Inventiones math. 22, 211 – 319 (1973)

Variation of Hodge Structure: The Singularities of the Period Mapping

Wilfried Schmid* (New York)

Table of Contents

§ 1. Introduction

By a classical construction, to each polarized Abelian variety and each compact Riemann surface, one canonically attaches a point in a certain modular variety. Because of Torelli's theorem in the case of Riemann surfaces, and for elementary reasons in the case of Abelian varieties, this invariant determines the Riemann surface or Abelian variety in question up to isomorphism. For a family of polarized Abelian varieties or Riemann surfaces, by associating to every fibre its invariant, one obtains a mapping of the parameter space of the family into the modular variety. This so-called period mapping reflects various geometric properties of the family.

In his study of the moduli of algebraic manifolds, Griffiths has extended the construction of the period mapping. If one considers the various ways of turning a fixed compact C^∞ manifold into an algebraic manifold, the Hodge decomposition of the cohomology groups becomes an invariant of the algebraic structure. To make it depend on the isomorphism class of the algebraic manifold alone, one must identify any two Hodge decompositions which are related by a diffeomorphism of the underlying C^∞ manifold. In this manner, Griffiths assigns to each

* Supported in part by an Alfred P. Sloan Memorial Fellowship and NSF contract GP32843.

polarized algebraic manifold[1] a point in a "classifying space for Hodge structures", modulo the action of a discrete group. Just as before, a family of polarized algebraic manifolds gives rise to a period mapping. Although the analogue of Torelli's theorem is known only for some very special types of algebraic manifolds, the period mapping again reflects a number of aspects of the geometry of the situation.

In the classical cases, the Siegel upper half plane plays the role of the classifying space, which is therefore a bounded symmetric domain. In general, the classifying space is not at all a bounded domain; nevertheless, those holomorphic mappings into it which have a certain property shared by all period mappings behave vaguely like mappings into bounded domains. One can thus investigate the period mapping of a family of algebraic manifolds by function-theoretic methods, and derive geometric results by analytic arguments.

When families of algebraic manifolds come up in algebraic geometry, they usually have some singular fibres. The period mapping is then defined not on the entire parameter space, but only on the complement of the subvariety corresponding to the singular fibres; along the subvariety, the period mapping may become singular. It is the object of this paper to study the singularities which can occur, and to discuss the geometric consequences of the resulting description of the singularities.

For relatively simple reasons, if a period mapping is defined outside a subvariety of codimension at least two, it can be continued across the subvariety, so that the singularities are removable [13]. I shall therefore consider only subvarieties of codimension one. According to Hironaka, a suitable modification turns the ambient space into a manifold, and the subvariety in question into a divisor with at most normal crossings. Thus, localizing the problem, one arrives at the following situation: the period mapping is defined on a polycylinder, from which some coordinate hyperplanes have been removed; in other words, on a product of punctured discs and discs. The mapping, by its very definition, takes values in the quotient of a classifying space for Hodge structures modulo a discrete group of automorphisms. However, passing to the universal covering of the product of punctured discs and discs, one can lift the mapping to the classifying space itself. As the first of the two main theorems of this paper assert, the lifted mapping is asymptotic to the orbit of a certain nilpotent Lie group in the classifying space; moreover, the approximating orbit inherits several of the properties of the original mapping.

There is a close relationship between this "nilpotent orbit theorem" and the regularity theorem for the Gauss-Manin connection [8]. The

[1] The polarization, i.e. the cohomology class of a projective embedding, is necessary for technical reasons.

regularity theorem implies the one-variable version of the nilpotent orbit theorem fairly directly, although it does not seem to give the several variables case. Conversely, with some effort, the regularity of the Gauss-Manin connection can be deduced from the nilpotent orbit theorem.

The nilpotent orbit theorem reduces problems concerning the singularities of the period mapping to questions about the nilpotent orbit which approximates the mapping. To make the theorem useful in various applications, one still needs detailed information on the nilpotent orbits which may occur. This, incidentally, appears to be a deeper matter than the nilpotent orbit theorem itself. For most applications, an understanding of these orbits in the case of a single variable suffices entirely. Also, the general case, of more than one variable, cannot be treated by the same arguments and causes major additional difficulties. I have therefore limited myself to the one-variable situation. However, I intend to take up the general case in a future continuation of this paper.

The nilpotent orbits which can come up in the one-variable version of the nilpotent orbit theorem are described by the second major theorem, the "SL_2-orbit theorem". Roughly speaking, each such orbit in turn is asymptotic to an equivariantly embedded copy of the upper half plane, which lies in the classifying space for Hodge structures in a special way. The precise statement of the theorem is complicated, but it seems to be sharp, and it works in various applications.

For both theorems, it is irrelevant whether the period mapping actually comes from a family of polarized algebraic manifolds; the proofs only depend on certain properties which are common to all period mappings. Consequently, both theorems hold in a more general, abstract setting, for which Griffiths has coined the term "variation of Hodge structure". Although this generalization is of little interest in itself, some of the geometric consequences of the theorems can be derived more easily in the wider context of a variation of Hodge structure.

According to a conjecture of Griffiths, there should be an analogue of the Satake-Borel-Baily compactification, for any quotient of a classifying space for Hodge structures by the action of an arithmetic subgroup of its automorphism group. However, in general one ought not to expect it to be a compact space. Such a partial compactification should have the property that period mappings, defined on punctured discs, can be continued across the punctures, as mappings into this larger space; also, relative to the extension property, it should be essentially minimal. The SL_2-orbit theorem suggests how one might try to construct a partial compactification. Once the partial compactifications are known to exist, the two orbit theorems may well enter the proof of the extension property.

This is speculation, of course. Among the more immediate applications of the two main theorems are a sharpened version of Landman's monodromy theorem [21]; more detailed information about the monodromy transformation around a singular variety; a result which relates the action of the monodromy transformation to the growth of cohomology classes; and an affirmative answer to a conjecture of Deligne. As the conjecture asserts, if a one-parameter family of polarized algebraic manifolds degenerates to a singular variety, the Hodge structures of the regular fibres approach what Deligne has called a "mixed Hodge structure". According to Deligne's Hodge theory for algebraic varieties [7], the cohomology groups of the singular fibre also carry natural mixed Hodge structures. By relating these two types of mixed Hodge structures, one may hope to get some information on how a nonsingular algebraic variety can degenerate into a singular one. To me, of the potential consequences of the two orbit theorems, this seems the most important. A forthcoming joint paper of Clemens and myself will be devoted to such questions.

Although the two main theorems are of a local nature, they have also global implications. In [11], Griffiths proved certain global statements about the period mapping of a family of algebraic manifolds, and more generally, of an abstract variation of Hodge structure. His arguments are differential-geometric, and they depend on the compactness of the base space. Since then, Deligne has given algebraic-geometric proofs of the same results, valid for any algebraic family with quasi-projective base [6]. By using the two orbit theorems, one can make Griffiths' original arguments go through for an arbitrary variation of Hodge structure, provided the base is Zariski open in some compact variety.

As for the organization of this paper, the two sections following the introduction establish notation and review the basic definitions and constructions. The nilpotent orbit theorem is stated in section four. In order to clarify its meaning, I have given a separate statement of the one-variable version of the theorem. Section four also discusses the connection between the regularity of the Gauss-Manin connection and the nilpotent orbit theorem; some of the details, which go beyond the framework of this paper, are omitted. Section five is devoted to the SL_2-orbit theorem and related questions, such as the rationality of the orbits which occur, and their behavior with respect to Siegel sets. The local and global consequences of the two main theorems are taken up in sections six and seven. In a conversation with Griffiths, it became apparent that the computations in [11] can be somewhat simplified. For this reason, and also for the sake of completeness, section seven contains complete arguments, instead of referring to [11]. Since the proofs of the two orbit theorems are lengthy, I have separated them from the statements of the theorems; they can be found in the final two sections.

The main results of this paper have been announced and discussed in [14]. There is also a proof of the one-variable version of the nilpotent orbit theorem, using the regularity of the Gauss-Manin connection, and an outline of the proof of the SL_2-orbit theorem.

I wish to record my indebtedness to Phillip Griffiths and Pierre Deligne. I had numerous discussions with Griffiths, which have influenced this paper. In particular, he helped me understand the relationship between the regularity of the Gauss-Manin connection and the nilpotent orbit theorem. Deligne read an earlier version of the proofs of the two main theorems, and he suggested several improvements. Lemma 8.17 in its present form it due to him. He also extracted the statement of Lemma 9.33, which was only implicit in the original argument, and simplified its proof. From the two theorems, he independently deduced most of the consequences mentioned in section six, as will be pointed out in that section.

§ 2. Hodge Structures

As is well-known, the complex cohomology groups of a nonsingular complex projective variety V have a decomposition according to Hodge type,

$$H^k(V,\mathbb{C})=\oplus H^{p,q}(V,\mathbb{C}),\qquad p,q\geqq 0,\ p+q=k; \tag{2.1}$$

$H^{p,q}(V,\mathbb{C})$ and $H^{q,p}(V,\mathbb{C})$ are then complex conjugate to each other, relative to $H^k(V,\mathbb{R})$. Now let $\eta\in H^2(V,\mathbb{Z})$ be the cohomology class of the projective embedding, i.e. the Chern class of any hyperplane section of V. Left multiplication by η determines a linear map

$$L:\ H^k(V,\mathbb{Q})\to H^{k+2}(V,\mathbb{Q}).$$

For $0\leqq k\leqq n=\dim_{\mathbb{C}} V$, the $(n-k)$-th power of this Kähler operator is injective. The primitive part of $H^k(V,\mathbb{Q})$, which will be denoted by $P^k(V,\mathbb{Q})$, is defined as the kernel of the next higher power of L, namely L^{n-k+1}. If $k>n$, one defines $P^k(V,\mathbb{Q})=0$. One then obtains the decomposition

$$H^k(V,\mathbb{Q})=\oplus L^j P^{k-2j}(V,\mathbb{Q}),\qquad j\geqq \max(k-n,0)$$

(cf. [25]). Tensoring with $\mathbb{C}$ gives the corresponding statement

$$H^k(V,\mathbb{C})=\oplus L^j P^{k-2j}(V,\mathbb{C}),\qquad j\geqq \max(k-n,0). \tag{2.2}$$

Here L and $P^k(V,\mathbb{C})$ are defined in analogy to the situation before. Since the image of η in $H^2(V,\mathbb{C})$ has Hodge type (1, 1), the two decompositions (2.1) and (2.2) are compatible: let $P^{p,q}(V,\mathbb{C})$ denote the intersection of $H^{p,q}(V,\mathbb{C})$ and $P^{p+q}(V,\mathbb{C})$; then

$$\begin{aligned} P^k(V,\mathbb{C})&=\oplus P^{p,q}(V,\mathbb{C}), && p,q\geqq 0,\ p+q=k,\\ H^{p,q}(V,\mathbb{C})&=\oplus L^j P^{p-j,q-j}(V,\mathbb{C}), && j\geqq \max(p+q-n,0). \end{aligned} \tag{2.3}$$

The operator L is defined over $\mathbb{Q}$, so that $P^{p,q}(V,\mathbb{C})$ and $P^{q,p}(V,\mathbb{C})$ are again complex conjugates.

The primitive cohomology groups carry a nondegenerate bilinear form, the so-called Hodge bilinear form,

$$S(c_1,c_2)=(-1)^{k(k-1)/2}(L^{n-k}c_1\wedge c_2)[V], \qquad c_i\in P^k(V,\mathbb{C})$$

($[V]$ = fundamental cycle of V), which is evidently defined over $\mathbb{Q}$. Depending on whether k is even or odd, S is symmetric or skew. The Hodge bilinear relations [25] assert that

$$(2.4)\qquad \begin{aligned} &S\big(P^{p,q}(V,\mathbb{C}),P^{r,s}(V,\mathbb{C})\big)=0 \quad \text{unless } p=s,\ q=r,\\ &i^{p-q}S(c,\bar{c})>0 \quad \text{for } c\in P^{p,q}(V,\mathbb{C}),\ c\neq 0 \end{aligned}$$

(barring designates complex conjugation).

This state of affairs has been codified by Deligne in the following definitions [5]: Let $H_{\mathbb{R}}$ be a finite dimensional real vector space with a $\mathbb{Q}$-structure defined by a lattice $H_{\mathbb{Z}}\subset H_{\mathbb{R}}$, and let $H_{\mathbb{C}}$ denote the complexification of $H_{\mathbb{R}}$. A *Hodge structure* is a decomposition

$$(2.5)\qquad H_{\mathbb{C}}=\oplus H^{p,q}, \qquad \text{with } H^{q,p}=\overline{H}^{p,q}.$$

It is not specifically assumed that $H^{p,q}=0$ unless $p,q\geqq 0$. The integers $h^{p,q}=\dim H^{p,q}$ are the *Hodge numbers*. The Hodge structure (2.5) is said to have *weight* k if the subspaces $H^{p,q}$ are nonzero only when $p+q=k$. To each Hodge structure of weight k one assigns the *Hodge filtration*

$$(2.6)\qquad H_{\mathbb{C}}\supset\cdots\supset F^{p-1}\supset F^p\supset F^{p+1}\supset\cdots\supset 0,$$

where

$$(2.7)\qquad F^p=\oplus_{i\geqq p}H^{i,k-i}.$$

This filtration has the property

$$(2.8)\qquad H_{\mathbb{C}}=F^p\oplus\overline{F}^{k-p+1}, \qquad \text{for each } p.$$

Conversely, every decreasing filtration with the property (2.8) determines a Hodge structure $\{H^{p,q}\}$, namely

$$(2.9)\qquad H^{p,q}=F^p\cap\overline{F}^q \qquad (p+q=k).$$

In this manner, weighted Hodge structures and Hodge filtrations correspond to each other bijectively.

A *morphism* of weighted Hodge structures is a rationally defined linear map between two vector spaces with Hodge structures of the same weight, which preserves the Hodge decompositions (or equivalently, which preserves the Hodge filtrations). More generally, a linear map

$$\lambda\colon H_{1,\mathbb{C}}\to H_{2,\mathbb{C}}$$

between two Hodge structures of weight k_1 and k_2 shall be called a *morphism of type* (r, r) if it is defined over $\mathbb{Q}$, if $k_2 = k_1 + 2r$, and if $\lambda H_1^{p,q} \subset H_2^{p+r,q+r}$ for all p, q. This last condition is again equivalent to $\lambda F_1^p \subset F_2^{p+r}$, for all p.

A polarization for a Hodge structure of weight k consists of the datum of a bilinear form S on $H_{\mathbb{C}}$, which is defined over $\mathbb{Q}$, and which is symmetric for even k, skew for odd k, such that

$$(2.10) \qquad \begin{aligned} S(H^{p,q}, H^{r,s}) &= 0 \quad \text{unless } p=s,\ q=r, \\ i^{p-q} S(v, \bar{v}) &> 0 \quad \text{if } v \in H^{p,q},\ v \neq 0. \end{aligned}$$

In particular, such a bilinear form must be nondegenerate. For each Hodge structure $\{H^{p,q}\}$, the Weil operator $C: H_{\mathbb{C}} \to H_{\mathbb{C}}$ is defined by

$$C v = i^{p-q} v, \qquad \text{if } v \in H^{p,q}.$$

In terms of the Hodge filtration, the relations (2.10) become equivalent to

$$(2.11) \qquad \begin{aligned} S(F^p, F^{k-p+1}) &= 0 \quad \text{for all } p, \\ S(Cv, \bar{v}) &> 0 \quad \text{if } v \in H_{\mathbb{C}},\ v \neq 0. \end{aligned}$$

As the definitions have been arranged, the decomposition (2.1) describes a Hodge structure on $H^k(V, \mathbb{C})$, of weight k. The sub-Hodge structure $\{P^{p,q}(V, \mathbb{C})\}$ on $P^k(V, \mathbb{C})$ is polarized with respect to the Hodge bilinear form. In view of (2.3), this sub-Hodge structure completely determines the full Hodge structure (2.1).

There are some obvious functorial constructions which can be performed with Hodge structures. Let H_1, H_2 be two finite dimensional complex vector spaces, each equipped with a $\mathbb{Q}$-structure and a Hodge structure of weight k_i, $i = 1, 2$. Then $H_1 \otimes H_2$ inherits a Hodge structure from the two factors, of weight $k_1 + k_2$: the tensor product of a vector $v_1 \in H_1^{p_1,q_1}$ and a vector $v_2 \in H_2^{p_2,q_2}$ is assigned the Hodge type $(p_1 + p_2, q_1 + q_2)$. If each of the factors H_1 and H_2 carries a bilinear form which polarizes the respective Hodge structure, then, as can be checked directly, the natural bilinear form on $H_1 \otimes H_2$ polarizes the product Hodge structure. In the case of a single vector space $H_{\mathbb{C}}$ with a Hodge structure of weight k, the n-th symmetric product of $H_{\mathbb{C}}$ carries a unique Hodge structure of weight nk, such that the symmetrization map from the n-th tensor product to the n-th symmetric product becomes a morphism of Hodge structures. In case the Hodge structure of $H_{\mathbb{C}}$ is polarized by a bilinear form S, the induced bilinear form on the symmetric product will polarize the product Hodge structure. Similarly, the n-th tensor product $\otimes^n H_{\mathbb{C}}$ induces a Hodge structure of weight nk on its subspace $\Lambda^n H_{\mathbb{C}}$. The dual space $H_{\mathbb{C}}^*$ of $H_{\mathbb{C}}$ has a unique Hodge structure of weight

$-n$, such that the natural pairing from $H_{\mathbb{C}}^* \otimes H_{\mathbb{C}}$ to $\mathbb{C}$ becomes a morphism when $\mathbb{C}$ is given the trivial Hodge structure of weight zero. Finally, since $\mathrm{Hom}(H_1, H_2) \simeq H_1^* \otimes H_2$, $\mathrm{Hom}(H_1, H_2)$ carries a natural Hodge structure whenever H_1 and H_2 do. All these constructures are again compatible with any polarizations which may be present. For future reference

(2.12) Hodge structures and polarized Hodge structures are compatible with the operations of tensor products, symmetric products, exterior products, Hom, and duality.

Griffiths [10] first considered Hodge structures in order to study the following geometric situation: Let $\mathscr{V}$ and M be connected complex manifolds,

$$\pi\colon \mathscr{V} \to M$$

a surjective, proper, holomorphic map with connected fibres, which is everywhere of maximal rank. For each $t \in M$, the fibre

$$V_t = \pi^{-1}(t) \subset \mathscr{V}$$

is thus a compact, complex submanifold. The fibres are assumed to have the structure of polarized algebraic variety, i.e. each comes equipped with the cohomology class $\eta_t \in H^2(V_t, \mathbb{Z})$ of a projective embedding; as a final hypothesis, the polarizations η_t shall fit together, to give a section of the direct image sheaf $R^2 \pi_*(\mathbb{Z})$. I shall refer to this geometric situation as a *family of polarized algebraic manifolds*, parameterized by M. Typically, it arises as follows: Let $\pi\colon \bar{\mathscr{V}} \to \bar{M}$ be a surjective algebraic mapping between complex projective varieties, whose generic fibre is smooth, let $M \subset \bar{M}$ be the Zariski open subset of the set of nonsingular points of M, over which π has smooth fibres, and let $\mathscr{V}$ be the inverse image of M; if each fibre is given the polarization of a particular projective embedding of $\bar{\mathscr{V}}$, then all the hypotheses are met.

One can regard $\mathscr{V} \to M$ as a C^∞ fibre bundle. Hence, for each k between 0 and $2n$ ($n = \dim_{\mathbb{C}} V_t$), there exists a flat complex vector bundle $\mathbf{H}_{\mathbb{C}}^k \to M$, whose sheaf of germs of flat sections is the direct image sheaf $R^k \pi_*(\mathbb{C})$. For $t \in M$, the fibre of $\mathbf{H}_{\mathbb{C}}^k$ over t can be naturally identified with $H^k(V_t, \mathbb{C})$. The flat bundle $\mathbf{H}_{\mathbb{C}}^k$ contains a flat real subbundle $\mathbf{H}_{\mathbb{R}}^k$, whose fibres correspond to the subspaces $H^k(V_t, \mathbb{R}) \subset H^k(V_t, \mathbb{C})$; and $\mathbf{H}_{\mathbb{R}}^k$, in turn, contains a flat lattice bundle $\mathbf{H}_{\mathbb{Z}}^k$, whose fibres are the images on $H^k(V_t, \mathbb{Z})$ in $H^k(V_t, \mathbb{R})$. From harmonic theory with variable coefficients [19] or Grauert's coherence theorem [9], it follows that the integers $h_t^{p,q} = \dim H^{p,q}(V_t, \mathbb{C})$ depend upper semicontinuously on t. Since their sum, extended over all p and q with $p+q=k$, remains constant, the individual summands must also stay constant. Thus one can again quote [9] or [19], to conclude: there exist C^∞ subbundles $H^{p,q} \subset$

$\mathbf{H}_{\mathbb{C}}^{k}$ $(p+q=k)$, with fibre $H^{p,q}(V_t, \mathbb{C})$ over $t\in M$. For $0\leqq p\leqq k$, $\mathbf{F}^p=\bigoplus_{i\geqq p}\mathbf{H}^{i,k-i}$ is then also a C^∞ subbundle of $\mathbf{H}_{\mathbb{C}}^{k}$.

Let $\mathbf{T}^*\to M$ be the holomorphic contangent bundle, and

$$\nabla\colon \mathcal{O}(\mathbf{H}^k)\to\mathcal{O}(\mathbf{H}^k\otimes\mathbf{T}^*)$$

the canonical flat connection $\big(\mathcal{O}(\ldots)=$sheaf of germs of holomorphic section of $\ldots\big)$. It is now possible to state Griffiths' most basic theorem on the variation of Hodge structures; proofs can be found in [10] or [5].

(2.13) **Theorem** (Griffiths). *The subbundles $\mathbf{F}^p\subset\mathbf{H}_{\mathbb{C}}^{k}$ are holomorphic subbundles. Furthermore, for each p, $\nabla\mathcal{O}(\mathbf{F}^p)\subset\mathcal{O}(\mathbf{F}^{p-1}\otimes\mathbf{T}^*)$.*

The first of these two statements asserts that the Hodge filtration varies holomorphically with $t\in M$. The second becomes vacuous if $k=1$; for higher k, it has turned out to be a crucial ingredient of many arguments.

