4,5$ and indeed $12, 24, 60$ is the order of $A_4$,$S_4$,$A_5$.*

## 2.3 The Euler integral revisited

Let us fix four rational parameters $\mu_0, \mu_1, \mu_2, \mu_3 \in (0,1)$ with $\sum \mu_j = 2$. In addition choose four distinct complex variables $z_0, z_1, z_2, z_3$. If we clear denominators and write $\mu_j = m_j/m$ with $m, m_j \in \mathbb{N}$ (so $\sum m_j = 2m$) and $\gcd(m, m_0, m_1, m_2, m_3) = 1$ then the multivalued differential

$$\omega = \frac{dx}{y} \ , \ \ y = \prod (x - z_j)^{\mu_j}$$

on $\mathbb{P} - \{z_0, z_1, z_2, z_3\}$ becomes a univalued holomorphic differential on the Riemann surface

$$C : \ y^m = \prod (x - z_j)^{m_j}$$

lying above $\mathbb{P}$ as a $m$-fold ramified covering via the map $(x, y) \mapsto x$. This covering map is just the quotient map for the action of the group $C_m$ of the order $m$ roots of unity on $C$ (by multiplication in the variable $y$). Upto a multiplicative scalar the holomorphic differential $\omega$ on $C$ is unique characterized by the transformation behaviour $\omega \mapsto \zeta^{-1}\omega$ if $y \mapsto \zeta y$ for $\zeta \in C_m$. Integrals of the form

$$\pi = \int_{z_i}^{z_j} \omega$$

along suitable curves on $C$ (whose projection on $\mathbb{P}$ apart from begin and end points avoids $z_0, z_1, z_2, z_3$) are called period integrals.

An element of $\mathrm{Aut}(\mathbb{P})$ transforms the quadruple $z_0, z_1, z_2, z_3$ and the corresponding Riemann surface $C$ into isomorphic objects. Without loss of generality we can take $z_0 = 0, z_1 = z, z_2 = 1, z_3 = \infty$ with $z \in \mathbb{P} - \{0, 1, \infty\}$. If we integrate from 0 to 1 then the period integral becomes

$$\int_0^1 \omega \ , \ \omega = \frac{dx}{x^{\mu_0}(x - z)^{\mu_1}(x - 1)^{\mu_2}}$$

and apart from $\Gamma$-factors, a factor $(-z)^{\mu_1}(-1)^{\mu_2}$ and a substition $z \mapsto 1/z$ this becomes the Euler integral with parameters

$$\mu_0 = 1 - \beta, \ \mu_1 = \alpha, \ \mu_2 = 1 + \beta - \gamma$$

as functions of the parameters $\alpha, \beta, \gamma$ in the Euler integral. Since $\sum \mu_j = 2$ we find $\mu_3 = \gamma - \alpha$ and therefore

$$(1 - \mu_0 - \mu_1) = (\beta - \alpha), (1 - \mu_0 - \mu_2) = (\gamma - 1), (1 - \mu_0 - \mu_3) = (\alpha + \beta - \gamma) \ .$$

Hence the Schwarz integrality conditions

$$\kappa = |\gamma - 1|, \lambda = |\alpha + \beta - \gamma|, \mu = |\alpha - \beta| \in 1/\mathbb{N}$$

amount in the new parameters $\mu_0, \cdots, \mu_3$ to

$$(1 - \mu_i - \mu_j) \in 1/\mathbb{N}$$

for all $i \neq j$ with $\mu_i + \mu_j < 1$.

Let us assume that the Schwarz integrality conditions do hold. If we write

$$\pi_1(z) = \int_0^1 \omega \ , \ \pi_2(z) = \int_1^z \omega$$

then the Schwarz projective evaluation map

$$z \mapsto \mathrm{Per}(z) = \pi_1(z)/\pi_2(z)$$

becomes a period map, which we emphasize by writing Per instead of Pev. This period map is a locally biholomorphic injection from the moduli space $\mathcal{M}_{0,4} \cong \mathbb{P} - \{0, 1, \infty\}$ (or better from the moduli space of Riemann surfaces of the form $y^m = \prod (x - z_j)^{m_j}$ with fixed exponents $m, m_1, \cdots, m_4$ and $\sum m_j = 2m$) to the period domain $\Gamma \backslash \mathbb{G}$.

Suppose $z_0, \cdots, z_{n+2}$ are $n + 3$ distinct points on the projective line $\mathbb{P}$ and $\mu_0, \cdots, \mu_{n+2} \in (0, 1)$ rational parameters with $\sum \mu_j = 2$. Let us write $\mu_j = m_j/m$ with $m, m_j \in \mathbb{N}$ and $\gcd(m, m_0, \cdots, m_{n+2})$ so that $\sum m_j = 2m$. The multivalued algebraic differential

$$\omega = \frac{dx}{\prod (x - z_j)^{\mu_j}}$$

on $\mathbb{P}$ becomes a univalued holomorphic differential on the Riemann surface $C : y^m = \prod (x - z_j)^{m_j}$. The Appell–Lauricella hypergeometric function $F_D$ is defined by the periods

$$\pi = F_D(\mu; z) = \int_{z_j}^{z_k} \omega$$

with $z_0 = 0, z = (z_1, \cdots, z_n), z_{n+1} = 1, z_{n+2} = \infty$. This function was introduced by Appell for $n = 2$ and in 1893 generalized to arbitrary $n$ by Lauricella [11]. The results of Schwarz and Klein on the biholomorphic isomorphism $\mathrm{Pev}^+ : \mathbb{P} \to \Gamma \backslash \mathbb{G}$ were generalized for the Appell–Lauricella hypergeometric function by Deligne and Mostow [6], [13].

## 2.4 Exercises

**Exercise 2.1.** *Prove the Kummer relations*

$$z^{-\alpha}F(\alpha,\alpha-\gamma+1,\alpha-\beta+1;1/z) = (z-1)^{-\alpha}F(\alpha,\gamma-\beta,\alpha-\beta+1;1/(1-z))$$

$$z^{-\beta}F(\beta,\beta-\gamma+1,\beta-\alpha+1;1/z) = (z-1)^{-\beta}F(\beta,\gamma-\alpha,\beta-\alpha+1;1/(1-z))$$

*using the language of Riemann schemes. Hint: Because* $z \mapsto 1/z$ *maps* 0,1,$\infty$ *to* $\infty$,1,0 *respectively the formal series*

$$F(\alpha,\alpha-\gamma+1,\alpha-\beta+1;1/z)$$

*has Riemann scheme*

| | | |
|---|---|---|
| $0$ | $1$ | $\infty$ |
| $\alpha$ | $0$ | $0$ |
| $\alpha-\gamma+1$ | $\gamma-(\alpha+\beta)$ | $\beta-\alpha$ |

*which in turn implies that that the formal series*

$$z^{-\alpha}F(\alpha,\alpha-\gamma+1,\alpha-\beta+1;1/z)$$

*has Riemann scheme*

| | | |
|---|---|---|
| $0$ | $1$ | $\infty$ |
| $0$ | $0$ | $\alpha$ |
| $1-\gamma$ | $\gamma-(\alpha+\beta)$ | $\beta$ |

*Hence* $z^{-\alpha}F(\alpha,\alpha-\gamma+1,\alpha-\beta+1;1/z)$ *is the upto scalar unique solution of the original hypergeometric equation around* $z=\infty$ *with exponent* $\alpha$*. A similar argument works for the right hand side*

$$(z-1)^{-\alpha}F(\alpha,\gamma-\beta,\alpha-\beta+1;1/(1-z))$$

*which proves the first Kummer relation.*

**Exercise 2.2.** *Prove the Kummer relation*

$$F(\alpha,\beta,\gamma;z) = (1-z)^{-\alpha}F(\alpha,\gamma-\beta,\gamma;z/(z-1))$$

