Invent. math. 92, 187–216 (1988)

Inventiones mathematicae

Werte hypergeometrischer Funktionen

Jürgen Wolfart

Mathematisches Seminar der Universität, Robert-Mayer-Str. 6–10, D-6000 Frankfurt 1, Bundesrepublik Deutschland

Summary. A partial result on a problem of Siegel is given: Hypergeometric functions with rational parameters have transcendental values in almost all algebraic points – up to some natural exceptions; these exceptions are the well-known algebraic functions and an (unexpected) second class of examples related to certain Shimura-curves.

Inhalt

§1. Ergebnis und Methode

Die klassischen hypergeometrischen Funktionen

$$F(a,b,c;z)=1+\frac{a\,b}{c}z+\frac{a(a+1)\,b(b+1)}{2!\,c(c+1)}z^2+\dots \tag{1}$$

mit $a,b,c\in\mathbb{Q}$, $c\neq 0,-1,-2,\dots$ und $|z|<1$ gehören zu den interessantesten Beispielen von G-Funktionen, und in seiner Untersuchung der algebraischen Unabhängigkeit der Werte der Besselfunktionen ([Sie 1], dem Beginn der heute so genannten Siegel-Shidlovsky-Methode der Transzendenztheorie) hat Siegel die Frage nach der algebraischen Natur der Werte der hypergeometrischen Funktionen an algebraischen Stellen $z\neq 0$ aufgeworfen. Für die allgemeine Theorie der

G-Funktionen sei auf Bombieri [Bo] verwiesen, für die Geschichte des Problems und neuere Resultate auf die Einleitung von [BW]. Für allgemeine G-Funktionen sind bisher nur Irrationalitätsresultate bekannt; die Erwartung, daß – wie bei den E-Funktionen – eigentlich Transzendenzaussagen gelten müßten, konnte bisher nur für die Klasse der nichtalgebraischen Integralfunktionen

$$f(z) := \int_0^z g(t)\,dt$$

von algebraischen, über $\overline{\mathbb{Q}}$ definierten und im Einheitskreis regulären $g(t)$ bestätigt werden (Wüstholz [Wü]). Hierbei ist allerdings nicht die Siegel-Shidlovsky-Methode verwendet worden, sondern eine u.a. von Wüstholz entwickelte weitreichende Verallgemeinerung der Gelfond-Schneiderschen Transzendenzmethode.

Dies gilt auch für die hier gewählte Behandlung der Funktionenklasse (1). Das Resultat ist aber kein reines Transzendenzresultat: Die im folgenden Theorem unter (B) genannten Ausnahmen erklären in natürlicher Weise die Schwierigkeiten, die bei der Behandlung allgemeiner G-Funktionen durch die Siegel-Shidlovsky-Methode auftreten müssen.

Theorem. *Seien a, b, c rationale Parameterwerte, $c \neq 0, -1, -2, \ldots$ und $F(z) := F(a, b, c; z)$ wie in* (1) *definiert. Die Monodromiegruppe der Differentialgleichung von F sei mit Δ bezeichnet, und*

$$E := \{\xi \in \overline{\mathbb{Q}} \mid F(\xi) \in \overline{\mathbb{Q}}\}$$

sei die Menge der Ausnahmeargumente von F. Dann sind drei Fälle zu unterscheiden:

(A) *$E = \overline{\mathbb{Q}}$ genau dann, wenn $F(z)$ algebraische Funktion ist.*

(B) *E und $\overline{\mathbb{Q}} - E$ liegen dicht in $\overline{\mathbb{Q}}$ genau dann, wenn Δ arithmetisch definiert ist und*

$$c < 1, \quad 0 < a < c, \quad 0 < b < c,$$
$$1 - c + |a - b| + |c - a - b| < 1$$

erfüllt ist.

(C) *E ist endlich in allen anderen Fällen.*

Bemerkungen. 1) Die Voraussetzung $|z| < 1$, die man für allgemeine G-Funktionen machen muß, ist hier deswegen unwesentlich, weil die Aussage des Theorems für alle analytischen Fortsetzungen von F richtig bleibt: Bei der im Beweis des Satzes verwendeten Integraldarstellung wirkt sich die analytische Fortsetzung nur auf die Wahl des Integrationswegs aus.

2) Die endlich vielen Ausnahmefunktionen $F(a, b, c; z)$, welche Teil (B) des Theorems erfüllen, sind explizit bekannt: Als Monodromiegruppen hypergeometrischer Differentialgleichungen treten ausschließlich Dreiecksgruppen auf, und durch die Arbeiten von Takeuchi [Ta 1], [Ta 2] weiß man, daß bis auf Isomorphie genau 85 arithmetisch definierte Dreiecksgruppen existieren und kennt ihre Signaturen (p, q, t), $p, q, t \in \mathbb{N} \cup \{\infty\}$ mit $\frac{1}{p} + \frac{1}{q} + \frac{1}{t} < 1$. Mit Ausnahme der Gruppe

mit der Signatur $(p,q,t)=(\infty,\infty,\infty)$ (isomorph zur Hauptkongruenzuntergruppe der Stufe 2 der elliptischen Modulgruppe) gibt es zu jeder arithmetischen Dreiecksgruppe Δ mindestens eine hypergeometrische Funktion $F(a,b,c;z)$, die Δ als Monodromiegruppe besitzt und deren Parameter die in (B) geforderten Ungleichungen erfüllen; man wähle zur Signatur (p,q,t), o.B.d.A. mit $p\neq\infty$, die Parameter z.B. so, daß

$$\lambda=1-c=\frac{1}{p}, \qquad |\mu|=|a-b|=\frac{1}{q}, \qquad |\nu|=|c-a-b|=\frac{1}{t}$$

gelten. In manchen Fällen lassen sich a, b, c auch noch auf andere Weise wählen[1].

3) Zu jeder solchen Ausnahmefunktion vom Typ (B) läßt sich die Menge E der Ausnahmeargumente explizit beschreiben: Zu $F(z)$ gehört, wie in § 8, Satz 7 erläutert werden wird, ein über $\mathbb{Q}$ abelscher Zahlkörper L, und die Monodromiegruppe Δ läßt sich so normieren, daß sie als Untergruppe von $PGL_2 L$ auf einer geeigneten Kreisscheibe K diskontinuierlich operiert. Dazu findet man eine Δ-automorphe Funktion j, eine „Shimura-Abbildung", die auf $K\cap L$ algebraische Werte annimmt [Shi 2], und genau diese Werte bilden die Ausnahmemenge

$$E=j(K\cap L).$$

Da $K\cap L$ dicht liegt in K, muß E dicht liegen in $\bar{\mathbb{Q}}$; andererseits können diese Werte von j nicht in beliebigen algebraischen Zahlkörpern liegen, darum ist auch $\bar{\mathbb{Q}}-E$ dicht in $\bar{\mathbb{Q}}$. Beispiele mit der Monodromiegruppe $\Delta\cong PSL_2\mathbb{Z}$ führen natürlich auf Funktionen j, die bis auf gebrochen-rationale Transformationen die klassische „absolute Invariante" sind; L ist hier je nach Wahl der Parameter $\mathbb{Q}(\sqrt{-1})$ oder $\mathbb{Q}(\sqrt{-3})$, und anstelle des Satzes von Shimura genügt hier die klassische Theorie der komplexen Multiplikation. In [BW] ist dieser Fall näher ausgeführt. Allgemeiner kann man in einfachen Fällen j als meromorphe Fortsetzung der Umkehrfunktion einer „Schwarzschen Dreiecksfunktion" konstruieren, die wir in § 2 der hypergeometrischen Funktion F zuordnen werden.

4) Mit der Bestimmung der Ausnahmeargumente $\xi\in E$ ist noch nichts gesagt über die Natur der algebraischen Werte $F(\xi)$ der hypergeometrischen Funktionen von Typ (B), und der hier geführte Beweis gibt kaum Hinweise darauf, wie $F(\xi)$ explizit bestimmt werden könnte. Frits Beukers hat dieses Problem gelöst, und mit seiner freundlichen Genehmigung sei hier die entscheidende Idee erwähnt, daß nämlich – mit den oben eingeführten Bezeichnungen – die Funktion $F\circ j$ sich wie eine Δ-automorphe Form vom Gewicht 1 verhält, allerdings mit Singularitäten und Multiplikatorsystem. Im Fall $\Delta\cong PSL_2\mathbb{Z}$ kann man diese mit Hilfe bekannter Modulformen näher identifizieren und $F(\xi)$ mit Hilfe von Modulargleichungen berechnen [BW], so z.B.

$$F\left(\frac{1}{12},\frac{5}{12},\frac{1}{2};\frac{1323}{1331}\right)=\frac{3}{4}\sqrt[4]{11}.$$

[1] Wenn Δ nämlich durch Transformationen erzeugt werden kann, die die üblichen Relationen erfüllen, aber keine Nachbartransformationen des kanonischen Fundamentalbereichs sind (vgl. § 8 und [Kn]). Diese Möglichkeit habe ich in der Fußnote zu Seite IX.5 in [Wo] irrtümlicherweise nicht berücksichtigt

5) Für hypergeometrische Funktionen vom allgemeinen Typ (C) läßt sich – mit möglicher Ausnahme der in §3 diskutierten Sonderfälle[2] – die Ausnahmemenge ebenfalls explizit bestimmen: Sie ist in den j-Bildern der Fixpunkte einer endlichen Erweiterung $\bar{\Delta}$ von Δ enthalten (§9, Satz 8 und §10). Da Dreiecksgruppen generisch maximal sind, kommen als Ausnahmeargumente im allgemeinen Fall nur 0 und allenfalls 1 in Frage. Für den Fall nicht-maximaler Δ, die dann auftretenden Ausnahmeargumente und die Berechnung der algebraischen Ausnahmewerte $F(\xi)$ vgl. [BW], Theorem 4.

6) Nach wie vor ungelöst ist Siegels Frage nach Sätzen über algebraische Unabhängigkeit von $F(\xi)$ und $F'(\xi)$ an algebraischen Stellen.

Der *Beweis* des Theorems beruht auf der Integraldarstellung der hypergeometrischen Funktion, meistens – bis auf die in §3 abgehandelten Sonderfälle – als Quotient zweier Periodenintegrale 2. Art auf gewissen glatten algebraischen Kurven, die beide über $\bar{\mathbb{Q}}$ definiert sind, wenn $z \in \bar{\mathbb{Q}}$. Diese Perioden muß man interpretieren als Perioden $\neq 0$ auf den einfachen Faktoren der Jacobivarietäten dieser Kurven. Dann sagt ein Satz von Wüstholz [WW, Satz 2], daß solche Quotienten von Perioden höchstens in dem offensichtlichen Fall algebraisch sein können, wenn die beteiligten einfachen, über $\bar{\mathbb{Q}}$ definierten abelschen Varietäten isogen sind. Deswegen ist der Schlüssel zu dem hier geführten Beweis die Zerlegung gewisser Jacobivarietäten und der Vergleich ihrer einfachen Faktoren anhand ihrer Endomorphismenringe.

Die Nennerperioden sind Werte der Betafunktion an rationalen Stellen; für die zugehörigen einfachen abelschen Varietäten A_2 kann man sich auf Untersuchungen von Rohrlich und Koblitz über die Jacobivarietäten von Fermatkurven berufen ([Gr], [KR]). Diese werden kurz in §2 referiert, zusammen mit einigen anderen Hilfsmitteln, die für die Behandlung der Zählerperioden benötigt werden.

Die abelschen Varietäten A_1, die zu den Zähler-Perioden gehören, sind Faktoren von $\mathrm{Jac}\, X(N,z)$, wobei $X(N,z)$ eine Familie glatter projektiver algebrascher Kurven ist (definiert über $\bar{\mathbb{Q}}$ für $z \in \bar{\mathbb{Q}}$), deren explizite Gestalt fast unmittelbar an der klasssischen Integraldarstellung für $F(z)$ abgelesen werden kann ($N \in \mathbb{N}$ ist Hauptnenner von a, b, c, vgl. §2). $\mathrm{Jac}\, X(N,z)$ wird als komplexer Torus bestimmt, indem zunächst eine Basis der Differentiale 1. Art auf $X(N,z)$ ermittelt wird (§4) und dann über die Homologie von $X(N,z)$ integriert wird (§5); eine Basis für diese Homologie läßt sich nach Felix Klein [Kl] in Form geeigneter „Pochhammer-Zykeln" auf $X(N,z)$ wählen. Nach Abspalten uninteressanter Faktoren aus $\mathrm{Jac}\, X(N,z)$ bleibt eine abelsche Varietät T der Dimension $\Phi(N)$ übrig (§5, Satz 1; Φ die Eulersche Funktion), die immer noch von z abhängt, und deren Endomorphismenalgebra den Kreisteilungskörper $\mathbb{Q}(\zeta_N)$ enthält (§5, Satz 2).

Alle polarisierten abelschen Varietäten des gleichen Endomorphietyps werden nach [Shi1] oder [Sie2] durch ein komplexes symmetrisches Gebiet G

[2] Nachtrag (Okt. 87): Herr T. Matali-aho hat mit Hilfe geschickter Summationsverfahren für (1) bzw. unter Verwendung der Gaußschen Nachbarschaftsrelationen folgende Beispiele für Ausnahmeargumente und -werte in den Sonderfällen gefunden: Für alle rationalen $|c|<|b|$ ist $F\left(2, b+1, c+2; \frac{c}{b}\right) = b\frac{c+1}{b-c}$

parametrisiert, dessen Dimension r explizit durch (24) mit (21) und (3) berechnet werden kann. Die Koordinaten, durch welche T in diesem Gebiet beschrieben wird, sind Werte $D_\omega(z)$ gewisser Dreiecksfunktionen D_ω, die die obere Halbebene $\mathfrak{H}$ auf ein hyperbolisches Kreisbogendreieck abbilden (§ 5, Satz 3). Es ist daher nicht verwunderlich, daß sich die Frage, ob A_2 isogen zu einem Faktor A_1 von T ist, umformulieren läßt in eine Rationalitätsbedingung an die $D_\omega(z)$ (§ 6, Satz 5). Hilfsmittel bei dieser Umformulierung ist ein genaues Studium jener Endomorphismen, die T eventuell noch außer $\mathbb{Q}(\zeta_N)$ besitzt (§ 6, Satz 4).