Because of technical reasons, which will become apparent in the following, it is necessary to consider the polarized Hodge structures (2.3) on the primitive parts of the cohomology groups of the fibres V_t, rather than the Hodge structures (2.1). Since these sub-Hodge structures completely determine the full Hodge structures (2.1), no information is lost in doing so. According to the definition of a family of polarized algebraic manifolds, the polarizations form a flat section of $\mathbf{H}_{\mathbb{Z}}^{2}$. It follows that the subspaces $P^k(V_r, \mathbb{C})\subset H^k(V_t, \mathbb{C})$ constitute the fibres of a flat complex subbundle $\mathbf{P}_{\mathbb{C}}^{k}\subset\mathbf{H}_{\mathbb{C}}^{k}$, which is the complexification of the flat real subbundle $\mathbf{P}_{\mathbb{R}}^{k}=\mathbf{P}_{\mathbb{C}}^{k}\cap\mathbf{H}_{\mathbb{R}}^{k}$; $\mathbf{P}_{\mathbb{R}}^{k}$ contains the flat bundle of lattices $\mathbf{P}_{\mathbb{Z}}^{k}=\mathbf{P}_{\mathbb{R}}^{k}\cap\mathbf{H}_{\mathbb{Z}}^{k}$. With respect to any local flat trivialization of $\mathbf{H}_{\mathbb{C}}^{k}$, the vector spaces $P^{p,q}(V_t, \mathbb{C})$ become the intersections of a fixed vector space with the continuously varying spaces $H^{p,q}(V_t, \mathbb{C})$, so that their dimensions must depend lower semi-continuously on t. In view of (2.3), this forces the dimensions to stay constant. Thus the vector spaces $P^{p,q}(V_t, \mathbb{C})$ are the fibres of C^∞ subbundles $\mathbf{P}^{p,q}\subset\mathbf{P}_{\mathbb{C}}^{k}$ $(p+q=k)$. With a change of notation, I now set $\mathbf{F}^p=\bigoplus_{i\geqq p}\mathbf{P}^{i,k-i}$. As a reformulation of Griffiths' theorem, one finds:

(2.14) the subbundles $\mathbf{F}^p\subset\mathbf{P}_{\mathbb{C}}^{k}$ are holomorphic, and

$$\nabla\mathcal{O}(\mathbf{F}^p)\subset\mathcal{O}(\mathbf{F}^{p-1}\otimes\mathbf{T}^*),\qquad \text{for each } p.$$

In the case of an algebraic family $\pi\colon \mathscr{V}\to M$, the above constructions can be performed in the algebraic category. More concretely, let $\mathscr{V}$ and M be nonsingular quasi-projective varieties over $\mathbb{C}$, and $\pi\colon \mathscr{V}\to M$ a surjective algebraic mapping with smooth, connected fibres. Then $\mathscr{V}\to M$ may be viewed as a family of polarized algebraic manifolds. The bundles $\mathbf{P}_{\mathbb{C}}^{k}$, $\mathbf{F}^p$ all have natural algebraic structures, and the flat connection ∇ is also algebraic. In this context, Grothendieck has called ∇ the *Gauss-*

Manin connection. A more detailed discussion can be found in [8] and [17].

It is sometimes convenient to consider collections of data with the properties mentioned above, which may not come directly from a family of polarized algebraic manifolds. For this purpose, one introduces the notion of a *variation of Hodge structure* (cf. [11]). The ingredients are:

a) a connected complex manifold M;

b) a flat complex vector bundle $\mathbf{H}_{\mathbb{C}} \to M$ with a flat real structure $\mathbf{H}_{\mathbb{R}} \subset \mathbf{H}_{\mathbb{C}}$, and with a flat bundle of lattices $\mathbf{H}_{\mathbb{Z}} \subset \mathbf{H}_{\mathbb{R}}$;

c) an integer k;

d) a flat, nondegenerate bilinear form $\mathbf{S}$ on $\mathbf{H}_{\mathbb{C}}$, which is rational with respect to the lattice bundle $\mathbf{H}_{\mathbb{Z}}$, and which is symmetric or skew, depending on whether k is even or odd;

e) and a decreasing filtration

$$\mathbf{H}_{\mathbb{C}} \supset \cdots \supset \mathbf{F}^{p-1} \supset \mathbf{F}^{p} \supset \mathbf{F}^{p+1} \supset \cdots \supset 0$$

of $\mathbf{H}_{\mathbb{C}}$ by holomorphic subbundles.

These objects are to satisfy the following two conditions:

i) For each point $t \in M$, the fibres $\mathbf{F}_t^p$ of the bundles $\mathbf{F}^p$ constitute the Hodge filtration of a Hodge structure of weight k on the fibre of $\mathbf{H}_{\mathbb{C}}$ at t, and $\mathbf{S}$ polarizes this Hodge structure.

ii) For each p, $\nabla \mathcal{O}(\mathbf{F}^p) \subset \mathcal{O}(\mathbf{F}^{p-1})$ (∇ = flat connection of $\mathbf{H}_{\mathbb{C}}$).

Now let $\{M, \mathbf{H}_{\mathbb{C}}, \mathbf{F}^p\}$ be a variation of Hodge structure. In view of (2.7) and (2.9), the fibres of the intersection $\mathbf{F}^p \cap \bar{\mathbf{F}}^{k-p}$ are isomorphic to those of $\mathbf{F}^p/\mathbf{F}^{p+1}$, and thus have constant dimension. This makes the intersection a C^∞ vector bundle. For future reference,

$$\text{(2.15)} \quad \begin{aligned} &\mathbf{H}^{p,k-p} = \mathbf{F}^p \cap \bar{\mathbf{F}}^{k-p} && \text{is a } C^\infty \text{ subbundle of } \mathbf{H}_{\mathbb{C}}, \text{ and} \\ &\mathbf{F}^p = \oplus_{i \geq p} \mathbf{H}^{i,k-i}, && \text{as } C^\infty \text{ vector bundle.} \end{aligned}$$

By definition, the fibres of the bundles $\mathbf{H}^{p,q}$ at a point $t \in M$ are the Hodge (p, q)-spaces of the Hodge structure corresponding to t. A section of $\mathbf{H}_{\mathbb{C}}$ is said to have Hodge type (p, q) at t if its value at t lies in the fibre of $\mathbf{H}^{p,q}$.

According to the explanations above (2.14), if $\pi: \mathcal{V} \to M$ is a family of polarized algebraic manifolds, the bundle of primitive k-th cohomology groups $\mathbf{P}_{\mathbb{C}}^k \to M$ carries a variation of Hodge structure in the sense of the definition.

As was mentioned already, when a variation of Hodge structure arises from algebraic geometry, the base space M will usually lie as a Zariski open subset in some larger space $\bar{M}$. The central technical problem of this paper is the question of the behavior of the bundles $\mathbf{F}^p$ near the subvariety $\bar{M}-M$. The problem becomes more manageable if one represents the bundles $\mathbf{F}^p$ as the pullback to M of certain universal bundles over a classifying space for Hodge structures, as Griffiths [10] has done by means of his period mapping. The behavior of the bundles $\mathbf{F}^p$ near $\bar{M}-M$ can then be described in terms of the singularities of the period mapping.

§ 3. Classifying Spaces for Hodge Structures

I shall briefly recall Griffiths' construction of the classifying spaces for polarized Hodge structures and of his period mapping. For this purpose, I consider a finite dimensional real vector space $H_{\mathbb{R}}$ with complexification $H_{\mathbb{C}}$, and containing a lattice $H_{\mathbb{Z}} \subset H_{\mathbb{R}}$. Also fixed throughout the discussion will be an integer k and a collection of non-negative integers $\{h^{p,q}\}$ which satisfy $h^{p,q}=h^{q,p}$, $h^{p,q} \neq 0$ only if $p+q=k$, $\sum h^{p,q}=\dim H_{\mathbb{R}}$. The first objective is to put a natural complex structure on the set of all Hodge structures of weight k on $H_{\mathbb{C}}$, having the integers $h^{p,q}$ as Hodge numbers. Let $\check{\mathscr{F}}$ be the set of all decreasing filtrations (2.6), such that $\dim F^p = \sum_{i \geqq p} h^{i,k-i}$. In the obvious manner, $\check{\mathscr{F}}$ forms a subvariety of a product of Grassman manifolds. As such it inherits the structure of complex projective variety. The general linear group of $H_{\mathbb{C}}$ operates transitively and holomorphically on $\check{\mathscr{F}}$, so that $\check{\mathscr{F}}$ is in fact a nonsingular complex projective variety. Those filtrations which satisfy (2.8) form an open (in the Hausdorff topology) subset $\mathscr{F} \subset \check{\mathscr{F}}$. Via the correspondence (2.7) between Hodge filtrations and weighted Hodge structures, the complex manifold $\mathscr{F}$ parameterizes exactly the Hodge structures of weight k on $H_{\mathbb{C}}$, which have the $h^{p,q}$ as Hodge numbers.

Now let S be a nondegenerate bilinear form on $H_{\mathbb{C}}$, symmetric or skew depending on whether k is even or odd, and defined over $\mathbb{Q}$, relative to the lattice $H_{\mathbb{Z}}$. I shall denote by $\check{D}$ the subset of all those filtrations in $\check{\mathscr{F}}$ which satisfy the first of the two conditions in (2.11); then $\check{D} \subset \check{\mathscr{F}}$ is a subvariety. The orthogonal group of the bilinear form S is a linear algebraic group, defined over $\mathbb{Q}$. The group of its $\mathbb{C}$-rational points

$$G_{\mathbb{C}} = \{g \in Gl(H_{\mathbb{C}}) \mid S(g\,u, g\,v) = S(u,v) \text{ for all } u, v \in H_{\mathbb{C}}\} \tag{3.1}$$

acts on $\check{\mathscr{F}}$ and $\check{D}$; this action will simply be denoted by juxtaposition. As can be checked by elementary arguments [10],

(3.2) $G_{\mathbb{C}}$ operates transitively on $\check{D}$.

In particular, $\check{D} \subset \check{\mathscr{F}}$ must be a nonsingular subvariety. Those filtrations in $\check{D}$ which also obey the second condition in (2.11) (equivalently, the second condition in (2.10)) automatically satisfy (2.8), and are therefore

Hodge filtrations. The subset $D \subset \check{D}$ of points corresponding to such filtrations is open in the Hausdorff topology of $\check{D}$. Thus D inherits the structure of complex manifold. By its very definition, D parameterizes all Hodge structures of weight k, which are polarized by S and have the $h^{p,q}$ as Hodge numbers.

The group of real points in $G_{\mathbb{C}}$, namely

(3.3) $$G_{\mathbb{R}} = \{g \in Gl(H_{\mathbb{R}}) \mid S(gu, gv) = S(u, v) \text{ for all } u, v \in H_{\mathbb{R}}\},$$

acts as a group of automorphisms on D. Again one can use simple arguments in linear algebra, to conclude:

(3.4) $G_{\mathbb{R}}$ acts transitively on D.

In order to exhibit $\check{D}$ and D as quotients of $G_{\mathbb{C}}$ and $G_{\mathbb{R}}$, I choose a particular Hodge structure $\{H_0^{p,q}\}$, corresponding to a point $o \in D$; this point will be called the *reference point* or *base point*. Let $\{F_0^p\}$, be the Hodge filtration determined by the reference Hodge structure $\{H_0^{p,q}\}$. A linear transformation $g \in G_{\mathbb{C}}$ keeps the base point o fixed precisely when $g F_0^p = F_0^p$ for all p. This gives the identification

(3.5) $$\check{D} \cong G_{\mathbb{C}}/B, \quad \text{where } B = \{g \in G_{\mathbb{C}} \mid g F_0^p = F_0^p \text{ for all } p\};$$

under this identification, the identity coset and the base point correspond to each other. As a quotient of a complex Lie group by a closed complex Lie subgroup, $G_{\mathbb{C}}/B$ has the structure of complex manifold; evidently (3.5) describes a complex analytic isomorphism.

In view of (3.4), one obtains an analogous identification

(3.6) $$D \cong G_{\mathbb{R}}/V, \quad \text{with } V = G_{\mathbb{R}} \cap B.$$

The embedding $D \subset \check{D}$ then corresponds to the inclusion $G_{\mathbb{R}}/V = G_{\mathbb{R}}/G_{\mathbb{R}} \cap B \subset G_{\mathbb{C}}/B$. As follows from (2.9), each $g \in V$ preserves not only the reference filtration $\{F_0^p\}$, but also the individual subspaces $H_0^{p,q}$, as well as the Weil operator C_0 of the Hodge structure $\{H_0^{p,q}\}$. Hence V leaves invariant a positive definite Hermitian form (cf. (2.11)). Moreover, as the intersection of closed subgroups of $Gl(H_{\mathbb{C}})$, V is also closed, so that

(3.7) V is compact.

The bilinear form S was assumed to take rational values on the lattice $H_{\mathbb{Z}}$. In particular, then,

(3.8) $$G_{\mathbb{Z}} = \{g \in G_{\mathbb{R}} \mid g H_{\mathbb{Z}} = H_{\mathbb{Z}}\}$$

lies in $G_{\mathbb{R}}$ as an arithmetic subgroup. Since $G_{\mathbb{R}}$ operates on D with compact isotropy group, and since $G_{\mathbb{Z}}$ is discrete in $G_{\mathbb{R}}$, the action of $G_{\mathbb{Z}}$

on D must be properly discontinuous. Hence, for any subgroup $\Gamma \subset G_{\mathbb{Z}}$, the quotient of the complex structure of D by Γ turns $\Gamma \backslash D$ into a complex analytic variety. The analogous statement about classifying spaces for weighted Hodge structures without polarization fails, and this is a major reason for considering polarized Hodge structures.

A most important feature of the spaces $\check{D}$ and D is the existence of a distinguished, group invariant tangent subbundle. In order to define it, it is necessary to mention some properties of the Lie algebras of $G_{\mathbb{C}}$ and $G_{\mathbb{R}}$; these will also be of use independently. The Lie algebra $\mathfrak{g}$ of the complex Lie group $G_{\mathbb{C}}$ can be described by the infinitesimal version of (3.1):

(3.9) $$\mathfrak{g} = \{X \in \operatorname{End}(H_{\mathbb{C}}) \mid S(Xu, v) + S(u, Xv) = 0 \text{ for all } u, v \in H_{\mathbb{C}}\}.$$

It is a simple complex Lie algebra, which contains

(3.10) $$\mathfrak{g}_0 = \{X \in \mathfrak{g} \mid X H_{\mathbb{R}} \subset H_{\mathbb{R}}\}$$

as a real form; i.e. $\mathfrak{g}_0$ is a real subalgebra such that $\mathfrak{g} = \mathfrak{g}_0 \oplus i\,\mathfrak{g}_0$. Via the containment $G_{\mathbb{R}} \subset G_{\mathbb{C}}$, $\mathfrak{g}_0$ becomes the Lie algebra of $G_{\mathbb{R}}$. The reference Hodge structure $\{H_0^{p,q}\}$ of $H_{\mathbb{C}}$ induces a Hodge structure of weight zero on

$$\operatorname{End}(H_{\mathbb{C}}) = \operatorname{Hom}(H_{\mathbb{C}}, H_{\mathbb{C}})$$

(cf. (2.12)). It can be checked that the – rationally defined – subspace $\mathfrak{g} \subset \operatorname{End}(H_{\mathbb{C}})$ carries a sub-Hodge structure. Hence

(3.11) $$\mathfrak{g} = \oplus_p \mathfrak{g}^{p,-p} \quad \text{is a Hodge structure of weight zero,}$$
$$\text{with } \mathfrak{g}^{p,-p} = \{X \in \mathfrak{g} \mid X H^{r,s} \subset H^{r+p,s-p} \text{ for all } r, s\}.$$

As a consequence of the naturality of the definition,

(3.12) $$[\ ,\ ]\colon \mathfrak{g} \otimes \mathfrak{g} \to \mathfrak{g} \quad \text{is a morphism of Hodge structures;}$$
$$\text{i.e., } [\mathfrak{g}^{p,-p}, \mathfrak{g}^{q,-q}] \subset \mathfrak{g}^{p+q,-(p+q)}.$$

The Lie algebra $\mathfrak{b}$ of B consists of all those $X \in \mathfrak{g}$ which preserve the reference Hodge filtration $\{F_0^p\}$; equivalently,

(3.13) $$\mathfrak{b} = \oplus_{p \geq 0}\, \mathfrak{g}^{p,-p}.$$

Let $\mathfrak{v}_0$ be the Lie algebra of $V = G \cap B$. Then

(3.14) $$\mathfrak{v}_0 = \mathfrak{g}_0 \cap \mathfrak{b} = \mathfrak{g}_0 \cap \mathfrak{b} \cap \bar{\mathfrak{b}} = \mathfrak{g}_0 \cap \mathfrak{g}^{0,0}.$$

As a final observation in this context, one should notice that

(3.15) $$\operatorname{Ad} g(\mathfrak{g}^{p,-p}) \subset \oplus_{i \geq p}\, \mathfrak{g}^{i,-i} \quad \text{for } g \in B.$$

The holomorphic tangent space of $\check{D} \cong G_{\mathbb{C}}/B$ at the base point eP is naturally isomorphic to $\mathfrak{g}/\mathfrak{b}$. Under this isomorphism, the action of the isotopy group B on the tangent space corresponds to the adjoint action of B on $\mathfrak{g}/\mathfrak{b}$. Consequently, the holomorphic tangent bundle $\mathbf{T} \to \check{D}$ coincides with the vector bundle associated to the holomorphic principal bundle

$$B \to G_{\mathbb{C}} \to G_{\mathbb{C}}/B \cong \check{D}$$

by the adjoint representation of B on $\mathfrak{g}/\mathfrak{b}$. Because of (3.15),

$$(\mathfrak{b} \oplus \mathfrak{g}^{-1,1})/\mathfrak{b} \subset \mathfrak{g}/\mathfrak{b}$$

defines an $\operatorname{Ad} B$-invariant subspace. By left translation via $G_{\mathbb{C}}$, it gives rise to a $G_{\mathbb{C}}$-invariant holomorphic subbundle of the holomorphic tangent bundle. It will be denoted by $\mathbf{T}_h(\check{D})$, and will be referred to as the holomorphic *horizontal tangent subbundle*. One can check that this construction does not depend on the particular choice of the base point $o \in D$; indeed, this is essentially the statement of Lemma 3.18 below. A holomorphic mapping Ψ: $M \to \check{D}$ of a complex manifold M into $\check{D}$ is said to be *horizontal* if at each point of M the induced map between the holomorphic tangent spaces takes values in the appropriate fibre of $\mathbf{T}_h(\check{D})$. The horizontal tangent subbundle, by restriction to D, determines a subbundle $\mathbf{T}_h(D)$ of the holomorphic tangent bundle $\mathbf{T}(D)$ of D. The $G_{\mathbb{C}}$-invariance of $\mathbf{T}_h(\check{D})$ implies the $G_{\mathbb{R}}$-invariance of $\mathbf{T}_h(D)$. The notion of a horizontal mapping into D is defined just as in the case of $\check{D}$; in other words, a mapping into D is horizontal precisely when it is horizontal, considered as a mapping into $\check{D}$.

It should be pointed out that the definition of a horizontal mapping which was given above is stricter than that in [13]. In fact, $\mathbf{T}_h(D)$ is a usually proper subbundle of the "horizontal distribution" of [13]. In order to see this, I define, in terms of the reference Hodge structure $\{H_0^{p,q}\}$,

$$H_{\text{even}} = \oplus_{p \text{ even}} H_0^{p,k-p}, \qquad H_{\text{odd}} = \oplus_{p \text{ odd}} H_0^{p,k-p},$$

so that $H_{\mathbb{C}} = H_{\text{even}} \oplus H_{\text{odd}}$. When k is even, these two subspaces are defined over $\mathbb{R}$ and orthogonal, relative to S. For odd k, they are mutually conjugate and coincide with their own annihilators. In both cases, the Hermitian form $i^k S(u, \bar{v})$ is positive definite on one of the two subspaces, negative definite on the other. As can be checked directly, and as will also be argued in some detail at the beginning of § 8 below, this makes

$$K = \{g \in G_{\mathbb{R}} \mid g\, H_{\text{even}} = H_{\text{even}}\}$$

a maximal compact subgroup of $G_{\mathbb{R}}$. Clearly K contains the isotropy group $V \subset G_{\mathbb{R}}$, and has

$$\mathfrak{k}_0 = \{X \in \mathfrak{g}_0 \mid X H_{\text{even}} \subset H_{\text{even}}\}$$
$$= \mathfrak{g}_0 \cap \oplus_{p \text{ even}} \mathfrak{g}^{p,-p}$$

as its Lie algebra. The adjoint action of K preserves

$$\mathfrak{p}_0 = \mathfrak{g}_0 \cap \oplus_{p \text{ odd}} \mathfrak{g}^{p,-p},$$

so that $\mathfrak{g}_0 = \mathfrak{k}_0 \oplus \mathfrak{p}_0$ turns out to be a Cartan decomposition [16]. Let $\mathfrak{p}$ be the complexification of $\mathfrak{p}_0$. Under the natural identification of $\mathfrak{g}/\mathfrak{b}$ with the holomorphic tangent space to $D \cong G_{\mathbb{R}}/V$ at the identity coset, the fibre of the "horizontal distribution" of [13] corresponds to $\mathfrak{b} \oplus \mathfrak{p}/\mathfrak{b}$, whereas the fibre of the horizontal tangent subbundle corresponds to $\mathfrak{b} \oplus \mathfrak{g}^{-1,1}/\mathfrak{b}$. Since $\mathfrak{p}$ contains $\mathfrak{g}^{-1,1}$, and because of the homogeneity of both bundles, the horizontal tangent subbundle lies inside the "horizontal distribution".

In view of this containement, Theorem 9.1 of [13] gives the following lemma. A more self-contained argument can be found in [5].

(3.16) **Lemma.** *There exists a $G_{\mathbb{R}}$-invariant Hermitian metric on D whose holomorphic sectional curvatures in the directions of $\mathbf{T}_h(D)$ are negative and bounded away from zero.*

Since $G_{\mathbb{R}}$ operates transitively, any two $G_{\mathbb{R}}$-invariant metrics on D are mutually bounded. Thus the previous lemma, together with standard arguments in hyperbolic complex analysis (see [18] or [26]), implies a very crucial property of horizontal holomorphic mappings:

(3.17) **Corollary** (cf. (9.3) in [13]). *For any horizontal holomorphic mapping of the complex upper half plane into D, the induced mapping between the holomorphic tangent spaces is uniformly bounded, relative to the Poincaré metric on the upper half plane and any given $G_{\mathbb{R}}$-invariant Hermitian metric for D. The value of the bound depends only on the normalization of the metrics.*

It will be useful to have an alternate description of the horizontal tangent subbundle and the notion of a horizontal mapping. Since $G_{\mathbb{C}}$ operates on $\check{D}$, its Lie algebra $\mathfrak{g}$ may be viewed as a Lie algebra of holomorphic vector fields, via infinitesimal translation. The transitivity of $G_{\mathbb{C}}$ ensures that $\mathfrak{g}$, by evaluation, maps onto the holomorphic tangent space at each point of $\check{D}$. At the base point, the evaluation mapping has kernel $\mathfrak{b}$. Thus one again obtains the familiar isomorphism between $\mathfrak{g}/\mathfrak{b}$ and the holomorphic tangent space at the base point.

(3.18) **Lemma.** *At a point $c \in \check{D}$, corresponding to the filtration $\{F^p(c)\}$, a vector field $X \in \mathfrak{g}$ takes its value in the fibre of $\mathbf{T}_h(\check{D})$ if and only if X, regarded as an endomorphism of $H_{\mathbb{C}}$, maps $F^p(c)$ into $F^{p-1}(c)$, for each p.*

Proof. At the base point, this amounts to a reformulation of the definition of the horizontal tangent subbundle. Elsewhere it then follows, because all of the constructions and identifications which are involved are preserved by the action of $G_{\mathbb{C}}$.