*using the language of Riemann schemes, as in the previous exercise.*

**Exercise 2.3.** *Let* $(\partial+A)F=0$ *and* $(\partial^n+a_1\partial^{n-1}+\cdots+a_{n-1}\partial+a_n)f=0$ *be a linear system and a scalar linear equation related as in Theorem 1.1. Show that for a solution matrix* $F=(F_{ij})$ *of the linear system the function* $\det(F)$ *is a solution of the differential equation* $(\partial+\mathrm{tr}(A))\det(F)=0$ *and conclude that for solutions* $f_1,\cdots,f_n$ *the Wronskian* $W(f_1,\cdots,f_n)$ *as defined by the determinant*

$$\begin{vmatrix} f_1 & f_2 & f_3 & \cdots & f_n \\ \partial f_1 & \partial f_2 & \partial f_3 & \cdots & \partial f_n \\ \partial^2 f_1 & \partial^2 f_2 & \partial^2 f_3 & \cdots & \partial^2 f_n \\ \vdots & \vdots & \vdots & & \vdots \\ \partial^{n-1} f_1 & \partial^{n-1} f_2 & \partial^{n-1} f_3 & \cdots & \partial^{n-1} f_n \end{vmatrix}$$

*is a solution of the differential equation* $(\partial+a_1)\,W(f_1,\cdots,f_n)=0$*. In turn conclude that the Wronskian vanishes if and only if the solutions* $f_1,\cdots,f_n$ *are linearly dependent.*

**Exercise 2.4.** *Show that for given* $2\leq k\leq l\leq m$ *the projective monodromy group* $\Gamma(k,l,m)=\Pi/\Pi(k,l,m)$ *is finite if and only if* $k=l=2, m\geq 2$ *or* $k=2, l=3, m=3,4,5$*.*

**Exercise 2.5.** *The modular group* $\mathrm{PSL}_2(\mathbb{Z})$ *acts on the upper halfplane* $\mathbb{H}$ *by fractional linear transformations. Using that* $\mathrm{PSL}_2(\mathbb{Z})$ *is generated by the two transformations*

$$S: z\mapsto -1/z\;,\; T: z\mapsto z+1$$

*show that* $\mathrm{PSL}_2(\mathbb{Z})\cong\Gamma(3,2,\infty)$*. Hint: Consider the Schwarz triangle*

$$\{z; -1/2\leq \Re z\leq 0, |z|\geq 1\}$$

*with vertices in the extended upper half plane* $\mathbb{H}\sqcup\mathbb{Q}\sqcup\{\infty\}$ *at the points* $\omega=(-1+\sqrt{-3})/2$*,* $i=\sqrt{-1}$ *and* $\infty$ *with corresponding angles* $\pi/3$*,* $\pi/2$ *and* $0$*.*

# 3 The Clausen–Thomae hypergeometric function

## 3.1 The hypergeometric function of Clausen–Thomae

The generalized hypergeometric series was introduced for $n = 3$ by Clausen [4] and in general by Thomae [23]. Let

$$\alpha = (\alpha_1, \cdots, \alpha_n) \in \mathbb{C}^n \ , \ \beta = (\beta_1, \cdots, \beta_{n-1}) \in \mathbb{C}^{n-1}$$

be complex parameters, and let us assume that $\beta_j \notin -\mathbb{N}$ for all $j$. The power series

$$F(\alpha, \beta; z) = {}_nF_{n-1}(\alpha, \beta; z) = \sum_{k=0}^{\infty} \frac{(\alpha_1)_k \cdots (\alpha_n)_k}{(\beta_1)_k \cdots (\beta_n)_k} z^k$$

with $\beta_n = 1$ is called the Clausen–Thomae hypergeometric series. The $\alpha_j$ are the numerator parameters and the $\beta_j$ the denominator parameters. Since $\beta_n = 1$ we get $(\beta_n)_k = k!$ and for $n = 2$ one recovers the Euler–Gauss hypergeometric series.

The hypergeometric series converges on the unit disc $\mathbb{D}$ and is a solution of the hypergeometric equation

$$[z(\theta + \alpha_1) \cdots (\theta + \alpha_n) - (\theta + \beta_1 - 1) \cdots (\theta + \beta_n - 1)]f = 0$$

with $\theta = zd/dz$ as before. This equation has regular singular points at $0, 1, \infty$ with local exponents given by the Riemann scheme

| $0$ | $1$ | $\infty$ |
|---|---|---|
| $1 - \beta_j$ | $0, 1, \cdots, n-2, \gamma$ | $\alpha_j$ |

by a direct computation. Here $\gamma = -1 + \sum_1^n (\beta_j - \alpha_j)$ and so the sum of all exponents equals $n(n-1)/2$ as should for a Fuchsian equation of order $n$ with regular singularities at $0, 1, \infty$. The point $z = 1$ is a remarkable singular point, because there is a codimension one subspace of holomorphic solutions around $z = 1$, corresponding to the local exponents $0, 1, \cdots, n-2$. This follows from the next theorem (after a substitution $z \mapsto z - 1$), which goes back to Pochhammer.

**Theorem 3.1.** *If we have given the differential equation*

$$(\partial^n + a_1 \partial^{n-1} + \cdots + a_{n-1}\partial + a_n)f = 0$$

*with $z \mapsto za_j(z)$ holomorphic on the unit disc $\mathbb{D}$ for all $j$ then there exist $n - 1$ linearly independent holomorphic solutions on $\mathbb{D}$.*

*Proof.* Let $A(z)$ be the matrix valued holomorphic function

$$\begin{pmatrix} 0 & -1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & -1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & -1 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & -1 \\ a_n & a_{n-1} & a_{n-2} & a_{n-3} & \cdots & a_2 & a_1 \end{pmatrix}$$

and rewrite the scalar differential equation in matrix form $(\partial + A)F = 0$ as in Theorem 1.1. This vector solution is of the form $F = (f_1, \cdots, f_n)^t$ with $f_{j+1} = \partial f_j$ and $f = f_1$ a solution of the original scalar differential equation of order $n$. Multiplication by $z$ gives the matrix form $(\theta + B)F = 0$ with $B(z) = zA(z)$ holomorphic on $\mathbb{D}$. If $B(z) = \sum B_k z^k$ then $\mathrm{Ker}(B_0)$ has dimension $n-1$, and so by Proposition 1.24 corresponds to a solution space of holomorphic solutions of dimension $n-1$, unless $-B_0$ has a positive integral eigenvalue. However, in that case the dimension of this holomorphic solution space goes down by one, but one finds an additional holomorphic solution with exponent corresponding to the positive integral eigenvalue of $-B_0$. □

The Clausen–Thomae hypergeometric function has the following integral representation.