Zur Behandlung dieser Rationalitätsbedingung ist eine Fallunterscheidung nötig: Erstens ist der Fall $\operatorname{Dim} G = r = 0$ möglich, d.h. daß keine Bedingung an irgendein $D_\omega(z)$ zu stellen ist, um zu garantieren, daß A_2 isogen zu einem Faktor von T ist. Hier läßt sich zeigen, daß Δ diskrete Untergruppe einer kompakten Gruppe und somit endlich ist. Daraus folgt nach einem klassischen Resultat von H.A. Schwarz [Schw], daß $F(z)$ algebraische Funktion von z ist (§ 7, Satz 6).

Zweitens ist $r = 1$ möglich oder allgemeiner die folgende Situation: Wegen linearer Abhängigkeiten zwischen den verschiedenen D_ω werden alle fraglichen T durch einen eindimensionalen linearen Unterraum des komplexen symmetrischen Gebiets G parametrisiert. Man befindet sich also in der Theorie der Shimurakurven, Δ spielt die Rolle der zugehörigen Modulgruppe, und D_ω hat als Umkehrabbildung eine als Shimura-Abbildung normierte Δ-automorphe Funktion j_ω oder eine algebraische Funktion von j_ω. Die schon erwähnten algebraischen Eigenschaften der Werte $z = j_\omega(\tau)$ an CM-Punkten $\tau = D_\omega(z)$ führen dazu, daß die Rationalitätsbedingungen $D_\omega(z) \in L$ auf einer Menge E von Ausnahmeargumenten des Typs (B) erfüllt sind (§ 8, Satz 7).

Drittens wird dann in § 9 die Situation betrachtet, daß die verschiedenen D_ω nicht linear, wohl aber analytisch voneinander abhängen. Das heißt: Die hier auftretenden abelschen Varietäten T werden durch eine nichtlineare analytische Untermenge des komplexen symmetrischen Gebiets G parametrisiert, und es genügt zu zeigen, daß diese Untermenge nur sehr wenige CM-Punkte des hier gefragten Typs besitzt. Dazu geht man so vor: Wenn T in abelsche Varietäten vom Typ CM zerfällt, existieren außer $\mathbb{Q}(\zeta_N)$ noch weitere Endomorphismen. Mit deren Hilfe wird Δ zu einer Gruppe $\bar{\Delta}$ erweitert; mit wenigen Ausnahmen ist dies eine unendliche Erweiterung von Δ, und dann kann man mit Hilfe der gebrochen-linearen Operation von $\bar{\Delta}$ auf $D_\omega(z)$ eine dichte Menge von Argumenten z' konstruieren, in denen die gleiche Rationalitätsbedingung für die $D_\omega(z')$ wie für die $D_\omega(z)$ gilt. Diese Rationalitätsbedingungen sagen nun im einzelnen (§ 6, Satz 5), daß die verschiedenen Koordinaten $D_\omega(z')$ durch algebraische Konjugationen ineinander übergeführt werden können, und eben dies erweist sich als Widerspruch zur analytischen Abhängigkeit der verschiedenen D_ω.

In § 10 wird schließlich die Voraussetzung des Wüstholzschen Satzes verifiziert, daß für $z \in \bar{\mathbb{Q}}$ die Zählerperiode $\neq 0$ ist, was außer für algebraische Funktionen $F(z)$ stets erfüllt ist, außerdem wird folgendes Problem behandelt: Auch wenn T einen Faktor A_1 isogen zu A_2 besitzt, ist noch nicht gesichert, daß Zählerperiode und Nennerperiode über $\bar{\mathbb{Q}}$ linear abhängig sind. Dies gilt genau dann, wenn auch noch die beteiligten Differentiale zueinander „passen". Es wird gezeigt, daß – wieder von Ausnahmen des Typs (C) abgesehen (Bem. 5) – diese

Bedingung an die Differentiale genau dann erfüllt ist, wenn Zähler- und Nennerperiode von erster Art sind und dem Zähler eine Dreiecksfunktion mit hyperbolischem $D_{\omega}(\mathfrak{H})$ zugeordnet ist. Expliziter läßt sich diese Bedingung in Form jener Ungleichungen für die Parameter a, b, c formulieren, die im Theorem unter (B) genannt sind.

Herrn Matthias Flach verdanke ich den Hinweis, daß die Eigenraumzerlegung in §4 ebenso wie die Dimensionsformel (21) auch aus einer Arbeit von Chevalley und Weil [CW] folgt. Er hat mich außerdem auf einen Fehler in einer früheren Version von Satz 4 aufmerksam gemacht. Frits Beukers verdanke ich die endgültige Formulierung der Parameterbedingungen unter (B) sowie ausführliche und sehr fruchtbare Diskussionen über mögliche Verallgemeinerungen des hier bewiesenen Theorems. Daniel Bertrand hat mich auf die Möglichkeit hingewiesen, Satz 6 auch mit Hilfe eines Kriteriums von Dwork und Katz [Ka] zu beweisen; diesen Beweis habe ich in [Wo] näher ausgeführt[3]. Gisbert Wüstholz und der Referent dieser Arbeit haben mich auf ein Problem bei der Definition der Jacobivarietäten von $X(N, z)$ aufmerksam gemacht. Henri Cohen verdanke ich die Kenntnis, daß einige spezielle $F \circ j$ auch von A.O.L. Atkin als Modulformen identifiziert worden sind (vgl. Bem. 4).

§ 2. Hilfsmittel

1. Die Parameter a, b, c seien im folgenden stets rationale Zahlen, dabei $c \neq 0, -1, -2, \ldots$. Es gibt zwar Fälle, in denen die Definition (1) auch für $1-c \in \mathbb{N}$ sinnvoll ist (wenn nämlich $1-a$ bzw. $1-b \in \mathbb{N}$ und a bzw. $b > c$ ist), da diese Definition jedoch auf Polynome, d.h. auf Ausnahmefälle vom Typ (A) führt, sei sie ebenfalls ausgeschlossen. Die Definition (1) der hypergeometrischen Funktionen und ihre Differentialgleichung werden wird vor allem implizit in nützlichen Identitäten verwenden – z.B. daß man die Parameter a und b vertauschen darf –, explizit wird jedoch stets mit der Integraldarstellung der hypergeometrischen Funktionen gearbeitet:

$$F(a, b, c; z) = \frac{1}{B(b, c-b)} \int_0^1 x^{b-1}(1-x)^{c-b-1}(1-zx)^{-a}\,dx \tag{2}$$

ist bei der üblichen Integration zwischen 0 und 1 zwar nur für $c > b > 0$ definiert; sieht man von trivialen Fällen ab (wie z.B. exakten Differentialen), so kann man diese Einschränkung jedoch fallenlassen, wenn die Integration auf einem geschlossenen Weg durchgeführt und ein algebraischer Faktor $\neq 0$ angefügt wird [Kl §§ 18ff.]. Genauer gesagt repräsentiert dieser Integrationsweg einen „Pochhammer-Zykel" (den Kommutator zweier Umläufe um 0 bzw. 1, wenn b, c nicht beide $\in \mathbb{Z}$ sind), d.i. eine Homologieklasse auf der algebraischen Kurve

$$y^N = x^A(1-x)^B(1-zx)^C;$$

dabei ist N kleinster natürlicher Hauptnenner von a, b, c und

$$A := (1-b)N, \quad B := (b+1-c)N, \quad C := aN, \tag{3}$$

[3] Nachtrag (Okt. 87): Das Dwork-Katz-Kriterium wurde erstmals von Landau und Stridsberg 1911 formuliert. Der Nachweis, daß es notwendig und hinreichend für die Algebraizität von $F(z)$ ist, wurde damals allerdings unter wesentlicher Verwendung der Ergebnisse von H.A. Schwarz [Schw] geführt, vgl. A. Errera: Zahlentheoretische Lösung einer funktionentheoretischen Frage, Rend. Circ. Mat. Palermo 35, 107–144 (1913)

so daß also in (2) über das Differential

$$\eta_1 = \frac{dx}{y}$$

integriert wird. Wenn $z \in \bar{\mathbb{Q}}$ fest gewählt ist, was wir meistens voraussetzen werden, dann sind die Kurve ebenso wie das Differential η_1 über $\bar{\mathbb{Q}}$ definiert. – Die Voraussetzung $|z| < 1$ ist für die Integraldarstellung (2) ersetzbar durch

$$z \neq 0 \quad \text{und} \quad z \neq 1.$$

Die Kurve (genauer: ihre projektive Vervollständigung) kann man dann vermöge der Abbildung

$$(x, y) \mapsto x$$

als N-blättrige Überlagerung der Riemannschen Zahlenkugel bzw. der projektiven Geraden ansehen, verzweigt in den Punkten $x = 0, 1, \frac{1}{z}$ und ∞. Da in diesen Verzweigungspunkten i.allg. Singularitäten der Kurve liegen, und da wir nachher die Perioden als Perioden abelscher Varietäten zu interpretieren haben, ist die Kurve durch eine geeignete nichtsinguläre Kurve zu ersetzen:

Definition. $X(N, z)$ sei eine projektive, nichtsinguläre algebraische Kurve, deren Funktionenkörper isomorph zum Funktionenkörper der Kurve

$$y^N = x^A (1-x)^B (1-zx)^C \tag{4}$$

ist. Für $z \in \bar{\mathbb{Q}}$ sei $X(N, z)$ über $\bar{\mathbb{Q}}$ definiert.

Da die ursprüngliche Kurve nur Spitzensingularitäten besitzt, ist ihre Desingularisierung (die für $z \in \bar{\mathbb{Q}}$ natürlich auch über $\bar{\mathbb{Q}}$ definiert werden kann) zu ihr biholomorph äquivalent; man überzeugt sich leicht, daß für alle in der Folge erforderlichen Untersuchungen (Differentiale, Homologie, Automorphismen) $X(N, z)$ durch (4) ersetzt werden kann.

2. Außer dem genannten Differential $\eta_1 = \frac{dx}{y}$ werden wir allgemeiner Differentiale

$$\eta = \eta(n, s, t, u) = \frac{x^s (1-x)^t (1-zx)^u \, dx}{y^n} \tag{5}$$

auf $X(N, z)$ zu betrachten haben ($n, s, t, u \in \mathbb{Z}$). Aus ihnen lassen sich alle Differentiale linear kombinieren, sie sind holomorph $\neq 0$ außerhalb der genannten Verzweigungspunkte, und in diesen haben sie die Ordnung

$$\begin{aligned}
\mathrm{Ord}_0\, \eta &= N(s+1) - nA - 1 \\
\mathrm{Ord}_1\, \eta &= N(t+1) - nB - 1 \\
\mathrm{Ord}_{1/z}\, \eta &= N(u+1) - nC - 1 \\
\mathrm{Ord}_\infty\, \eta &= n(A+B+C) - N(1+s+t+u) - 1,
\end{aligned} \tag{6}$$

wie man mit Hilfe der lokalen Variablen $x^{1/N}$, $(1-x)^{1/N}$, $(1-zx)^{1/N}$ bzw. $(1/x)^{1/N}$ ausrechnet. Ebenso überlegt man sich, daß alle Residuen von η verschwinden, d.h. daß η von 2. Art ist, wenn N kein Teiler von nA, nB, nC und $n(A+B+C)$ ist.

Die Perioden der Differentiale aus (5) lassen sich analog durch hypergeometrische Funktionen ausdrücken, bis auf Faktoren aus $\bar{\mathbb{Q}}^*$ nämlich durch

$$F(na-u, nb-n+1+s, nc-2n+2+t+s; z)\cdot B(nb, nc-nb), \tag{7}$$

jedenfalls soweit dieser Ausdruck definiert ist. An dieser Stelle sei auch daran erinnert, daß für je drei hypergeometrische Funktionen F_1, F_2, F_3, deren Parameter sich nur um ganzrationale Zahlen unterscheiden, eine „relatio inter functiones contiguas" (Gauß)

$$0=R_1(z)\,F_1(z)+R_2(z)\,F_2(z)+R_3(z)\,F_3(z)$$

besteht mit teilerfremden Koeffizientenpolynomen $R_i(z)\in\bar{\mathbb{Q}}[z]$ [Kl]. In unserem Beweis des Theorems drückt sich das darin aus, daß die in § 10 eingeführten Eigenräume W_n der de-Rham-Cohomologie von $X(N,z)$ zweidimensional sind. Man beachte, daß sich die Werte der Betafunktionen, die in den Integraldarstellungen der F_i auftreten, höchstens um rationale Faktoren unterscheiden.

3. Für die hier auftretenden Werte der Betafunktion $B(\alpha,\beta)$ an rationalen Stellen α,β verwenden wir folgende bekannte Tatsachen (vgl. Rohrlich und Koblitz [KR], [Gr]):

$$B(\alpha,\beta)\begin{cases}\in\mathbb{Q} \text{ für } \alpha \text{ oder } \beta\in\mathbb{N}\\ =\gamma\pi \text{ mit } \gamma\in\bar{\mathbb{Q}}^* \text{ für } \alpha+\beta\in\mathbb{Z},\ \alpha,\beta\notin\mathbb{Z}\\ =\text{Periodenintegral } \oint\eta_2 \text{ zweiter Art in allen anderen Fällen.}\end{cases}$$

Genauer ist η_2 ein Differential mit Koeffizienten in $\bar{\mathbb{Q}}$ auf der Fermatkurve F_N, wo N der Hauptnenner von α und β ist, und wenn man o.B.d.A. α und $\beta\in]0,1[$ annimmt, so ist η_2 von erster Art genau dann, wenn $\alpha+\beta<1$. Bezeichnen wir mit $\langle\gamma\rangle$ den gebrochenen Anteil $\gamma-[\gamma]$ von γ, so definiert man ein Repräsentantensystem von $(\mathbb{Z}/N\mathbb{Z})^*/\{\pm1\}$ durch

$$T_{\alpha,\beta}:=\{n\in(\mathbb{Z}/N\mathbb{Z})^* \mid \langle n\alpha\rangle+\langle n\beta\rangle<1\}. \tag{8}$$

Wenn σ_n die durch $\zeta_N:=\exp\left(\frac{2\pi i}{N}\right)\mapsto\zeta_N^n$ bestimmte Einbettung von $\mathbb{Q}(\zeta_N)$ in $\mathbb{C}$ bezeichnet, bildet

$$\{(\sigma_n(u)\,B(\langle n\alpha\rangle,\langle n\beta\rangle))_{n\in T_{\alpha,\beta}} \mid u\in\mathbb{Z}[\zeta_N]\}$$

das Periodengitter erster Art einer abelschen Varietät A_2^m der Dimension $\Phi(N)/2$ mit komplexer Multiplikation durch $\mathbb{Z}[\zeta_N]$ im Sinne von Shimura-Taniyama [ST]. Faßt man via der Identifikation $n\leftrightarrow\sigma_n$ die Menge $T_{\alpha,\beta}$ als Menge von Einbettungen $\mathbb{Q}(\zeta_N)\hookrightarrow\mathbb{C}$ auf, so ist $T_{\alpha,\beta}$ der „CM-Typ" von A_2^m, durch den A_2^m bis auf Isogenie eindeutig bestimmt ist. A_2^m zerfällt in m isogene einfache Faktoren A_2 mit komplexer Multiplikation durch einen CM-Körper $K\subset\mathbb{Q}(\zeta_N)$ mit m

$=[\mathbb{Q}(\zeta_N):K]$, den man als Fixkörper der Untergruppe

$$H_{\alpha,\beta}:=\{h\in(\mathbb{Z}/N\mathbb{Z})^* \mid h\,T_{\alpha,\beta}=T_{\langle h^{-1}\alpha\rangle,\langle h^{-1}\beta\rangle}=T_{\alpha,\beta}\}$$

von $(\mathbb{Z}/N\mathbb{Z})^*\cong \mathrm{Gal}(\mathbb{Q}(\zeta_N)/\mathbb{Q})$ gewinnt. Wenn $\mathcal{O}$ der Ring der ganzen Zahlen von K ist, so ist das Periodengitter erster Art von A_2 isogen zu

$$\{(\sigma_n(u)\,B(\langle n\alpha\rangle,\langle n\beta\rangle))_{n\in T_{\alpha,\beta}/H_{\alpha,\beta}} \mid u\in\mathcal{O}\}.$$

Das duale Gitter der Perioden 2. Art modulo 1. Art gewinnt man daraus, indem man jeweils n durch $-n$ ersetzt. Nach [WW] ist A_2 und damit auch A_2^m durch jede ihrer Perioden $B(\alpha,\beta)$ mod $\mathbb{Q}^*$ bis auf Isogenie eindeutig bestimmt.