By construction, $\check{D}$ carries a trivial complex vector bundle $\mathbf{H}_{\mathbb{C}}(\check{D})$, with fibre $H_{\mathbb{C}}$, and a filtered family of holomorphic subbundles

$$(3.19) \qquad \mathbf{H}_{\mathbb{C}}(D) \supset \cdots \supset \mathbf{F}^{p-1}(\check{D}) \supset \mathbf{F}^p(\check{D}) \supset \mathbf{F}^{p+1}(\check{D}) \supset \cdots \supset 0,$$

whose fibres over any point constitute the filtration of $H_{\mathbb{C}}$ which describes the point in question. Let $\mathbf{T}^*(\check{D})$ be the holomorphic cotangent bundle, and

$$\nabla\colon \mathcal{O}(\mathbf{H}_{\mathbb{C}}(\check{D})) \to \mathcal{O}(\mathbf{H}_{\mathbb{C}}(\check{D}) \otimes \mathbf{T}^*(\check{D}))$$

the flat connection. For every integer p, there is a natural quotient mapping

$$q^p\colon \mathbf{H}_{\mathbb{C}}(\check{D}) \to \mathbf{H}_{\mathbb{C}}(\check{D})/\mathbf{F}^p(\check{D}).$$

As one can check readily, the composition

$$q^p \circ \nabla\colon \mathcal{O}(\mathbf{F}^p(\check{D})) \to \mathcal{O}(\mathbf{H}_{\mathbb{C}}(\check{D})/\mathbf{F}^p(\check{D}) \otimes \mathbf{T}^*(\check{D}))$$

is linear over $\mathcal{O}$, the sheaf of germs of holomorphic functions. Hence $q^p \circ \nabla$ defines a bundle map

$$(3.20) \qquad \sigma^p\colon \mathbf{F}^p(\check{D}) \to \mathbf{H}_{\mathbb{C}}(\check{D})/\mathbf{F}^p(\check{D}) \otimes \mathbf{T}^*(\check{D}),$$

the *second fundamental form* of $\mathbf{F}^p(D)$ in $\mathbf{H}_{\mathbb{C}}(\check{D})$. If X is a holomorphic vector field, one can compose ∇ with the operation of contraction with X; notation: $\nabla(X)$. Similarly, I shall write $\sigma^p(X)$ for $q^p \circ \nabla(X)$. Then $\sigma^p(X)$ is a bundle map from $\mathbf{F}^p(\check{D})$ to $\mathbf{H}_{\mathbb{C}}(\check{D})/\mathbf{F}^p(\check{D})$. An element $X \in \mathfrak{g}$ may be regarded either as a holomorphic vector field on $\check{D}$, or as an endomorphism of $H_{\mathbb{C}}$, and thus as a bundle map from $\mathbf{H}_{\mathbb{C}}(\check{D})$ to itself.

(3.21) **Lemma.** *For every $X \in \mathfrak{g}$, regarded as holomorphic vector field, $\sigma^p(X)$ is equal to the composition $q^p \circ X$; in this latter expression, X is considered as a bundle map.*

Proof. Let f be a local holomorphic section of $\mathbf{F}^p(\check{D})$ over an open set $U \subset \check{D}$. Equivalently, f can be viewed as a holomorphic $H_{\mathbb{C}}$-valued function on U, whose value at any point $a \in U$ lies in the fibre of $\mathbf{F}^p(\check{D})$ over a. For $a \in U$, $X \in \mathfrak{g}$, $t \in \mathbb{R}$, the $\exp(-tX)$-translate of the fibre of $\mathbf{F}^p(\check{D})$ over $\exp(tX)a$ coincides with the fibre over a. Hence $\exp(-tX) f((\exp tX)a)$ takes values in the fibre over a, for all sufficiently small $t \in \mathbb{R}$. Differentiating with respect to t at $t=0$, one finds that q^p annihilates $(\nabla(X) f - X \circ f)(a)$, as the lemma asserts.

Next I shall consider a holomorphic mapping $\Psi: M \to \check{D}$ of a complex manifold M into $\check{D}$. The vector bundles (3.19) pull back to trivial complex vector bundle $\mathbf{H}_{\mathbb{C}} \to M$, with a family of holomorphic subbundles

$$\mathbf{H}_{\mathbb{C}} \supset \cdots \supset \mathbf{F}^{p-1} \supset \mathbf{F}^p \supset \mathbf{F}^{p+1} \supset \cdots \supset 0.$$

Just as in the case of $\check{D}$, one can define the flat connection ∇ on $\mathbf{H}_{\mathbb{C}}$ and the second fundamental form σ^p of $\mathbf{F}^p$ in $\mathbf{H}_{\mathbb{C}}$. This second fundamental form then becomes the pullback via Ψ of the second fundamental form of $\mathbf{F}^p(\check{D})$ in $\mathbf{H}_{\mathbb{C}}(\check{D})$. Hence, by combining (3.18) with (3.21), one obtains a criterion for horizontality of maps into $\check{D}$:

(3.22) **Corollary.** *A holomorphic map $\Psi: M \to D$ is horizontal if and only if $\nabla \mathcal{O}(\mathbf{F}^p) \subset \mathcal{O}(\mathbf{F}^{p-1} \otimes \mathbf{T}^*)$, for all p.*

I now consider a variation of Hodge structure $\{M, \mathbf{H}_{\mathbb{C}}, \mathbf{F}^p\}$. The case of primary interest, of course, is that of a variation of Hodge structure arising from the k-th primitive cohomology groups of the fibres of a family of polarized algebraic manifolds. The universal covering of the base space M will be denoted by $\tilde{M}$. As a flat vector bundle, $\mathbf{H}_{\mathbb{C}} \to M$ is associated to the principal bundle

$$\pi_1(M) \to \tilde{M} \to M$$

by a representation

$$\varphi: \pi_1(M) \to Gl(H_{\mathbb{C}}); \tag{3.23}$$

here $H_{\mathbb{C}}$ denotes the fibre of the canonically trivial bundle $\tilde{\mathbf{H}}_{\mathbb{C}} \to \tilde{M}$, which is obtained by pulling $\mathbf{H}_{\mathbb{C}}$ back to $\tilde{M}$. Thus $\mathbf{H}_{\mathbb{C}}$ can be realized as the quotient of $\tilde{M} \times H_{\mathbb{C}}$ by the product of the obvious action of $\pi_1(M)$ on $\tilde{M}$ and the action φ of $\pi_1(M)$ on $H_{\mathbb{C}}$. The objects $\mathbf{H}_{\mathbb{R}}$, $\mathbf{H}_{\mathbb{Z}}$, $\mathbf{S}$ correspond to a real form $H_{\mathbb{R}} \subset H_{\mathbb{C}}$, a lattice $H_{\mathbb{Z}}$ in $H_{\mathbb{R}}$, and a rationally defined bilinear form S on $H_{\mathbb{C}}$. The subbundles $\mathbf{F}^p \subset \mathbf{H}_{\mathbb{C}}$ pull back to subbundles $\tilde{\mathbf{F}}^p$ of the trivial vector bundle $\tilde{M} \times H_{\mathbb{C}}$. At each point of $\tilde{M}$, the fibres of the $\tilde{\mathbf{F}}^p$ constitute a Hodge filtration on the fixed vector space $H_{\mathbb{C}}$. Since the Hodge numbers $h^{p,k-p}$ are equal to the ranks of the vector bundles $\mathbf{F}^p/\mathbf{F}^{p+1}$, they must remain constant. Thus one is led to consider the classifying space D for weighted Hodge structures on $H_{\mathbb{C}}$, which are polarized by S and have the collection of Hodge numbers $\{h^{p,k-p}\}$. Each point of $\tilde{M}$ then determines such a Hodge structure; in this way, one obtains a mapping $\tilde{\Phi}: \tilde{M} \to D$. According to the definition of a variation of Hodge structure, together with (3.22), $\tilde{\Phi}$ is holomorphic and horizontal.

I recall the definition of the groups $G_{\mathbb{C}}$, $G_{\mathbb{R}}$, $G_{\mathbb{Z}}$ in (3.1), (3.3), (3.8). Because of the flatness of $\mathbf{H}_{\mathbb{Z}}$ and $\mathbf{S}$, the action (3.23) preserves both S and the lattice $H_{\mathbb{Z}}$, so that its image lies in $G_{\mathbb{Z}}$. The subgroup

(3.24) $$\Gamma=\varphi(\pi_1(M))\subset G_{\mathbb{Z}}$$

is called the *monodromy group* of the variation of Hodge structure. In the case of a variation of Hodge structure coming from a family of polarized algebraic manifolds, Γ represents the action of $\pi_1(M)$ on the cohomology of the fibres. By construction of $\tilde{\Phi}$, if two points of $\tilde{M}$ are related by some $\sigma\in\pi_1(M)$, the corresponding Hodge structures are related by $\varphi(\sigma)$. Explicitely,

(3.25) $$\tilde{\Phi}(\sigma\tilde{t})=\varphi(\sigma)\circ\tilde{\Phi}(\tilde{t}),\quad \textit{for } \tilde{t}\in\tilde{M},\ \sigma\in\pi_1(M).$$

Hence $\tilde{\Phi}$ drops to a mapping

(3.26) $$\Phi\colon\ M\to\Gamma\backslash D$$

of M into the analytic space $\Gamma\backslash D$. This is *Griffiths' period mapping* for the variation of Hodge structure in question.

A holomorphic mapping into the quotient of a complex manifold D by the action of a properly discountinuous group of automorphisms Γ is said to be *locally liftable* if its restriction to some neighborhood of any given point in the domain can be factored through the quotient map $D\to\Gamma\backslash D$. If this is the case, any two local liftings, provided they are defined on a common connected open set, are related by an element of Γ. As a direct consequence of its construction, the period mapping is locally liftable. Thus:

(3.27) **Theorem** (Griffiths). *The period mapping* (3.26) *is holomorphic, locally liftable, and the local liftings are horizontal.*

§ 4. The Nilpotent Orbit Theorem

Throughout this section, D will be a classifying space for Hodge structures, $G_{\mathbb{Z}}$ the arithmetic subgroup of the automorphism group defined by (3.8), and Γ a subgroup of $G_{\mathbb{Z}}$. I shall consider a holomorphic mapping Φ: $M\to\Gamma\backslash D$ of a complex manifold M into $\Gamma\backslash D$, which is locally liftable and has horizontal liftings. It will also be assumed that M is given as a Zariski open subset of a (reduced) analytic space. The subject of this section is a discussion of the possible singularities of Φ along the complement of M.

According to Hironaka, one can embed M as a Zariski open set into a *manifold*, which is a modification of the original ambient space. I shall suppose that this modification has already been made, so that M

lies as a Zariski open subset in a complex manifold $\bar{M}$. First, I consider the case when $\bar{M}-M$ has codimension at least two. Every point of $\bar{M}-M$ possesses a simply connected neighborhood $\mathscr{U}$ in $\bar{M}$. For dimension reasons, $\mathscr{U} \cap M$ will then also be simply connected. On a simply connected manifold, a locally liftable map can be lifted globally (cf. Lemma 9.6 of [13]; when Φ arises as the period mapping of a variation of Hodge structure, this follows directly from the construction). Hence, restricted to $\mathscr{U} \cap M$, Φ has a lifting $\tilde{\Phi}$: $\mathscr{U} \cap M \to D$. By elementary arguments involving the Kobayashi pseudometric [18], together with (3.15), it can then be shown that $\tilde{\Phi}$ extends continuously, and thus holomorphically, to all of $\mathscr{U}$. Since $\mathscr{U}$ was an arbitrary simply connected open subset of $\bar{M}$, this proves:

(4.1) **Proposition** (cf. 9.8 in [13]). *If the codimension of $\bar{M}-M$ in $\bar{M}$ is at least two, Φ has a holomorphic, locally liftable continuation to $\bar{M}$.*

In the case of codimension one, the matter of the singularities of Φ along $\bar{M}-M$ becomes considerably more complicated. Let Δ denote the open unit disc in $\mathbb{C}$, and Δ^* the punctured open disc, i.e. $\Delta^* = \Delta - \{0\}$. Again the results of Hironaka make it possible to simplify the situation: by modifying $\bar{M}$ along $\bar{M}-M$, it can be arranged that $\bar{M}-M$ has no singularities worse than normal crossings. Hence, if $\bar{M}$ is replaced by some other suitable ambient manifold, which will also be denoted by $\bar{M}$, every point of $\bar{M}$ lies in some polycylindrical coordinate neighborhood $\mathscr{U} \cong \Delta^k$, such that $\mathscr{U} \cap M \cong \Delta^{*l} \times \Delta^{k-l}$, with $0 \leqq l \leqq k$. Since the problem of describing the singularities is a local one, I then may as well assume that $M \cong \Delta^{*l} \times \Delta^{k-l}$, $\bar{M} = \Delta^k$. The simplest nontrivial case occurs for $k=l=1$; I shall discuss it separately, both because it suffices for most applications, and because it makes the eventual result more transparent.

I thus let Φ: $\Delta^* \to \Gamma \backslash D$ be holomorphic, locally liftable, with horizontal local liftings. Via the mapping

$$\tau: U \to \Delta^*, \quad \tau(z) = e^{2\pi i z}, \tag{4.2}$$

the upper half plane $U = \{z \in \mathbb{C} \mid \operatorname{Im} z > 0\}$ becomes the universal covering of Δ^*. Viewed as a transformation group of U, the fundamental group $\pi_1(\Delta^*)$ is generated by the translation $z \mapsto z+1$. Since $\Phi \circ \tau$ is locally liftable and U simply connected, there exists a global lifting $\tilde{\Phi}$: $U \to D$, making the diagram

$$\begin{array}{ccc} U & \xrightarrow{\tilde{\Phi}} & D \\ {\scriptstyle\tau}\downarrow & & \downarrow \\ \Delta^* & \xrightarrow{\Phi} & \Gamma \backslash D \end{array} \tag{4.3}$$

commutative. One can choose an element $\gamma \in \Gamma \subset G_{\mathbb{Z}}$, such that

(4.4) $$\tilde{\Phi}(z+1)=\gamma \circ \tilde{\Phi}(z), \quad \text{for all } z \in U.$$

I shall refer to γ as the *monodromy transformation* of Φ around the puncture of Δ^*. The definite article "the" should not deceive: there may be more than one $\gamma \in \Gamma$ with the property (4.4). If Φ is the period mapping of a variation of Hodge structure with base Δ^* (or more concretely, of a family of polarized algebraic manifolds parameterized by Δ^*), the image of the generator $z \mapsto z+1$ of $\pi_1(\Delta^*)$ under the representation φ of (3.23) is one possible choice for γ. However, in what follows, γ may be any element of $G_{\mathbb{Z}}$ for which (4.4) holds.

The next lemma, as well as its proof, is due to Borel:

(4.5) **Lemma** (Borel). *All eigenvalues of* γ *are roots of unity.*

Proof. Computed in the Poincaré metric $y^{-2}(dx^2+dy^2)$ on U, the points $i \cdot n$ and $i \cdot n+1$, $n \in \mathbb{N}$, have distance $1/n$. According to (3.17), if a $G_{\mathbb{R}}$-invariant Riemannian distance function d on D is suitably renormalized, $\tilde{\Phi}$ will not increase distances. When the identification $D \cong G_{\mathbb{R}}/V$ is made as in (3.6), each of the image points $\tilde{\Phi}(i \cdot n)$, $n \in \mathbb{N}$, has a representation as a coset $g_n V$, with $g_n \in G_{\mathbb{R}}$. Because of (4.4), $\tilde{\Phi}(i \cdot n+1)=\gamma \tilde{\Phi}(i \cdot n)$ corresponds to the coset $\gamma g_n V$. Thus

$$\begin{aligned} d(g_n^{-1} \gamma g_n V, e V) &= d(\gamma g_n V, g_n V) \\ &= d\big(\tilde{\Phi}(i \cdot n+1), \tilde{\Phi}(i \cdot n)\big) \leqq \frac{1}{n}. \end{aligned}$$

It follows that the conjugacy class of γ in $G_{\mathbb{R}}$ has a point of accumulation in the compact subgroup $V \subset G_{\mathbb{R}}$. This forces the eigenvalues of γ to have absolute value one. On the other hand, since γ lies in an arithmetically defined group of matrices, the eigenvalues are algebraic integers. Hence, by a theorem of Kronecker, the eigenvalues must be roots of 1.

In view of the lemma, if $\gamma=\gamma_s \gamma_u$ is the Jordan decomposition of γ into its semisimple and unipotent parts, γ_s has finite order. I let m be the least positive integer such that $\gamma_s^m=1$, and I define

(4.6) $$\begin{aligned} N &= \log \gamma_u = \textstyle\sum_{k \geqq 1} (-1)^{k+1} \frac{1}{k} (\gamma_u - 1)^k \\ &= \frac{1}{m} \log \gamma^m = \frac{1}{m} \textstyle\sum_{k \geqq 1} (-1)^{k+1} \frac{1}{k} (\gamma^m - 1)^k; \end{aligned}$$

then N is a nilpotent and rational element of the Lie algebra $\mathfrak{g}_0$ of $G_{\mathbb{R}}$. The complexification $\mathfrak{g}$ of $\mathfrak{g}_0$, it should be recalled, is the Lie algebra of the complex Lie group $G_{\mathbb{C}}$. Also, $G_{\mathbb{C}}$ acts on $\check{D}$, which contains D

as a $G_{\mathbb{R}}$-orbit, as was discussed in §3. Let exp: $\mathfrak{g} \to G_{\mathbb{C}}$ be the exponential mapping of $G_{\mathbb{C}}$. Then

$$(4.7)\qquad \tilde{\Psi}(z) = \exp(-m z N) \circ \tilde{\Phi}(m z)$$

describes a holomorphic map of U into $\check{D}$, which has the transformation property

$$(4.8)\qquad \tilde{\Psi}\left(z + \frac{1}{m}\right) = \exp(-m z N)\, \gamma_u^{-1} \gamma \circ \tilde{\Phi}(m z) = \gamma_s \circ \tilde{\Psi}(z).$$

This makes $\tilde{\Psi}$ invariant under the translation $z \mapsto z+1$, so that $\tilde{\Psi}$ drops to a map Ψ: $\Delta^* \to \check{D}$, with $\Psi(e^{2\pi i z}) = \tilde{\Psi}(z)$.

(4.9) **Nilpotent Orbit Theorem (One-Variable Version).** *The mapping* Ψ *can be continued holomorphically over the puncture of* Δ^*. *The point* $a = \Psi(0) \in \check{D}$ *is a fixed point of* γ_s. *For a suitable constant* $\alpha \geqq 0$, $\operatorname{Im} z > \alpha$ *implies* $\exp(zN) \circ a \in D$. *Perhaps after increasing the constant* α, *and for a suitable choice of* $\beta \geqq 0$, $\operatorname{Im} z > \alpha$ *also implies the inequality*

$$d\big(\exp(zN) \circ a, \tilde{\Phi}(z)\big) \leqq (\operatorname{Im} z)^{\beta} e^{-2\pi m^{-1} \operatorname{Im} z};$$

here d *denotes a* $G_{\mathbb{R}}$*-invariant Riemannian distance function on* D. *The mapping* $z \mapsto \exp(zN) \circ a$ *of* $\mathbb{C}$ *into* $\check{D}$ *is horizontal.*

In order to explain the meaning of this statement, I shall consider the *multiple valued* mapping

$$t \mapsto \tilde{\Phi}\left(\frac{1}{2\pi i} \log t\right)$$

of Δ^* into D, which is a lifting of Φ: $\Delta^* \to \Gamma \setminus D$. According to the theorem, near $t=0$, this map behaves asymptotically like the – also multiple valued – mapping

$$(4.10)\qquad t \mapsto \exp\left(\frac{1}{2\pi i} \log t N\right) \circ a.$$

Since N is nilpotent, the entries of $\exp\left(\frac{1}{2\pi i} \log t N\right)$ are polynomial functions of $\log t$. For these reasons, I shall call (4.10) the *principal part* of the singularity of Φ. From Φ, the principal part inherits the property of being horizontal.

The proof of the theorem is technical and will be postponed until §8. As Griffiths has pointed out to me, in the case of a period mapping coming from algebraic geometry, (4.9) can be deduced from the regularity of the Gauss-Manin connection. However, the general version of the theorem, for more than a single variable, does not seem to be a consequence of the regularity theorem. Conversely, the general nilpotent orbit

Theorem (4.12) implies the regularity of the Gauss-Manin connection. These matters will be taken up below the statement of (4.12), at the end of this section.

If the monodromy transformation γ in (4.9) happens to be of finite order, it coincides with γ_s, so that $N=0$. According to the theorem, the point a must then lie in D, Ψ takes values in D, and the mapping $t \mapsto \Phi(t^m)$ of Δ^* into $\Gamma \setminus D$ can be covered by a holomorphic map from Δ to D, namely Ψ. Thus $\Phi: \Delta^* \to \Gamma \setminus D$ extends holomorphically to Δ, but not necessarily as a locally liftable map (unless $m=1$, i.e. $\gamma=1$). Griffiths proved this result in [11], using the existence of a discrete subgroup of G with compact quotient.

(4.11) **Corollary** (Griffiths). *If the monodromy transformation γ of $\Phi: \Delta^* \to \Gamma \setminus D$ has finite order, then Φ continues holomorphically to Δ.*

In order to extend (4.9) beyond the case of a single variable, I now consider a holomorphic, locally liftable map with horizontal local liftings

$$\Phi: \Delta^{*l} \times \Delta^{k-l} \to \Gamma \setminus D,$$

where $l \geqq 1$, $k \geqq l$. By going to the universal covering $U^l \times \Delta^{k-l}$, one can lift Φ to a mapping

$$\tilde{\Phi}: U^l \times \Delta^{k-l} \to D.$$

Corresponding to each of the first l variables, I choose a monodromy transformation $\gamma_i \in \Gamma$, so that

$$\tilde{\Phi}(z_1, \ldots, z_i+1, \ldots, z_l, w_{l+1}, \ldots, w_k) = \gamma_i \circ \tilde{\Phi}(z_1, \ldots, z_l, w_{l+1}, \ldots, w_k)$$

holds identically in all the variables. I shall assume that the γ_i commute with each other. If Φ arises as a period mapping, one can take as γ_i the image of the generator of the fundamental group of the i-th copy of Δ^*, under the representation

$$\varphi: \pi_1(\Delta^{*l} \times \Delta^{k-l}) \to \Gamma$$

(cf. (3.23)); since the fundamental group of $\Delta^{*l} \times \Delta^{k-l}$ is Abelian, the γ_i certainly do commute in this situation. Let $\gamma_i = \gamma_{i,s} \gamma_{i,u}$ be the Jordan decomposition of γ_i. Applying Borel's result (4.5) to each of the first l variables separately, one finds that each $\gamma_{i,s}$ has finite order m_i, for some $m_i \in \mathbb{N}$. Let $N_i \in \mathfrak{g}_0$ be the logarithm of $\gamma_{i,u}$. Since the γ_i commute, the set composed of all $\gamma_{i,s}, \gamma_{i,u}, N_i$ is also commutative.