**Theorem 3.2.** *For $\Re(\beta_i) > \Re(\alpha_i) > 0$ the Clausen–Thomae hypergeometric function $F(\alpha, \beta; z)$ is given by*

$$\prod_{i=1}^{n-1} \frac{\Gamma(\beta_i)}{\Gamma(\alpha_i)\Gamma(\beta_i - \alpha_i)} \int_0^1 \cdots \int_0^1 \frac{\prod_1^{n-1} t_i^{\alpha_i - 1}(1-t_i)^{\beta_i - \alpha_i - 1}}{(1 - zt_1 \cdots t_{n-1})^{\alpha_n}} dt_1 \cdots dt_{n-1}$$

*which for $n = 2$ boils down to the Euler integral formula.*

*Proof.* Substitute the binomial series

$$(1 - zt_1 \cdots t_{n-1})^{-\alpha_n} = \sum_{k=0}^{\infty} \frac{\Gamma(\alpha_n + k)}{\Gamma(\alpha_n)} (t_1 \cdots t_{n-1})^k z^k / k!$$

in the integral formule and use the Euler Beta integral

$$\int_0^1 t_i^{\alpha_i + k - 1}(1 - t_i)^{\beta_i - \alpha_i - 1} dt_i = \Gamma(\alpha_i + k)\Gamma(\beta_i - \alpha_i)/\Gamma(\beta_i + k)$$

to conclude that

$$F(\alpha,\beta;z)=\sum_{k=0}^{\infty}\frac{\Gamma(\alpha_n+k)}{\Gamma(\alpha_n)k!}\prod_{i=1}^{n-1}\frac{\Gamma(\alpha_i+k)\Gamma(\beta_i)}{\Gamma(\alpha_i)\Gamma(\beta_i+k)}z^k$$

and the result follows. □

So the three standard ways of introducing the classical hypergeometric function of Euler and Gauss, by means of a power series, as a solution of a differential equation or by means of an integral formula, generalize in a natural way for the Clausen–Thomae hypergeometric function.

## 3.2 The monodromy according to Levelt

Let $V(\alpha,\beta)$ be the local solution space at the base point $\frac{1}{2}$ and consider the monodromy representation

$$M(\alpha,\beta):\Pi\to\mathrm{GL}(V(\alpha,\beta))$$

with $\Pi$ the fundamental group of $\mathbb{M}=\mathbb{P}-\{0,1,\infty\}$ with generators $g_0,g_1,g_\infty$ and relation $g_\infty g_1 g_0=1$ as before. The monodromy group is the image of $\Pi$ under the monodromy representation. It is generated by the elements

$$h_0=M(\alpha,\beta)(g_0)\ ,\ h_1=M(\alpha,\beta)(g_1)\ ,\ h_\infty=M(\alpha,\beta)(g_\infty)$$

satisfying the relation $h_\infty h_1 h_0=1$. The local exponents at $z=0$ and $z=\infty$ in the Riemann scheme imply that

$$\det(t-h_\infty)=(t-a_1)\cdots(t-a_n)\ ,\ \det(t-h_0^{-1})=(t-b_1)\cdots(t-b_n)$$

with

$$a_j=\exp(2\pi i\alpha_j)\ ,\ b_j=\exp(2\pi i\beta_j)$$

while the linear map

$$(h_1-\mathrm{Id})\in\mathrm{End}(V(\alpha,\beta))$$

has rank at most one by Theorem 3.1.

**Theorem 3.3.** *Let $n\geq 2$ and $H<\mathrm{GL}(\mathbb{C}^n)$ be a subgroup generated by two matrices $A,B$ such that $\mathrm{rk}(A-B)\leq 1$. Then $H$ acts irreducibly on $\mathbb{C}^n$ if and only if $A$ and $B$ have disjoint sets of eigenvalues.*

*Proof.* Suppose $H$ acts reducibly on $\mathbb{C}^n$. Let $V_1$ be a nontrivial invariant subspace of $\mathbb{C}^n$ and let $V_2$ be $\mathbb{C}^n/V_1$. Since $\mathrm{rk}(A-B) \le 1$ it follows that $A$ and $B$ coincide either on $V_1$ or on $V_2$. Hence $A$ and $B$ have a common eigenvalue.

Suppose $A$ and $B$ have a common eigenvalue $\lambda$. If $W = \mathrm{Ker}(A-B)$ then $\dim W \ge (n-1)$ by assumption. If $\dim W = n$ then $A = B$ and $H$ acts reducibly on $\mathbb{C}^n$. Therefore we may assume that $\dim W = (n-1)$. If $A$ has an eigenvector in $W$ then it must also be an eigenvector for $B$, since $A$ and $B$ coincide on $W$. This common eigenvector generates an invariant subspace in $\mathbb{C}^n$ of dimension one. Hence $H$ acts reducibly on $\mathbb{C}^n$. Therefore we may assume that neither $A$ nor $B$ have an eigenvector in $W$.

We claim that $V = (A-\lambda)\mathbb{C}^n$ is an invariant subspace for $H$. Clearly $AV = A(A-\lambda)\mathbb{C}^n = (A-\lambda)A\mathbb{C}^n = (A-\lambda)\mathbb{C}^n = V$ and so $V$ is invariant under $A$. Since $\mathrm{Ker}(A-\lambda)$ is nontrivial and has trivial intersection with the codimension one subspace $W$ the dimension of $\mathrm{Ker}(A-\lambda)$ is one. Hence the dimension of $V$ is $n-1$ and so $V = (A-\lambda)W$. Since $A$ and $B$ coincide on $W$ we get $V = (B-\lambda)W$ and by a similar argument as for $A$ we get $V = (B-\lambda)\mathbb{C}^n$ and $V$ is invariant under $B$. Hence $V$ is a nontrivial invariant subspace for $H$ and the representation of $H$ on $\mathbb{C}^n$ becomes reducible. □

The next algebraic characterization of the monodromy group of the Clausen–Thomae hypergeometric equation is due to Levelt [12].

**Theorem 3.4.** *Suppose that* $a_1, \cdots, a_n, b_1, \cdots, b_n \in \mathbb{C}^\times$ *with* $a_i \ne b_j$ *for all* $1 \le i, j \le n$. *Up to simultaneous conjugation in* $\mathrm{GL}(n,\mathbb{C})$ *there exist unique elements* $A, B \in \mathrm{GL}(n,\mathbb{C})$ *with*

$$\det(t-A) = \prod_{j=1}^{n}(t-a_j)\ ,\ \det(t-B) = \prod_{j=1}^{n}(t-b_j)$$

*while the matrix* $A - B$ *has rank one.*

*Proof.* First we shall prove the existence of $A$ and $B$. We have to find matrices $A, B$ in $\mathrm{GL}(n,\mathbb{C})$ with

$$\prod_{j=1}^{n}(t-a_j) = t^n + A_1 t^{n-1} + \cdots + A_n$$
$$\prod_{j=1}^{n}(t-b_j) = t^n + B_1 t^{n-1} + \cdots + B_n$$

as their characteristic polynomials. If we take

$$A = \begin{pmatrix} 0 & 0 & \cdots & 0 & -A_n \\ 1 & 0 & \cdots & 0 & -A_{n-1} \\ 0 & 1 & \cdots & 0 & -A_{n-2} \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -A_1 \end{pmatrix} \quad B = \begin{pmatrix} 0 & 0 & \cdots & 0 & -B_n \\ 1 & 0 & \cdots & 0 & -B_{n-1} \\ 0 & 1 & \cdots & 0 & -B_{n-2} \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -B_1 \end{pmatrix}$$

then an easy computation shows

$$\det(t - A) = t^n + A_1 t^{n-1} + \cdots + A_n \ , \ \det(t - B) = t^n + B_1 t^{n-1} + \cdots + B_n$$

and $\mathrm{rk}(A - B) = 1$ holds trivially. This proves the existence of $A$ and $B$.