4. Schließlich werden noch einige klassische Tatsachen über die Schwarzschen Dreiecksfunktionen verwendet (vgl. hierzu [Kl], [Schw] oder die Übersicht in [Wo], §4): Das Differential aus (5) sei dazu von erster Art und o.B.d.A. in der Form η_1 gegeben, was man notfalls durch Abänderung von N, A, B, C erreicht, d.h. es sei

$$\omega:=\eta_1=\frac{dx}{y}=x^{b-1}(1-x)^{c-b-1}(1-zx)^{-a}\,dx$$

holomorph auf $X(N,z)$, was wir nach (6) auch als

$$\begin{gathered} A,B,C<N\\ A+B+C>N \end{gathered} \tag{9}$$

formulieren können. Wie in (2) lesen wir die beiden Integrale

$$\int_0^1 \omega = B(b,c-b)\,F(a,b,c;z) \tag{10}$$

und

$$\begin{aligned}\int_{1/z}^{\infty}\omega&=(-1)^{1+a+b-c}z^{1-c}\int_0^1 x^{a-c}(1-x)^{-a}(1-zx)^{c-b-1}\,dx\\ &=(-1)^{1-\nu}z^{\lambda}B(1+a-c,1-a)\,F(1+b-c,1+a-c,2-c;z)\\ &=(-1)^{1-\nu}z^{\lambda}B(1+a-c,1-a)\,F(1+a-c,1+b-c,2-c;z)\end{aligned} \tag{11}$$

als Periodenintegrale (obwohl sie wegen der Homomorphie von ω hier auch in gewöhnlichem Sinn existieren), wobei wir

$$\begin{aligned} \lambda&:=1-c &&=\frac{A+B}{N}-1\\ \mu&:=b-a &&=1-\frac{A+C}{N}\\ \nu&:=c-b-a&&=1-\frac{B+C}{N}\end{aligned} \tag{12}$$

definieren. Beide Integrale (10) und (11) sind wieder Perioden 1. Art von $X(N, z)$, und als Funktionen von z sind sie bekanntlich linear unabhängige Lösungen der gleichen hypergeometrischen Differentialgleichung, wenn nur $\lambda \neq 0$ ist [Kl]. Aus (9) folgt leicht, daß außerdem

$$|\lambda|, |\mu|, |\nu| < 1 \tag{13}$$

sind, und daß darum die Schwarzsche Dreiecksfunktion

$$D_\omega(z) = D(\lambda, \mu, \nu; z) := \int_{1/z}^{\infty} \omega \Big/ \int_0^1 \omega$$

eine biholomorphe Abbildung der oberen Halbebene $\mathfrak{H}$ auf ein Kreisbogendreieck vermittelt (vgl. [Kl] oder [Schw]). Dieses Kreisbogendreieck $D_\omega(\mathfrak{H}) \subset \bar{\mathbb{C}}$ hat die Eckpunkte $D_\omega(0)$, $D_\omega(\infty)$ und $D_\omega(1)$ und in diesen die Winkel $|\lambda|$, $|\mu|$, bzw. $|\nu|$. Die Lage dieses Dreiecks in $\mathbb{C}$ bestimmt sich eindeutig daraus, daß die reellen Intervalle $]0, 1[$ und $]-\infty, 0[$ auf Geradenstücke in $\mathbb{C}$ abgebildet werden, und daß für positive λ

$$\begin{aligned} D_\omega(0) &= 0 \\ D_\omega(1) &= (-1)^{1-\nu} \frac{B(1+a-c, c-a-b)}{B(b, c-a-b)} = (-1)^{1-\nu} \frac{\Gamma(1+a-c)\,\Gamma(c-a)}{\Gamma(1-b)\,\Gamma(b)} \\ &= (-1)^{1-\nu} \frac{\sin b\pi}{\sin(c-a)\pi} \\ D_\omega(\infty) &= (-1)^{\lambda+1-\nu} \frac{\Gamma(1+a-c)\,\Gamma(c-a)}{\Gamma(1+b-c)\,\Gamma(c-b)} = (-1)^{\lambda+1-\nu} \frac{\sin(c-b)\pi}{\sin(c-a)\pi} \end{aligned} \tag{15}$$

gilt; zum Beweis verwendet man die Gaußsche Formel für die Werte der hypergeometrischen Reihen im Punkt $z=1$, die Paritätsrelation für die Werte der Gammafunktion, (10), (11) und die Kummersche Identität

$$F(a, b, c; z) = F\left(a, c-b, c; \frac{z}{z-1}\right)(1-z)^{-a}.$$

Ähnlich rechnet man aus, wie sich die Dreiecksfunktion bei Vorzeichenänderung der Parameter λ, μ, ν verhält:

$$D(\lambda, \mu, \nu; z) = D(\lambda, -\mu, \nu; z) \frac{\sin(c-b)\pi}{\sin(c-a)\pi} \frac{\sin b\pi}{\sin a\pi}; \tag{16}$$

aus der Kummerschen Identität

$$F(a, b, c; z) = (1-z)^{c-a-b} F(c-a, c-b, c; z)$$

folgt

$$D(\lambda, \mu, \nu; z) = (-1)^{-2\nu} D(\lambda, -\mu, -\nu; z); \tag{17}$$

und schließlich erhält man aus (11) und (14)

$$D(\lambda, \mu, \nu; z) = (-1)^{-2\nu} D(-\lambda, -\mu, \nu; z)^{-1}. \tag{18}$$

Wenn $|\lambda|+|\mu|+|\nu|<1$ gilt, handelt es sich bei dem Kreisbogendreieck $D_\omega(\mathfrak{H})$ um ein hyperbolisches Dreieck. Die zugehörige hyperbolische Ebene ist eine Kreisscheibe $K_\omega \subset \mathbb{C}$, welche für $\lambda>0$ den Mittelpunkt 0 und einen Radius $R_\omega \in \mathbb{Q}(\zeta_{2N})$ besitzt, den man aus (15) nach bekannten Formeln der hyperbolischen Trigonometrie berechnen kann.

§ 3. Sonderfälle: Ganzzahlige Parameter und Parameterdifferenzen

Fall 1. Wenn alle drei Parameter $a, b, c \in \mathbb{Z}$ sind, ist $F(a, b, c; z)$ bekanntlich eine rationale Funktion von z und $\log(1-z)$ mit Koeffizienten in $\mathbb{Q}$, so daß man sich entweder im Ausnahmefall (A) befindet oder für $z \in \overline{\mathbb{Q}}$ transzendente Funktionswerte mit höchstens endlich vielen Ausnahmen erhält. In den anderen beiden Fällen seien darum a, b, c nicht alle $\in \mathbb{Z}$.

Fall 2. Seien $a, b, c-a$ oder $c-b \in \mathbb{Z}$. Dann dürfen wir wieder mit Hilfe der Kummerschen Identitäten und eventueller Vertauschung von a und b voraussetzen, daß $c-a \notin \mathbb{Z}$ und $b \in \mathbb{Z}$, aber o.E. $\neq 0$, und daß a oder $c-b \notin \mathbb{Z}$ sind. Dann ist $\frac{dx}{y}$ entweder exakt, was zu einem Ausnahmefall vom Typ (A) führt, oder ein $\overline{\mathbb{Q}}$-rationales Differential auf $X(N, z)$, der zugehörige Wert der Betafunktion ist rational, und nach [Wü] sind nichtverschwindende Perioden solcher Differentiale stets transzendent.

Fall 3. Sei $c \in \mathbb{Z}$, nicht aber $a, b, c-a, c-b$. Hier ist der Wert von (2) Quotient einer Periode zweiter Art durch π, nach [WW, Satz 2] also transzendent oder 0.

§ 4. Die Differentiale erster Art

Alle in § 3 diskutierten Fälle seien von nun an ausgeschlossen, d.h. wir wollen im folgenden voraussetzen, daß

$$a, b, c, a-c, b-c \notin \mathbb{Z}$$

sind, was man nach (3) formulieren kann als

$$N \nmid A,\ B,\ C,\ A+B,\ A+B+C \quad \text{und} \quad \lambda \neq 0. \tag{19}$$

Unter diesen Voraussetzungen soll so explizit wie möglich eine Basis der Differentiale erster Art auf $X(N, z)$ konstruiert werden.

Auf $X(N, z)$ operiert die Abbildung

$$\kappa:\ x \mapsto x, \qquad y \mapsto \zeta_N^{-1}\, y$$

als Automorphismus der Ordnung N, und dieser induziert auf dem Vektorraum $H^0(X(N, z), \Omega)$ der Differentiale erster Art einen Automorphismus κ^*, dessen Eigenräume V_n, $n \in \mathbb{Z}/N\mathbb{Z}$, von Differentialen der Gestalt (5) erzeugt werden. Deren Exponenten erfüllen wegen (6) die folgenden vier Ungleichungen (n sei nun ein fest gewählter Repräsentant seiner Restklasse mod N):

$$\begin{aligned} s &\geqq N^{-1}(nA+1)-1 \\ t &\geqq N^{-1}(nB+1)-1 \\ u &\geqq N^{-1}(nC+1)-1 \\ s+t+u &\leqq N^{-1}(n(A+B+C)-1)-1. \end{aligned} \tag{20}$$

Man sieht leicht, daß es zu jedem n höchstens vier Lösungen $(s,\ t,\ u)\in\mathbb{Z}^3$ von (20) gibt, und zwar erstens die Minimallösung

$$s_0=-\left[1-\frac{nA+1}{N}\right] \quad \text{etc.}$$

(soweit diese mit der vierten Ungleichung verträglich ist) sowie eventuell (wieder mit der gleichen Einschränkung) drei weitere Lösungen, bei denen man jeweils s_0, t_0 oder u_0 um 1 vergrößert; bis auf Linearkombinationen mit dem zu (s_0, t_0, u_0) gehörigen Differential erhält man dabei allerdings nur ein neues Differential. Die Dimension des Eigenraums V_n ist demnach die Anzahl der ganzen Zahlen $s+t+u$, die zwischen den durch (20) abgesteckten Grenzen liegen, also

$$r_n=\left[n\frac{A+B+C}{N}-\frac{1}{N}\right]+\left[1-\frac{nA+1}{N}\right]+\left[1-\frac{nB+1}{N}\right]+\left[1-\frac{nC+1}{N}\right].$$

Im Fall $n\in(\mathbb{Z}/N\mathbb{Z})^*$ läßt sich wegen (19) diese Dimension folgendermaßen durch gebrochene Anteile ausdrücken (vgl. die Konvention in § 2, Abschn. 3):

$$r_n:=\dim V_n=\left\langle n\frac{A}{N}\right\rangle+\left\langle n\frac{B}{N}\right\rangle+\left\langle n\frac{C}{N}\right\rangle-\left\langle n\frac{A+B+C}{N}\right\rangle. \tag{21}$$

§ 5. Die Jacobivarietät und ihre Endomorphismen

Fassen wir die Jacobivarietät $\operatorname{Jac} X(N,z)$ als komplexen Torus auf, so ist zu ihrer genaueren Kennzeichnung ihr Periodengitter erster Art zu berechnen. Die in § 4 gefundenen Basisdifferentiale sind über die Homologie $H_1(X(N,z),\mathbb{Z})$ zu integrieren. Nach [Kl, § 18] kann man eine Basis dieser Homologie aus Wegen auswählen, deren Bilder unter der natürlichen Projektion

$$X(N,z)\to\mathbb{C}:\quad (x,y)\mapsto x$$

aus Strecken zusammengesetzt wird, die zwischen den Verzweigungspunkten 0, 1, $1/z$, ∞ dieser Projektion verlaufen. Da wir es hier mit Differentialen erster Art zu tun haben, ist die Integration über diese Strecken unbedenklich. Aus Kleins Überlegungen folgt genauer, daß sich alle Perioden eines solchen Differentials ω bis auf einen gemeinsamen Faktor in $\mathbb{Q}(\zeta_N)$ als $\mathbb{Z}[\zeta_N]$-Linearkombinationen von

$$\int_0^1\omega \quad \text{und} \quad \int_{1/z}^{\infty}\omega$$

schreiben lassen; streng genommen wird dabei nicht über ω, sondern über einen geeignet gewählten Pullback von ω unter der Projektion $X(N,z)\to\mathbb{C}$ integriert. Die Auswahl dieses Pullbacks entspricht dabei der Auswahl bestimmter Zweige von $x^{1/N}$, $(1-x)^{1/N}$ und $(1-zx)^{1/N}$. Wenn $\omega\in V_n$ mit $n\in(\mathbb{Z}/N\mathbb{Z})^*$, so folgt aus der Existenz der schon in §4 benutzten Automorphismen von $X(N,z)$, daß die Koeffizienten von $\int_0^1 \omega$ und $\int_{1/z}^{\infty} \omega$ in diesen Linearkombinationen ein Gitter vom Maximalrang in $\mathbb{Q}(\zeta_N)$ durchlaufen.