For the remainder of this section, I shall let (z) denote a typical l-tuple $(z_1, \ldots, z_l) \in U^l$, and (w) a typical $(k-l)$-tuple $(w_{l+1}, \ldots, w_k) \in \Delta^{k-l}$; for any $(z) \in U^l$, $(m\,z)$ will be shorthand for the l-tuple

$$(m_1 z_1, m_2 z_2, \ldots, m_l z_l).$$

Because of the commutativity of the N_i, and because $\exp(m_i N_i)$ equals $\gamma_i^{m_i}$, the mapping $\tilde{\Psi}: U^l \times \Delta^{k-l} \to \check{D}$, which is defined by

$$\tilde{\Psi}(z, w) = \exp\left(-\sum_{i=1}^{l} m_i z_i N_i\right) \circ \tilde{\Phi}(m z, w),$$

remains invariant under the translation $z_i \mapsto z_i + 1$, $1 \leqq i \leqq l$. It follows that $\tilde{\Psi}$ drops to a mapping

$$\Psi: \Delta^{*l} \times \Delta^{k-l} \to \check{D}.$$

(4.12) **Nilpotent Orbit Theorem.** *The map Ψ extends holomorphically to Δ^k. For $(w) \in \Delta^{k-l}$, the point*

$$a(w) = \Psi(0, w) \in \check{D}$$

is left fixed by $\gamma_{i,s}$, $1 \leqq i \leqq l$. For any given number η with $0 < \eta < 1$, there exist constants $\alpha, \beta \geqq 0$, such that under the restrictions

$$\operatorname{Im} z_i \geqq \alpha, \quad 1 \leqq i \leqq l, \quad \text{and} \quad |w_j| \leqq \eta, \quad l+1 \leqq j \leqq k,$$

the point $\exp\left(\sum_{i=1}^{l} z_i N_i\right) \circ a(w)$ lies in D and satisfies the inequality

$$d\left(\exp\left(\sum_{i=1}^{l} z_i N_i\right) \circ a(w), \tilde{\Phi}(z, w)\right) \\ \leqq \left(\prod_{i=1}^{l} \operatorname{Im} z_i\right)^{\beta} \sum_{i=1}^{l} \exp(-2\pi m_i^{-1} \operatorname{Im} z_i);$$

here d again denotes a $G_{\mathbb{R}}$-invariant Riemannian distance function on D. Finally, the mapping

$$(z, w) \mapsto \exp\left(\sum_{i=1}^{l} z_i N_i\right) \circ a(w)$$

is horizontal.

The proof of this statement, which evidently contains (4.9), will be found in §8.

It was mentioned already that the nilpotent orbit theorem, for period mappings arising from algebraic geometry, is closely related to the regularity of the Gauss-Manin connection. A proof of the one-variable version, using the regularity theorem, is given in §9a of [14]. In rough outline, the argument proceeds as follows. Let $\{\Delta^*, \mathbf{H}_{\mathbb{C}}, \mathbf{F}^p\}$ be a one-parameter variation of Hodge structure, coming from algebraic geometry, and localized near a singularity. When an algebraic section of $\mathbf{H}_{\mathbb{C}}$ is expressed in terms of a multiple-valued, flat frame over Δ^*, the coefficient functions grow at most like a negative power of the local parameter, on every angular sector; this is the regularity theorem. This estimate still holds if the algebraic section is twisted by

$$\exp\left(-\frac{1}{2\pi i} \log t N\right),$$

since $t \log t \to 0$ as $t \to 0$, on every angular sector. It follows that the Plücker coordinates of the single valued map Ψ can be made to have only poles; for this purpose, $\check{D}$ should be viewed as a subvariety of a product of Grassmannians, as described in §3. Since Ψ takes values in a projective variety, this can only happen if Ψ has a holomorphic extension over the puncture. The remaining assertions of (4.9) can now be deduced relatively easily.

For more than a single variable, the argument which was just sketched breaks down: it can only prove the existence of a meromorphic extension of the mapping Ψ. On the other hand, the nilpotent orbit theorem implies the regularity of the Gauss-Manin connection. In fact, it implies the following theorem of Griffiths, whose original proof is based on Nevanlinna theory and L^2 estimates, giving a more general result than stated below.

(4.13) **Theorem** (Griffiths). *Let* $\{M, \mathbf{H}_{\mathbb{C}}, \mathbf{F}^p\}$ *be a variation of Hodge structure with quasi-projective base* M. *Then:*

a) *The bundle* $\mathbf{H}_{\mathbb{C}}$ *carries a unique algebraic structure such that the flat connection* ∇ *becomes algebraic, and such that* ∇ *has regular singular points of infinity, relative to any smooth compactification of* M. *With respect to this structure, the subbundles* $\mathbf{F}^p \subset \mathbf{H}_{\mathbb{C}}$ *are algebraic.*

b) *If the variation of Hodge structure comes from the cohomology of the fibres of an algebraic family of polarized algebraic manifolds, these algebraic structures on* $\mathbf{H}_{\mathbb{C}}$ *and the subbundles* $\mathbf{F}^p$ *coincide with the intrinsic algebraic structures. In particular, the Gauss-Manin connection has regular singular points.*

Part a) is a straightforward consequence of (4.12) and the comparison theorems of GAGA [23]. The second part depends on a nontrivial fact, concerning the relationship between algebraic and de Rham cohomology, which occurs also in Griffiths' proof. Without it, but *assuming* the regularity of the Gauss-Manin connection, one finds trivially that the two algebraic structures agree.

Omitting this one major detail, I shall show how the nilpotent orbit theorem leads to (4.13). I may assume that M lies as a Zariski open set in a nonsingular projective variety $\bar{M}$, so that $\bar{M} - M$ is a divisor with at most normal crossings. Every point at infinity then has a coordinate neighborhood $\mathscr{U}$, with

$$\text{(4.14)} \qquad \mathscr{U} \cong \Delta^k, \qquad \mathscr{U} \cap M \cong \Delta^{*l} \times \Delta^{k-l}.$$

Restricting the variation of Hodge structure to such a neighborhood at infinity, I now use the notation of (4.12). As before, I identify the universal covering of $\mathscr{U} \cap M$ with $U^l \times \Delta^{k-l}$. By pulling back $\mathbf{H}_{\mathbb{C}}$ from

$\mathscr{U} \cap M$ to its universal covering, one obtains a canonically trivial bundle, whose fibre shall be denoted by $H_{\mathbb{C}}$. In a 1:1 manner, the holomorphic sections of $\mathbf{H}_{\mathbb{C}}$ over $\mathscr{U} \cap M$ correspond to the holomorphic, $H_{\mathbb{C}}$-valued functions on $U^l \times \Delta^{k-l}$, which have the transformation property

$$(4.15)\qquad \begin{aligned} &v(z_1, \ldots, z_i+1, z_{i+1}, \ldots, z_l, w_{l+1}, \ldots, w_k) \\ &\qquad = \gamma_i \circ v(z_1, \ldots, z_l, w_{l+1}, \ldots, w_k), \qquad 1 \leqq i \leqq l. \end{aligned}$$

For the moment, I assume that the monodromy transformations γ_i are unipotent. If one composes the functions v in (4.15) with $\exp(-\sum_{i=1}^{l} z_i N_i)$, they become invariant under $z_i \mapsto z_i+1$, and thus drop to holomorphic, $\mathbf{H}_{\mathbb{C}}$-valued functions on $\mathscr{U} \cap M$. This now gives an isomorphism of $\mathcal{O}$-modules

$$(4.16)\qquad \mathcal{O}_{\mathscr{U} \cap M}(\mathbf{H}_{\mathbb{C}}) \cong \mathcal{O}_{\mathscr{U} \cap M} \otimes_{\mathbb{C}} H_{\mathbb{C}},$$

which leads to a distinguished continuation of $\mathbf{H}_{\mathbb{C}}$ to a vector bundle over all of $\mathscr{U}$. It can be characterized uniquely as follows: Let s be a section of $\mathbf{H}_{\mathbb{C}}$ on $\mathscr{U} \cap M$, and $s_1, \ldots, s_N$ a multiple-valued, flat frame, so that $s = \sum f_j s_j$, with multiple-valued coefficient functions f_j. Then

(4.17) s extends holomorphically to $\mathscr{U}$ if and only if the f_j have at most logarithmic singularities.

A similar argument, applied to the subbundles $\mathbf{F}^p$, and using the fact that the mapping Ψ of (4.12) extends holomorphically to Δ^k, gives continuations of these bundles as well. Again, a statement analogous to (4.17) characterizes the continuations uniquely. If the monodromy transformations are not unipotent, the constructions above have to be modified slightly: the bundles no longer extend as vector bundles, but rather as coherent sheaves.

On overlapping coordinate neighborhoods at infinity, the description (4.17) of the extendable sections is consistent. Hence the local continuations of $\mathbf{H}_{\mathbb{C}}$ and its subbundles $\mathbf{F}^p$ fit together, as global coherent sheaves over all of $\bar{M}$. According to GAGA [23], these global coherent sheaves have unique algebraic structures, which induce algebraic structures on the bundles $\mathbf{H}_{\mathbb{C}}$ and $\mathbf{F}^p$. Again I consider a coordinate neighborhood at infinity, $\mathscr{U}$, as in (4.14), and a multiple-valued, flat frame $s_1, \ldots, s_N$ for $\mathbf{H}_{\mathbb{C}}$ over $\mathscr{U} \cap M$. Let $s = \sum f_j s_j$ be a holomorphic section of $\mathbf{H}_{\mathbb{C}}$ on $\mathscr{U} \cap M$, with multiple-valued coefficient functions f_j. As follows from (4.17),

(4.18) s is meromorphic along $\bar{M} - M$, relative to the algebraic structure of $\mathbf{H}_{\mathbb{C}}$, if and only if the f_j are bounded by some polynomial in $|t_1|^{-1}, \ldots, |t_l|^{-1}$, on the intersection of M with any given compact subset of $\mathscr{U}$

(t_i=local parameter on the i-th copy of Δ^*). On the one hand, (4.18) uniquely determines the algebraic structure of $\mathbf{H}_{\mathbb{C}}$; on the other hand, it is equivalent to the regularity of the flat connection along $\bar{M}-M$. This proves part a) of (4.13).

I now assume that the variation of Hodge structure comes from an algebraic family of polarized algebraic manifolds. The bundles $\mathbf{H}_{\mathbb{C}}, \mathbf{F}^p$ thus have an intrinsic algebraic structure. In order to be able to refer to the algebraic structure given by (4.13a), I shall call it the *extrinsic* structure. It must be shown that the two coincide. For this, it is enough that the two induced structures on $\mathbf{F}^p/\mathbf{F}^{p+1}$ agree, for all p: the extension class of the sequence

$$0 \to \mathbf{F}^{p+1} \to \mathbf{F}^p \to \mathbf{F}^p/\mathbf{F}^{p+1} \to 0$$

is algebraic, both extrinsicly and intrinsicly; hence one can identify the two structures on each $\mathbf{F}^p$, by induction on p, once the two structures are known to coincide on each of the quotients.

Let s be a holomorphic section of $\mathbf{F}^p/\mathbf{F}^{p+1}$, over a set of the form $\mathscr{U} \cap M$, with $\mathscr{U}$ as in (4.14). With respect to the extrinsic algebraic structure

(4.19) s is meromorphic along $\bar{M}-M$, if and only if $i^k \mathbf{S}(s,\bar{s})$ can be bounded by a polynomial in $|t_1|^{-1}, \ldots, |t_l|^{-1}$, on the intersection of M with any given compact subset of $\mathscr{U}$.

Here $i^k \mathbf{S}(s,\bar{s})$ denotes the value of s on the Hermitian form induced by the polarization. The statement (4.19) follows from (4.18), together with the flatness of the polarization and the definiteness of the induced Hermitian form on $\mathbf{F}^p/\mathbf{F}^{p+1}$. In order to finish the proof of (4.13), one only needs the "only if" part of (4.19), but for the *intrinsic* algebraic structure on $\mathbf{F}^p/\mathbf{F}^{p+1}$. An argument which establishes this fact was shown to me by Griffiths. Its inclusion, however, would lead too far afield.

§ 5. The SL_2-Orbit Theorem

In order to make the nilpotent orbit theorem useful in applications, it is necessary to have detailed information about the nilpotent orbits which can occur. For most purposes, the one variable case suffices; also, the case of several variables involves major additional difficulties. I shall therefore limit myself to the consideration of a single variable. I intend to take up the general situation, of more than one variable, in a future paper. The main statement of this section is the SL_2-orbit Theorem (5.13); its proof will be given in §9.

Let then a be a point of $\check{D}$, α a fixed positive constant, and $N \in \mathfrak{g}_0$ a nonzero nilpotent element, such that

(5.1) a) $z \mapsto \exp(zN) \circ a$ is a horizontal mapping;
b) $\exp(zN) \circ a \in D$ for $\operatorname{Im} z > \alpha$.

This is precisely the situation arising in (4.9). For $z = x + iy$, $\exp(zN) = \exp(xN)\exp(iyN)$; since the first factor on the right lies in $G_{\mathbb{R}}$, and since $G_{\mathbb{R}}$ leaves $D \subset \check{D}$ invariant, (5.1 b) becomes equivalent to

$$\exp(iyN) \circ a \in D \quad \text{if } y \in \mathbb{R},\ y > \alpha. \tag{5.2}$$

In order to motivate the statement (5.13) below, it may be helpful to look at the simplest possible case. If D happens to be the upper half plane U, then $SL(2,\mathbb{R})$ plays the role of $G_{\mathbb{R}}$, and the Riemann sphere that of $\check{D}$. There are exactly two conjugacy classes of nonzero nilpotent elements in the Lie algebra $\mathfrak{sl}(2,\mathbb{R})$. Only for choices of N in the conjugacy class of

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \in \mathfrak{sl}(2,\mathbb{R}) \tag{5.3}$$

does the mapping $\exp(zN) \circ a$ take values in D, for all large positive values of $\operatorname{Im} z$. Thus, under the hypotheses (5.1), if the mapping $z \mapsto \exp(zN) \circ a$ is composed with a suitable automorphism of the upper half plane, it will be of the form

$$z \mapsto a + z \quad (z \in \mathbb{C}). \tag{5.4}$$

In general, for an arbitrary classifying space for Hodge structures D and an arbitrary mapping of the type (5.1) into D, there exists an equivariantly and horizontally embedded copy of the upper half plane U in D, such that the mapping $z \mapsto \exp(zN) \circ a$ asymptotically approaches a mapping of the form (5.4) into this copy of U in D; moreover, N turns out to be the image of the element (5.3) under the homomorphism $\mathfrak{sl}(2,\mathbb{R}) \to \mathfrak{g}_0$, which corresponds to the equivariant embedding.

Before these statements can be made more precise, some preliminary remarks are needed. In the usual manner, I shall think of the Riemann sphere $\mathbb{P}^1$ as the one point compactification $\mathbb{C} \cup \{\infty\}$ of $\mathbb{C}$. The complex Lie group $SL(2,\mathbb{C})$ operates transitively on $\mathbb{P}^1$, with isotropy subgroup

$$L = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2,\mathbb{C}) \,\middle|\, a - d = i(b - c) \right\}$$

at the point $i \in \mathbb{P}^1$; this gives the identification $\mathbb{P}^1 \cong SL(2,\mathbb{C})/L$. The $SL(2,\mathbb{R})$-orbit of i is then the upper half plane $U \subset \mathbb{P}^1$. For future reference, I shall record the identity

$$z = \exp\left(-\tfrac{1}{2}\log(-iz)\,Y\right) \circ i, \quad \text{if } z \in U; \tag{5.5}$$

the branch of the function $\log(-iz)$ on U is to be chosen so that its value at $z=i$ is zero, and Y denotes the element

$$Y=\begin{pmatrix}-1 & 0\\ 0 & 1\end{pmatrix}\in\mathfrak{sl}(2,\mathbb{R}). \tag{5.6}$$

I recall the definition of the subspaces $\mathfrak{g}^{p,-p}\subset\mathfrak{g}$ and of the subgroups $B\subset G_{\mathbb{C}}$, $V\subset G_{\mathbb{R}}$. A homomorphism of complex Lie groups

$$\psi\colon SL(2,\mathbb{C})\to G_{\mathbb{C}}, \quad \text{with Image}\,\psi\not\subset B,\ \psi(L)\subset B, \tag{5.7}$$

determines a holomorphic, equivariant embedding

$$\tilde{\psi}\colon \mathbb{P}^1\to\check{D}, \quad \text{with } \tilde{\psi}(g\circ i)=\psi(g)\circ o, \quad \text{for } g\in SL(2,\mathbb{C}) \tag{5.8}$$

(o = base point in D). The condition

$$\psi(SL(2,\mathbb{R}))\subset G_{\mathbb{R}} \tag{5.9}$$

insures that $\tilde{\psi}(U)\subset D$. The embedding (5.8) is horizontal precisely when the infinitesimal homomorphism of Lie algebras

$$\psi_*\colon \mathfrak{sl}(2,\mathbb{C})\to\mathfrak{g}$$

satisfies the condition

$$\psi_*(\mathfrak{sl}(2,\mathbb{C}))\subset\mathfrak{b}\oplus\mathfrak{g}^{-1,1}.$$

Indeed, because of the equivariance, it suffices to check the horizontality at the single point $i\in\mathbb{P}^1$; and the induced mapping between the tangent spaces of $\mathbb{P}^1$ at i and of $\check{D}$ at o corresponds to

$$\psi_*\colon \mathfrak{sl}(2,\mathbb{C})/\mathfrak{l}\to\mathfrak{g}/\mathfrak{b}$$

($\mathfrak{l}$ = Lie algebra of L).

The elements

$$Z=\begin{pmatrix}0 & -i\\ i & 0\end{pmatrix},\quad X_+=\tfrac{1}{2}\begin{pmatrix}-i & 1\\ 1 & i\end{pmatrix},\quad X_-=\tfrac{1}{2}\begin{pmatrix}i & 1\\ 1 & -i\end{pmatrix} \tag{5.10}$$

span $\mathfrak{sl}(2,\mathbb{C})$. The subalgebra $\mathfrak{l}\subset\mathfrak{sl}(2,\mathbb{C})$ is spanned by Z and X_-. By defining

$$\begin{aligned}\mathfrak{sl}(2,\mathbb{C})^{-1,1}&=\mathbb{C}X_+\\ \mathfrak{sl}(2,\mathbb{C})^{0,0}&=\mathbb{C}Z\\ \mathfrak{sl}(2,\mathbb{C})^{1,-1}&=\mathbb{C}X_-,\end{aligned} \tag{5.11}$$

one obtains a Hodge structure on $\mathfrak{sl}(2,\mathbb{C})$, relative to the real form $\mathfrak{sl}(2,\mathbb{R})\subset\mathfrak{sl}(2,\mathbb{C})$, which plays the same role with respect to U and the reference point $i \in U$ as the Hodge structure (3.11) does with respect to D and the base point o. If ψ_* is a mapping of type (0, 0) i.e. if

(5.12) $$\psi_*(X_+) \in \mathfrak{g}^{-1,1}, \quad \psi_*(Z) \in \mathfrak{g}^{0,0}, \quad \psi_*(X_-) \in \mathfrak{g}^{1,-1},$$

the embedding (5.8) will certainly be horizontal. When D happens to be Hermitian symmetric, (5.12) together with (5.9) is equivalent to saying that $\tilde{\psi}$ embeds the upper half plane U totally geodesicly in D.

One should keep in mind that the subspaces $\mathfrak{g}^{p,-p} \subset \mathfrak{g}$ and the isotropy subgroups B and V depend on the choice of the base point o – or, equivalently, of the reference Hodge structure $\{H_0^{p,q}\}$. I can now state the main technical result of this section; for this purpose, I assume that the point $a \in \check{D}$, the nonzero nilpotent element $N \in \mathfrak{g}_0$, and the positive constant α satisfy the condition (5.2).

(5.13) **Theorem.** *It is possible to choose*

i) *a homomorphism of complex Lie groups* $\psi: SL(2, \mathbb{C}) \to G_{\mathbb{C}}$,

ii) *a holomorphic, horizontal, equivariant embedding* $\tilde{\psi}: \mathbb{P}^1 \to \check{D}$, *which is related to* ψ *by* (5.8),

iii) *and a holomorphic mapping* $z \mapsto g(z)$ *of a neighborhood* $\mathscr{W} \subset \mathbb{P}^1$ *of* ∞ *into the complex Lie group* $G_{\mathbb{C}}$,

with all the following properties:

a) $\exp(zN) \circ a = g(-iz)\tilde{\psi}(z)$ *for* $z \in \mathscr{W} - \{\infty\}$;

b) $\psi(SL(2, \mathbb{R})) \subset G_{\mathbb{R}}$, *and* $\tilde{\psi}(U) \subset D$;

c) ψ_* *is a mapping of type* (0, 0) (cf. (5.12));

d) $g(y) \in G_{\mathbb{R}}$ *for* $iy \in \mathscr{W} \cap i\mathbb{R}$;

e) $\operatorname{Ad} g(\infty)^{-1}(N)$ *is the image under* ψ_* *of* $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \in \mathfrak{sl}(2, \mathbb{R})$;

f) *for* $iy \in \mathscr{W} \cap i\mathbb{R}$, $y > 0$, *let* $h(y)$ *be defined by*

$$h(y) = g(y) \exp(-\tfrac{1}{2} \log y \, \psi_*(Y))$$

(cf. (5.6)); *then*

$$h(y)^{-1} \frac{d}{dy} h(y) \in (\mathfrak{g}^{1,-1} \oplus \mathfrak{g}^{-1,1}) \cap \mathfrak{g}_0;$$

g) *the linear transformation* $\psi_*(Y) \in \operatorname{Hom}(H_{\mathbb{C}}, H_{\mathbb{C}})$ *operates semi-simply, with integral eigenvalues; let*

$$g(z) = g(\infty)(1 + g_1 z^{-1} + g_2 z^{-2} + \cdots + g_k z^{-k} + \cdots)$$

$$g(z)^{-1} = (1 + f_1 z^{-1} + f_2 z^{-2} + \cdots + f_k z^{-k} + \cdots) g(\infty)^{-1}$$

be the power series expansions of $g(z)$ *and* $g(z)^{-1}$ *around* $z = \infty$; *then, for* $n \geq 1$, g_n *and* f_n *map the* l*-eigenspace of* $\psi_*(Y)$ *into the linear span of the eigenspaces corresponding to eigenvalues less than or equal to* $l + n - 1$.

If the base point $o \in D$ *is suitably chosen, it can be arranged, moreover, that*

h) $g(\infty)=1$, *and* N *is the* ψ_**-image of* $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$; *in this situation, with the notation of* g), *for* $n \geqq 1$, $(\operatorname{Ad} N)^{n+1} g_n = 0$ *and* $(\operatorname{ad} N)^{n+1} f_n = 0$.