In order to prove the uniqueness of $A, B \in \mathrm{GL}(n, \mathbb{C})$ up to a simultaneous conjugation let $W = \mathrm{Ker}(A - B) \subset \mathbb{C}^n$. By assumption $W$ has codimension one in $\mathbb{C}^n$. Hence

$$V = W \cap A^{-1}W \cap \cdots \cap A^{-(n-2)}W$$

has dimension at least one. For $v \in V$ a nonzero vector the elements $A^i v$ for $i = 0, 1, \cdots, n-2$ all lie in $W$, which in turn implies that $A^i v = B^i v$ for $i = 0, 1, \cdots, n-1$. By the Cayley–Hamilton theorem the linear span of $A^i v = B^i v$ for $i = 0, 1, \cdots, n-1$ is invariant under the group generated by $A$ and $B$. Since $a_i \neq b_j$ the action of this group on $\mathbb{C}^n$ is irreducible by the previous theorem, which in turn implies that $A^i v = B^i v$ for $i = 0, 1, \cdots, n-1$ is a basis of $\mathbb{C}^n$. Relative to this basis the matrices of $A$ and $B$ have the above form. $\square$

Under the irreducibility condition $a_i \neq b_j$ the monodromy group of the Clausen–Thomae hypergeometric equation is obtained by

$$h_\infty = A \ , \ h_0 = B^{-1} \ , \ h_1 = A^{-1}B$$

and we will denote this monodromy group by $H(a, b)$. Indeed the linear map $h_1 - \mathrm{Id} = A^{-1}(B - A)$ has rank one. A linear transformation $h \in \mathrm{GL}(\mathbb{C}^n)$ is called a (complex) reflection if $h - \mathrm{Id}$ has rank one. The distinguished property of the Clausen–Thomae hypergeometric equation is that under the irreducibility condition $a_i \neq b_j$ the monodromy $h_1 = M(\alpha, \beta)(g_1)$ around the point 1 is a reflection. This makes the Clausen–Thomae hypergeometric equation a rigid equation, in the sense that it is characterized among all Fuchsian equation of order $n$ with regular singular points $\{0, 1, \infty\}$ by its local exponents at the three singular points. For rigid Fuchsian equations the monodromy group should be determined in linear algebra terms by the characteristic polynomials of the monodromy operators around the various regular singular points. For other examples of rigid equations we refer to work by Simpson [19]. A description in algebraic geometric terms of all rigid equations is due to Katz [9].

**Corollary 3.5.** *Under the irreducibility condition $a_i \neq b_j$ the monodromy group $H(a,b)$ of the Clausen–Thomae hypergeometric equation is defined in a suitable basis by matrices with entries in the ring $\mathbb{Z}[A_i, B_j, 1/A_n, 1/B_n]$.*

This is clear from the proof of the above theorem since $\det A = \pm A_n$ and $\det B = \pm B_n$. The rigidity of the Clausen–Thomae hypergeometric equation enables one to derive certain results by just looking at the Riemann schemes. In the next example the proof of Clausen's formula gives an illustration of this idea.

**Example 3.6.** *Clausen's formula says that*

$$ {}_2F_1(\alpha,\beta,\alpha+\beta+1/2;z)^2 = {}_3F_2(2\alpha,2\beta,\alpha+\beta,2\alpha+2\beta,\alpha+\beta+1/2;z) $$

*with the ${}_2F_1$ on the left hand side a second order Euler–Gauss hypergeometric function and the ${}_3F_2$ on the right hand side a third order Clausen–Thomae hypergeometric function. Clausen's formula can be proved by comparison of the two Riemann schemes. The Riemann scheme for the ${}_2F_1$ is given by*

| $0$ | $1$ | $\infty$ |
|---|---|---|
| $0$ | $0$ | $\alpha$ |
| $1/2-(\alpha+\beta)$ | $1/2$ | $\beta$ |

*while the Riemann scheme for the ${}_3F_2$ equals*

| $0$ | $1$ | $\infty$ |
|---|---|---|
| $0$ | $0$ | $2\alpha$ |
| $1/2-(\alpha+\beta)$ | $1$ | $2\beta$ |
| $1-2(\alpha+\beta)$ | $1/2$ | $\alpha+\beta$ |

*Observe that the latter Riemann scheme is just the second symmetric square of the former Riemann scheme. Moreover near the point $z=1$ there is a two dimensional subspace of holomorphic solutions, corresponding to the local exponents $0,1$. This proves Clausen's formula, and this particular third order Clausen–Thomae hypergeometric equation is just the second symmetric square of this particular second order Euler–Gauss hypergeometric equation.*

**Example 3.7.** *The symmetric group $S_{n+1}$ on $n+1$ letters acts on $\mathbb{C}^{n+1}$ by permutations of the coordinates. This action is reducible, but the restriction to the invariant linear subspace $V$ of vectors with zero sum of coordinates is an irreducible representation, called the reflection representation of $S_{n+1}$. The nearest neighbour transpositions $s_i = (i\, i+1)$ for $i = 1, \cdots, n$ generate the symmetric group $S_{n+1}$. It is easy to see that the symmetric group $S_{n+1}$ is also generated by the elements*

$$h_\infty = A = s_1 \cdots s_n \ , \ h_0^{-1} = B = s_1 \cdots s_{n-1} \ , \ h_1 = A^{-1} B = s_n$$

*considered as elements of $\mathrm{GL}(V)$ and the relation $h_\infty h_1 h_0 = \mathrm{Id}$ is trivial. Moreover $A = (1\ 2\ \cdots n+1)$ and $B = (1\ 2 \cdots n)$ implies*

$$\det(t - A) = t^n + t^{n-1} + \cdots + t + 1 \ , \ \det(t - B) = t^n - 1$$

*and so the symmetric group $S_{n+1}$ acting on $V$ by the reflection representation is an example of a hypergeometric group.*

## 3.3 The criterion of Beukers–Heckman

Throughout this section we have given $a_1, \cdots, a_n, b_1, \cdots, b_n \in \mathbb{C}^\times$ for which the irreducibility condition

$$a_i \neq b_j$$

of the previous section holds. Let $H(a,b) < \mathrm{GL}(n,\mathbb{C})$ be the corresponding hypergeometric group acting irreducibly on $\mathbb{C}^n$. In this section we will discuss a criterion for finiteness of the hypergeometric group $H(a,b)$ due to Beukers and Heckman [1]. Independently similar results were obtained around the same time by Kontsevich, but they remained unpublished after he learned about our preprint.

**Theorem 3.8.** *There exists a nondegenerate Hermitian form of $\mathbb{C}^n$ which is invariant under the hypergeometric group $H(a,b)$ if and only if both sets $\{a_1, \cdots, a_n\}$ and $\{b_1, \cdots, b_n\}$ are invariant under the substitution $z \mapsto \overline{1/z}$.*

*Proof.* An nonzero invariant Hermitian form for $H(a,b)$ is automatically nondegenerate, because $H(a,b)$ acst irreducibly on $\mathbb{C}^n$. Indeed, the kernel of such a form would be a proper invariant linear subspace of $\mathbb{C}^n$, and therefore is equal to 0. A nondegenerate invariant Hermitian form gives rise to and is determined by an isomorphism $H(a,b) \to H(\overline{1/a}, \overline{1/b})$ of hypergeometric groups. Here we denote

$$a = \{a_1, \cdots, a_n\} \ , \ b = \{b_1, \cdots, b_n\}$$

and likewise

$$\overline{1/a} = \{\overline{1/a_1}, \cdots \overline{1/a_n}\} \ , \ \overline{1/b} = \{\overline{1/b_1}, \cdots , \overline{1/b_n}\}$$

for the various parameter sets. Hence the result is obvious. □

**Theorem 3.9.** *Suppose that* $\{a_1, \cdots , a_n\}$ *and* $\{b_1, \cdots , b_n\}$ *have all modulus one. Suppose the exponents* $\alpha_i$ *and* $\beta_j$ *are contained in* $(0, 1]$ *and numbered by increasing argument. Let* $m_j = \#\{k; \alpha_k < \beta_j\}$ *for* $j = 1, \cdots , n$*. Then the signature* $(p, q)$ *of the invariant Hermitian form for the hypergeometric group* $H(a, b)$ *is given by*

$$|p - q| = |\sum_{j=1}^{n} (-1)^{j+m_j}| \ .$$

*In particular the invariant Hermitian form is definite if and only if the two sets*

$$\{a_1, \cdots , a_n\} \ , \ \{b_1, \cdots , b_n\}$$

*interlace on the unit circle.*

*Proof.* For the complete proof we refer to our paper [1]. Here we shall prove how the interlacing property of the two eigenvalue sets implies definiteness of the invariant Hermitian form. It is clear that the signature of the invariant Hermitian form does not change as long as $a_i$ and $b_j$ vary continuously over the unit circle while $a_i \neq b_j$ throughout the variation. Hence it is sufficent to have just one example of two sets interlacing on the unit circle, for which the invariant Hermitian form is definite. But from Example 3.7 we know that the invariant Hermitian form is definite for the case

$$\det(t - A) = t^n + t^{n-1} + \cdots + t + 1 \ , \ \det(t - B) = t^n - 1$$

of the reflection representation of the symmetric group $S_{n+1}$. Clearly these two parameter sets interlace on the unit circle. □

The next theorem gives an arithmetic criterion for finiteness of the hypergeometric group $H(a, b)$.