Die übrigen Basisdifferentiale $\omega\in V_n$ mit $(n,N)=\frac{N}{D}\neq 1$ sind für uns weniger interessant, weil ihre Perioden bereits als Perioden der Kurve $X(D,z)$ auftreten (Desingularisierung von

$$y^D = x^A(1-x)^B(1-zx)^C,$$

A, B, C wie in $X(N,z)$ gewählt), m.a.W. weil eine natürliche Überlagerungsabbildung

$$X(N,z)\to X(D,z)$$

für alle Teiler D von N existiert und weil darum $\operatorname{Jac} X(D,z)$ in $\operatorname{Jac} X(N,z)$ enthalten ist. Eine andere Möglichkeit besteht darin, $\operatorname{Jac} X(D,z)$ als jene abelsche Untervarietät zu definieren, die von κ^D elementweise festgelassen wird, wenn man den in §4 eingeführten Automorphismus κ von $X(N,z)$ auf $\operatorname{Jac} X(N,z)$ ausdehnt. – Die $\omega\in V_n$ mit $n\in(\mathbb{Z}/N\mathbb{Z})^*$ erzeugen also genau die Perioden des Komplements aller $\operatorname{Jac} X(D,z)$ in $\operatorname{Jac} X(N,z)$, wobei D alle echten Teiler von N durchläuft. Wählt man die erwähnten Zweige von $x^{1/N}$ etc. unabhängig von n und ω aus und berücksichtigt man das Verhalten unter den Automorphismen von $X(N,z)$, so ergibt sich für die Jacobivarietät die im folgenden Satz 1 beschriebene einfache Gestalt. Dabei sei S ein maximales System linear unabhängiger Differentiale erster Art auf $X(N,z)$ von der Form (5)

$$\omega=\frac{x^s(1-x)^t(1-zx)^u\,dx}{y^n}$$

mit $(n,N)=1$, also eine Basis von $\sum_{(n,N)=1}\oplus V_n$; zu jedem $\omega\in V_n$ ist eine Einbettung $\sigma_\omega:=\sigma_n\colon \mathbb{Q}(\zeta_N)\hookrightarrow\mathbb{C}$ durch $\zeta_N\mapsto\zeta_N^n$ eindeutig bestimmt.

Satz 1. *Die Jacobivarietät* $\operatorname{Jac} X(N,z)$ *ist isogen zu einer Zerlegung*

$$\operatorname{Jac} X(N,z)\triangleq T\oplus\sum_{\substack{D\mid N\\ D\neq N}}\operatorname{Jac} X(N,z),$$

so daß der wesentliche Anteil T eine (für $z\in\bar{\mathbb{Q}}$ über $\bar{\mathbb{Q}}$ definierte) abelsche Varietät der Dimension $\Phi(N)$ ist, isogen zu dem Torus $\mathbb{C}^{\Phi(N)}/\Lambda$ mit dem Periodengitter

$$\Lambda:=\left\{\left(\sigma_\omega(u)\int_0^1\omega+\sigma_\omega(v)\int_{1/z}^{\infty}\omega\right)_{\omega\in S}\,\middle|\, u,v\in\mathbb{Z}[\zeta_N]\right\}.$$

Die Multiplizität r_σ, mit der eine vorgegebene Einbettung $\sigma \in \mathrm{Gal}\,\mathbb{Q}(\zeta_N)/\mathbb{Q}$ unter den σ_ω vorkommt, errechnet sich aus (21): Wenn nämlich $\sigma(\zeta_N)=\zeta_N^n$, so ist

$$r_\sigma := \#\{\omega \in S \mid \sigma=\sigma_\omega\} = r_n = \left\langle n\frac{A}{N}\right\rangle + \left\langle n\frac{B}{N}\right\rangle + \left\langle n\frac{C}{N}\right\rangle - \left\langle n\frac{A+B+C}{N}\right\rangle.$$

Bezeichnet $\bar\sigma$ die zu σ konjugiert komplexe Einbettung, so ist offenbar $r_{\bar\sigma}=r_{-n}$, und es gilt für alle $\sigma \in \mathrm{Gal}\,\mathbb{Q}(\zeta_N)/\mathbb{Q}$ bzw. $n \in (\mathbb{Z}/N\mathbb{Z})^*$

$$r_\sigma + r_{\bar\sigma} = r_n + r_{-n} = 2. \tag{22}$$

Offensichtlich folgt aus Satz 1 sofort

Satz 2. *Die Endomorphismenalgebra* $\mathrm{End}_0\, T = \mathbb{Q} \otimes \mathrm{End}\, T$ *enthält* $\mathbb{Q}(\zeta_N)$, *und zwar operiert jedes* $\alpha \in \mathbb{Z}[\zeta_N]$ *auf* T *vermöge*

$$\alpha: \begin{cases} \mathbb{C}^{\Phi(N)}/\Lambda \to \mathbb{C}^{\Phi(N)}/\Lambda \\ (z_\omega)_{\omega \in S} \bmod \Lambda \mapsto (\sigma_\omega(\alpha)\cdot z_\omega)_{\omega \in S} \bmod \Lambda \end{cases}$$

Für manche Zwecke ist es praktisch, Λ durch ein (über $\mathbb{C}$, nicht über $\mathbb{Q}$) isomorphes Gitter

$$\Lambda^* := \{(\sigma_\omega(u) + \sigma_\omega(v)\, D_\omega(z))_{\omega \in S} \mid u, v \in \mathbb{Z}[\zeta_N]\} \tag{23}$$

zu ersetzen, wo D_ω die nach (14) zu ω gehörige Dreiecksfunktion $\int_{1/z}^{\infty} \omega \Big/ \int_0^1 \omega$ bezeichnet.

Daß abelsche Varietäten dieses Endomorphietyps $(\mathbb{Q}(\zeta_N), (\sigma_\omega)_{\omega \in S})$ existieren, weiß man seit den Arbeiten von Albert [A]. Shimura [Shi 1] und Siegel [Sie 2] haben unabhängig voneinander bewiesen, daß die Gesamtheit aller abelschen Varietäten dieses Endomorphietyps durch ein symmetrisches komplexes Gebiet der Dimension

$$r = \sum_{n \in (\mathbb{Z}/N\mathbb{Z})^*/\{\pm 1\}} r_n r_{-n} = \sum r_\sigma r_{\bar\sigma} \tag{24}$$

parametrisiert wird; in der zweiten Summe wird über alle Einbettungen $\mathbb{Q}(\zeta_N) \hookrightarrow \mathbb{C}$ modulo komplexer Konjugation summiert. Bei der Beschreibung der hier auftretenden komplexen Gebiete spielen die $\omega \in S$, für welche $D_\omega(\mathfrak{H})$ ein hyperbolisches Dreieck ist, eine besondere Rolle. Es gilt

Satz 3. *Sei* $n \in (\mathbb{Z}/N\mathbb{Z})^*$ *und* $\omega \in S \cap V_n$. *Dann ist*

a) $D_\omega(\mathfrak{H})$ *sphärisches Dreieck genau dann, wenn* $r_n = 2$,

b) $D_\omega(\mathfrak{H})$ *hyperbolisches Dreieck genau dann, wenn* $r_n = 1$.

In diesem Fall gilt außerdem: Wenn $D_\omega(z) = D(\lambda, \mu, \nu; z)$, *so gilt für das (ebenfalls eindeutig bestimmte)*[4] $\bar\omega \in S \cap V_{-n}$

$$D_{\bar\omega}(z) = D(-\lambda, -\mu, -\nu; z).$$

[4] Achtung: $\bar\omega$ ist nicht komplex konjugiert zu ω, wohl aber σ_ω zu $\sigma_{\bar\omega}$

Zum Beweis nehmen wir wieder o.B.d.A. an, daß $n=1$ und $\omega=\frac{dx}{y}$ ist, was man durch Abänderung von A, B, C immer erreichen kann; es gelten die Ungleichungen (9) und wir dürfen A, B, C als positiv annehmen. Wenn $r_n=2$ ist, muß ein weiteres Differential $\omega' \in S$ in V_1 liegen, nach § 4 dürfen wir z.B.

$$\omega'=\frac{(1-zx)\,dx}{y}$$

wählen, d.h. in (20) ist $s=t=0$ und $u=1$ zu setzen. Nach der letzten Ungleichung (20) folgt $2N<A+B+C$. Aus (9) und (19) erhalten wir damit die Ungleichungen

$$A+B,\ A+C,\ B+C>N,$$

d.h.

$$\lambda>0, \quad \mu<0, \quad \nu<0$$

(vgl. (12)). Die (durch π dividierte) Winkelsumme von $\mathrm{D}_\omega(\mathfrak{H})$ ist somit

$$\lambda-\mu-\nu=2\frac{A+B+C}{N}-3>1,$$

d.h. $\mathrm{D}_\omega(\mathfrak{H})$ ist sphärisch. Die Rechnung für $D_{\omega'}$ verläuft analog, man hat lediglich C durch $C-N$ zu ersetzen, μ durch $\mu+1>0$, ν durch $\nu+1>0$, so daß also $D_{\omega'}(z)=D(\lambda,\mu+1,\nu+1;z)$. Die Winkelsumme ist hier ebenfalls wieder

$$\lambda+\mu+\nu+2=3-2\frac{C}{N}>1.$$

Den Fall $r_n=1$ behandelt man analog mit dem Unterschied, daß hier kein zweites Differential $\omega' \in S \cap V_1$ existieren kann und darum

$$N<A+B+C<2N$$

sein muß. Eine Fallunterscheidung nach den Vorzeichen der λ, μ, ν führt direkt zum Ziel: Zum Beispiel

$$\lambda<0, \quad \mu\geqq 0, \quad \nu\geqq 0$$

$$\Rightarrow \quad -\lambda+\mu+\nu=3-2\frac{A+B+C}{N}<1.$$

Das zu $\omega=\frac{dx}{y}$ gehörige Differential $\bar{\omega}\in S\cap V_{-1}$ hat nach (20) notwendig die Gestalt

$$\bar{\omega}=\frac{y\,dx}{x(1-x)(1-zx)},$$

und die Berechnung von $D_{\bar{\omega}}$ führt man am einfachsten so durch, daß man in Gl. (4) der Definition von $X(N,z)$ die Exponenten A, B, C durch $N-A$, $N-B$ bzw. $N-C$ ersetzt.

Nun zurück zur Beschreibung der komplexen symmetrischen Gebiete, die wir – unserem Zweck entsprechend – etwas anders normieren als in [Shi 1] bzw. [Sie 2]. Sei

$$R := \{\omega \in S \mid D_\omega(\mathfrak{H}) = D(\lambda, \mu, \nu; \mathfrak{H}) \text{ hyperbolisch mit } \lambda > 0\}, \tag{25}$$

also das maximale System von Basisdifferentialen, die zu eindimensionalen Eigenräumen V_n gehören (allein diese tragen zur Summe (24) einen Summanden bei) und für welche das Dreieck $D_\omega(\mathfrak{H})$ in der Kreisscheibe K_ω mit Mittelpunkt 0 liegt. Die Differentiale $\bar{\omega}$ aus den dazu komplementären Eigenräumen V_{-n} führen auf Dreiecksfunktionen

$$D_{\bar{\omega}}(z) = \frac{c_\omega}{D_\omega(z)} \tag{26}$$

mit einem Faktor $c_\omega \in \mathbb{Q}(\zeta_N)$, den man nach Satz 3 und (16) bis (18) berechnen kann. Die fraglichen symmetrischen Gebiete haben nun die Form

$$G = \prod_{\omega \in R} K_\omega .$$

Wenn R leer ist, was durchaus vorkommen kann – s. § 7 – bestehe G aus einem Punkt. Jedem Punkt $\tau = (\tau_\omega)_{\omega \in R} \in G$ entspricht ein Periodengitter Λ_τ nach folgender Vorschrift: Man ergänzt die Definition der τ_ω zunächst auf ganz S, indem man für die eindimensionalen Eigenräume V_n und V_{-n}, soweit $\tau_\omega \neq 0$ ist,

$$\tau_\omega \tau_{\bar{\omega}} = c_\omega \tag{27}$$

vorschreibt und für $\tau_\omega = 0$ in der folgenden Definition (28) die $\bar{\omega}$-Koordinate durch $\sigma_{\bar{\omega}}(v)$ ersetzt; für die zweidimensionalen V_n mit $V_n \cap S = \{\omega, \omega'\}$ wählt man beliebige komplexe Zahlen $\tau_\omega \neq \tau_{\omega'}$. Dann setzt man

$$\Lambda_\tau := \{(\sigma_\omega(u) + \sigma_\omega(v)\,\tau_\omega)_{\omega \in S} \mid u, v \in \mathbb{Z}[\zeta_N]\}, \tag{28}$$

so daß unsere in (23) gefundenen Gitter gerade den Punkten

$$\tau = (D_\omega(z))_{\omega \in R}$$

entsprechen.

§ 6. Die Nullteiler des Endomorphismenrings

Für die Zerlegung von T in einfache abelsche Varietäten gibt es zwei Alternativen, je nachdem wieviele Endomorphismen von T mit der in Satz 2 gefundenen Endomorphismenalgebra $\mathbb{Q}(\zeta_N)$ kommutieren. Die erste Alternative wird diskutiert im folgenden

Satz 4. *T sei wie in Satz 1 definiert. Wenn* $\mathrm{Comm}(\mathbb{Q}(\zeta_N), \mathrm{End}_0\, T)$ *keine Nullteiler besitzt, ist T isogen zu einer direkten Summe m einfacher isogener abelscher Varietäten $A_1 \oplus A_1 \oplus \ldots \oplus A_1$ mit* $\dim A_1 = \frac{1}{m}\,\Phi(N)$ *und Endomorphismenalgebra* $D = \mathrm{End}_0\, A_1$. *Wenn $D \subset \mathbb{Q}(\zeta_N)$, so ist*

$$[D : \mathbb{Q}] = \dim A_1,$$

d.h. A_1 ist nicht vom Typ CM im Sinne von Shimura-Taniyama.