According to a), if h) also holds, the two mappings $z \mapsto \exp(zN) \circ a$ and $z \mapsto \tilde{\psi}(z)$ agree to first order at $z=\infty$, and g) then gives more specific information about the degree of proximity. Again if h) holds, the two mappings have the same invariance property under the translation $z \mapsto z+1$: $\tilde{\psi}(z+1) = \exp N \circ \tilde{\psi}(z)$,

$$\exp((z+1)N) \circ a = \exp N \exp(zN) \circ a.$$

The conditions b) and c) assert that the embedding $\tilde{\psi}$ is compatible with various other structures which are present. In a number of applications it is important to know a specific lifting of the mapping

$$y \mapsto \exp(iyN) \circ a \in D \cong G_{\mathbb{R}}/V, \quad y > \alpha, \tag{5.14}$$

from $G_{\mathbb{R}}/V$ to $G_{\mathbb{R}}$ — primarily because the action of $G_{\mathbb{R}}$ on D lifts to the family of Hodge subspaces $\{H^{p,q}\}$ corresponding to the points of D, whereas the action of $G_{\mathbb{C}}$ on $\check{D}$ only lifts to the family of Hodge flags $\{F^p\}$. Since $\psi_*(Y) \in \mathfrak{g}_0$, and since $g(y)$ takes values in $G_{\mathbb{R}}$ for $y \in \mathbb{R}$, according to d), the mapping $y \mapsto h(y)$ goes into $G_{\mathbb{R}}$; in view of a) and of (5.5), it is indeed a lifting of the map (5.14) to $G_{\mathbb{R}}$. The differential equation stated in f) distinguishes this lifting, in a way which will become apparent in the proof of (5.13). The property g) embodies the information on $g(y)$, and hence also on $h(y)$, which is crucial to all applications.

The proof of Theorem (5.13) is somewhat involved and will be postponed until §9. I shall draw several conclusions from this theorem, together with Theorem (4.9), in §6 and §7. For the remainder of this section, I shall look at the behavior near infinity of the mappings $z \mapsto \exp(zN) \circ a$ and of the mappings $\tilde{\Phi}$ which occur in §4; in particular, I shall study how the behavior at infinity is related to the arithmetic subgroup $G_{\mathbb{Z}} \subset G_{\mathbb{R}}$.

In the situation of §4, the monodromy transformation γ lies in the arithmetic group $G_{\mathbb{Z}}$. It follows that N, which was defined as the logarithm of the unipotent part of γ, preserves the rational structure $H_{\mathbb{Q}} \subset H_{\mathbb{R}}$ ($H_{\mathbb{Q}} = \mathbb{Q} \otimes_{\mathbb{Z}} H_{\mathbb{Z}}$). Hence, from now on, I assume that

$$N \in \mathfrak{g}_0 \text{ is an endomorphism of } H_{\mathbb{Q}}. \tag{5.15}$$

One may be tempted to ask, if, in this situation, the homomorphism ψ: $SL(2, \mathbb{Q}) \to G_{\mathbb{C}}$ becomes a homomorphism of $\mathbb{Q}$-groups, relative to the standard $\mathbb{Q}$-structure of $SL(2, \mathbb{C})$. Even the simplest example – namely $D = U$, $G_{\mathbb{C}} = SL(2, \mathbb{C})$ – shows that this is asking for too much, as long as one insists on (5.13h). However, if this requirement is dropped, one can, in fact, choose the homomorphism ψ so that it becomes defined over $\mathbb{Q}$, as will be argued below. Let

(5.16) $$\mathfrak{c}_0 = \operatorname{Im}\{\operatorname{ad} N\colon \mathfrak{g}_0 \to \mathfrak{g}_0\} \cap \operatorname{Ker}\{\operatorname{ad} N\colon \mathfrak{g}_0 \to \mathfrak{g}_0\}.$$

One can verify directly that $\mathfrak{c}_0$ is a Lie subalgebra of $\mathfrak{g}_0$. As will be pointed out in the proof of (5.17), it is in fact a nilpotent subalgebra.

(5.17) **Lemma.** *There exists a morphism of $\mathbb{Q}$-groups $\psi_1\colon SL(2, \mathbb{C}) \to G_{\mathbb{C}}$, and an element $g_1 \in G_{\mathbb{R}}$, such that $\psi_1 = \operatorname{Ad} g_1 \circ \psi$, and $g_1 g(\infty)^{-1} \in \exp \mathfrak{c}_0$.*

Proof. It is necessary to refer to some results of Kostant [20]. Let K be a field of characteristic zero, $\mathfrak{b}$ a semisimple Lie algebra over K, and $X \in \mathfrak{b}$ a nonzero nilpotent element. Let $\mathfrak{c}$ be the intersection of the kernel and the image of $\operatorname{Ad} X\colon \mathfrak{b} \to \mathfrak{b}$. Then $\mathfrak{c}$ is a nilpotent subalgebra of $\mathfrak{b}$, so that one can define the exponential map from $\mathfrak{c}$ into the adjoint group of $\mathfrak{b}$. Also, $\exp \mathfrak{c}$ centralizes X. The set of K-homomorphisms

$$\rho\colon \mathfrak{sl}(2, K) \to \mathfrak{b}, \qquad \text{with } \rho\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = X,$$

is nonempty, and under Ad, $\exp \mathfrak{c}$ acts on this set transitively. In [20], these results are stated for $K = \mathbb{C}$; however, they remain valid for any field of characteristic zero, and the proofs carry over without modification. In the case at hand, I let

$$\mathfrak{g}_{\mathbb{Q}} = \{X \in \mathfrak{g}_0 \mid X H_{\mathbb{Q}} \subset H_{\mathbb{Q}}\}$$

play the role of $\mathfrak{b}$, and $\mathfrak{c}_0$ the role of $\mathfrak{c}$. Since the nondegenerate bilinear form S, which was used in the definition of $\mathfrak{g}_0$, assumes integral values on $H_{\mathbb{Z}}$, $\mathfrak{g}_{\mathbb{Q}}$ is a semisimple Lie algebra over $\mathbb{Q}$, with

$$\mathfrak{g}_0 = \mathfrak{g}_{\mathbb{Q}} \otimes_{\mathbb{Q}} \mathbb{R}.$$

Moreover, because of (5.15), $N \in \mathfrak{g}_{\mathbb{Q}}$. According to Kostant's results, there exists a homomorphism

$$\rho\colon \mathfrak{sl}(2, \mathbb{R}) \to \mathfrak{g}_0,$$

which is defined over $\mathbb{Q}$, such that ρ maps

(5.18) $$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \in \mathfrak{sl}(2, \mathbb{R})$$

onto N. It follows from standard results on algebraic groups that there exists a morphism of $\mathbb{Q}$-groups $\psi_1\colon SL(2, \mathbb{C}) \to G_{\mathbb{C}}$, lying above ρ. Both ψ_{1*} and $\operatorname{Ad} g(\infty) \circ \psi_*$ map the element (5.18) to N. Hence, by Kostant's conjugacy statement, they are related by $\mathrm{Ad}\, g_0$, for some $g_0 \in \exp \mathfrak{c}_0$. Then $g_1 = g_0\, g(\infty)$ has all the desired properties.

In the notation of Theorem (5.13), replacing the original base point o by g_1 o for some $g_1 \in G_{\mathbb{R}}$, has the effect of changing ψ to $\mathrm{Ad}\, g_1 \circ \psi$, $g(z)$ to $g(z)\, g_1^{-1}$, $g(\infty)^{-1}$ to $g_1\, g(\infty)^{-1}$, and $\mathfrak{g}^{p,-p}$ to $\mathrm{Ad}\, g_1(\mathfrak{g}^{p,-p})$. With g_1 as in (5.17), since $\exp \mathfrak{c}_0$ commutes with N, this gives

(5.19) **Corollary.** *In the statement of Theorem* (5.13), *if the base point* o *is suitably chosen,* ψ *will be a morphism of* $\mathbb{Q}$*-groups,* ψ_* *maps the element* (5.18) *to* N, *and* $g(\infty) \in \exp \mathfrak{c}_0$. *This choice of base point does not destroy any of the properties listed in* (5.13), *except for* (5.13 h).

For the remainder of this section, I assume that the base point o$\in D$ has been chosen in accordance with (5.19). Then ψ maps the diagonal subgroup of $SL(2, \mathbb{C})$ onto a 1-dimensional $\mathbb{Q}$-split torus in $G_{\mathbb{C}}$. One can thus choose a maximal $\mathbb{Q}$-split subtorus T of $G_{\mathbb{C}}$ which contains this 1-dimensional torus. I let $\mathfrak{t}$ denote the Lie algebra of T, and Π the set of nonzero roots of $(\mathfrak{g}, \mathfrak{t})$. Via exponentiation, Π can be identified with the set of nontrivial $\mathbb{Q}$-roots of $(G_{\mathbb{C}}, T)$. It is possible to select a system of positive roots $\Pi_+ \subset \Pi$, such that

(5.20) $$\sigma \in \Pi, \quad \langle \sigma, \psi_*(Y) \rangle < 0, \qquad \text{implies } \sigma \in \Pi_+ .$$

For $\sigma \in \Pi$, $\mathfrak{g}_0$ will designate the σ-rootspace. Then

(5.21) $$\mathfrak{r} = \oplus_{\sigma \in \Pi_+} \mathfrak{g}_\sigma$$

is a nilpotent subalgebra of $\mathfrak{g}$. Since N lies in the (-2)-eigenspace of $\psi_*(Y)$,

(5.22) $$\mathfrak{r} \text{ contains } N.$$

Let $R = \exp \mathfrak{r}$; then R is a unipotent $\mathbb{Q}$-subgroup of $G_{\mathbb{C}}$. The centralizer of T can be expressed as TM, where M is anisotropic over $\mathbb{Q}$. One knows that

(5.23) $$P = RTM$$

forms a minimal $\mathbb{Q}$-parabolic subgroup of $G_{\mathbb{C}}$, with unipotent radical R.

The isotropy subgroup V of $G_{\mathbb{R}}$ at the reference point o is compact. One can therefore choose a maximal compact subgroup K of $G_{\mathbb{R}}$ which contains V. Let $P_{\mathbb{R}}$ denote the group of real points of P; then

(5.24) $$G_{\mathbb{R}} = P_{\mathbb{R}} K .$$

In fact, this last statement holds for any maximal compact subgroup of $G_{\mathbb{R}}$ (cf. § 11 of [1]).

The following lemma could be derived directly from Theorem (5.13). However, it is more economical to deduce it from the proof of the theorem.

I shall do so at the end of §9. Independently, Deligne drew a similar conclusion from Theorem (5.13).

(5.25) **Lemma.** *There exist functions $r(y)$, $t(y)$, $m(y)$, $k(y)$, which take values in, respectively, $R \cap G_{\mathbb{R}}$, $T \cap G_{\mathbb{R}}$, $M \cap G_{\mathbb{R}}$, K, and which are defined on some ray $\{y \in \mathbb{R} \mid y > \beta\}$, with the following properties:*

a) *for $y > \beta$, the point $\exp(i\,y N) \circ a$ coincides with the translate of the reference point o by $r(y)\, t(y)\, m(y)\, k(y)$;*

b) *the functions $r(y)$, $\exp(\frac{1}{2}\log y\, \psi_*(Y))\, t(y)$, $m(y)$, and $k(y)$ are real analytic functions of the variable $y^{-\frac{1}{2}}$ at $y = \infty$;*

c) $r(\infty) \in \exp \mathfrak{c}_0$, $\lim_{y\to\infty} \exp(\frac{1}{2}\log y\, \psi_*(Y))\, t(y) = 1$, $m(\infty) = 1$, and $k(\infty) = 1$.

Now let Γ be a subgroup of $G_{\mathbb{Z}}$, and $\tilde{\Phi}: \Delta^* \to \Gamma \setminus D$ a locally liftable, holomorphic mapping with horizontal local liftings. As in (4.3), I choose a lifting $\tilde{\Phi}: U \to D$ and some $\gamma \in G_{\mathbb{Z}}$, such that (4.4) holds. The semisimple part γ_s of γ has finite order m. I denote the logarithm of the unipotent part γ_u by N; then N has the property (5.15). With $a \in \check{D}$ as in the statement of the Nilpotent Orbit Theorem (4.9), the hypotheses at the beginning of this section are met. Hence the group P, D, etc. can be defined as above.

(5.26) **Theorem.** *It is possible to select functions $r(x,y)$, $t(x,y)$, $m(x,y)$, $k(x,y)$, with values in, respectively, $R \cap G_{\mathbb{R}}$, $T \cap G_{\mathbb{R}}$, $M \cap G_{\mathbb{R}}$, K, which are defined and real analytic on a set of the form $\{(x,y) \in \mathbb{R}^2 \mid y > \beta\}$, such that:*

a) *for $y > \beta$, the point $\tilde{\Phi}(x + i\,y)$ coincides with the translate of o by $r(x,y)\, t(x,y)\, m(x,y)\, k(x,y)$;*

b) *as $y \to \infty$, the limits of $r(x,y)$, $\exp(\frac{1}{2}\log y\, \psi_*(Y))\, t(x,y)$, $m(x,y)$, and $k(x,y)$ exist uniformly in x;*

c) *in the case of $r(x,y)$, this limit is a continuous function of x, with values in $R \cap G_{\mathbb{R}}$;*

d) $\lim_{y\to\infty} \exp(\frac{1}{2}\log y\, \psi_*(Y))\, t(x,y) = 1$, $\lim_{y\to\infty} m(x,y) = 1$, $\lim_{y\to\infty} k(x,y) = 1$.

Proof. According to Theorem (4.9), for some $\varepsilon > 0$,

$$d(\exp((x+i\,y)\,N) \circ a, \tilde{\Phi}(x+i\,y)) = O(e^{-\varepsilon y}), \quad \text{uniformly in } x.$$

In the notation of (5.25),

$$\exp((x+i\,y)\,N) \circ a = \exp(xN)\, r(y)\, t(y)\, m(y)\, k(y) \circ o.$$

Since the distance function d is $G_{\mathbb{R}}$-invariant, the distance, relative to d, between o and

$$k(y)^{-1}\, m(y)^{-1}\, t(y)^{-1}\, r(y)^{-1} \exp(-xN) \circ \tilde{\Phi}(x+i\,y) \tag{5.27}$$

is bounded by a constant multiple of $e^{-\varepsilon y}$. Locally, any two Riemannian metrics are mutually bounded. Also, the principal bundle $V \to G_{\mathbb{R}} \to G_{\mathbb{R}}/V \cong D$ has local sections. Using these statements, one can find a real analytic, $G_{\mathbb{R}}$-valued function $g(x, y)$, defined for $y \gg 0$, such that the point (5.27) can be represented as the $g(x, y)$-translate of o, and such that

$$\|g(x, y) - 1\| = O(e^{-\varepsilon y}), \quad \text{uniformly in } x;$$

the double bars denote the norm as a linear transformation. Since K is compact, this also gives

$$\|\mathrm{Ad}\, k(y)(g(x, y)) - 1\| = O(e^{-\varepsilon y}),$$

again uniformly in x. Because

$$G_{\mathbb{R}} = (R \cap G_{\mathbb{R}})(T \cap G_{\mathbb{R}})(M \cap G_{\mathbb{R}})K,$$

the function $\mathrm{Ad}\, k(y)(g(x, y))$ can be expressed as a product

$$\mathrm{Ad}\, k(y)(g(x, y)) = r_1(x, y)\, t_1(x, y)\, m_1(x, y)\, k_1(x, y);$$

here the factors are real analytic functions, defined for $y \gg 0$, with values in the obvious groups, and they can be chosen so that they satisfy the estimates

$$\|r_1(x, y) - 1\| = O(e^{-\varepsilon y}), \quad \text{uniformly in } x, \tag{5.28}$$

and similarly for $t_1(x, y)$, $m_1(x, y)$, $k_1(x, y)$.

The groups T and M commute, and they both normalize R. Hence, if I define

$$r(x, y) = \exp(xN)\, r(y)\, t(y)\, m(y)\, r_1(x, y)\big(t(y)\, m(y)\big)^{-1},$$
$$t(x, y) = t(y)\, t_1(x, y), \qquad m(x, y) = m(y)\, m_1(x, y),$$
$$k(x, y) = k_1(x, y)\, k(y),$$

these functions are real analytic, they take values in the appropriate groups, and they satisfy the statement a). As $y \to \infty$, the matrix entries of $t(y)$ remain bounded by polynomials in y, and $m(y) \to 1$, all according to (5.25). Together with (5.28), this gives the statements about $r(x, y)$ which were still missing. The corresponding statement about the other functions follow similarly.

By a Siegel set in D, I shall mean a set of the form

$$\mathfrak{S} = \{r\,t\,m\,k \circ o \in D \mid r \in \omega_1,\ m \in \omega_2,\ k \in K,\ t \in T \cap G_{\mathbb{R}}, \text{ and } e^{\sigma}(t) > c \text{ for all } \sigma \in \Pi_+\}$$

where ω_1 is a fixed compact subset of $R \cap G_{\mathbb{R}}$, ω_2 a fixed compact subset of $M \cap G_{\mathbb{R}}$, and c a positive constant. Aside from ω_1, ω_2, and c, the definition depends on the rational structure of $G_{\mathbb{C}}$, the choice of the maximal $\mathbb{Q}$-split torus T, of the minimal $\mathbb{Q}$-parabolic subgroup $P \subset G_{\mathbb{C}}$, and the choice of the base point o. Roughly speaking, the Siegel sets are fundamental sets for the action of $G_{\mathbb{Z}}$ on D (cf. [1]; the apparent difference in the definition is explained, of course, by the left – rather than right – action of $G_{\mathbb{Z}}$ in the present context). Because of the property (5.20) of the system of positive roots Π_+, and because of (5.26), for each positive root σ, $e^{\sigma}(t(x, y))$ either tends to the value one or has a positive infinite limit as $y \to \infty$, uniformly in x. Combined with the other statements in (5.26), this shows: for any constant $C > 0$, there exists a Siegel set $\mathfrak{S}$ and a constant $\alpha > 0$, such that $\tilde{\Phi}(z) \in \mathfrak{S}$ if $|\mathrm{Re}\, z| \leqq C$, $\mathrm{Im}\, z > \alpha$. The interiors of Siegel sets exhaust all of D; hence any compact set can be enclosed in a Siegel set. This proves:

(5.29) **Corollary.** *For any given constants $C > 0$ and $\eta > 0$, there exists a Siegel set $\mathfrak{S}$ in D, with $\tilde{\Phi}(z) \in \mathfrak{S}$ whenever $|\mathrm{Re}\, z| \leqq C$, $\mathrm{Im}\, z \geqq \eta$.*

§ 6. Monodromy and the Weight Filtration

In this section, I shall draw some conclusions from the nilpotent orbit theorem and the SL_2-orbit theorem, concerning the local behavior of a variation of Hodge structure near a singularity. Since the primary case of interest is that of a geometric variation of Hodge structure, I shall phrase the various statements in terms of the cohomology of a family of polarized algebraic manifolds. All of these statements have analogues for an abstract variation of Hodge structure, which are more or less obvious. I shall explicitely state the results for the abstract setting only in the few cases that will be referred to later.

Let then $\pi\colon \mathscr{V} \to \Delta^*$ be a family of polarized algebraic manifolds, with the punctured disc Δ^* as parameter space. As was pointed out already in §2, when this situation arises in practice, it is usually possible to continue π to a mapping $\bar{\mathscr{V}} \to \Delta$, with $\mathscr{V}$ Zariski open in the complex manifold $\bar{\mathscr{V}}$, by inserting a possibly singular fibre over the origin in Δ. However, for most of the arguments and statements below, the existence of a central fibre turns out to be quite irrelevant.

Forgetting about the complex structure, one can think of $\mathscr{V} \to \Delta^*$ as a C^{∞} fibre bundle. Thus $\pi_1(\Delta^*)$ acts on the cohomology groups of a general fibre $V_t = \pi^{-1}(t)$. Let k be an integer between 0 and $2n$ ($n = \dim_{\mathbb{C}} V_t$), and $\gamma \in Gl(H^k(V_t, \mathbb{Q}))$ the action of a generator of $\pi_1(\Delta^*)$. Griffiths has called γ the *Picard-Lefschetz* or *monodromy transformation* of the family. As before, I let $\gamma = \gamma_s \gamma_u$ be the Jordan decomposition of γ into its semisimple and its unipotent part.

(6.1) **Monodromy Theorem** ([21]). *The eigenvalues of γ are m-th roots of unity, for a suitable positive integer m, so that $\gamma_s^m = 1$. Let l be the largest number of successive nonzero Hodge subspaces of $H^k(V_t, \mathbb{C})$. In* *other words, l is the largest integer such that, for some p, $H^{i,k-i}(V_t, \mathbb{C}) \neq 0$ if $p \leqq i < p+l$; in particular, $l \leqq \min(k, 2n-k)+1$. Then $(\gamma_u - 1)^l = 0$, and hence $(\gamma^m - 1)^l = 0$.*

The proof will actually put a slightly stronger bound on the index of unipotency l: the statement remains correct if l is replaced by the maximum of the integers $l_0, l_1, \ldots, l_k$, where now l_j denotes the largest number of successive nonzero subspaces $P^{j-i,i}(V_t, \mathbb{C})$ of the primitive part of the j-th cohomology group.

The original version of the monodromy theorem is Landman's [21]. In his survey paper [12; pp. 235–236, 294], Griffiths discusses various proofs of the theorem, and he sketches a conjectured outline of the argument which I shall give below. Implicit in this argument – indeed, already in Theorem (5.13) – is an affirmative answer to conjecture 8.4′ of [12]. It should perhaps be remarked that the proof does not give any information about the integer m, as do the geometric proofs of Landman [21] and others. This is the price one must pay for not using the existence of a central fibre, over the puncture of Δ^*. On the other hand, the estimate on the index of unipotency is a little stronger than in previous versions of the theorem.

The statement, as well as its proof, carries over immediately to an abstract variation of Hodge structure with base Δ^*. Indeed, let γ be the image of a generator of $\pi_1(\Delta^*)$ under the representation (3.23), and l the largest number of successive nonzero Hodge bundles $\mathbf{H}^{p,q}$. The semisimple part γ_s is again of finite order, and the unipotent part satisfies $(\gamma_u - 1)^l = 0$. Of course, unless $\mathbf{H}^{p,q} = 0$ for $p < 0$ and $q < 0$, the inequality $l \leqq k+1$ need not hold.

Proof of (6.1). First of all, I may replace the coefficient field $\mathbb{Q}$ by $\mathbb{C}$. Secondly, since the polarizing classes are assumed to be $\pi_1(\Delta^*)$-invariant, the decompositions (2.2) are preserved by the action of $\pi_1(\Delta^*)$; hence it suffices to consider, for each k, the action of a generator of $\pi_1(\Delta^*)$ on $P^k(V_t, \mathbb{C})$. If γ is assigned this new meaning, I must show that γ_s has finite order, and that $(\gamma_u - 1)^{l_k} = 0$; the integer l_k is defined below (6.1). I shall now consider the period mapping corresponding to the k-th primitive cohomology groups of the fibers. Let τ: $U \to \Delta^*$ be the universal covering (4.2), and $\tilde{\Phi}$: $U \to D$ the lifting of the period mapping, composed with τ, to D. According to (3.25), γ lies in $G_{\mathbb{Z}}$ and has the property (4.4). Thus (4.5), (4.9), and (5.13) all apply. In particular, γ_s is of finite order, by Borel's lemma. Let $N = \log \gamma_u$; because of (5.13e), N is conjugate to the image under ψ_* of the element (5.18) of $\mathfrak{sl}(2, \mathbb{R})$. In $\mathfrak{sl}(2, \mathbb{C})$, any two nonzero nilpotent elements are conjugate. Hence, under the adjoint action of $G_{\mathbb{C}}$ on $\mathfrak{g}$, N becomes conjugate to $\psi_*(X_+)$ (cf. (5.10)). Also, because of (5.13c), $\psi_*(X_+) \in \mathfrak{g}^{-1,1}$. Any element of

$\mathfrak{g}^{-1,1}$ maps the i-th subspace in the reference Hodge filtration into the $(i+1)$-st subspace, and is therefore nilpotent of index at most l_k. It follows that N, which is known to be conjugate to an element of $\mathfrak{g}^{-1,1}$, has also index of nilpotency at most l_k. Since $\gamma_u = \exp N$, this proves $(\gamma_u - 1)^{l_k} = 0$, as desired.