**Theorem 3.10.** *Suppose the parameters* $a_1, \cdots , a_n, b_1, \cdots , b_n$ *are roots of unity, and say*

$$\mathbb{Z}[a_1, \cdots , a_n, b_1, \cdots , b_n] = \mathbb{Z}[\exp(2\pi i/h)]$$

*for some* $h \in \mathbb{N}$. *Then the group* $H(a,b)$ *is finite if and only if for each* $k \in (\mathbb{Z}/h\mathbb{Z})^\times$ *the two sets*

$$a^k = \{a_1^k, \cdots, a_n^k\}\ ,\ b^k = \{b_1^k, \cdots, b_n^k\}$$

*interlace on the unit circle.*

*Proof.* For $\zeta_h = \exp(2\pi i/h)$ the ring of integers $\mathbb{Z}[\zeta_h]$ of the cyclotomic field $\mathbb{Q}(\zeta_h)$ is a free $\mathbb{Z}$-module of rank $m = \varphi(n)$ with basis $\zeta_h^k$ for $k \in (\mathbb{Z}/h\mathbb{Z})^\times$. For $k \in (\mathbb{Z}/h\mathbb{Z})^\times$ and $\zeta^h = 1$ the Galois automorphism

$$\sigma_k(\zeta) = \zeta^k$$

identifies the Galois group with $(\mathbb{Z}/h\mathbb{Z})^\times$.

By Corollary 3.5 we have $H(a,b) < \mathrm{GL}(n, \mathbb{Z}[\zeta_h])$. Multiplication by an algebraic integer $\zeta \in \mathbb{Z}[\zeta_h]$ in the basis $\zeta_h^k$ has a square matrix of size $m$ with rational integral coefficients, whose eigenvalues are the Galois conjugates $\sigma_k(\zeta)$. In this way the diagonal embedding

$$\prod_{k \in (\mathbb{Z}/h\mathbb{Z})^\times} \sigma_k : H(a,b) \to \prod_{k \in (\mathbb{Z}/h\mathbb{Z})^\times} H(a^k, b^k)$$

realizes $H(a,b)$ as a subgroup of $\mathrm{GL}(mn, \mathbb{Z})$. Since a subgroup of $\mathrm{GL}(N, \mathbb{Z})$ is finite if and only if it leaves invariant a positive definite Hermitian form the theorem follows. □

**Remark 3.11.** *By the same method one can prove that the hypergeometric group* $H(a,b) < \mathrm{GL}(n, \mathbb{Z}[\zeta_h])$ *is a discrete subgroup of* $\mathrm{GL}(n, \mathbb{C})$ *if the sets* $a^k$ *and* $b^k$ *interlace on the unit circle for all* $k \in (\mathbb{Z}/h\mathbb{Z})^\times$ *with* $1 < k < h/2$. *For a discussion of the geometric representation of algebraic integers in algebraic number theory we refer to [20].*

**Example 3.12.** *The image under* $s \mapsto \exp(2\pi i s)$ *of the two sets*

$$\{\alpha_j\} = \{1/30, 7/30, 11/30, 13/30, 17/30, 19/30, 23/30, 29/30\}$$

$$\{\beta_j\} = \{6/30, 10/30, 12/30, 15/30, 18/30, 20/30, 24/30, 30/30\}$$
$$= \{1/5, 1/3, 2/5, 1/2, 3/5, 2/3, 4/5, 1\}$$

*interlace on the unit circle. Note that the characteristic polynomials*

$$\prod_{j=1}^{8}(t - a_j) = \Phi_{30}(t)\ ,\ \prod_{j=1}^{8}(t - b_j) = \Phi_1(t)\Phi_2(t)\Phi_3(t)\Phi_5(t)$$

*are defined over $\mathbb{Z}$. Here $\Phi_m(t)$ is the $m^{th}$ cyclotomic polynomial of degree $\varphi(m)$. Hence both sets $\{a_j\}$ and $\{b_j\}$ are stable under raising to the power $k \in (\mathbb{Z}/30\mathbb{Z})^\times$. Therefore the group $H(a,b)$ is finite (with order* 696.729.600*) and so the Clausen–Thomae hypergeometric function with these parameters is an algebraic function.*

Let me now explain the origin of this example.

### 3.4 Intermezzo on Coxeter groups

Suppose $M = (m_{ij})_{1 \le i,j \le n}$ is a Coxeter matrix which means that $m_{ii} = 1$ for all $i$ and $m_{ij} = m_{ji} \in \mathbb{N}_{\ge 2}$ for all $i \ne j$.

**Definition 3.13.** *The Coxeter group $W = W(M)$ associated to the Coxeter matrix $M$ is given by the presentation*

$$W = \langle s_i; i = 1, \cdots, n \rangle / \{(s_i s_j)^{m_{ij}} = 1\}$$

*so in particular $s_i^2 = 1$, hence $s_i$ is an involution.*

**Definition 3.14.** *The Coxeter diagram associated to the Coxeter matrix $M$ is a marked graph, with nodes indexed by $i = 1, \cdots, n$. The $i^{th}$ and the $j^{th}$ node are connected if $m_{ij} \ge 3$, and the edge is marked $m_{ij}$ if $m_{ij} \ge 4$. So an unmarked edge between the $i^{th}$ and $j^{th}$ node means $m_{ij} = 3$, while no edge between the $i^{th}$ and $j^{th}$ node means $m_{ij} = 2$.*

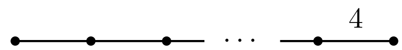
A Coxeter diagram is called crystallographic if $m_{ij} \in \{1, 2, 3, 4, 6\}$ for all $i, j$. Finite Coxeter groups corresponding to crystallographic Coxeter diagrams are also called Weyl groups. Both the symmetric group $S_{n+1}$ and the hyperoctahedral group $C_2^n \rtimes S_n$ are Weyl groups.