Weil $[\mathbb{Q}(\zeta_N):\mathbb{Q}] = \dim T$ ist und $\mathrm{Comm}(\mathbb{Q}(\zeta_N), \mathrm{End}_0 T)$ nullteilerfrei, muß nach [Be] (vgl. § 3, Cor. 1 und den Beweis von Lèmme 1) $\mathrm{Comm}(\mathbb{Q}(\zeta_N), \mathrm{End}_0 T)$ ein Körper sein, und zwar $\mathbb{Q}(\zeta_N)$ oder eine quadratische Erweiterung F von $\mathbb{Q}(\zeta_N)$; in beiden Fällen ist T isogen zu einer direkten Summe von m einfachen abelschen Varietäten A_1 mit Endomorphismenalgebra $D = \mathrm{End}_0 A_1$, also $\mathrm{End}_0 T \cong M_m(D)$ (vgl. [Be], § 1, Exemple 3). Wenn außerdem $D \subset \mathbb{Q}(\zeta_N)$, so ist D das Zentrum von $\mathrm{End}_0 T$, und nach einem bekannten Satz der Algebrentheorie [Ja, Ch. 5, Thm. 19] ist dann

$$m^2 = [\mathrm{End}_0 T : D] = [\mathrm{Comm}(\mathbb{Q}(\zeta_N), \mathrm{End}_0 T) : D] \cdot [\mathbb{Q}(\zeta_N) : D],$$

was im Fall $\mathrm{Comm}(\mathbb{Q}(\zeta_N), \mathrm{End}_0 T) = F$ mit $[F : \mathbb{Q}(\zeta_N)] = 2$ auf den Widerspruch

$$m^2 = 2[\mathbb{Q}(\zeta_N) : D]^2$$

führt und im Fall $\mathrm{Comm}(\mathbb{Q}(\zeta_N), \mathrm{End}_0 T) = \mathbb{Q}(\zeta_N)$ auf

$$m^2 = [\mathbb{Q}(\zeta_N) : D]^2 = \frac{\Phi(N)^2}{[D:\mathbb{Q}]^2},$$

also

$$[D:\mathbb{Q}] = \frac{\Phi(N)}{m} = \dim A_1.$$

Die zweite Alternative führt letztlich auf die algebraischen Ausnahmewerte der hypergeometrischen Funktionen vom Typ (A) oder (B): Wenn es Endomorphismen ε in $\mathrm{End}_0 T$ gibt, die Nullteiler sind und mit $\mathbb{Q}(\zeta_N)$ kommutieren, so muß das Bild von ε eine echte abelsche Untervarietät $U \subset T$ mit $\mathbb{Q}(\zeta_N) \subset \mathrm{End}_0 U$ sein. Wegen

$$\Phi(N) = [\mathbb{Q}(\zeta_N) : \mathbb{Q}] \mid 2 \dim U < 2 \dim T = 2\Phi(N)$$

[ST] muß $\Phi(N) = 2 \dim U$ sein, U ist also abelsche Varietät vom Typ CM; dazu muß eine weitere abelsche Untervarietät U' von T vom Typ CM mit $U \cap U' = \{0\}$ existieren, so daß T isogen zu $U \oplus U'$ ist. ε läßt sich als Projektion von T auf U wählen, so daß U' die Zusammenhangskomponente der 0 von $\mathrm{Kern}\,\varepsilon$ ist. Die rationale Darstellung von ε auf dem Periodengitter Λ^* aus (23), das wir in natürlicher Weise mit $\mathbb{Z}[\zeta_N] \oplus \mathbb{Z}[\zeta_N]$ identifizieren, muß Λ^* auf ein Gitter abbilden, das zugleich $\mathbb{Z}[\zeta_N]$-Modul vom Rang 1 ist, während die komplexe Darstellung von ε auf $\mathbb{C}^{\Phi(N)}$ eine Einbettung

$$\varepsilon^* : H^0(U, \Omega) \to H^0(T, \Omega)$$

induziert, die die $\mathbb{Q}(\zeta_N)$-Eigenräume respektiert. Wegen der bekannten Eigenraum-Zerlegung für abelsche Varietäten vom Typ CM (s. [ST]) enthält das Bild von ε^* von je zwei eindimensionalen Eigenräumen $V_n, V_{-n} \subset H^0(T, \Omega)$ genau einen, etwa V_{-n}. Sei also $r_n = r_{-n} = 1$ und $\{\omega\} = S \cap V_n$, $\{\bar{\omega}\} = S \cap V_{-n}$; dann heißt das zunächst, daß die ω-Koordinaten des Bildgitters $\varepsilon\Lambda^*$ nur aus Nullen bestehen, die $\bar{\omega}$-Koordinaten jedoch einen $\mathbb{Z}[\zeta_N]$-Modul $\cong \mathbb{Z}[\zeta_N]$ durchlaufen. Das ist offenbar genau dann möglich, wenn $D_\omega(z) \in \mathbb{Q}(\zeta_N)$ (nach (26) ist dann auch $D_{\bar{\omega}}(z) \in \mathbb{Q}(\zeta_N)$). Die rationale Darstellung von ε muß darum bis auf Faktoren aus $\mathbb{Q}(\zeta_N)^*$ von der Form

$$\mathbb{Z}[\zeta_N]\oplus\mathbb{Z}[\zeta_N]\to\mathbb{Q}\otimes(\mathbb{Z}[\zeta_N]\oplus\mathbb{Z}[\zeta_N])$$
$$(\sigma_\omega(u),\sigma_\omega(v))\mapsto(-D_\omega(z)\,\sigma_\omega(v),\sigma_\omega(v))$$

sein oder einfacher

$$(u,v)\mapsto(-\sigma_\omega^{-1}(D_\omega(z))\,v,v).$$

Für jedes weitere $V_{n'}$ mit $r_{n'}=1=r_{-n'}$ muß aber ebenfalls Bild ε^* entweder $V_{n'}$ oder $V_{-n'}$ enthalten, und nach der gleichen Überlegung schließt man aus (23), daß mit $\{\omega'\}=S\cap V_{n'}$ und $\{\bar{\omega}'\}=S\cap V_{-n'}$ dann jeweils

$$\sigma_\omega^{-1}(D_\omega(z))=\sigma_{\omega'}^{-1}(D_{\omega'}(z))\quad\text{oder}\quad\sigma_\omega^{-1}(D_\omega(z))=\sigma_{\bar{\omega}'}^{-1}(D_{\bar{\omega}'}(z))\tag{29}$$

erfüllt ist. Dem Leser sei der Beweis der Umkehrung überlassen: Wenn $D_\omega(z)\in\mathbb{Q}(\zeta_N)$ und (29) für alle eindimensionalen V_n erfüllt ist, existiert dazu ein $\varepsilon\in\mathrm{End}_0\,T$, das die angegebene rationale Darstellung besitzt. Man beachte: Wenn gar kein eindimensionaler Eigenraum existiert, tritt jede $\mathbb{Q}(\zeta_N)$-bilineare Abbildung von $\mathbb{Q}\otimes(\mathbb{Z}[\zeta_N]\oplus\mathbb{Z}[\zeta_N])$ in sich als rationale Darstellung eines $\varepsilon\in\mathrm{End}_0\,T$ auf!

Wenn mehr als zwei eindimensionale Eigenräume V_n von $H^0(T,\Omega)$ existieren, ist die Bedingung (29) nichtleer; welche der beiden dort genannten Möglichkeiten ist dann für unsere Zwecke relevant? Der in der hypergeometrischen Funktion (2) auftretende Wert $B(b,c-b)$ der Betafunktion ist bis auf algebraische Faktoren $B\left(\left\langle-\frac{A}{N}\right\rangle,\left\langle-\frac{B}{N}\right\rangle\right)$ und tritt als Periode eines Eigendifferentials η_0 einer einfachen abelschen Varietät A_2 vom Typ CM auf (vgl. §2.3). Eine zweite Periode $\oint\eta_0'\neq0$ auf einer anderen einfachen abelschen Varietät A über $\bar{\mathbb{Q}}$ mit $\eta_0'\in H_{DR}^1(A)$ erfüllt genau dann

$$\oint\eta_0/\oint\eta_0'\in\bar{\mathbb{Q}}^*,$$

wenn eine Isogenie $\varphi: A\to A_2$ so existiert, daß $\varphi^*(\eta_0)$ ein algebraisches Vielfaches von η_0' ist; dies tritt genau dann ein, wenn auch A komplexe Multiplikation besitzt, und zwar vom gleichen CM-Typ wie A_2 ([WW], Satz 2 und Prop. 2). Entsprechend ist folgendes richtig: $B(b,c-b)=\oint\eta_2$ ist Periode eines $\mathbb{Q}(\zeta_N)$-Eigendifferentials auf einer – bis auf Isogenie eindeutigen – abelschen Varietät A_2^m mit komplexer Multiplikation durch $\mathbb{Q}(\zeta_N)$. „Eindeutig" heißt dabei genauer, daß eine Periode $\oint\eta'\neq0$ von $\eta'\in H_{DR}^1(A')$ auf einer anderen abelschen Varietät A' mit komplexer Multiplikation durch $\mathbb{Q}(\zeta_N)$ genau dann

$$\oint\eta'/\oint\eta_2\in\bar{\mathbb{Q}}^*$$

erfüllt, wenn eine Isogenie $\varphi: A'\to A_2^m$ existiert, die bis auf algebraische Faktoren $\varphi^*(\eta_2)=\eta'$ befriedigt. Auch dies läßt sich wieder am CM-Typ ablesen: Wie η_2 muß auch η' ein $\mathbb{Q}(\zeta_N)$-Eigendifferential sein; normieren wir die Operation von $\mathbb{Q}(\zeta_N)$ auf A' und A_2^m so, daß jedes α auf η_2 und η' in gleicher Weise durch Multiplikation

$$\eta_2\mapsto\alpha\,\eta_2$$
$$\eta'\mapsto\alpha\,\eta'$$

operiert, so ist die Isogenie φ mit der Operation von $\mathbb{Q}(\zeta_N)$ vertauschbar, und A' besitzt genau den CM-Typ

$$T_{\langle b\rangle,\langle c-b\rangle}=\{n\in(\mathbb{Z}/N\mathbb{Z})^*\,|\,\langle nb\rangle+\langle n(c-b)\rangle<1\}$$

von A_2^m (vgl. (8)), d.h. $H^0(A',\Omega)$ besitzt eine Eigenbasis von Differentialen ω_n 1. Art, $n\in T_{\langle b\rangle,\langle c-b\rangle}$, auf denen $\mathbb{Q}(\zeta_N)$ vermöge

$$\alpha\colon\ \omega_n\mapsto\sigma_n(\alpha)\cdot\omega_n$$

operiert. – Auch für mehr als zwei Perioden dieser Art gilt nach [WW, Satz 2], daß $\bar{\mathbb{Q}}$-lineare Abhängigkeiten nur trivial auf Grund solcher Isogenien bzw. Übereinstimmung der so normierten CM-Typen möglich sind.

Denkt man sich nun unsere abelsche Varietät T als isogen zu einer direkten Summe einfacher abelscher Varietäten geschrieben, so wird

$$\oint\eta_1=\int_0^1\frac{dx}{y}$$

eine $\bar{\mathbb{Q}}$-Linearkombination von Perioden 1. bzw. 2. Art auf den einfachen Faktoren von T. Nach [WW, Satz 2] kann also höchstens dann

$$z\in\bar{\mathbb{Q}}\quad\text{und}\quad F(z)=\oint\eta_1/B(b,c-b)=\oint\eta_1/\oint\eta_2\in\bar{\mathbb{Q}}^*$$

sein, wenn A_2 isogen zu einem einfachen Faktor von T ist. Da A_2 aber komplexe Multiplikation durch einen Körper $D\subset\mathbb{Q}(\zeta_N)$ besitzt, muß nach Satz 4 $\mathrm{Comm}(\mathbb{Q}(\zeta_N),\mathrm{End}_0\,T)$ Nullteiler haben, T ist also isogen zur direkten Summe zweier abelscher Untervarietäten $U\oplus U'$ mit komplexer Multiplikation durch $\mathbb{Q}(\zeta_N)$, und $\oint\eta_1$ wird eine $\bar{\mathbb{Q}}$-Linearkombination von Perioden 1. bzw. 2. Art $\oint\eta$ und $\oint\eta'$ auf U und U'. Da die Zerlegung von T in U und U' mit der Operation von $\mathbb{Q}(\zeta_N)$ verträglich ist, sind η und η' ebenso wie η_1 Eigendifferentiale zu gleichen Eigenwerten, d.h. wir dürfen annehmen, daß jedes $\alpha\in\mathbb{Q}(\zeta_N)$ auf η, η', η_1 und η_2 in gleicher Weise operiert, nämlich durch Multiplikation

$$\eta\mapsto\alpha\cdot\eta$$

etc.; für η_1 stimmt diese Normierung der Operation mit der in Satz 2 verwendeten Normierung überein, wie man anhand der Definition der V_n in §4 bzw. (5) nachprüft. $z\in\bar{\mathbb{Q}}$ und $F(z)\in\bar{\mathbb{Q}}^*$ impliziert also $\oint\eta_1=\oint\eta+\oint\eta'$ mit Eigendifferentialen η und η' auf U und U', und wir dürfen o.B.d.A. voraussetzen, daß $\oint\eta'\neq 0$, nach unserer Vorüberlegung also U' isogen zu A_2^m ist. Diese Bedingung läßt sich, wie oben erläutert, am CM-Typ ablesen: Dieser muß $T_{\langle b\rangle,\langle c-b\rangle}$ sein; nach (8) ist aber

$$n\in T_{\langle b\rangle,\langle c-b\rangle}\Leftrightarrow\left\langle -n\frac{A}{N}\right\rangle+\left\langle -n\frac{B}{N}\right\rangle<1\Leftrightarrow\left\langle n\frac{A}{N}\right\rangle+\left\langle n\frac{B}{N}\right\rangle>1.$$

Für zweidimensionale V_n ist diese Bedingung nach (21) trivial erfüllt. Für eindimensionale V_n ist sie genau dazu äquivalent, daß für $\{\omega\}=V_n\cap S$ der zu D_ω gehörige Winkel $\lambda>0$ ist, daß also $\omega\in R$ ist (vgl. Satz 3 und (25)).