Deligne has used the Picard-Lefschetz transformation to put an additional structure on the cohomology groups of the fibres V_t, $t \in \Delta^*$. In order to make the construction more transparent, it may be helpful to start with a discussion of the representation theory of the Lie algebra $\mathfrak{sl}_2$. Let K be a field of characteristic zero, and $\mathfrak{s}$ the three dimensional Lie algebra over K, with generators Z, X_+, X_-, which satisfy the commutation relations

(6.2) $\quad [Z, X_+] = 2X_+, \quad [Z, X_-] = -2X_-, \quad [X_+, X_-] = Z.$

A proof of the following assertion can be found in [24], for example.

(6.3) **Fact.** *Every finite dimensional representation of $\mathfrak{s}$ is fully reducible. Next, let ψ: $\mathfrak{s} \to \mathrm{End}(V)$ be an irreducible representation of $\mathfrak{s}$ on an $(n+1)$-dimensional vector space V. Then $\psi(Z)$ acts semisimply, with eigenvalues $n, n-2, n-4, \ldots, -n$, each with multiplicity one. By $\psi(X_+)$, the l-eigenspace of $\psi(Z)$ gets mapped onto the $(l+2)$-eigenspace, except when $l = -n-2$. Similarly, for $l \neq n+2$, $\psi(X_-)$ maps the l-eigenspace onto the $(l-2)$-eigenspace.*

I shall consider a linear transformation N: $V \to V$ on a finite dimensional vector space over a field of characteristic zero, which satisfies $N^{k+1} = 0$, for a given positive integer k.

(6.4) **Lemma** (cf. [12], pp. 255–256). *There exists a unique filtration*

$$0 \subset W_0 \subset W_1 \subset \cdots \subset W_{2k-1} \subset W_{2k} = V,$$

such that $N(W_l) \subset W_{l-2}$, and such that

$$N^l\colon\ Gr_{k+l}(W_*) \to Gr_{k-l}(W_*)$$

is an isomorphism, for each $l \geqq 0$ $\big(Gr_l(W_*) = W_l / W_{l-1}\big)$. If $l \geqq k$, let $\mathscr{P}_l \subset Gr_l(W_*)$ be the kernel of

$$N^{l-k+1}\colon\ Gr_l(W_*) \to Gr_{2k-l-2}(W_*),$$

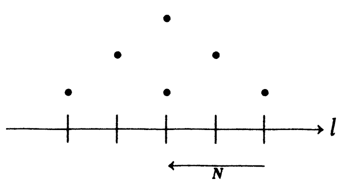
and set $\mathscr{P}_l = 0$ if $l < k$. Then one has the decomposition

$$(*) \qquad Gr_l(W_*) = \oplus_i N^i(\mathscr{P}_{l+2i}), \quad i \geqq \max(k-l, 0).$$

If N is an infinitesimal isometry of a nondegenerate symmetric or skew-symmetric form S on V, i.e. if $S(Nu, v) + S(u, Nv) = 0$ for all $u, v \in V$, the filtration becomes self-dual, in the sense that each W_l is the orthogonal *complement of W_{2k-l-1}. In this situation, moreover, the spaces $Gr_l(W_*)$ carry nondegenerate bilinear forms S_l, which are uniquely determined by the following requirements: if $l \geq k$, and if $u, v \in W_l$ represent $\tilde{u}, \tilde{v} \in Gr_l(W_*)$, $S_l(\tilde{u}, \tilde{v}) = S(u, N^{l-k} v)$; if $l < k$, N^{k-l} is to be an isometry from $Gr_{2k-l}(W_*)$ to $Gr_l(W_*)$. The decomposition (*) then becomes orthogonal with respect to S_l. Whenever S is symmetric and $k-l$ even, or S skew and $k-l$ odd, S_l is symmetric; S_l is skew in the remaining two cases. Finally, if ψ is a representation of the three dimensional Lie algebra $\mathfrak{s}$ on V, with $\psi(X_-) = N$, each W_l coincides with the linear span of the eigenspaces of $\psi(Z)$ which belong to eigenvalues less than or equal to $l-k$; $\mathscr{P}_l$ is the isomorphic image in $Gr_l(W_*)$ of the kernel of $\psi(X_-)^{l-k+1}$ on the $(l-k)$ eigenspace of $\psi(Z)$.*

Before giving an indication of the proof, I would like to make some observations about this statement. If ψ is a representation of $\mathfrak{s}$ on V, with $\psi(X_-) = N$, the last assertion of the lemma suggests how the filtration $\{W_l\}$ should be constructed. Such a representation always exists: according to the Jacobson-Morosov theorem, in a semisimple Lie algebra over a field of characteristic zero, every nonzero nilpotent element can be embedded in a copy of $\mathfrak{sl}_2$; applied to the Lie algebra $\mathfrak{sl}(V)$, or the Lie algebra of infinitesimal isometries of a bilinear form S, the theorem gives the existence of a representation ψ with the desired property. In order to deduce the lemma from the Jacobson-Morosov theorem and the representation theory of $\mathfrak{sl}_2$, one needs to know that the resulting filtration depends only on N, not on the particular representation ψ. This follows from a theorem of Kostant, which was quoted in the proof of (5.17). Thus (6.4) becomes a consequence of the representation theory of $\mathfrak{sl}_2$, the Jacobson-Morosov theorem, and the result of Kostant. However, since (6.4) is an elementary statement, I shall give the outline of an elementary proof.

In this connection with the theorems of Jacobson-Morosov and Kostant, a certain schematic diagram may help to clarify the statement (6.4). For definiteness, I shall assume that $k=2$, and thus $N^3=0$. Let ψ be a particular representation of $\mathfrak{s}$ on V, with $\psi(X_-) = N$, whose existence is guaranteed by the Jacobson-Morosov theorem. Since $N^3=0$, each $\mathfrak{s}$-irreducible subspace of V has dimension one, two, or three (cf. (6.3)). In the diagram below,



each row shall represent the span of all $\mathfrak{s}$-irreducible subspaces of dimension, respectively, one, two, and three, reading from top to bottom. Within a given row, a dot stands for the l-eigenspace of $\psi(Z)$. The operation of N preserves the rows and shifts l by -2. The subspaces corresponding to the various dots depend on the choice of ψ, of course. Certain combinations, however, have canonical meaning: the left edge represents the kernel of N, the right edge the cokernel, and all dots on, or to the left of, the vertical line through l give W_l.

On the cohomology groups of a compact Kähler manifold, the Kähler operator L, is adjoint Λ, and their commutator $B=[L,\Lambda]$ satisfy the relations

$$[B,L]=2L, \quad [B,\Lambda]=-2\Lambda, \quad [L,\Lambda]=B.$$

As Serre has pointed out, these three operators therefore span a Lie algebra isomorphic to $\mathfrak{s}$. Applied in this setting, the statement (6.3) proves the existence of the decomposition (2.2). In particular, (2.2) can be viewed as a special case – with reversed indices – of Lemma (6.4). This explains the formal analogy between the Kähler operator L and the logarithm of the unipotent part of the Picard-Lefschetz transformation, which was commented upon in [12].

Proof of (6.4). Proceeding inductively, I may assume that subspaces $W_i \subset V$ have already been found for $i \leqq l-1$ and $i \geqq 2k-l$, where l is an integer between 0 and k, starting with $W_{2k}=V$ and $W_{-1}=0$, subject to the following conditions:

a) $N(W_i) \subset W_{i-2}$ and $W_i \subset W_j$, whenever both sides of the containment are defined, and $i<j$;

b) $W_{k-i}=N^i W_{k+i}$ for $i \geqq k-l+1$;

c) $W_{k+i}=\{v \in V \mid N^{i+1} v \in W_{k-i-2}\}$ for $i \geqq k-l$;

d) if $i \geqq k-l$, W_{k+i} is spanned by $N(W_{k+i-2})$ and the kernel of N^{i+1};

e) $\ker N^{i+1} \cap N(W_{k+i+2}) \subset W_{k+i-1}$ for $i \geqq k-l+1$.

At the next step, the statements of the lemma require W_l and W_{2k-l-1} to be consistent with b) and c), which determines these subspaces completely. This description of the two subspaces, together with the inductive assumptions, gives a), d), and e) at the next stage. At the final step, when $l=k$, the containments $N(W_{k+1}) \subset W_{k-1}$ and $N(W_k) \subset W_{k-2}$ are automatic. Existence and uniqueness of the filtration now follow. Since N^{l-k} identifies $Gr_{2k-l}(W_*)$ with $Gr_l(W_*)$, the decomposition (*) only needs to be constructed for $l \geqq k$; in view of d) and e), this can be done.

If N is skew adjoint with respect to a nondegenerate bilinear form S, then the dual filtration $\{{}^tW_l\}$, with ${}^tW_l = W_{2k-l-1}^{\perp}$, also satisfies a)–c) whenever $0 \leqq l \leqq k$. Hence $\{{}^{\prime}W_l\}$ has all the properties which characterize the filtration $\{W_l\}$, and therefore the two filtrations coincide. In particular, the bilinear form S_l on $Gr_l(W_*)$, $l \geqq k$, is well defined. The various properties of S_l also follow from the self-dual nature of the filtration.

According to (6.3), if ψ is a representation of $\mathfrak{s}$ on V, $\psi(Z)$ must be diagonalizable, with integral eigenvalues. Also, $\psi(X_-)$ maps the l-eigenspace into the $(l-2)$-eigenspace, and for $l>0$, $\psi(X_-)^l$ maps the l-eigenspace isomorphically onto the $(-l)$-eigenspace. Hence, if $\psi(X_-)=N$, and if $\tilde{W}_l$ is defined as the span of all eigenspaces of $\psi(Z)$ corresponding to eigenvalues less than or equal to $l-k$, the filtration $\{\tilde{W}_l\}$ has all properties which define $\{W_l\}$, so that $W_l = \tilde{W}_l$. Again because of (6.3), the l-eigenspace of $\psi(Z)$ decomposes into the kernel of $\psi(X_-)^{l+1}$ and the image under $\psi(X_-)$ of the $(l+2)$-eigenspace, if $l \geqq 0$. This gives the last assertion of the lemma.

I shall return to the geometric situation of a family $\pi\colon \mathscr{V} \to \Delta^*$. According to (6.1), the unipotent part γ_u of the Picard-Lefschetz transformation

$$\gamma\colon H^k(V_t, \mathbb{Q}) \to H^k(V_t, \mathbb{Q}), \quad t \in \Delta^*,$$

is unipotent of index at most $k+1$, so that $N^{k+1}=0$, where

$$N = \log \gamma_n = \sum_{l=1}^{k} (-1)^{l+1} \frac{1}{l} (\gamma_u - 1)^l .$$

The filtration

$$0 \subset W_0 \subset W_1 \subset \cdots \subset W_{2k-1} \subset W_{2k} = H^k(V_t, \mathbb{Q}),$$

constructed from N in the manner described by Lemma (6.4), is Deligne's *monodromy weight filtration.* For reasons of notational economy, I shall refer to its complexification in $H^k(V_t, \mathbb{C})$ by the same letters. Relative to the flat structure on the vector bundle $\mathbf{H}^k_{\mathbb{C}} \to \Delta^*$ (cf. §2), the Picard-Lefschetz transformation γ, and therefore also γ_s, γ_u, and N, are flat. Hence the weight filtration is flat, and it can be transferred to a filtration

$$(6.5) \qquad 0 \subset W_0 \subset W_1 \subset \cdots \subset W_{2k-1} \subset W_{2k} = H^k_{\mathbb{C}},$$

where $H^k_{\mathbb{C}}$ denotes the fibre of the pullback of $\mathbf{H}^k_{\mathbb{C}} \to \Delta^*$ to the universal covering space of Δ^*. On the quotients $Gr_l(W)$, the actions of γ and γ_s coincide and are of finite order. Also, the weight filtration is compatible with the decomposition (2.2), because the polarization was assumed to be $\pi_1(\Delta^*)$-invariant.

As an additional assumption on the family $\pi\colon \mathscr{V} \to \Delta^*$, I shall require that the total space $\mathscr{V}$ carries a Kähler metric, corresponding to an integral cohomology class $\eta \in H^2(\mathscr{V}, \mathbb{C})$. The polarizing classes η_t are to be the restrictions of η to each V_t. The Kähler metric then puts a distinguished inner product on $H^k(V_t, \mathbb{C})$, for every $t \in \Delta^*$. I shall denote the resulting norm by $\| \ \|_t$. If $t_0 \in \Delta^*$ is fixed and $v \in H^k(V_{t_0}, \mathbb{C})$ some cohomology class, v determines a multiple-valued, flat section of $\mathbf{H}^k_{\mathbb{C}} \to \Delta^*$. Thus $t \mapsto \|v\|_t$ may be regarded as a multiple-valued function on Δ^*. On each radical ray $t = r e^{i\theta}$, $0 < r < 1$, or more generally on a proper angular sector, one can choose a single-valued, continuous branch of this function. Deligne first conjectured the following statement, and later independently deduced it from Theorem (4.9) and an earlier version of Theorem (5.13).

(6.6) **Theorem.** *Let $t_0 \in \Delta^*$ and $v \in H^k(V_{t_0}, \mathbb{C})$. Then v belongs to W_l if and only if, along the radial ray through t_0,*

$$\|v\|_t = O\big((-\log|t|)^{(l-k)/2}\big) \quad \text{as } |t| \to 0.$$

Moreover, this estimate holds uniformly on every angular sector.

Implicit in the theorem is the assertion that any two branches of the function $t \mapsto \|v\|_t$ on a common angular sector must be mutually bounded.

If $v \in H^k(V_{t_0}, \mathbb{C})$ happens to be an invariant cohomology class, i.e. $\gamma v = v$, $t \mapsto \|v\|_t$ becomes a single valued function on Δ^*. According to the theorem, this function is bounded near the puncture precisely when $v \in W_k$. On the other hand, for $l > k$, N^{l-k} determines an isomorphism between $Gr_l(W_*)$ and $Gr_{2k-l}(W_*)$; consequently, W_k contains the kernel of N, and thus also all invariant cohomology classes. Hence:

(6.7) **Corollary.** *An invariant cohomology class has bounded norm near the puncture of Δ^*.*

In Section 7, in order to prove certain global statements about the period mapping of an algebraic family, it will be helpful to have (6.7) not just in the geometric setting, but also for an abstract variation of Hodge structure. Let then $\{\mathbf{H}_{\mathbb{C}}, \mathbf{F}^p\}$ be a variation of Hodge structure of weight k, with base space Δ^*. The weight filtration, which can be defined as before, filters $\mathbf{H}_{\mathbb{C}}$ by a family of flat subbundles $\{\mathbf{W}_l\}$. However, unless $\mathbf{H}^{p,q} = 0$ for $p < 0$ and $q < 0$, the length of the weight filtration may exceed $2k$. The Weil operators of the Hodge structures corresponding to the various points $t \in \Delta^*$ fit together, to give a C^∞ bundle map $\mathbf{C}\colon \mathbf{H}^{p,q} \to \mathbf{H}^{p,q}$ with $\mathbf{C}$ equal to multiplication by i^{p-q} on $\mathbf{H}^{p,q}$. Since $\mathbf{S}$ polarizes the Hodge structures,

$$\mathbf{S}(\mathbf{C}u, \bar{v})$$

defines a positive definite Hermitian metric on the fibres of $\mathbf{H}_{\mathbb{C}}$. Let v be a, possibly multiple-valued, flat section of $\mathbf{H}_{\mathbb{C}}$. The proof of (6.6) below also proves the following abstract version:

(6.6′) **Theorem.** *The (multiple-valued) flat section v lies in $\mathbf{W}_l$ if and only if, along a radial ray,*

$$\mathbf{S}(\mathbf{C}\,v,\bar{v})=O\big((-\log|t|)^{l-k}\big) \quad \text{as } |t|\to 0.$$

Moreover, this estimate holds uniformly on every angular sector.

Just as (6.6) implies (6.7), (6.6′) leads to

(6.7′) **Corollary.** *For any single-valued, flat section v of $\mathbf{H}_{\mathbb{C}}$, the function $\mathbf{S}(\mathbf{C}\,v,\bar{v})$ is bounded near the puncture.*

Proof of (6.6). As was pointed out before, the weight filtration is compatible with the decomposition (2.2). Also, according to standard facts in Kähler geometry [25], the decomposition (2.2) is orthogonal, and the operators L^j in (2.2) are isometries. It therefore suffices to consider primitive cohomology classes. Let $\mathbf{P}^k_{\mathbb{C}}\to\Delta^*$ be the flat bundle whose fibres are the k-th primitive cohomology groups of V_t, $t\in\Delta^*$. The universal covering $\tau\colon U\to\Delta^*$, defined in (4.2), pulls $\mathbf{P}^k_{\mathbb{C}}$ back to a trivial bundle on U, whose fibre I shall denote by $P^k_{\mathbb{C}}$. As in § 3, D will refer to the classifying space for Hodge structures on $P^k_{\mathbb{C}}$, with the appropriate Hodge numbers. The period mapping, composed with τ, lifts to a map $\tilde{\Phi}\colon U\to D$. For each $z\in U$, I choose $m(z)\in G_{\mathbb{R}}$, such that the V-coset of $m(z)$ represents $\tilde{\Phi}(z)\in D\cong G_{\mathbb{R}}/V$. As follows from the basic Kähler identities,

$$\|v\|_t^2=i^{p-q}\,S(v,\bar{v}) \qquad \text{for all } v\in P^{p,q}(V_t,\mathbb{C}). \tag{6.8}$$

Hence, in order to compute $\|v\|_t^2$ for $t=\tau(z)$ and $v\in P^k$, I may split v into its Hodge components relative to the Hodge structure at $\tilde{\Phi}(z)$, compute the square lengths of these Hodge components by means of (6.8), and add them up. Since $G_{\mathbb{R}}$ preserves S and complex conjugation, I may equivalently take the Hodge components of $m(z)^{-1}$ relative to the reference Hodge structure, which corresponds to the base point $o\in D$, and sum up the squares of their lengths. Since τ maps vertical lines to radial rays, and vertical strips to sectors, this reduces the theorem to the following assertion:

(6.9) $v\in P^k_{\mathbb{C}}$ belongs to W_l if and only if

$$\|m(z)^{-1}v\|=O\big((\operatorname{Im} z)^{(l-k)/2}\big) \qquad \text{as } \operatorname{Im} z\to\infty,$$

and the estimate holds uniformly on vertical strips;

here $\|\ \|$ stands for an arbitrary V-invariant linear norm on $P^k_{\mathbb{C}}$.

Because of (4.9), for some $\varepsilon>0$, the $G_{\mathbb{R}}$-invariant distance between $\tilde{\Phi}(z)$ and $\exp(zN)\circ a$ is bounded by $e^{-\varepsilon \operatorname{Im} z}$ as $\operatorname{Im} z\to\infty$. If $z=x+iy$, in the notation of (5.13),

$$\exp(zN)\circ a=\exp(xN)\,g(y)\exp\big(-\tfrac{1}{2}\log y\,\psi_*(Y)\big)\circ o.$$

Since all of the factors on the right lie in $G_{\mathbb{R}}$, the distance between o and the point

$$(6.10) \qquad \exp\left(\tfrac{1}{2}\log y\, \psi_*(Y)\right) g(y)^{-1} \exp(-xN) \circ \tilde{\Phi}(z)$$

has order $e^{-\varepsilon \operatorname{Im} z}$, as $\operatorname{Im} z \to \infty$. One can therefore choose a $G_{\mathbb{R}}$-valued function $s(z)$, defined for $\operatorname{Im} z \gg 0$, such that the V-coset of $s(z)$ represents the point (6.10), and such that

$$(6.11) \qquad \|s(z)-1\| = O(e^{-\varepsilon \operatorname{Im} z}), \qquad \text{as } \operatorname{Im} z \to \infty.$$

Clearly, I can now arrange that

$$m(z)^{-1} = s(z)^{-1} \exp\left(\tfrac{1}{2}\log y\, \psi_*(Y)\right) g(y)^{-1} \exp(-xN).$$

Since $\exp(-xN)$ operates trivially on the successive quotients of the filtration, and in view of (6.11), this makes (6.9) equivalent to

(6.12) $v \in P^k_{\mathbb{C}}$ belongs to W_l if and only if, as $y \to \infty$,

$$\left\|\exp\left(\tfrac{1}{2}\log y\, \psi_*(Y)\right) g(y)^{-1} v\right\| = O(y^{(l-k)/2}).$$

In particular, the estimate in (6.9) will be uniform on vertical strips.

As in (5.13h), I assume $g(\infty)=1$. The elements

$$\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \in \mathfrak{sl}(2, \mathbb{R})$$

satisfy the same commutation relations as Z, X_+, X_- in (6.2); also ψ_* maps the first of these to $\psi_*(Y)$ and the last to N. Hence, according to the next to last assertion of (6.4),

(6.13) W_l is the span of all eigenspaces of $\psi_*(Y)$ which correspond to eigenvalues less than or equal to $l-k$.

I consider an element v of the $(l-k)$-eigenspace of $\psi_*(Y)$. Because of (6.13g) and the preceeding description of the filtration $\{W_l\}$, $g(y)^{-1}v$ has a power series expansion

$$g(y)^{-1} v = v + \textstyle\sum_{j\geq 1} v_j\, y^{-j}, \qquad \text{with } v_j \in W_{l+j-1}.$$

For each coefficient v_j, $\exp\left(\tfrac{1}{2}\log y\, \psi_*(Y)\right) v_j$ has order of growth at most $y^{(l-k+j-1)/2}$. Hence

$$\exp\left(\tfrac{1}{2}\log y\, \psi_*(Y)\right) g(y)^{-1} v = v\, y^{(l-k)/2} + O(y^{(l-k-2)/2}),$$

which verifies (6.12) and concludes the proof.