**Example 3.15.** *The symmetric group $W = S_{n+1}$ has a Coxeter presentation with generators $s_1 = (12), \cdots, s_n = (n\ n+1)$ the nearest neighbour transpositions. The Coxeter matrix is given by*

$$m_{ij} = \begin{cases} 1 & \text{if } i = j \\ 2 & \text{if } |i-j| \ge 2 \\ 3 & \text{if } |i-j| = 1 \end{cases}$$

*So the Coxeter diagram with nodes numbered from left to right is of the form*

**Example 3.16.** *The hyperoctahedral group* $W = C_2^n \rtimes S_n$ *has a Coxeter presentation with generators* $s_1 = (12), \cdots, s_{n-1} = (n-1\ n) \in S_n$ *and* $s_n = (1, \cdots, 1, -1) \in C_2^n$. *The Coxeter diagram with nodes numbered from left to right is of the form*




Consider a Euclidean vector space $V$ with basis $e_1, \cdots, e_n$ and with inner product given by the Gram matrix

$$\langle e_i, e_j \rangle = -2\cos(\pi/m_{ij})$$

for all $i, j$. Define the orthogonal reflection

$$s_i : V \to V\ ,\ s_i(v) = v - \langle v, e_i \rangle e_i$$

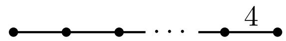
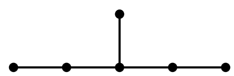
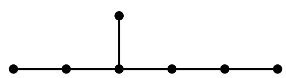
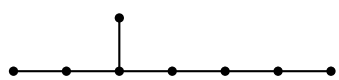
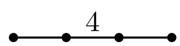
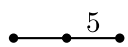
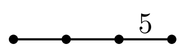
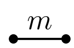
with mirror the orthogonal complement of $e_i$. It is easy to check that this assignment extends to a homomorphism $W \to \mathrm{O}(V)$. This is called the reflection representation of the Coxeter group $W$. The inner product on $V$ is positive definite if and only if the Coxeter group $W$ is finite. From now on we assume that the Coxeter group $W$ is a finite group.

It turns out that for finite Coxeter groups the Coxeter diagram has no loops. The Coxeter element is the product of the involutive generators taken in some order. One can show that all Coxeter element are conjugated in $W$. Suppose in addition that the Coxeter diagram is connected. The order of a Coxeter element is called the Coxeter number, usally denoted $h$. The eigenvalues of a Coxeter element in the reflection representation are $\exp(2\pi i m_j/h)$ with

$$1 = m_1 \leq m_2 \leq \cdots \leq m_n = (h-1)$$

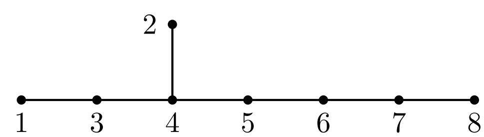
the sequence of exponents.

Here is a list of the finite Coxeter groups, which are irreducible in the sense that the Coxeter diagram is connected, or equivalently for which the reflection representation is irreducible. This classification can be found in various text books [2], [8]. In the first column we have the Cartan symbol, with the subindex $n$ for the number of nodes of the Coxeter diagram. In the second column we have the Coxeter diagram, in the third column the Coxeter number, and in the last column the exponents. This classification of finite Coxeter groups is one of the most fundamental classifications in mathematics. For example it plays a crucial role in the classification of the simple algebraic groups.

| Name | Coxeter diagram | $h$ | Exponents |
|---|---|---|---|
| $A_n$ |  | $n+1$ | $1, 2, \cdots, n$ |
| $B_n$ |  | $2n$ | $1, 3, 5, \cdots, 2n-1$ |
| $D_n$ |  | $2n-2$ | $1, 3, \cdots, 2n-3, n-1$ |
| $E_6$ |  | 12 | $1, 4, 5, 7, 8, 11$ |
| $E_7$ |  | 18 | $1, 5, 7, 9, 11, 13, 17$ |
| $E_8$ |  | 30 | $1, 7, 11, 13, 17, 19, 23, 29$ |
| $F_4$ |  | 12 | $1, 5, 7, 11$ |
| $H_3$ |  | 10 | $1, 5, 9$ |
| $H_4$ |  | 30 | $1, 11, 19, 29$ |
| $I_2(m)$ |  $m \geq 5$ | $m$ | $1, m-1$ |

With this basic knowledge of Coxeter groups in mind it is easy to see that the Weyl group $W(E_8)$ can occur as a hypergeometric group as shown in Example 3.12.

**Example 3.17.** *Consider the Coxeter diagram of type $E_8$ with the nodes traditionally numbered by*




*In Example 3.12 the monodromy group $H(a, b)$ is the subgroup of the Coxeter group $W(E_8)$ generated by*

$$h_\infty = s_2 s_1 s_3 s_5 s_6 s_7 s_8 s_4 \ , \ h_1 = s_4 \ , \ h_0 = s_8 s_7 s_6 s_5 s_3 s_1 s_2$$

*for which the topological relation* $h_\infty h_1 h_0 = 1$ *indeed is true. So* $h_\infty$ *is a Coxeter element of type* $E_8$ *and* $h_0$ *is a Coxeter element of type* $A_1+A_2+A_4$*. The element* $h_1$ *is indeed a reflection. The fact that* $H(a,b) = W(E_8)$ *follows from the fact that* $W(E_8) \cong \mathrm{O}_8^+(2)$ *and* $\mathrm{O}_8^+(2) = 2.G.2$ *with* $G \cong \mathrm{PSO}_8^+(2)$ *a simple group of order* 174.182.400*.*

**Example 3.18.** *Recall Clausen's formula*

$$_2F_1(\alpha,\beta,\alpha+\beta+1/2;z)^2 = {}_3F_2(2\alpha,2\beta,\alpha+\beta,2\alpha+2\beta,\alpha+\beta+1/2;z)$$

*as discussed in Example 3.6, and look at the particular example*

$$_2F_1(1/4,-1/12,2/3;z)^2 = {}_3F_2(1/2,-1/6,1/6,1/3,2/3;z)$$

*with* $\alpha = 1/4, \beta = -1/12$ *and* $\gamma = \alpha+\beta+1/2 = 2/3$*. The function on the left hand side is algebraic with projective monodromy group the tetrahedral group* $A_4$*. Indeed the exponent differences are* $\frac{1}{2},\frac{1}{3},\frac{1}{3}$ *at the points* $1,0,\infty$ *respectively. So the right hand side is again an algebraic function. The monodromy group of the latter is* $C_2 \times A_4 \simeq C_2^3 \rtimes A_3$ *(of index* 2 *in* $W(B_3) = C_2^3 \rtimes S_3$*) in its three dimensional reflection representation. Indeed the eigenvalues match for*

$$h_\infty = -(234)\ ,\ h_1 = -(12)(34)\ ,\ h_0 = (123)$$

*and* $h_\infty h_1 h_0 = 1$ *as should.*

## 3.5 Prime Number Theorem after Tchebycheff

In this section we discuss the proof by Tchebycheff of a weak version of the Prime Number Theorem. His proof is very elegant. See also page 622 of the interview from 2005 with Selberg [3]. It was pointed out by Rodriguez-Villegas [16] that a crucial step in this proof of Tchebycheff is the same interlacing property that we encountered in Example 3.12.

Let $\pi(x) = \#\{p; p \le x\}$ denote the standard prime counting function. Introduce the numbers

$$A = \log\frac{2^{\frac{1}{2}}3^{\frac{1}{3}}5^{\frac{1}{5}}}{30^{\frac{1}{30}}} = 0.92129022\cdots\ ,\ B = 6A/5 = 1.105550428\cdots$$

which enter in the argument below. In 1852 Tchebycheff proved in an elementary way the following result towards the Prime Number Theorem [22].

**Theorem 3.19.** *We have*

$$\frac{Ax}{\log x}(1+o(x)) < \pi(x) < \frac{Bx}{\log x}(1+o(x))$$

Introduce the following three prime counting functions for $x > 0$

$$\pi(x) = \sum_{p \le x} 1 \ , \ \theta(x) = \sum_{p \le x} \log p \ , \ \psi(x) = \sum_{p^m \le x} \log p$$

with $p$ always denoting a prime number, and $m = 1, 2, 3, \cdots$ denoting a positive integer. It is obvious that

$$\psi(x) = \sum_{p \le x} \left[ \frac{\log x}{\log p} \right] \log p$$

with $[\log x / \log p]$ the largest integer $m$ with $p^m \le x$. In turn

$$\psi(x) = \theta(x) + \theta(x^{\frac{1}{2}}) + \theta(x^{\frac{1}{3}}) + \theta(x^{\frac{1}{4}}) + \cdots$$

is clear as well.