Im Beweis von (29) war ε als Projektion von T auf U so gewählt worden, daß ε^* gerade $H^0(U,\Omega)$ in $H^0(T,\Omega)$ einbettet, d.h. für eindimensionale V_n gilt $V_n \cap \text{Bild } \varepsilon^* = \{0\}$ genau dann, wenn V_n zu $H^0(U',\Omega)$ gehört, also wenn $n \in T_{\langle b \rangle, \langle c-b \rangle}$ ist bzw. $\omega \in R$. In (29) muß also

$$\sigma_\omega^{-1}(D_\omega(z)) = \sigma_{\omega'}^{-1}(D_{\omega'}(z)) \qquad \text{für alle } \omega, \omega' \in R$$

sein, und wenn umgekehrt (29) in dieser Form erfüllt ist, ist $T_{\langle b \rangle, \langle c-b \rangle}$ der CM-Typ von U'. Zusammenfassend können wir daher sagen:

Satz 5. *Für folgende Aussagen gelten die Implikationen* (i) $\Leftrightarrow$ (ii) $\Leftrightarrow$ (iii) $\Leftarrow$ (iv) $\Leftrightarrow$ (v) $\Leftarrow$ (vi):

(i) $\mathrm{Comm}(\mathbb{Q}(\zeta_N), \mathrm{End}_0\, T)$ *hat Nullteiler.*

(ii) *T ist isogen zum Produkt zweier abelscher Varietäten mit komplexer Multiplikation durch* $\mathbb{Q}(\zeta_N)$.

(iii) *Für je zwei $\omega, \omega' \in R$ sind $D_\omega(z)$ und $D_{\omega'}(z) \in \mathbb{Q}(\zeta_N)$ und* $\sigma_\omega^{-1}(D_\omega(z)) = \sigma_{\omega'}^{-1}(D_{\omega'}(z))$ *oder* $\sigma_\omega^{-1}(D_\omega(z)) = \sigma_{\bar\omega'}^{-1}(D_{\bar\omega'}(z))$.

(iv) *T enthält eine abelsche Varietät U' mit komplexer Multiplikation durch $\mathbb{Q}(\zeta_N)$ vom CM-Typ* $T_{\langle b \rangle, \langle c-b \rangle}$.

(v) $\begin{cases} 1) & D_\omega(z) \in \mathbb{Q}(\zeta_N) \textit{ für alle } \omega \in R. \\ 2) & \sigma_\omega^{-1}(D_\omega(z)) \textit{ ist unabhängig von } \omega \in R. \end{cases}$

(vi) *$z \in \bar{\mathbb{Q}}$ und $F(z) \in \bar{\mathbb{Q}}^*$.*

§ 7. Endliche Monodromiegruppen

$r=0$ heißt, daß nach (22) und (24) alle Eigenräume V_n die Dimension 0 oder 2 besitzen, daß also das in (25) gewählte Repräsentantensystem leer ist und somit das symmetrische Gebiet G aus nur einem Punkt besteht. Die Bedingungen (iii) und (v) aus Satz 5 sind leer, unsere abelsche Varietät T aus Satz 1 zerfällt in diesem Fall in zwei abelsche Varietäten vom Typ CM, die beide isogen zu dem oben erwähnten A_2^m sind (wohlbekannt nach [Sie 2] und [Shi 1]). Daß man hier im Fall (A) des Theorems ist, zeigt

Satz 6. *Der Fall $r=0$ tritt genau dann ein, wenn die Monodromiegruppe Δ der hypergeometrischen Differentialgleichung für $F(z)$ endlich ist, d.h. wenn deren Lösungen algebraische Funktionen in z sind.*

Die Sonderfälle aus § 3 sind weiterhin ausgeschlossen; daß dann die beiden letzten Aussagen äquivalent sind, ist ein bekannter Satz von H.A. Schwarz aus dem Jahre 1873 [Schw]. – Da ich in [Wo] zwei ausführliche Beweise für Satz 6 gegeben habe, beschränke ich mich hier auf eine kurze Skizze des ersten Beweises. Er beruht darauf, daß Δ auf verschiedenen Räumen operiert: Erstens als lineare Gruppe auf dem Lösungsraum der Differentialgleichung, zweitens gebrochen-linear auf den zugehörigen Dreiecksfunktionen D, geometrisch gesehen als Dreiecksgruppe, d.h. Untergruppe vom Index 2 in der Gruppe, die von allen Spiegelungen an den Seiten des Kreisbogendreiecks $D(\mathfrak{H})$ erzeugt wird. Drittens läßt sie sich mit Hilfe der Integraldarstellung der hypergeometrischen Funktionen als Automorphismengruppe der Homologie $H_1(X(N,z),\mathbb{Z})$ auffassen [Kl, § 19]. Dies führt viertens zu einer natürlichen Operation von Δ auf dem Periodengitter Λ bzw. Λ^* von T; so kann Δ in natürlicher Weise in $GL_2\mathbb{Z}[\zeta_N]$ eingebettet werden. Jedenfalls operiert Δ simultan auf allen D_ω mit $\omega\in S$, und zwar so, daß jede Matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \gamma\in\Delta\subset GL_2\mathbb{Z}[\zeta_N]$$

als (komponentenweise konjugierte) $\sigma_\omega\gamma$ auf D_ω wirkt, d.h. in der Form

$$D_\omega(z)\mapsto\frac{\sigma_\omega(a)\,D_\omega(z)+\sigma_\omega(b)}{\sigma_\omega(c)\,D_\omega(z)+\sigma_\omega(d)}.$$

Wenn nun $r>0$ ist, d.h. wenn ein $\omega\in R$ existiert, so sagt Satz 3b, daß Δ als hyperbolische Dreiecksgruppe auf einem D_ω operiert, und hyperbolische Dreiecksgruppen sind unendlich. Im Fall $r=0$ ist andererseits Δ diskrete Untergruppe (weil Automorphismengruppe eines Gitters) eines über S erstreckten Produkts von kompakten (weil nach Satz 3a sphärischen) Bewegungsgruppen und muß daher endlich sein.

Für spätere Zwecke sei noch eine fünfte Operation von Δ erwähnt: Läßt man für $r>0$ die Differentiale nicht ganz S, sondern nur R durchlaufen, so operiert $\gamma\in\Delta$ in der Form $(\sigma_\omega\gamma)_{\omega\in R}$ komponentenweise gebrochen-linear auf $G=\prod_{\omega\in R}K_\omega$ und ist auf diese Weise eine Untergruppe der zu G gehörigen Modulgruppe (vgl. [Shi 1], [Wo, § 5]).

§ 8. Arithmetische Monodromiegruppen

Sei zunächst $r=1$, d.h. R bestehe aus genau einem Differential ω 1. Art. Dann ist das komplexe symmetrische Gebiet G die Kreisscheibe K_ω, und die gebrochen-lineare Operation der Monodromiegruppe Δ auf K_ω muß diskontinuierlich sein, da Δ Untergruppe der zugehörigen Modulgruppe ist, sogar von endlichem Index, weil Δ ebenso wie die Modulgruppe endliches Covolumen hat. Δ ist somit „arithmetisch definiert". Das Kriterium (v) aus Satz 5 reduziert die Frage, ob T eine abelsche Untervarietät U' vom CM-Typ $T_{\langle b\rangle,\langle c-b\rangle}$ enthält, auf die Bedingung $D_\omega(z)\in\mathbb{Q}(\zeta_N)$, die natürlich leicht erfüllbar ist. Für unsere Zwecke ist aber nur interessant, ob *gleichzeitig*

$$z\in\overline{\mathbb{Q}} \quad\text{und}\quad D_\omega(z)\in\mathbb{Q}(\zeta_N)$$

sein können. Die folgende Überlegung zeigt, daß in der Tat für sämtliche $\tau\in\mathbb{Q}(\zeta_N)\cap K_\omega$ die D_ω-Urbilder in $\overline{\mathbb{Q}}$ liegen:

Aus dem hyperbolischen Dreieck $D_\omega(\mathfrak{H})$ erhält man durch analytische Fortsetzung von D_ω über die reellen Intervalle $]-\infty,0[$ oder $]0,1[$ oder $]1,\infty[$ ein Bild $D_\omega(\mathfrak{H}_-)$ der unteren Halbebene als hyperbolisch gespiegeltes Dreieck, und die abgeschlossene Vereinigung

$$F_\Delta:=\overline{D_\omega(\mathfrak{H})\cup D_\omega(\mathfrak{H}_-)}$$

ist in den meisten Fällen – wenn nämlich die Dreieckswinkel von der Form $\frac{\pi}{p}, \frac{\pi}{q}, \frac{\pi}{t}$ mit $p, q, t \in \mathbb{N} \cup \{\infty\}$ sind – ein Fundamentalbereich für die Operation von Δ auf K_ω. Weitere sukzessive Anwendung des Schwarzschen Spiegelungsprinzips auf D_ω und eine naheliegende Analyse des Verhaltens in den Singularitäten 0, 1, ∞ zeigen, daß die Umkehrfunktion j_ω von D_ω eindeutig und meromorph auf ganz K_ω fortsetzbar ist, und zwar als Δ-automorphe Funktion, die eine biholomorphe Abbildung der (eventuell kompaktifizierten) „Shimurakurve" $\Delta \backslash K_\omega$ auf $\mathbb{C}$ induziert. j_ω ist so normiert, daß auf den Fixpunkten von Δ die Werte 0, 1, ∞ angenommen werden, ist also eine „Shimura-Abbildung", die jedenfalls dann algebraische Werte annimmt, wenn dem Argument τ im Sinne von (28) eine abelsche Varietät T entspricht, die isogen zum Produkt zweier abelscher Varietäten vom Typ CM ist [Shi 2]. Das ist nach Satz 5 insbesondere dann der Fall, wenn $\tau = D_\omega(z) \in \mathbb{Q}(\zeta_N)$ ist; dann ist also automatisch $z = j_\omega(\tau) \in \overline{\mathbb{Q}}$, d.h. für alle

$$z \in E := j_\omega(\mathbb{Q}(\zeta_N) \cap K_\omega) \subset \overline{\mathbb{Q}}$$

ist T über $\overline{\mathbb{Q}}$ definiert und erfüllt Satz 5(v).

Wenn die Signatur (p, q, t) von Δ von der Form $(2, 3, t)$ oder $(2, q, t)$ ist, besteht allerdings die Möglichkeit [Kn], daß $\overline{D_\omega(\mathfrak{H}) \cup D_\omega(\mathfrak{H}_-)}$ aus mehreren Δ-Fundamentalbereichsbildern zusammengesetzt ist. Die Umkehrfunktion f von D_ω besitzt aber auch hier nur endlich viele Zweige in K_ω und (algebraische) Verzweigungsstellen ausschließlich in den Fixpunkten von Δ. Wenn wir j_ω wie oben als Shimuraabbildung für Δ normieren, so ist f algebraische Funktion von j_ω, und zwar genauer algebraisch über $\overline{\mathbb{Q}}(j_\omega)$; die Abhängigkeit von f und j_ω läßt sich auf verschiedene Weise herleiten und ist in der klassischen Theorie im Rahmen der „quadratischen und höheren Transformationen" der hypergeometrischen Funktionen untersucht und explizit bekannt (vgl. [EMOT] und die dort zitierte Literatur). D_ω ist also Umkehrfunktion einer über $\overline{\mathbb{Q}}(j_\omega)$ algebraischen Funktion, damit bleibt das oben gewonnene Ergebnis über die Ausnahmemenge E unverändert auch hier gültig.

Arithmetisch definierte Monodromiegruppen treten aber nicht nur im Fall $r=1$ auf, sondern unter besonderen Bedingungen auch in den Fällen $r=2$ oder $r=4$. Wir erläutern dies zunächst an einem

Beispiel. $(\lambda, \mu, \nu) = (\frac{1}{4}, \frac{1}{5}, \frac{1}{5})$ führt nach (12) mit $N=40$ auf

$$\begin{aligned} a &= \tfrac{1}{2}(1-\lambda-\mu-\nu) = \tfrac{7}{40}; & A &= \tfrac{1}{2}(1+\lambda-\mu+\nu)N = 25 \\ b &= \tfrac{1}{2}(1-\lambda+\mu-\nu) = \tfrac{3}{8}; & B &= \tfrac{1}{2}(1+\lambda+\mu-\nu)N = 25 \\ c &= 1-\lambda = \tfrac{3}{4}; & C &= \tfrac{1}{2}(1-\lambda-\mu-\nu)N = 7. \end{aligned}$$

Die abelsche Varietät T hat also die Dimension $\Phi(40) = 16$, und die Dimensionen der Eigenräume V_n, $n \in (\mathbb{Z}/40\mathbb{Z})^*$ sind nach (21)

$$r_1 = r_{-1} = r_9 = r_{-9} = 1,$$

$$r_3 = r_{11} = r_{17} = r_{19} = r_{-7} = r_{-13} = 2$$

und

$$r_{-3} = r_{-11} = r_{-17} = r_{-19} = r_7 = r_{13} = 0.$$

Es ist also $r=2$, und R besteht aus $\omega_1 \in V_1$ und $\omega_9 \in V_9$. Zu ω_1 gehört das eingangs genannte Winkeltripel, zu ω_9 nach (12) das Winkeltripel

$$\lambda = \left\langle 9\frac{A}{N}\right\rangle + \left\langle 9\frac{B}{N}\right\rangle - 1 = \tfrac{1}{4}$$

$$\mu = 1 - \left\langle 9\frac{A}{N}\right\rangle - \left\langle 9\frac{C}{N}\right\rangle = -\tfrac{1}{5}$$

$$\nu = 1 - \left\langle 9\frac{B}{N}\right\rangle - \left\langle 9\frac{C}{N}\right\rangle = -\tfrac{1}{5}.$$

Nach (17) unterscheiden sich die beiden Dreiecksfunktionen

$$D_{\omega_9}(z) = \gamma\, D_{\omega_1}(z) =: \gamma\, D(z)$$

nur um eine fünfte Einheitswurzel γ als Faktor. Bezeichnen wir mit σ kurz jenen Automorphismus σ_{ω_9} von $\mathbb{Q}(\zeta_{40})$, der

$$\sigma(\zeta_{40}) = \zeta_{40}^9$$

erfüllt (sein Fixkörper ist ein CM-Körper L vom Grad $\frac{1}{2}\Phi(40)=8$), so ist zu untersuchen, ob und wann für ein $z \in \overline{\mathbb{Q}}$ die Bedingung (v) aus Satz 5

$$D(z) \in \mathbb{Q}(\zeta_{40}) \quad \text{und} \quad \gamma = \frac{\sigma(D(z))}{D(z)} \tag{30}$$

gültig ist (der Fall $D(z)=0 \Leftrightarrow z=0$ ist trivial). Es gibt eine primitive 40. Einheitswurzel ζ mit $\gamma = \dfrac{\zeta}{\sigma(\zeta)}$, so daß (30) äquivalent ist zu

$$\zeta\, D(z) \in \mathbb{Q}(\zeta_{40}) \quad \text{und} \quad \zeta\, D(z) = \sigma(\zeta\, D(z)),$$

also einfach $\zeta\, D(z) \in L$. In der Tat ist auch das $\Phi(40)$-Tupel der r_n invariant unter dem (involutorischen) Automorphismus $n \mapsto 9n$ von $(\mathbb{Z}/40\mathbb{Z})^*$. Daraus folgt, daß in den uns interessierenden Fällen (30) T aus zwei isogenen abelschen Varietäten $T_1 \oplus T_1$ besteht, die einen zu T analogen Endomorphietyp besitzen, nur mit L anstelle von $\mathbb{Q}(\zeta_{40})$ und $r=1$ anstelle von 2. Nun ist wieder ζD durch eine Shimura-Abbildung j (automorphe Funktion zur arithmetisch definierten Dreiecksgruppe der Signatur (4, 5, 5)) umkehrbar, so daß wie im Fall $r=1$

$$\tau = \zeta\, D(z) \in L \Rightarrow z = j(\tau) \in \overline{\mathbb{Q}}$$

gesichert ist.