In order to put the next result into perspective, I shall recall a definition of Deligne. Let $H_{\mathbb{R}}$ be a finite dimensional real vector space with complexification $H_{\mathbb{C}}$, and with a $\mathbb{Q}$-structure coming from a lattice

$H_{\mathbb{Z}} \subset H_{\mathbb{R}}$. A *mixed Hodge structure* on this vector space consists of an increasing filtration

$$0 \subset \cdots \subset W^{l-1} \subset W^l \subset W^{l+1} \subset \cdots \subset H_{\mathbb{C}},$$

which is defined over $\mathbb{Q}$ (the "weight filtration"), and a decreasing filtration

$$H_{\mathbb{C}} \supset \cdots \supset F^{p-1} \supset F^p \supset F^{p+1} \supset \cdots \supset 0$$

(the "Hodge filtration"), such that for each l, the filtration $\{F^p(Gr_l(W_*))\}$, with

$$F^p(Gr_l(W_*)) = F^p \cap W_l / F^p \cap W_{l-1},$$

constitutes a Hodge filtration of pure weight l on $Gr_l(W_*)$. The notion of a Hodge structure of pure weight k can be viewed as a special case: as F-filtration, one takes the Hodge filtration of the weighted Hodge structure, W_k is set equal to the full vector space, and W_{k-1} is set equal to the zero subspace. If $H_{\mathbb{C}}$ and $H'_{\mathbb{C}}$ are vector spaces with mixed Hodge structures, a linear transformation T: $H_{\mathbb{C}} \to H'_{\mathbb{C}}$ is said to be a morphism of mixed Hodge structures of type (n, n), with $n \in \mathbb{Z}$, if it preserves the rational structure, and if

$$T F^p \subset F'^{\,p+n}, \qquad T W_l \subset W'_{l+2n},$$

for all indices p and l. A morphism of type $(0, 0)$ will simply be called a morphism.

The importance of mixed Hodge structures stems from the following theorem of Deligne:

Theorem (Deligne [7]). *The complex cohomology groups of a complex projective variety carry mixed Hodge structures which are functorial. In the case of a nonsingular variety, these mixed Hodge structures reduce to the ordinary Hodge structures of pure weight.*

Again, I consider a family $\pi: \mathscr{V} \to \Delta^*$, parameterized by the punctured disc. Let $\mathbf{H}^k_{\mathbb{C}} \to \Delta^*$ be the flat bundle whose fibre over $t \in \Delta^*$ is $H^k(V_t, \mathbb{C})$. Via the universal covering τ: $U \to \Delta^*$, $\mathbf{H}^k_{\mathbb{C}}$ pulls back to a trivial bundle on U, with fibre $H^k_{\mathbb{C}}$. For each $z \in U$, there is a natural identification between $H^k_{\mathbb{C}}$ and $H^k(V_t, \mathbb{C})$, with $t = \tau(z)$. By transferring the Hodge filtration of $H^k(V_t, \mathbb{C})$ to $H^k_{\mathbb{C}}$ via this identification, one obtains a Hodge filtration

$$H^k_{\mathbb{C}} = F^0_z \supset F^1_z \supset \cdots \supset F^{k-1}_z \supset F^k_z \supset 0 \tag{6.14}$$

which depends holomorphically on z. Also, if γ: $H^k_{\mathbb{C}} \to H^k_{\mathbb{C}}$ denotes the Picard-Lefschetz transformation, $F^p_{z+1} = \gamma F^p_z$. Let $\gamma = \gamma_s \gamma_u$ be the Jordan decomposition of γ, m the order of γ_s, and $N = \log \gamma_u$. As in § 4, one finds that

$$z \mapsto \exp(-zN) F^p_z,$$

considered as a mapping of U into an appropriate Grassmann variety, is invariant under the translation $z \mapsto z+m$. Because the decomposition (2.2) is compatible with the Hodge decomposition and with the action of γ, (4.9) guarantees the existence of

$$F_\infty^p = \lim_{\operatorname{Im} z \to \infty} \exp(-zN) F_z^p,$$

uniformly in $\operatorname{Re} z$. The resulting filtration

$$H_{\mathbb{C}}^k = F_\infty^0 \supset F_\infty^1 \supset \cdots \supset F_\infty^{k-1} \supset F_\infty^k \supset 0 \tag{6.15}$$

need not be a Hodge filtration, of course.

It should be pointed out that the filtration (6.15) depends on the choice of the coordinate t on the disc Δ. To be precise, the construction becomes canonical only after the value of the differential dt at $t=0$ is fixed. Passing from one local coordinate to another has the effect of replacing the filtration $\{F_\infty^p\}$ by $\{\exp(\lambda N) F_\infty^p\}$, for some $\lambda \in \mathbb{C}$. However, the filtrations which $\{F_\infty^p\}$ induces on the kernel and the cokernel of N, and on the quotients of the weight filtration, are completely canonical. It is these induced filtrations which have geometric significance.

The Kähler operator $L: H_{\mathbb{C}}^k \to H_{\mathbb{C}}^{k+2}$ commutes with γ, and hence also with N, and since it maps the subspace F_z^p of $H_{\mathbb{C}}^k$ into the subspace F_z^{p+1} of $H_{\mathbb{C}}^{k+2}$, for each $z \in U$, it raises the index of the weight filtrations by two and the index of the filtrations (6.15) by one.

(6.16) **Theorem.** *The two filtrations $\{W_l\}$ of (6.5) and $\{F_\infty^p\}$ of (6.15) determine a mixed Hodge structure on $H_{\mathbb{C}}^k$. With respect to this mixed Hodge structure, N is a morphism of type $(-1,-1)$, and the Kähler operator $L: H_{\mathbb{C}}^k \to H_{\mathbb{C}}^{k+2}$ is a morphism of type $(1,1)$. In particular, the mixed Hodge structure of $H_{\mathbb{C}}^k$ restricts to one on the primitive part $P_{\mathbb{C}}^k \subset H_{\mathbb{C}}^k$. The induced Hodge structures of pure weight l on $Gr_l(P_{\mathbb{C}}^k \cap W_*)$ further restrict to Hodge structures on*

$$\mathscr{P}_l \subset Gr_l(P_{\mathbb{C}}^k \cap W_*), \qquad l \geqq k,$$

which are polarized with respect to the nondegenerate bilinear forms S_l on $\mathscr{P}_l$ (cf. Lemma (6.4)).

This statement had been conjectured by Deligne, and he also deduced it from an earlier version of Theorem (5.13).

As the theorem asserts, for each l, the filtration $\{F_\infty^p\}$ induces a Hodge structure of weight l on $Gr_l(W_*)$. Since N operates trivially on $Gr_l(W_*)$, $\{\exp(zN) F_\infty^p\}$ induces the same filtration on $Gr_l(W_*)$ as $\{F_\infty^p\}$, for every $z \in \mathbb{C}$. Hence the theorem is equivalent to saying that the two filtrations $\{W_l\}$ and $\{\exp(zN) F_\infty^p\}$ determine a mixed Hodge structure, for each $z \in \mathbb{C}$. According to (4.9), as $\operatorname{Im} z \to \infty$, the filtration $\{\exp(zN) F_\infty^p\}$

asymptotically approaches the filtration (6.14). One might therefore suspect, that for all $z \in U$ with sufficiently large imaginary part, $\{F_z^p\}$ and $\{W_l\}$ give a mixed Hodge structure on $H_{\mathbb{C}}^k$. Deligne had originally conjectured this stronger version of (6.16), which is listed as problem (9.17) in [12]. However, as Deligne has pointed out since, (6.16) is all that should be expected: if the Hodge filtration of a mixed Hodge structure is perturbed slightly, the result will usually not be a mixed Hodge structure any more; not even the horizontal nature of the period mapping suffices to remedy this problem. In some special cases, which include the periods of Abelian varieties and of $K3$ surfaces, the stronger version of (6.16) does become correct. This fact was observed by Deligne, and it may be worthwhile to state it explicitely:

(6.17) **Proposition.** *Let $P_{\mathbb{C}}^k \subset H_{\mathbb{C}}^k$ be the primitive part, and suppose that the classifying space for the Hodge structure on $P_{\mathbb{C}}^k$ happens to be Hermitian symmetric. Then, for every $z \in U$ with sufficiently large imaginary part, the two filtrations $\{F_z^p \cap P_{\mathbb{C}}^k\}$ and $\{W_l \cap P_{\mathbb{C}}^k\}$ determine a mixed Hodge structure on $P_{\mathbb{C}}^k$. The resulting Hodge structures of pure weight l on $Gr_l(W_* \cap P_{\mathbb{C}}^k)$, viewed as a function of z, have a limit as $\operatorname{Im} z \to \infty$; the limit coincides with the Hodge structure of weight l induced by the filtration $\{F_\infty^p \cap P_{\mathbb{C}}^k\}$.*

I shall prove the proposition at the end of this section, following the proof of Theorem (6.16).

I now suppose temporarily that $\pi\colon \mathscr{V} \to \Delta^*$ can be continued to a family over the disc Δ, by inserting a possibly singular fibre over the origin. More precisely, let $\mathscr{V}$ be Zariski open in a complex submanifold $\bar{\mathscr{V}}$ of some projective space, such that π extends to a proper, surjective, holomorphic mapping $\pi\colon \bar{\mathscr{V}} \to \Delta$. The polarization of the fibres V_t, $t \neq 0$, are to be the ones induced by the given projective embedding. The central fibre $V_0 = \pi^{-1}(0)$ the has the structure of a projective variety. According to Clemens [3], V_0 is a strong deformation retract of $\bar{\mathscr{V}}$, so that $H^*(V_0, \mathbb{C}) \cong H^*(\bar{\mathscr{V}}, \mathbb{C})$. Composing this isomorphism with the mapping which corresponds to the inclusion $V_t \hookrightarrow \bar{\mathscr{V}}$, one obtains a map $H^k(V_0, \mathbb{C}) \to H^k(V_t, \mathbb{C})$, $t \neq 0$. The image consists of cohomology classes which come by restriction from the total space to the fibre of the C^∞ bundle $V_t \to \mathscr{V} \to \Delta^*$, and these must be invariant under the action of the Picard-Lefschetz transformation γ. Since there is an identification between $H_{\mathbb{C}}^k$ and $H^k(V_t, \mathbb{C})$, which is distinguished up to the action of γ, this gives a well-defined map

(6.18) $$H^k(V_0, \mathbb{C}) \to H_{\mathbb{C}}^k.$$

According to Deligne's construction on the one hand and to Theorem (6.16) on the other, both the domain and the target of the mapping

(6.18) carry canonical mixed Hodge structures. It is thus natural to ask if the mapping is a morphism. An affirmative answer to this question and a discussion of its geometric significance, along with a number of related matters, will be contained in a forthcoming paper, jointly written with H. Clemens.

Now to the proof of (6.16)! Once the filtrations $\{F_\infty^p\}$ and $\{W_l\}$ are known to give a mixed Hodge structure on $H_{\mathbb{C}}^k$, the statements about N and L do not present any problems: as was pointed out above (6.16), L raises the index in the Hodge filtration by one, and in the weight filtration by two; according to the definition of the weight filtration, $N(W_l) \subset W_{l-2}$, and in view of the final statement in (4.9), combined with Lemma (3.18), N maps F_∞^p to F_∞^{p-1}. As for the main assertions of Theorem (6.16), since N and the Kähler operator commute, it suffices to consider the primitive part $P_{\mathbb{C}}^k$ of $H_{\mathbb{C}}^k$. To simplify the notation, I shall refer to the filtration $\{F_\infty^p \cap P_{\mathbb{C}}^k\}$ as $\{F_\infty^p\}$ from now on, and similarly to $\{W_l \cap P_{\mathbb{C}}^k\}$ as $\{W_l\}$. It should be recalled that the filtration $\{F_\infty^p\}$ corresponds to the point $a \in \check{D}$. According to (5.13),

$$a = \exp(-zN)\, g(-iz) \circ \tilde{\psi}(z).$$

On the other hand, $\tilde{\psi}(z)$ is the $\exp(zN)$-translate of the base point $o \in D$. Hence, whenever $g(-iz)$ is defined,

$$a = \operatorname{Ad} \exp(-zN)\big(g(-iz)\big) \circ o. \tag{6.19}$$

For reasons of convenience, I shall assume that the base point has been chosen as in (5.13h), so that $g(\infty) = 1$.

(6.20) **Lemma.** *The limit of* $\operatorname{Ad}\exp(-zN)\big(g(-iz)\big)$, *as z tends to infinity, exists in $G_{\mathbb{C}}$. Let g_∞ be the value of this limit; then g_∞ preserves the filtration $\{W_l\}$ and operates as the identity on $Gr_l(W_*)$, for each l.*

Because of (6.19), the point a coincides with the g_∞-translate of o. Hence:

(6.21) **Corollary.** *For each l, the reference Hodge filtration and the filtration $\{F_\infty^p\}$ determine the same filtration on $Gr_l(W_*)$.*

Proof of (6.20). It follows from the next to last statement in (6.4) that W_l can be described as the direct sum of those eigenspaces of $\psi_*(Y)$ which correspond to eigenvalues less than or equal to $l-k$; this was already observed in the proof of (6.6). Hence if $g(-iz)$ is expanded as

$$g(-iz) = 1 + \textstyle\sum_{n \geq 1} g_n (-iz)^{-n},$$

the coefficient g_n maps W_l into W_{l+n-1} (cf. (5.13g)). Consequently, $(\operatorname{Ad} N)^n g_n$ maps W_l into W_{l-n-1}; in particular, for all l and all $n \geq 1$,

$$(\operatorname{Ad} N)^n g_n(W_l) \subset W_{l-2}. \tag{6.22}$$

According to (5.13 h),

$$(\operatorname{Ad} N)^{n+1} g_n = 0,$$

which leads to the identity

$$\begin{aligned}&\operatorname{Ad}\exp(-zN)(g(-iz))\\ &\quad = 1 + \textstyle\sum_{n\geqq 1} \sum_{0\leqq t\leqq n} (-1)^t\, i^n (\operatorname{Ad} N)^t\, g_n\, z^{-(n-t)}.\end{aligned}$$

Hence $\lim_{z\to\infty} \operatorname{Ad}\exp(-zN)\, g(-iz)$ exists and equals

$$g_\infty = 1 + \textstyle\sum_{n\geqq 1} (-i)^n (\operatorname{Ad} N)^n g_n.$$

The remaining assertion now follows from (6.22).

In view of (6.21), as far as the proof of (6.16) is concerned, I may as well assume that $\{F_\infty^p\}$ is the reference Hodge filtration, and I shall now consider this special case. In the original definition of a weighted Hodge structure, the underlying vector space comes equipped with a $\mathbb{Q}$-structure and a lattice. In the following arguments, various $\mathbb{Q}$-structures and lattices can be dragged along; however, they would only be excess baggage. Hence, for the remainder of the proof, the underlying vector spaces of Hodge structures will be required only to carry an $\mathbb{R}$-structure.

In (5.10), I defined generators Z, X_+, X_- of $\mathfrak{sl}(2,\mathbb{C})$; they satisfy

$$[Z, X_+] = 2X_+, \quad [Z, X_-] = -2X_-, \quad [X_+, X_-] = Z$$
$$\bar{Z} = -Z, \quad \bar{X}_+ = X_-.$$

Now let $\{H^{p,q}\}$ be a Hodge structure of weight k on the vector space $H_\mathbb{C} = H_\mathbb{R} \otimes_\mathbb{R} \mathbb{C}$. A linear action of $\mathfrak{sl}(2,\mathbb{C})$ on $H_\mathbb{C}$ – for simplicity, I shall denote it by juxtaposition – will be called *horizontal* if it is defined over $\mathbb{R}$, and if

$$X_+ H^{p,q} \subset H^{p-1,q+1}, \quad X_- H^{p,q} \subset H^{p+1,q-1}, \quad Z H^{p,q} \subset H^{p,q}$$

for all p, q. Incidentally, the second and third of these relations are consequences of the first. When the Hodge structure happens to be polarized by a bilinear form S, I shall say that the $\mathfrak{sl}(2,\mathbb{C})$-action is *compatible with the polarization* if $\mathfrak{sl}(2,\mathbb{C})$ acts as a Lie algebra of infinitesimal isometries of S. A subspace $H_1 \subset H_\mathbb{C}$ is *invariant* with respect to the given Hodge structure and horizontal $\mathfrak{sl}(2,\mathbb{C})$-action, if it remains stable under complex conjugation, the $\mathfrak{sl}(2,\mathbb{C})$-action, and under the projections onto the Hodge subspaces $H^{p,q} \subset H_\mathbb{C}$. The subspace $H_1 \subset H_\mathbb{C}$ shall be said to be *irreducible*, again with respect to the given data, if it is a minimal nontrivial invariant subspace.

For the statement of the next lemma, I need a repertoire of basic examples. The one-dimensional complex vector space $\mathbb{C}$, with the obvious real structure, carries a unique Hodge structure of weight 2. Deligne denotes it by $H(1)$ and calls it the "Hodge structure of Tate". For $n \geqq 0$, $H(n)$ shall be the n-th symmetric power of $H(1)$, and $H(-n)$ the dual of $H(n)$ (cf. (2.12)). The trivial $\mathfrak{sl}(2, \mathbb{C})$ action is clearly horizontal with respect to $H(n)$. Also, each $H(n)$ has a natural polarization: the underlying vector space of $H(n)$ can be identified with $\mathbb{C}$; the nondegenerate bilinear form S on $\mathbb{C}$, which is normalized by $S(1,1)=1$, then gives the polarization. Next, let e_1, e_2 be the standard basis vectors of $\mathbb{C}^2$. For $p \neq q$, I define a Hodge structure $E(p,q)$ of weight $p+q$ on $\mathbb{C}^2$, with the natural real structure, by requiring that

$$v_+ = e_1 + i e_2$$

shall be of type (q,p), and

$$v_- = e_1 - i e_2$$

of type (p,q). Again the trivial $\mathfrak{sl}(2, \mathbb{C})$ action is horizontal, and with respect to it, the Hodge structure $E(p,q)$ becomes irreducible. The bilinear form S on $\mathbb{C}^2$, which is described by the identities

$$\text{(6.23)} \qquad \begin{array}{ll} S(v_+, v_+) = 0, & S(v_-, v_-) = 0, \\ S(v_+, v_-) = 2\, i^{p-q}, & S(v_-, v_+) = 2\, i^{q-p}, \end{array}$$

polarizes $E(p,q)$. Next, I let $S(1)$ denote the same Hodge structure as $E(1,0)$, but equipped with the standard $\mathfrak{sl}(2, \mathbb{C})$-action on the underlying vector space $\mathbb{C}^2$. One can check directly that this action is horizontal. With respect to the bilinear form S of (6.23), in the special case when $p=1$, $q=0$, $\mathfrak{sl}(2, \mathbb{C})$ acts as a Lie algebra of infinitesimal isometries, so that the action and the polarization become compatible. Finally, $S(n)$ shall be the n-th symmetric product of $S(1)$, as polarized Hodge structure (cf. (2.12)); from $S(1)$, $S(n)$ also inherits a horizontal $\mathfrak{sl}(2, \mathbb{C})$-action, compatible with the polarization. Since $S(n)$, for $n \geqq 0$, is irreducible even as an $\mathfrak{sl}(2, \mathbb{C})$-module – this follows from a comparison of the eigenvalues of Z with the statement (6.3) – it must certainly be irreducible with respect to the Hodge structure and $\mathfrak{sl}(2, \mathbb{C})$-action, considered as an entity. The following result was first used by Deligne:

(6.24) **Lemma.** *Let $H_{\mathbb{C}} = H_{\mathbb{R}} \otimes \mathbb{C}$ be a complex vector space with a Hodge structure of weight k and a horizontal $\mathfrak{sl}(2, \mathbb{C})$-action. Then $H_{\mathbb{C}}$ can be decomposed into a direct sum of subspaces which are invariant and irreducible with respect to the given structures. Every irreducible subspace is isomorphic – relative to the Hodge structure and horizontal action – to one of the following types: $H(k_1) \otimes S(k_2)$, with $k_1 \in \mathbb{Z}$, $k_2 \geqq 0$, and $k = 2k_1 + k_2$; or $E(p,q) \otimes S(k_1)$, with $p > q$, $k_1 \geqq 0$, and $k = p + q + k_1$. If the Hodge structure of $H_{\mathbb{C}}$ happens to be polarized, compatibly with the $\mathfrak{sl}(2, \mathbb{C})$-action, then the decomposition can be chosen to be orthogonal* *with respect to the polarization, and the isomorphisms between the irreducible constituents and the irreducible structures of special type can be chosen with the further restriction that they should preserve the polarizations.*

Proof. Since Z leaves the Hodge subspaces invariant and acts semisimply, with integral eigenvalues, one can define an $SO(2)$-action on $H_{\mathbb{C}}$ as follows: if $v \in H^{p,q}$ also lies in the l-eigenspace of Z, the element

$$\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$$

of $SO(2)$ shall operate on v as multiplication by $e^{i(l+p-q)\theta}$. The action of $SO(2)$ is orthogonal with respect to any polarization which $H_{\mathbb{C}}$ might carry. I may regard $SO(2)$ as the group of real points of an algebraic 1-torus T, which is defined and anisotropic over $\mathbb{R}$. The $SO(2)$-action extends to a representation of T over $\mathbb{R}$. The action of $\mathfrak{sl}(2,\mathbb{C})$ on $H_{\mathbb{C}}$ determines a representation, again defined over $\mathbb{R}$, of the algebraic group SL_2. Because of the horizontal property, the representations of T and SL_2 commute. Hence they determine a representation of the product $T \times SL_2$. A Z-stable subspace of $H_{\mathbb{C}}$ carries a sub-Hodge structure if and only if it is self-conjugate and invariant under the $SO(2)$-action. It follows that the subspaces which are invariant, respectively irreducible, with respect to the Hodge structure and $\mathfrak{sl}(2,\mathbb{C})$-action correspond bijectively to those invariant, respectively irreducible, subrepresentations of the representation of $T \times SL_2$, which are defined over $\mathbb{R}$. Because of the reductive nature of the product group $T \times SL_2$, this proves the first assertion of the lemma. If a bilinear form S on $H_{\mathbb{C}}$ polarizes the Hodge structure in question and is compatible with the $\mathfrak{sl}(2,\mathbb{C})$-action, then $T \times SL_2$ operates as a group of isometries; moreover, on each sub-Hodge structure, S must be nondegenerate. Hence the decomposition can be performed orthogonally with respect to S.

An irreducible representation over $\mathbb{R}$ of the group $T \times SL_2$ either remains irreducible under SL_2, in which case T acts trivially, or splits into two conjugate subspaces, each of which is T-stable and SL_2-irreducible, with T acting nontrivially. The first situation corresponds to an irreducible Hodge structure with $\mathfrak{sl}(2,\mathbb{C})$-action of the type $H(k_1) \otimes S(k_2)$; here k_2 is the dimension of $H_{\mathbb{C}}$ plus one, and k_1 is determined by $k = 2k_1 + k_2$. The second case corresponds to a Hodge structure of the type $E(p,q) \otimes S(k_1)$; now $H_{\mathbb{C}}$ has dimension $2(k_1+1)$, the integer $p-q$ is an invariant of the action of T, and $k = k_1 + p + q$; this determines k_1, p, q completely. The final details of the verification are left to the reader. For an irreducible representation of $T \times SL_2$ over $\mathbb{R}$, the space of bilinear forms over $\mathbb{R}$, which are preserved by the representation, has dimension one. Also, a real multiple of a polarization form is again a polarization precisely when the multiplicative constant is positive. Hence any two polarizations on an irreducible Hodge structure with horizontal $\mathfrak{sl}(2,\mathbb{C})$-action, as long as they are compatible with the action, must be related by a multiple of the identity transformation. Hence the lemma.