**Theorem 3.20.** *We have* $\psi(x) = \theta(x) + O(x^{\frac{1}{2}} \log^2 x)$.

*Proof.* Clearly $\theta(x^{\frac{1}{m}}) = 0$ if $x < 2^m$ or equivalently $\log x / \log 2 < m$. For $m \ge 2$ we have

$$\theta(x^{\frac{1}{m}}) \le (x^{\frac{1}{m}} \log x)/m \le x^{\frac{1}{2}} \log x$$

using $\theta(x) \le x \log x$, which in turn implies that

$$\sum_{m \ge 2} \theta(x^{\frac{1}{m}}) \le x^{\frac{1}{2}} \log x \cdot \frac{\log x}{\log 2} < 2x^{\frac{1}{2}} \log^2 x$$

using $2 \log 2 > 1$. Hence we have

$$\theta(x) \le \psi(x) \le \theta(x) + O(x^{\frac{1}{2}} \log^2 x)$$

which proves the theorem. □

The Prime Number Theorem is usually stated in the form

$$\pi(x) \sim \frac{x}{\log x}$$

but can be reformulated as

$$\psi(x) \sim x$$

and the proof of Tchebycheff will focus on the latter formulation.

**Theorem 3.21.** *We have* $T(x) \stackrel{\text{def}}{=} \sum_{k\geq 1} \psi(x/k) = \log([x]!)$.

*Proof.* For $n = [x]$ a natural number the numbers

$$1, 2, \cdots, n$$

include just $[n/p] = [x/p]$ multiples of $p$, and $[n/p^2] = [x/p^2]$ multiples of $p^2$, and so on. Hence

$$n! = \prod_p p^{k_p}\ ,\ k_p = \sum_{m\geq 1} [x/p^m]$$

which can be rewritten as

$$\log(n!) = \sum_p k_p \log p = \sum_{p,m} [x/p^m] \log p\ .$$

Observe that

$$[x/p^m] = l \geq 1 \Leftrightarrow x/kp^m \geq 1 \text{ exactly for } k = 1, 2, \cdots, l$$

and therefore (with the sum in the middle term over those triples $p, m, k$ with $p^m \leq x/k$)

$$\log(n!) = \sum_{p,m,k} \log p = \sum_{k\geq 1} \psi(x/k)$$

which proves the theorem. □

The problem is to turn the good asymptotic understanding of $T(x)$ by Stirling's formula into asymptotic understanding of $\psi(x)$. For this purpose Tchebycheff made the following crucial step. If we introduce the function

$$F(x) = T(x) + T(x/30) - T(x/2) - T(x/3) - T(x/5)$$

and use

$$T(x) = \sum_{k\geq 1} \psi(x/k)$$

then we can rewrite

$$F(x) = \sum_{k\geq 1} A_k \psi(x/k)$$

with

$$A_k = \begin{cases} +1 & \text{if } k \text{ is not divisible by } 2, 3, 5 \\ 0 & \text{if } k \text{ is divisible by exactly one number from } 2, 3, 5 \\ -1 & \text{if } k \text{ is divisible by at least two numbers from } 2, 3, 5 \end{cases}$$

For example if $k$ is divisble by 2 but not by $3, 5$ then the term $\psi(x/k)$ enters in $T(x)$ and in $-T(x/2)$, but does not enter in $T(x/30) - T(x/3) - T(x/5)$. Hence $A_k = 0$ in that case. A direct verification shows that

$$A_k = \begin{cases} +1 & \text{if } k \equiv 1, 7, 11, 13, 17, 19, 23, 29 \bmod 30 \\ -1 & \text{if } k \equiv 6, 10, 12, 15, 18, 20, 24, 30 \bmod 30 \\ 0 & \text{if else} \end{cases}$$

Observe that the two sequences of natural numbers

$$\{k; A_k = +1\} \qquad \{k; A_k = -1\}$$

interlace. It is the same interlacing property that we have seen in Example 3.12.

**Corollary 3.22.** *We can write*

$$F(x) = \psi(x) - \psi(x/6) + \psi(x/7) - \psi(x/10) + \psi(x/11) - \psi(x/12) + \cdots$$

*with alternating plus and minus signs, which in turn implies the key inequality*

$$\psi(x) - \psi(x/6) < F(x) < \psi(x)$$

*because* $\psi(x) = \sum_p [\log x / \log p] \log p$ *is monotonic increasing in* $x$.

Recall Stirling's formula

$$n! = \sqrt{2\pi n} \exp(n \log n - n + \theta/12n)$$

for some $0 < \theta < 1$.

**Corollary 3.23.** *Using* $T(x) = \log([x]!)$ *and Stirling's formula we have the inequalities*

$$\frac{1}{2}\log(2\pi) + x \log x - x - \frac{1}{2}\log x < T(x)$$

$$T(x) < \frac{1}{2}\log(2\pi) + x \log x - x + \frac{1}{2}\log x + 1/12$$

*as lower and upper bound for* $T(x)$.

**Corollary 3.24.** *Using* $F(x) = T(x) + T(x/30) - T(x/2) - T(x/3) - T(x/5)$ *we have the inequalities*

$$F(x) < Ax + \frac{5}{2}\log x - \frac{1}{2}\log(1800\pi) + 2/12 < Ax + \frac{5}{2}\log x$$

$$F(x) > Ax - \frac{5}{2}\log x + \frac{1}{2}\log(450/\pi) - 3/12 > Ax - \frac{5}{2}\log x$$

*with* $A = \frac{1}{2}\log 2 + \frac{1}{3}\log 3 + \frac{1}{5}\log 5 - \frac{1}{30}\log 30 = 0.92129022\cdots$.

Using the key inequality of Corollary 3.22

$$\psi(x) - \psi(x/6) < F(x) < \psi(x)$$

we get

$$Ax - \frac{5}{2}\log x < \psi(x)\ ,\ \psi(x) - \psi(x/6) < Ax + \frac{5}{2}\log x$$

and the second inequality can be iterated. Indeed

$$\psi(x) < Ax + \frac{5}{2}\log x + \psi(x/6)$$

$$< Ax(1 + 1/6) + \frac{5}{2}(2\log x - \log 6) + \psi(x/6^2)$$

$$< Ax(1 + 1/6 + 1/6^2) + \frac{5}{2}(3\log x - (1+2)\log 6) + \psi(x/6^3)$$

$$< Ax(1+1/6+\cdots+1/6^m) + \frac{5}{2}((m+1)\log x - \frac{1}{2}m(m+1)\log 6) + \psi(x/6^{m+1})$$

$$< \frac{6}{5}Ax + O(\log^2 x)$$

since

$$\psi(x/6^{m+1}) = 0 \Leftrightarrow x/6^{m+1} < 2 \Leftrightarrow (m+1) > \frac{\log(x/2)}{\log 6}$$

This ends our discussion of the proof of the following theorem of Tchebycheff.

**Theorem 3.25.** *We have*

$$Ax + O(\log x) < \psi(x) < Bx + O(\log^2 x)$$

*with* $A = 0.92129022\cdots$ *and* $B = 6A/5 = 1.105550428\cdots$.

Equivalently we arrive at

$$\frac{Ax}{\log x}(1 + o(x)) < \pi(x) < \frac{Bx}{\log x}(1 + o(x))$$

and so

$$\pi(x) \asymp \frac{x}{\log x}$$

which is Tchebycheff's weak version of the Prime Number Theorem.