Weitere Beispiele finden sich in [Wo] und [BW]. Allgemein gilt: Δ spielt genau dann die Rolle einer auf einem eindimensionalen, komplexen symmetrischen Gebiet operierenden arithmetischen Gruppe, wenn alle D_ω, $\omega \in R$, konstante Vielfache voneinander sind; genau dann nämlich durchlaufen die Parameter $\tau \in G$, die den hier auftretenden abelschen Varietäten T nach (25) bis (28) entsprechen, einen eindimensionalen linearen Unterraum, der aus G eine Kreisscheibe ausschneidet. Mit Hilfe von (16) und (17) läßt sich dies auch als Bedingung an die Winkel (λ, μ, ν) formulieren, die nach (12) den verschiedenen $\omega \in R$ entsprechen:

Satz 7. *Folgende Bedingungen sind äquivalent:*

(i) *Die Monodromiegruppe Δ der $\omega \in R$ ist eine arithmetisch definierte, auf einer Kreisscheibe diskontinuierlich operierende Dreiecksgruppe.*

(ii) *Die verschiedenen Dreiecksfunktionen D_ω, $\omega \in R$, sind konstante Vielfache voneinander.*

(iii) *Die den verschiedenen $\omega \in R$ zugeordneten Winkeltripel (λ, μ, ν) unterscheiden sich nur in den Vorzeichen von μ und ν.*

Unter diesen Bedingungen gehören zu den $\omega \in R$ Einbettungen $\sigma_\omega \in \mathrm{Gal}(\mathbb{Q}(\zeta_N)/\mathbb{Q})$ mit einem Fixkörper L; die Gruppe Δ und ihre zugehörige Shimura-Abbildung j lassen sich so normieren, daß $\Delta \subset PGL_2 L$ ist und j in den Fixpunkten von Δ die Werte 0, 1, ∞ annimmt. Dann gilt: Genau in den Punkten $z \in E := j(L) \subset \bar{\mathbb{Q}}$ erfüllt T die Bedingungen aus Satz 5(v).

§ 9. Nichtarithmetische Monodromiegruppen

In den verbleibenden Fällen dürfen wir demnach voraussetzen, daß ω und $\omega' \in R$ existieren, so daß $D_{\omega'} \circ D_\omega^{-1} =: f$ keine lineare Funktion ist. Sei o.B.d.A. $\sigma_\omega = \mathrm{id}$ und $\sigma_{\omega'} =: \sigma$. Nach Satz 5(v) ist zu untersuchen, wann für $w = D_\omega(z) \subset \mathbb{Q}(\zeta_N)$ die Bedingung

$$f(w) = \sigma(w) \tag{31}$$

erfüllt ist. Wenn (31) gilt, dann nicht nur für einen Punkt $w \in K_\omega$: Wenn wir vorerst voraussetzen, daß die Monodromiegruppe von Λ bzw. D_ω diskontinuierlich auf K_ω operiert, also eine (hier nichtarithmetische, hyperbolische) Dreiecksgruppe Δ ist, so läßt sich f holomorph auf ganz K_ω fortsetzen, da die Winkel des Kreisbogendreiecks $D_{\omega'}(\mathfrak{H})$ ganzzahlige Vielfache der entsprechenden Winkel von $D_\omega(\mathfrak{H})$ sind und somit f holomorph in die Fixpunkte von Δ fortsetzbar ist, in denen D_ω^{-1} Singularitäten besitzt (vgl. [Kl, § 63]). Dann ist zunächst klar, daß (31) auch in allen Punkten der Bahn $\Delta(w)$ richtig bleibt. Ferner ist w Fixpunkt einer gebrochen-linearen Transformation

$$\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{Aut}\, K_\omega \cap PGL_2 \mathbb{Q}(\zeta_N);$$

man wähle dazu eine geeignete $\mathbb{Q}(\zeta_N)$-Linearkombination der Identität mit dem in § 6 konstruierten Nullteiler-Endomorphismus (s. auch die Diskussion der nicht-generischen Punkte in [Shi 1]). γ vermittelt also eine Isogenie von Λ, und ebenso alle $\bar{\gamma}$ in der von γ und Δ erzeugten Untergruppe $\bar{\Delta} \subset \mathrm{Aut}\, K_\omega$. Es ist klar, daß (31) damit sogar in der $\bar{\Delta}$-Bahn von w erfüllt sein muß. Nun sind zwei Fälle zu unterscheiden:

1) $\bar{\Delta}$ ist endliche Erweiterung von Δ und damit ebenfalls Dreiecksgruppe; welche $\bar{\Delta}$ hier in Frage kommen, ist nach L. Greenberg [Gre] bekannt. w kann dann nur einer der – modulo Δ endlich vielen – $\bar{\Delta}$-Fixpunkte sein. In der Tat

kann es vorkommen, daß $w \in \mathbb{Q}(\zeta_N)$ und gleichzeitig $z = D_\omega^{-1}(w) \in \bar{\mathbb{Q}}$ ist, so daß man hier zusätzlich endlich viele algebraische Ausnahmewerte der zugehörigen hypergeometrischen Funktionen erhält. Für Beispiele vgl. [BW, Thm. 4].

2) $\bar{\Delta}$ ist unendliche Erweiterung von Δ und damit nicht diskontinuierlich. Daß dieser Fall nicht eintritt, kann man sich folgendermaßen klarmachen: Wenn $\delta = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ auf $w = D_\omega(z)$ in der Form

$$w \mapsto \frac{a w + b}{c w + d} = \delta(w)$$

operiert, so auf $f(w) = D_{\omega'}(z)$ in der Form [5]

$$\delta^\sigma\colon\ f(w) \mapsto \frac{\sigma(a) f(w) + \sigma(b)}{\sigma(c) f(w) + \sigma(d)} =: \delta^\sigma(f(w)).$$

Wenn (31) auf w und $\delta(w)$ zutrifft, ist dies gleichzeitig $\sigma(\delta(w))$, also

$$\delta^\sigma(f(w)) = f(\delta(w)),$$

und da $\bar{\Delta}(w)$ dicht in K_ω liegt, gilt $\delta^\sigma \circ f = f \circ \delta$ in ganz K_ω. Insbesondere werden δ-Fixpunkte durch f in δ^σ-Fixpunkte abgebildet.

Nach Satz 7 unterscheiden sich die Winkeltripel (λ, μ, ν) von ω und (λ', μ', ν') von ω' nicht nur in den Vorzeichen. Sei also etwa

$$\frac{q_1}{q} = \mu \neq \pm\mu' = \pm\frac{q_2}{q} \bmod 1$$

mit $(q_1, q) = (q_2, q) = 1$ und $\sigma(\zeta_q^{q_1}) = \zeta_q^{q_2} \neq \zeta_q^{\pm q_1}$. Da $\bar{\Delta}$ nicht diskontinuierlich ist, liegt auch die Bahn $\bar{\Delta}(D_\omega(\infty))$ dicht in K_ω und besteht aus $\bar{\Delta}$-Fixpunkten der Ordnung q. Man nehme also einen solchen $\bar{\Delta}$-Fixpunkt w_0, der nicht Δ-Fixpunkt ist und daher sogar $f'(w_0) \neq 0$ erfüllt; f ist also in w_0 lokal konform. Dazu gibt es ein $\delta \in \bar{\Delta}$ mit Fixpunkt w_0 und Eigenwert $\zeta_q^{q_1}$. Dieses δ bildet einen Punkt $w_1 \neq w_0$ unserer Bahn $\bar{\Delta}(w)$ in $\delta(w_1) = w_2$ so ab, daß das hyperbolische Dreieck $w_1 w_0 w_2$ in w_0 den Winkel $\frac{2\pi q_1}{q}$ hat. Durch f wird dieses wegen

$$f(w_2) = f(\delta(w_1)) = \delta^\sigma(f(w_1))$$

in ein Dreieck mit dem Winkel $\frac{2\pi q_2}{q}$ in $f(w_0)$ abgebildet. Da man w_1 beliebig nahe an w_0 wählen kann, ist f in w_0 nicht konform im Widerspruch zu unseren Annahmen.

Schließlich bleibt noch die Voraussetzung zu beseitigen, daß die zu D_ω gehörige Monodromiegruppe diskontinuierlich auf K_ω operiert. Wenn Δ die Signatur (p, q, t) hat, lassen sich in jedem Fall die durch π dividierten Winkel von $D_\omega(\mathfrak{H})$

[5] Um Verwechslungen mit der Konjugation der Bildpunkte zu vermeiden, wird hier nicht die Schreibweise $\sigma\delta$ aus § 7 gewählt

bzw. $D_{\omega'}(\mathfrak{H})$ schreiben als

$$\frac{p_1}{p}, \frac{q_1}{q}, \frac{t_1}{t} \quad \text{bzw.} \quad \frac{p_2}{p}, \frac{q_2}{q}, \frac{t_2}{t}$$

mit $p_i, q_i, t_i \in \mathbb{N}$ und $t_i = 1$ für $t = \infty$ etc., andernfalls

$$(p_i, p) = (q_i, q) = (t_i, t) = 1 \qquad \text{für } i = 1, \ 2.$$

Wenn D eine biholomorphe Abbildung von $\mathfrak{H}$ auf ein hyperbolisches Dreieck mit Winkeln $\frac{\pi}{p}, \frac{\pi}{q}, \frac{\pi}{t}$ in der Einheitskreisscheibe K_1 als hyperbolischer Ebene bezeichnet, dann ist jedenfalls D^{-1} als Δ-automorphe Funktion auf ganz K_1 fortsetzbar für eine auf K_1 diskontinuierliche Dreiecksgruppe der Signatur (p, q, t), die zur Monodromiegruppe von D_ω bzw. Δ isomorph ist. Die Funktionen

$$f_1 := D_\omega \circ D^{-1} \quad \text{bzw.} \quad f_2 := D_{\omega'} \circ D^{-1}$$

sind ebenfalls holomorph auf ganz K_1 fortsetzbar, mit möglicher Ausnahme der Fixpunkte von Δ sogar biholomorph.

$$f := f_2 \circ f_1^{-1}$$

läßt sich also als verzweigte, aber außerhalb der Verzweigungspunkte lokal biholomorphe Abbildung $K_\omega \to K_{\omega'}$ lesen, und die oben angewandte Schlußweise läßt sich auf den hier vorliegenden Fall übertragen. Zusammenfassend ergibt sich

Satz 8. *Wenn die Monodromiegruppe Δ der $\omega \in R$ keine arithmetisch definierte Dreiecksgruppe ist, können die Bedingungen* (v) *aus Satz 5 höchstens für jene endlich vielen z erfüllt sein, deren $D_\omega(z)$ Fixpunkte der maximalen Dreiecksgruppe $\bar{\Delta} \supset \Delta$ sind.*

§ 10. Nullstellen. Die Rolle der Parametergleichungen

Es bleibt erstens zu zeigen, daß die Nullstellen der hypergeometrischen Funktionen transzendent sind, soweit sie nicht zu den bisher schon behandelten Ausnahmen gehören. Da unser Differential $\eta_1 = \frac{dx}{y}$ auf $X(N, z)$ bzw. in $H^1_{DR}(T)$ Eigendifferential unter der Operation des Automorphismus κ^* ist, heißt

$$\int_0^1 \eta_1 = 0,$$

daß alle Perioden von η_1 auf einem $\mathbb{Z}[\zeta_N]$-invarianten Untergitter des $\mathbb{Z}$-Rangs $\Phi(N)$ von $H_1(T, \mathbb{Z})$ verschwinden (charakterisiert durch $v = 0$ in Satz 1 bzw. (23)); für $z \in \bar{\mathbb{Q}}$, also für über $\bar{\mathbb{Q}}$ definierte $X(N, z)$ und T, kann das nach [WW, Satz 2] nur dann eintreten, wenn dieses Untergitter zu einer abelschen Untervarietät

U von T gehört, die dann $\mathbb{Q}(\zeta_N)$-invariant ist und darum komplexe Multiplikation durch $\mathbb{Q}(\zeta_N)$ besitzen muß. $H_1(U, \mathbb{Z})$ müßte man dadurch erhalten, daß man in Satz 1 bzw. (23) $v=0$ einsetzt. Nun beachte man, daß Perioden der $\omega \in S$ in Satz 1 höchstens für $z=0$ verschwinden, wie man anhand der Dreiecksfunktionen D_ω einsieht (§ 2, Abschn. 4); die Anzahl der Einbettungen von $\mathbb{Q}(\zeta_N)$ in $\mathrm{End}_0\, U$ wäre darum nur in dem trivialen Ausnahmefall $r=0$ (Satz 6) gleich $\frac{1}{2}\Phi(N)$, wie man es von einer abelschen Varietät vom Typ CM erwartet.