**Remark 3.26.** *The proof of Tchebycheff has two main ideas. The first step is to work with the prime counting function $\psi(x)$ instead of the usual function $\pi(x)$, and to consider the function $T(x) = \sum \psi(x/k) = \log([x]!)$.*

*The second step is to turn good asymptotic understanding for $T(x)$ from Stirling's formula into good asymptotic understanding for $\psi(x)$. A first try might be to consider*

$$F(x) = T(x) - 2T(x/2) = \psi(x) - \psi(x/2) + \psi(x/3) - \psi(x/4) + \cdots$$

*which in turn implies that*

$$F(x) < \psi(x) < F(x) + \psi(x/2) \ .$$

*Now the same method of proof works in a simpler way leading to*

$$Ax + O(\log x) < \psi(x) < Bx + O(\log x)$$

*with $A = \log 2 = 0.693\cdots$ and $B = 2\log 2 = 1.386\cdots$. Having established this special case first it might be not unreasonable to try*

$$F(x) = T(x) + T(2x/n) - T(x/p) - T(x/q) - T(x/r)$$

*with $p \geq q \geq r \geq 2$ and $1/p + 1/q + 1/r = 1 + 2/n$. Of course, the case $(p,q,r,n) = (m,2,2,2m)$ gives back the previous case. There are just a few other possibilities*

| $p$ | $q$ | $r$ | $n$ |
|---|---|---|---|
| $m$ | 2 | 2 | $2m$ |
| $p$ | 3 | 2 | $12p/(6-p)$ |

*with $p = 3,4,5$ and $n = 12,24,60$. These numbers are also familiar from the classification of the Platonic solids.*

*For $q = 3, r = 2$ the coefficients $A_k$ for $p = 3$ are given by*

$$A_k = \begin{cases} +1 & \textit{if } k \equiv 1,5 \textit{ mod } 6 \\ -1 & \textit{if } k \equiv 3,6 \textit{ mod } 6 \\ 0 & \textit{if else} \end{cases}$$

*and for $p = 4$ become*

$$A_k = \begin{cases} +1 & \textit{if } k \equiv 1,5,7,11 \textit{ mod } 12 \\ -1 & \textit{if } k \equiv 4,6,8,12 \textit{ mod } 12 \\ 0 & \textit{if else} \end{cases}$$

*and so the same interlacing property holds for all these cases.*

*If* $q = 3, r = 2$ *and* $p = 3, 4, 5$ *and so* $n = 12p/(6-p) = 12, 24, 60$ *respectively one gets*

$$A = \tfrac{1}{2}\log 2 + \tfrac{1}{3}\log 3 + \tfrac{1}{p}\log p - \tfrac{2}{n}\log(n/2)$$

*which amounts to* $A = 0.780\cdots$ *for* $p = 3$ *and* $A = 0.852\cdots$ *for* $p = 4$. *Moreover* $B = 3A/2 = 1.171\cdots$ *for* $p = 3$ *and* $B = 4A/3 = 1.136\cdots$ *for* $p = 4$. *All in all, the method gives the sharpest bounds for the icosahedron with* $(p, q, r, n) = (5, 3, 2, 60)$, *and this is the case discussed by Tchebycheff.*

## 3.6 Exercises

**Exercise 3.1.** *The monodromy group of Example 2.6 is a hypergeometric group* $H(a, b)$ *with parameter sets*

$$a = \{\zeta_{12}\zeta_{2p}, \zeta_{12}/\zeta_{2p}\}\ ,\ b = \{\zeta_3^{-2}, 1\}$$

*and* $\zeta_k = \exp(2\pi i/k)$. *Show that for* $p \geq 7$ *all Galois conjugates of* $H(a, b)$ *different from the identity and complex conjugation have parameters that interlace on the unit circle if and only if*

$$p = 7, 8, 9, 10, 11, 12, 14, 16, 18, 24, 30, \infty\ .$$

*This list was found by Fricke and Klein, and extended by Takeuchi to a complete list of arithmetic triangle groups [7], [21].*

**Exercise 3.2.** *Show by the method of Example 3.17 that there are eight variations for obtaining the Weyl group* $W(E_8)$ *as a hypergeometric group.*

# References


[1] F. Beukers and G.J. Heckman, Monodromy for the hypergeometric function ${}_nF_{n-1}$, Invent. Math. 95 (1989), 325-354.

[2] N. Bourbaki, Éléments de Mathematiques, Groupes et Algèbres de Lie, Chapitres 4,5 et 6, Masson, Paris, 1981.

[3] N.A. Baas and Ch.F. Skau, The lord of the numbers, Atle Selberg. On his life and mathematics, Bulletin AMS 45 (2008), 617-649.

[4] T. Clausen, Über die Fälle, wenn die Reihe von die Form $y = 1 + \cdots$ ein Quadrat von der form $z = 1 + \cdots$ hat, J. Reine Angew. Math. 3 (1828), 89-91.

[5] H.S.M. Coxeter, Finite groups generated by unitary reflections, Abh. Math. Sem. Hamburg 31 (1967), 125-135.

[6] P. Deligne and G.D. Mostow, Monodromy of hypergeometric functions and non-lattice integral monodromy, Publ. Math. IHES 63 (1986), 1-89.

[7] K.E.R. Fricke and F. Klein, Vorlesungen über die Theorie der automorphen Funktionen I, Teubner, Leipzig, 1897.

[8] J.E. Humphreys, Reflection Groups and Coxeter Groups, Cambridge Studies in Advanced Mathematics 29, Cambridge University Press, Cambridge, 1990.

[9] N. Katz, Rigid Local Systems, Annals of Mathematics Studies 139, Princeton University Press, Princeton, 1995.

[10] F. Klein, Vorlesungen über die hypergeometrische Funktion, Springer Verlag, Berlin, 1933.

[11] G. Lauricella, Sulla funzioni ipergeometriche a pi variabili, Rend. Circ. Math. Palermo 7 (1893), 111-158.

[12] A.H.M. Levelt, Hypergeometric Functions, PhD, University of Amsterdam, 1961.

[13] E.J.N. Looijenga, Uniformization by Lauricella functions, an overview of the theory of Deligne–Mostow, Arithmetic and Geometry Around Hypergeometric Functions. Progress in Mathematics 260, Birkhauser, Basel (2007), 207-244.

[14] A. Pressley, Elementary Differential Geometry, Springer Undergraduate Mathematical Series, 2008.

[15] G.F.B. Riemann, Beiträge zur Theorie der durch die Gaussische hypergeometrische Reihe $F(\alpha, \beta, \gamma; x)$ darstellbaren Funktionen (1857), Gesammelte mathematische Werke, Teubner, Leipzig, 1892.

[16] F. Rodriguez-Villegas, Integral ratios of factorials and algebraic hypergeometric functions, Notes from a lecture at Oberwolfach, 2004, http://www.ma.utexas.edu/users/villegas/publications.html

[17] H.A. Schwarz, Über diejenigen Fälle in welchen die Gaussische hypergeometrische Reihe einer algebraische Funktion ihres vierten Elementes darstellt, Crelle J. 75 (1873), 292-335.

[18] G.C. Shephard and J.A. Todd, Finite unitary reflection groups, Canadian J. Math. 6 (1954), 274-304.

[19] C.T. Simpson, Product of Matrices, Canadian Mathematical Society Conference Proceeding, Volume 12 (1992), 157-185.

[20] I. Stewart and D.O. Tall, Algebraic Number Theory and Fermat's Last Theorem, Third Edition, Taylor and Francis, UK, 2001.

[21] K. Takeuchi, Arithmetic triangle groups, J. Math. Soc. Japan 29 (1977), 91-106.

[22] P.L. Tchebycheff, Mémoire sur les nombres premiers, J. Math. Pures et Appliquées XVII (1852), 366-390.

[23] J. Thomae, Über die höheren hypergeometrische Reihen, Math. Ann. 2 (1870), 427-444.

[24] E.C. Titchmarsh, Eigenfunction Expansions, Clarendon Press, Oxford, 1962.