Zweitens ist folgendes Problem zu behandeln: Nach Satz 5 ist es für $z \in \overline{\mathbb{Q}}$ und

$$0 \neq F(a, b, c; z) = \frac{1}{B(b, c-b)} \int_0^1 \eta_1 \in \overline{\mathbb{Q}}$$

zwar notwendig, daß T eine Untervarietät $U' \triangleq A_2^m$ vom CM-Typ $T_{\langle b\rangle, \langle c-b\rangle}$ enthält, aber noch keineswegs hinreichend, da $\int_0^1 \eta_1$ nicht notwendig Periode auf U' ist. Für eine zusätzliche hinreichende Bedingung kann man sich auf jene Fälle beschränken, in denen Bedingung (v) aus Satz 5 erfüllt ist, und man kann ferner die in Satz 8 genannten Ausnahmepunkte ausschließen; es handelt sich dabei nur um endlich viele Punkte, und in diesen läßt sich die algebraische Natur der Werte $F(z)$ einfacher ermitteln mit Hilfe von quadratischen und höheren Transformationen [EMOT], dem Gaußschen Satz über die Werte hypergeometrischer Funktionen an der Stelle 1, und mit Hilfe eines Satzes über die möglichen $\overline{\mathbb{Q}}$-linearen Abhängigkeiten zwischen Werten der Betafunktion an rationalen Stellen [WW, Satz 4]. Wir beweisen daher zunächst

Satz 9. *Die Monodromiegruppe Δ der hypergeometrischen Differentialgleichung von* $F(z) = F(a, b, c; z) = \dfrac{1}{B(b, c-b)} \displaystyle\int_0^1 \eta_1$ *sei arithmetisch definiert, es sei* $a, b, c, a-c, b-c \in \mathbb{Q} - \mathbb{Z}$, *$z$ gehöre zu der in Satz 7 definierten Ausnahmemenge* $E \subset \overline{\mathbb{Q}}$, *aber* $D_\omega(z)$, $\omega \in R$, *sei kein $\bar{\Delta}$-Fixpunkt für die maximale endliche Erweiterung $\bar{\Delta}$ von Δ. Dann gilt: $F(z)$ ist algebraisch genau dann, wenn $\eta_1 \in R$.*

Zum Beweis ist vorweg zu bemerken, daß alle abelschen Varietäten und alle Differentiale, um die es sich hier dreht, über $\overline{\mathbb{Q}}$ definiert sind. Die De-Rham-Cohomologiegruppen $H^1_{DR}(X(N, z))$ und $H^1_{DR}(T)$ zerfallen analog zu $H^0(X(N, z), \Omega) = \sum \oplus V_n$ in Eigenräume W_n unter der natürlichen Operation des Automorphismus κ^* (§ 4), und für alle $n \in (\mathbb{Z}/N\mathbb{Z})^*$ wird dabei $\dim W_n = 2$,

$$W_n \cap H^0(X(N, z), \Omega) = V_n$$

und

$$H^1_{DR}(T) = \sum_{n \in (\mathbb{Z}/N\mathbb{Z})^*} \oplus W_n.$$

Speziell liegt dabei η_1 in $W_1 \subset H^1_{DR}(T)$.

Eine zweite Zerlegung ergibt sich aus den Voraussetzungen des Satzes und aus den Sätzen 5 und 7: T ist isogen zu einer direkten Summe $U \oplus U'$ zweier abelscher Untervarietäten mit komplexer Multiplikation durch $\mathbb{Q}(\zeta_N)$. O.B.d.A. ist $A_2^m \triangleq U'$ vom CM-Typ $T_{\langle b\rangle, \langle c-b\rangle}$, vgl. den Beweis von Satz 5. Dieser Zerlegung entspricht eine Zerlegung der Differentiale

$$\begin{aligned} H^1_{DR}(T) &\cong H^1_{DR}(U) \oplus H^1_{DR}(U') \\ H^0(T,\Omega) &\cong H^0(U,\Omega) \oplus H^0(U',\Omega) \end{aligned} \tag{32}$$

und eine Zerlegung der Homologie

$$H_1(T,\mathbb{Q}) \cong H_1(U,\mathbb{Q}) \oplus H_1(U',\mathbb{Q}). \tag{33}$$

Zunächst zeigen wir – als Variation des Beweises von Satz 5 – die Eindeutigkeit von U und U': Der CM-Typ $T_{\langle b\rangle, \langle c-b\rangle}$ von U' besteht aus den $n \in (\mathbb{Z}/N\mathbb{Z})^*$ mit

$$\left\langle n\frac{A}{N}\right\rangle + \left\langle n\frac{B}{N}\right\rangle - \left\langle n\frac{A+B}{N}\right\rangle = 1,$$

vgl. (8) und den Beweis von Satz 5; da sich die Dimensionen der $\mathbb{Q}(\zeta_N)$-Eigenräume von $H^0(U,\Omega)$ und $H^0(U',\Omega)$, die jeweils 0 oder 1 sind, zu r_n aus (21) ergänzen müssen, und weil wegen $r \geqq 1$ nicht alle $r_n = 0$ oder 2 sein können, sind die CM-Typen (bzw. Eigenwertspektren bei Operation von ζ_N) von U und U' verschieden. Also sind die zugehörigen Nullteilerendomorphismen $\varepsilon, \varepsilon' \in \operatorname{End}_0 T$ mit

$$\operatorname{Bild} \varepsilon = U \quad \text{und} \quad \operatorname{Bild} \varepsilon' = U'$$

bis auf Multiplikation mit Endomorphismen aus $\mathbb{Q}(\zeta_N)$ eindeutig bestimmt, somit auch U und U' ebenso wie die Zerlegungen (32) und (33), sehr im Gegensatz zum Fall $r=0$. Wenn wir also wie im Beweis von Satz 5 $\oint \eta_1$ in zwei Perioden $\oint \eta$ und $\oint \eta'$ von U und U' zerlegen, so ist $F(z) = \oint \eta / B(b, c-b) \in \overline{\mathbb{Q}}^*$ genau dann, wenn $\oint \eta + \oint \eta'$ und $B(b, c-b)$ über $\overline{\mathbb{Q}}$ linear abhängig sind. Wegen der unterschiedlichen CM-Typen von U und U' ist das nach [WW, Satz 2] genau dann der Fall, wenn $\oint \eta = 0$, also wenn $\oint \eta_1 = \oint \eta'$ Periode auf U' ist.

Wenn nun $\eta_1 = \omega \in R$, so ist nach Definition (25) von R und nach Satz 3b) $r_1 = \dim V_1 = 1$ und nach (9) und (12)

$$0 < \lambda = 1 - c = \frac{A+B}{N} - 1 < 1,$$

also gehört V_1 in der Zerlegung (32) zu $H^0(U',\Omega)$, und η_1 ist darum algebraisches Vielfaches von $\varepsilon'^* \eta'$, dem Bild eines Differentials $\eta' \in H^0(U',\Omega)$, das als einzige Perioden algebraische Vielfache von $B(b, c-b)$ besitzt. Damit ist die eine Richtung von Satz 9 bewiesen.

Wegen der Eindeutigkeit von U' genügt es für die Umkehrung zu zeigen, daß η_1, wenn es nicht in R liegt, bei der Zerlegung (32) auch nicht in $H^1_{DR}(U')$ fallen kann, genauer gesagt, daß kein $\eta' \in H^1_{DR}(U')$ mit

$$\eta_1 = \varepsilon'^* \eta' \tag{34}$$

existiert. Wäre (34) erfüllt, so müßten für alle Zykeln $h \in H_1(T,\mathbb{Z})$ die Perioden

$$\oint_{\varepsilon h} \eta_1 = 0 \tag{35}$$

sein, wenn man den Nullteilerendomorphismus ε (mit Bild $\varepsilon = U$ und $U' \subset$ Kern ε) auf $H_1(T, \mathbb{Q})$ operieren läßt. Die Homologie von U ist aus §6 bekannt; (35) ist daher äquivalent zu

$$\sigma_\omega^{-1}(-D_\omega(z)\, v) \int_0^1 \eta_1 + \sigma_\omega^{-1}(v) \int_{1/z}^{\infty} \eta_1 = 0$$

für alle $v \in \mathbb{Z}[\zeta_N]$ und alle $\omega \in R$, mit anderen Worten also

$$\sigma_\omega^{-1}(D_\omega(z)) = D_{\eta_1}(z) := \int_{1/z}^{\infty} \eta_1 \Big/ \int_0^1 \eta_1 . \tag{36}$$

Die Integrale sind als Periodenintegrale wie in (2) und (14) zu lesen.

Wegen $\eta_1 \notin R$ dürfen wir wie in §9 voraussetzen, daß $D_{\eta_1} \circ D_\omega^{-1} = f$ eine meromorphe nichtlineare Funktion auf K_ω ist und (36) in der Form

$$\sigma_\omega^{-1}(w) = f(w) \tag{37}$$

schreiben. Wenn (37) für ein $w \in K_\omega \cap L$ erfüllt ist, das nicht Fixpunkt einer maximalen diskontinuierlichen Erweiterung von Δ ist, so gilt (37) auch in einer dichtliegenden Punktmenge, wie man ganz ähnlich wie in den §§ 8 und 9 einsieht: Mit Hilfe des Nullteiler-Endomorphismus ε konstruiert man eine Erweiterung $\bar{\Delta}$ von Δ, die auf K_ω nicht diskontinuierlich operiert, und deren Elemente auf $H_1(T, \mathbb{Z})$ als Isogenien operieren. Da für $z = D_\omega^{-1}(w)$ alle Perioden $\oint \eta_1$ von η_1 bis auf algebraische Faktoren gleich $B(b, c-b)$ sind, gilt (37) in gleicher Weise für alle $\bar{\Delta}$-Bilder von w. Ganz ähnlich wie in §9 schließt man, daß das für unser nichtlineares f nicht möglich ist.

Schließlich ist die Bedingung

$$\eta_1 = \frac{dx}{y} = \omega \in R$$

noch in eine handlichere Bedingung an die Parameter a, b, c umzuformulieren. Die Holomorphie von ω ist äquivalent zu (9) bzw.

$$b > 0, \quad c > b, \quad a < 1, \quad c - a < 1,$$

die Bedingung $\lambda > 0$ wird zu $c < 1$, und daß $r_1 = 1$ bzw. $D_\omega(\mathfrak{H})$ hyperbolisch sein soll, schreibt sich daher als

$$1 - c + |b - a| + |c - b - a| < 1. \tag{38}$$

Wäre nun etwa $a < 0$, so nähme die linke Seite von (38) die Form

$$1 - c + b - a + c - b - a = 1 - 2a > 1$$

an, was ebenso unmöglich ist wie $a > c > b > 0$ wegen

$$1 - c + a - b + a + b - c = 1 + 2a - 2c > 1.$$

Neben (38) gelten – symmetrisch in a und b – daher die Ungleichungen

$$0 < a < c, \quad 0 < b < c, \quad c < 1. \tag{39}$$

Aus (38) und (39) folgt trivialerweise wieder $\omega \in R$. Ebenso leicht ist einzusehen, daß (38) und (39) dazu äquivalent sind, daß $\oint \omega$ und $B(b, c-b)$ Perioden erster Art sind und $D_\omega(\mathfrak{H})$ ein hyperbolisches Dreieck ist.

Literatur

[A] Albert, A.A.: A solution of the principal problem in the theory of Riemann matrices. Ann. Math. **35**, 500–515 (1934)

[Be] Bertrand, D.: Endomorphismes de groupes algébriques; applications arithmétiques. In: Approximations diophantiennes et nombres transcendants, Luminy 1982, Progress in Math. 31, pp. 1–45. Basel-Boston: Birkhäuser 1983

[Bo] Bombieri, E.: On G-functions. In: Recent progress in analytic number theory, vol. 2, pp. 1–67. Durham 1979, London: Academic press 1981

[BW] Beukers, F., Wolfart, J.: Algebraic values of hypergeometric functions, erscheint in den Proceedings der Durham conference on transcendental numbers 1986

[CW] Chevalley, Cl., Weil, A.: Über das Verhalten der Integrale 1. Gattung bei Automorphismen des Funktionenkörpers. Abh. Hamburger Math. Sem. **10**, 358–361 (1934)

[EMOT] Erdélyi, A., Magnus, W., Oberhettinger, F., Tricomi, F.G.: Higher transcendental functions, Bateman manuscript project. New York: McGraw-Hill 1953

[Gre] Greenberg, L.: Maximal Fuchsian groups. Bull. Am. Math. Soc. **69**, 569–573 (1963)

[Gr] Gross, B.: On the periods of abelian integrals and a formula of Chowla and Selberg. Invent. Math. **45**, 193–208 (1978)

[Ja] Jacobson, N.: Theory of rings. Am. Math. Soc. (1943)

[Ka] Katz, N.M.: Algebraic solutions of differential equations (p-curvature and the Hodge filtration). Invent. Math. **18**, 1–118 (1972)

[Kl] Klein, F.: Vorlesungen über die hypergeometrische Funktion. Berlin: Springer 1933

[Kn] Knapp, M.: Doubly generated Fuchsian groups. Mich. Math. J. **15**, 289–304 (1968)

[KR] Koblitz, N., Rohrlich, D.: Simple factors in the Jacobian of a Fermat curve. Can. J. Math. **30**, 1183–1205 (1978)

[Schw] Schwarz, H.A.: Über diejenigen Fälle, in welchen die Gaußische hypergeometrische Reihe eine algebraische Funktion ihres vierten Elements darstellt. J. Reine Angew. Math. **75**, 292–335 (1873)

[ST] Shimura, G., Taniyama, Y.: Complex multiplication of abelian varieties and its applications to number theory. Publ. Math. Soc. Japan **6**, (1961)

[Shi 1] Shimura, G.: On analytic families of polarized abelian varieties and automorphic functions. Ann. Math. **78**, 149–192 (1963)

[Shi 2] Shimura, G.: Construction of class fields and zeta functions of algebraic curves. Ann. Math. **85**, 58–159 (1967)

[Sie 1] Siegel, C.L.: Über einige Anwendungen diophantischer Approximationen, Ges. Werke, Bd. I, pp. 209–266. Berlin-Heidelberg-New York: Springer 1966

[Sie 2] Siegel, C.L.: Lectures on Riemann matrices. Tata Institute, Bombay 1963

[Ta 1] Takeuchi, K.: Arithmetic triangle groups. J. Math. Soc. Japan **29**, 91–106 (1977)

[Ta 2] Takeuchi, K.: Commensurability classes of arithmetic triangle groups. J. Fac. Sci. Univ. Tokyo, Sec. IA, **24**, 201–212 (1977)

[WW] Wolfart, J., Wüstholz, G.: Der Überlagerungsradius gewisser algebraischer Kurven und die Werte der Betafunktion an rationalen Stellen. Math. Ann. **273**, 1–15 (1985)

[Wo] Wolfart, J.: Fonctions hypergéométriques, arguments exceptionels et groupes de monodromie, Publ. Math. Univ. Pierre Marie Curie No. 79, Problèmes diophantiens 1985–1986, IX.1–IX.24

[Wü] Wüstholz, G.: Recent progress in transcendence theory. In: Number theory Noordwijkerhout 1983. (Lect. Notes Math., vol. 1068, pp. 280–296). Berlin-Heidelberg-New York: Springer 1984

Oblatum 13-III-1986 & 1-XII-1986 & 9-IV-1987