In the following, $H_{\mathbb{C}}=H_{\mathbb{R}}\otimes_{\mathbb{R}}\mathbb{C}$ will be a complex vector space with a Hodge structure $\{H^{p,q}\}$ of weight k, and a horizontal $\mathfrak{sl}(2,\mathbb{C})$ action. The elements $Y=i(X_- - X_+)$,

$$N_+=\tfrac{1}{4}(X_+ + X_- -2iZ), \qquad N_-=\tfrac{1}{4}(X_+ + X_- +2iZ)$$

satisfy the same commutation relations as Z, X_+, X_-. Thus Y acts semisimply, with integral eigenvalues, and N_- maps the l-eigenspace of Y into the $(l-2)$-eigenspace. I let W_l denote the direct sum of all those eigenspaces of Y which correspond to eigenvalues less than or equal to $l-k$. Then $\{W_l\}$ forms an increasing filtration of $H_{\mathbb{C}}$, defined over $\mathbb{R}$. By $\mathcal{P}_l$, I denote the projection into $Gr_l(W_*)$ of the kernel of N_-^{l-k+1} on the $(l-k)$-eigenspace of Y.

(6.25) **Lemma.** *The Hodge filtration $\{F^p\}$ of the Hodge structure $\{H^{p,q}\}$, together with the filtration $\{W_l\}$, determines a mixed Hodge structure on $H_{\mathbb{C}}$. The induced weighted Hodge structure of $Gr_l(W_*)$ restricts to a Hodge structure on the subspace $\mathcal{P}_l$. If a bilinear form S polarizes the original Hodge structure, compatibly with the $\mathfrak{sl}(2,\mathbb{C})$-action, then S_l, defined as in* (6.4), *polarizes the induced Hodge structure on $\mathcal{P}_l$.*

Proof. If $H_{\mathbb{C}}$ can be decomposed into a direct sum of subspaces which are invariant with respect to the given data, and if the statement holds for each of the summands, then it holds for the entire space. Hence, and in view also of the preceeding lemma, I may assume that $H_{\mathbb{C}}$ is of type $H(k_1)\otimes S(k_2)$, or of type $E(p,q)\otimes S(k_1)$.

In the first situation, the tensor product with the one-dimensional Hodge structure $H(k_1)$ only has the effect of shifting the indices in the Hodge filtration by k_1, and in the weight filtration by $2k_1$. I may thus limit myself to the case when $k_1=0$, $k_2=k$. The underlying vector space $H_{\mathbb{C}}$ is now the k-th symmetric product of $\mathbb{C}^2$. The vectors e_1, e_2, v_+, v_- shall have the same meaning as in the definition of $S(1)$, or more accurately, as in the definition of $E(1,0)$. Then $H^{p,k-p}$ is spanned by $v_+^{k-p}v_-^p$; $\{e_1^{k-i}e_2^i \mid 2i\leq l\}$ forms a basis of W_l. Evidently $Gr_l(W_*)$ vanishes for all odd integers l. In the even case, I claim that

$$\text{(6.26)} \qquad \begin{aligned} &W_{2j}\subset W_{2j-2}+F^j, && \text{and}\\ &W_{2j}\not\subset W_{2j-2}+F^{j+1}, && \text{if } j\geqq 0. \end{aligned}$$

Because the $\mathfrak{sl}(2, \mathbb{C})$-action is known to be irreducible, N_- must map W_l onto W_{l-2} (unless $l-2=k$; cf. (6.3)). Hence it suffices to verify the containment for $j=k$, and the non-containment for $j=0$. The vector v_-^k spans F^k; moreover,

$$v_-^k = (e_1 - i\, e_2)^k \equiv (-i)^k\, e_2^k \mod W_{2k-2},$$

so that $W_{2k} = H_{\mathbb{C}} = W_{2k-2} + F^k$. On the other hand, e_1^k spans W_0, but does not lie in F^1:

$$e_1^k = 2^{-k}(v_+ + v_-)^k \notin F^1 = \text{span of } \{v_+^{k-p}\, v_-^p \mid p \geqq 1\}.$$

This gives (6.26). Consequently, for $0 \leqq j \leqq k$, $\{F^p\}$ induces a Hodge structure of weight $2j$ on $Gr_{2j}(W_*)$, which is isomorphic to $H(j)$. Except for $l=2k$, $\mathscr{P}_l$ is zero; in the remaining case, $\mathscr{P}_{2k} = Gr_{2k}(W_*)$. On $S(1)$, the polarization form S, which is described by (6.23), must satisfy

$$S(e_1, e_1) = S(e_2, e_2) = 0, \qquad S(e_1, e_2) = -S(e_2, e_1) = -1.$$

Consequently, for the resulting bilinear form on the k-th symmetric product of $\mathbb{C}^2$ – I shall also denote it by S – one has $S(e_2^k, e_1^k) = 1$. If e designates the image of e_2^k in $\mathscr{P}_{2k}$, according to the definition of S_{2k} in (6.4),

$$S_{2k}(e, e) = S(e_2^k, N_-^k\, e_2^k) = S(e_2^k, e_1^k) = 1.$$

Since e is a real basis vector for $\mathscr{P}_{2k}$, this shows that S_{2k} polarizes the Hodge structure of $\mathscr{P}_{2k}$.

The remaining situation, when the Hodge structure is of type $E(p, q) \otimes S(k_2)$, can be reduced to the case which was just treated. Since $\mathfrak{sl}(2, \mathbb{C})$ acts trivially on $E(p, q)$, the $(l+p+q)$-th quotient in the gradation of the tensor product is naturally isomorphic to the l-th quotient in the gradation of $S(k_2)$, tensored with $E(p, q)$. The isomorphism is also an isometry with respect to the various bilinear forms. Hence I may quote (2.12), to obtain the desired conclusion.

In order to deduce Theorem (6.16), I choose the base point $o \in D$ as in (5.13g). As was pointed out already, I may then assume that $\{F_\infty^p\}$ is the Hodge filtration of the reference Hodge structure. According to (5.13), the reference Hodge structure comes equipped with a horizontal $\mathfrak{sl}(2, \mathbb{C})$-action. Since the action is described by a homomorphism of $\mathfrak{sl}(2, \mathbb{C})$ into the Lie algebra of infinitesimal isometries of the polarization, the horizontal $\mathfrak{sl}(2, \mathbb{C})$-action and the polarization are compatible. The linear transformation N in (5.18) corresponds to N_-, as defined above (6.25); also Y, N_+, N_- satisfy the same commutation relations as Z, X_+, X_-. Hence the filtration $\{W_l\}$ and the subspaces $\mathscr{P}_l \subset Gr_l(W_*)$, which were introduced just before the statement of (6.25), coincide with the weight filtration and the spaces of primitive elements for the weight filtration; this may be inferred from the final assertions in Lemma (6.4). Lemma (6.25) now gives the theorem.

Proposition (6.17) still remains to be proven. Again in order to simplify the notation, I shall refer to the filtrations $\{F_\infty^p \cap P_{\mathbb{C}}^k\}$, $\{F_z^p \cap P_{\mathbb{C}}^k\}$, $\{W_l \cap P_{\mathbb{C}}^k\}$ of $P_{\mathbb{C}}^k$ simply as $\{F_\infty^p\}$, $\{F_z^p\}$, $\{W_l\}$. At this stage, one should recall Theorem (4.9). The filtration $\{F_\infty^p\}$ corresponds to the point

$$a = \lim_{\operatorname{Im} z \to \infty} \tilde{\Psi}(z)$$

of $\check{D}$, and the filtration $\{F_z^p\}$ to the point

$$\tilde{\Phi}(z) = \exp(zN) \circ \tilde{\Psi}(m^{-1} z).$$

For any $z \in \mathbb{C}$, $\exp(zN)$ preserves the weight filtration and operates trivially on the successive quotients. Hence, without altering the conclusion of (6.17), I may replace the filtration $\{F_z^p\}$ by the filtration corresponding to the point $\tilde{\Psi}(m^{-1} z)$. I choose the base point $o \in D$ as in (5.13g), and the element $g_\infty \in G_{\mathbb{R}}$ as in Lemma (6.20). Since g_∞ operates trivially on the quotients $Gr_l(W_*)$, I may further simplify the situation by translating both a and $\tilde{\Psi}$ by g_∞^{-1}. In other words, I can assume that a coincides with the base point o, and that

$$o = \lim_{\operatorname{Im} z \to \infty} \tilde{\Psi}(z).$$

Every filtration which is sufficiently close to a given Hodge filtration must again be a Hodge filtration. Hence the preceeding remarks, Theorem (6.16), and the following lemma together imply (6.17).

(6.27) **Lemma.** *Suppose that the classifying space for Hodge structures happens to be Hermitian symmetric. Then there exists a Zariski open subset $\mathscr{U}$ of $\check{D}$, which contains D, and which has the following property: the filtration $\{F_b^p\}$ corresponding to the points b of $\mathscr{U}$ induce filtrations on the quotients $Gr_l(W_*)$, which depend continuously on b.*

Proof. In the case of a Hermitian symmetry space, the isotropy group V at the base point must be maximal compact in $G_{\mathbb{R}}$. I shall therefore denote it by K. It is possible to choose an Iwasawa decomposition UAK of $G_{\mathbb{R}}$, such that the Lie algebra of A contains $\psi_*(Y)$ (notation of (5.13)), and such that the Lie algebra of U includes all eigenspaces of $\psi_*(Y)$ on $\mathfrak{g}_0$ which belong to negative eigenvalues. Let M be the centralizer of A in K. Then UAM is the group of real points of an $\mathbb{R}$-parabolic subgroup P of G. Since UA acts transitively on D, the $P_{\mathbb{C}}$-orbit of any point of D contains all of D. I denote the orbit by $\mathscr{U}$. As an orbit of a parabolic subgroup of $G_{\mathbb{C}}$, which contains an open set relative to the Hausdorff topology, $\mathscr{U}$ has to be Zariski open in $\check{D}$. The groups U, A, M all preserve the weigth filtration (A and M even preserve the gradation defined by $\psi_*(Y)$); hence so does $P_{\mathbb{C}}$. The action of $P_{\mathbb{C}}$ on the quotients $Gr_l(W_*)$ is certainly continuous. This gives the assertion of the lemma.

§ 7. Global Properties of the Period Mapping

By looking at the curvature of the Hodge bundles, Griffiths [11] established certain global properties of the period mapping of a variation of Hodge structure – and thus in particular for the period mapping of a family of polarized algebraic manifolds – under the assumption that the base space is compact. In rough outline, the arguments proceed by considering the square lengths of certain holomorphic sections of the Hodge bundles. These functions on the base space turn out to be plurisubharmonic, because of the nature of the curvature in the Hodge bundles. A compact manifold does not carry any nonconstant plurisubharmonic functions. Through further reasoning, this then leads to the conclusions about the period mapping.

In the geometric situations which usually occur, the base is not compact, but only Zariski open in a compact variety. Such manifolds do admit nonconstant plurisubharmonic functions, but only unbounded ones. Griffiths' arguments therefore extend to this more general setting, as soon as the plurisubharmonic functions which come up are known to be bounded. The results of §6, in particular (6.7), do give the boundedness of the functions in question, and hence they make it possible to apply the curvature arguments, even if the base is not necessarily compact. As was mentioned in the introduction, Deligne [6] had already proven Griffiths' theorems for algebraic families with quasi-projective base, by algebraic geometric methods. The transcendental arguments which are based on the curvature of the Hodge bundles have the minor advantage that they do not depend on the presence of any algebraic structures.

In this section, I shall prove Griffiths' theorems about the period mapping for families whose base is Zariski open in some compact analytic space, and more generally, for an arbitrary variation of Hodge structure with a base space of this type. Passing to the more general abstract setting has the advantage of actually simplifying the proofs. The starting point, of course, is the curvature of the Hodge bundles, which was computed by Griffiths in [11]. I shall include a derivation of the curvature properties; not only for the sake of completeness, but also because some simplifications are possible. These simplifications were the result of a conversation with Griffiths.

I begin by recalling some basic facts about connections. Let M be a complex manifold, and $\mathbf{E} \to M$ a holomorphic vector bundle. A connection for $\mathbf{E}$ is a $\mathbb{C}$-linear mapping

$$\nabla\colon\ C^{\infty}(\mathbf{E}) \to C^{\infty}(\mathbf{T}_{\mathbb{C}}^{*} \otimes \mathbf{E})$$

($\mathbf{T}_{\mathbb{C}}^{*}$ = complexified C^{∞} cotangent bundle), which satisfies the derivation rule

$$\nabla f e = df \otimes e + f \nabla e,$$

for every C^{∞} section e of $\mathbf{E}$ and every C^{∞} function f. If X is a C^{∞} vector field, $\nabla(X)$ will denote the operation of ∇, followed by contraction with X. The connection ∇ is said to be of type $(1,0)$ if $\nabla(X)\,e=0$, whenever e is a holomorphic section and X an antiholomorphic vector field. Let $\{e_1, \ldots, e_r\}$ be a local holomorphic frame for $\mathbf{E}$. The connection matrix of ∇, relative to the local frame, is the $r \times r$ matrix of 1-forms (φ_i^j), such that

$$\nabla e_i = \textstyle\sum_j \varphi_i^j \otimes e_j;$$

on its domain of definition, it determines the connection completely. The connection forms have type $(1,0)$ precisely when ∇ is a $(1,0)$-connection.

Next, I suppose that $\mathbf{E}$ carries a nondegenerate, but not necessarily positive definite, Hermitian pseudometric. A connection ∇ will be called compatible with the pseudometric if, for every vector field X and sections e_1, e_2 of $\mathbf{E}$,

$$\big(\nabla(X)\, e_1, e_2\big) + \big(e_1, \nabla(\bar{X})\, e_2\big) = X(e_1, e_2).$$

In terms of the connection matrix (φ_i^j) of a local holomorphic frame $\{e_1, \ldots, e_r\}$, this condition becomes equivalent to

$$\textstyle\sum_l \varphi_i^l h_{lj} + \sum_l h_{il} \bar{\varphi}_j^l = dh_{ij}, \tag{7.1}$$

where $h_{ij} = (e_i, e_j)$. When ∇ happens to be of type $(1,0)$, the first summand in (7.1) has type $(1,0)$, and the second has type $(0,1)$. Thus, equating terms of the same type, one finds $\sum_l \varphi_i^l h_{lj} = \partial h_{ij}$; or, in matrix notation,

$$(\varphi_i^j) = (\partial h_{ij})(h_{ij})^{-1}. \tag{7.2}$$

In particular, this proves:

(7.3) A holomorphic vector bundle with a nondegenerate Hermitian pseudometric carries a unique $(1,0)$-connection which is compatible with the pseudometric.

I shall refer to this connection as the *metric connection.*

To each connection ∇ of $\mathbf{E}$, there corresponds a dual connection ∇^* on the dual bundle $\mathbf{E}^*$, which is uniquely determined by the requirement that

$$\langle \nabla^* \lambda, e\rangle + \langle \lambda, \nabla e\rangle = d\langle \lambda, e\rangle,$$

for every section e of $\mathbf{E}$ and λ of $\mathbf{E}^*$. If (φ_i^j) is the connection matrix of ∇ relative to a local holomorphic frame $\{e_1, \ldots, e_r\}$ of $\mathbf{E}$, and if

$\{\lambda^1, \dots, \lambda^r\}$ is the dual frame,

(7.4) $$\nabla^* \lambda^i = -\sum_j \varphi^i_j \lambda^j,$$

as one checks directly. A Hermitian pseudometric of **E** induces one also on **E***. With respect to it, the matrix of inner products (λ^i, λ^j) coincides with the transposed inverse of the matrix (h_{ij}). In view of (7.2) and (7.4), the last remark implies that

(7.5) the dual connection of a metric connection is the metric connection corresponding to the dual pseudometric.

It will be necessary to study induced connections on subbundles and quotient bundles. For this purpose, I consider an exact sequence of holomorphic vector bundles

(7.6) $$0 \longrightarrow \mathbf{E}' \xrightarrow{\ j\ } \mathbf{E} \xrightarrow{\ q\ } \mathbf{E}'' \longrightarrow 0.$$

Let ∇ be a (1, 0)-connection on **E**. Every C^∞ splitting s: $\mathbf{E}'' \to \mathbf{E}$ of the sequence (7.6) determines a connection ∇' on E':

(7.7) $$\nabla' e = j^{-1}(1 - s\,q)\,\nabla j\,e, \quad e \in C^\infty(\mathbf{E}').$$

Evidently ∇' is again a (1, 0)-connection, even though the splitting s need not be holomorphic. If **E** comes equipped with a Hermitian pseudometric, whose restriction to **E**′ shall be nondegenerate, the C^∞ orthogonal decomposition $\mathbf{E} = \mathbf{E}' \oplus \mathbf{E}'^{\perp}$ defines a C^∞ splitting s: $\mathbf{E}'' \to \mathbf{E}$. In this situation, under the assumption that ∇ is the metric connection,

(7.8) the connection ∇' of (7.7) coincides with the metric connection of **E**′, relative to the restricted pseudometric.

In order to verify the statement (7.8), one only has to check that ∇' is compatible with the restricted pseudometric. Let e_1, e_2 be sections of **E**′, and X a vector field. Then

$$\nabla(X)\,j\,e_1 = j\,a_1 + s\,b_1, \quad \nabla(\bar{X})\,j\,e_2 = j\,a_2 + s\,b_2,$$

for suitable C^∞ sections a_i, b_i of, respectively, **E**′ and **E**″. By definition of ∇',

$$\nabla'(X)\,e_1 = a_1, \quad \nabla'(\bar{X})\,e_2 = a_2.$$

Since $s\mathbf{E}''$ is perpendicular to **E**′,

$$\begin{aligned}\big(\nabla'(X)\,e_1, e_2\big) + \big(e_1, \nabla'(\bar{X})\,e_2\big) &= (j\,a_1, j\,e_2) + (j\,e_1, j\,a_2)\\ &= (j\,a_1 + s\,b_1, j\,e_2) + (j\,e_1, j\,a_2 + s\,b_2)\\ &= \big(\nabla(X)\,j\,e_1, j\,e_2\big) + \big(j\,e_1, \nabla(\bar{X})\,j\,e_2\big)\\ &= X(j\,e_1, j\,e_2) = X(e_1, e_2),\end{aligned}$$

as was to be shown.

Again I assume that $\mathbf{E}$ carries a Hermitian pseudometric, whose restriction to $\mathbf{E}'$ is nondegenerate, and that ∇ is the corresponding metric connection. Because the pseudometric is nondegenerate on the subbundle $\mathbf{E}'$, it projects to a pseudometric on the quotient bundle. With respect to the projected pseudometric, the splitting $s\colon \mathbf{E}'' \to \mathbf{E}$ becomes an isometry. By dualizing the sequence (7.6), one obtains an exact sequence of vector bundles

$$0 \longrightarrow \mathbf{E}''^* \xrightarrow{\;{}^tq\;} \mathbf{E}^* \xrightarrow{\;{}^tj\;} \mathbf{E}'^* \longrightarrow 0$$

(${}^tj$ and ${}^tq$ are the bundle maps dual to j and q). As before, I let $s\colon \mathbf{E}'' \to \mathbf{E}$ denote the C^∞ splitting of the exact sequence (7.6) which corresponds to the pseudometric. Its dual ${}^ts$ coincides with the orthogonal projection of $\mathbf{E}^*$ onto $\mathbf{E}''^*$. According to (7.8), the metric connection of E^* is therefore given by the formula

$$\nabla''^* \lambda = {}^ts\, \nabla^*\, {}^tq\, \lambda \qquad (\lambda \in C^\infty(\mathbf{E}''^*)).$$

Dualizing again, and using (7.5), one finds:

(7.9) the metric connection ∇'' of $\mathbf{E}''$ satisfies $\nabla'' e = q\, \nabla\, s\, e$, for $e \in C^\infty(\mathbf{E}'')$.

In order to review the definition of the curvature form, I consider a holomorphic vector bundle $\mathbf{E}$ with a (1,0)-connection ∇. If X, Y are vector fields and e a section of $\mathbf{E}$, the expression

$$\Theta(X,Y)\,e = \big(\nabla(X)\,\nabla(Y) - \nabla(Y)\,\nabla(X) - \nabla([X,Y])\big)\,e$$

is linear over the ring of C^∞ functions in each of the three variables, and it is skew in X and Y. Thus Θ may be regarded as a differential 2-form with values in the vector bundle $\operatorname{Hom}(\mathbf{E},\mathbf{E})$; Θ is the curvature form of the connection. I let Θ^* denote the curvature form of the dual connection ∇^* on the dual bundle $\mathbf{E}^*$. If X, Y are vector fields, and e and λ sections of, respectively, $\mathbf{E}$ and $\mathbf{E}^*$, one finds

$$\begin{aligned}
\langle \Theta^*(X,Y)\,\lambda, e\rangle &= -\langle \nabla^*(Y)\,\lambda, \nabla(X)\,e\rangle + \langle \nabla^*(X)\,\lambda, \nabla(Y)\,e\rangle \\
&\quad + \langle \lambda, \nabla([X,Y])\,e\rangle + X\langle \nabla^*(Y)\,\lambda, e\rangle \\
&\quad - Y\langle \nabla^*(X)\,\lambda, e\rangle - [X,Y]\langle \lambda, e\rangle \\
&= \langle \lambda, \big(\nabla(Y)\,\nabla(X) - \nabla(X)\,\nabla(Y) + \nabla([X,Y])\big)\,e\rangle \\
&\quad - Y\langle \lambda, \nabla(X)\,e\rangle + X\langle \lambda, \nabla(Y)\,e\rangle \\
&\quad + X\langle \nabla^*(Y)\,\lambda, e\rangle - Y\langle \nabla^*(X)\,\lambda, e\rangle \\
&\quad - [X,Y]\langle \lambda, e\rangle \\
&= -\langle \lambda, \Theta(X,Y)\,e\rangle + XY\langle \lambda, e\rangle - YX\langle \lambda, e\rangle \\
&\quad - [X,Y]\langle \lambda, e\rangle \\
&= -\langle \lambda, \Theta(X,Y)\,e\rangle.
\end{aligned}$$

Hence $-\Theta^*(X, Y)$ coincides with the transpose of $\Theta(X, Y)$:

$$\Theta^*(X, Y) = -{}^t\Theta(X, Y). \tag{7.10}$$

In the definition of the curvature form, if both vector fields X and Y happen to be antiholomorphic, $\Theta(X, Y)\,e$ must vanish, because of the $(1,0)$-property of the connection. Consequently, the curvature form has no component of type $(0,2)$. I now suppose that ∇ is the metric connection corresponding to a Hermitian pseudometric. Then, for every pair of vector fields X, Y,

(7.11) $\Theta(X, Y)$ is adjoint to $-\Theta(\bar{X}, \bar{Y})$, relative to the pseudometric.

This can be verified by a computation which is formally identical to the computation preceeding (7.10), except for the fact that the pseudometric is conjugate linear in the second variable. In particular, since Θ has no $(0,2)$ component, it cannot have a $(2,0)$ component, either. Hence

(7.12) the curvature form of a metric connection is of type $(1,1)$.

One other observation should be made in this context. Let e_1, e_2 be holomorphic sections of the pseudo-Hermitian vector bundle **E**. The mapping

$$(X, Y) \mapsto (\nabla(X)\,e_1, \nabla(\bar{Y})\,e_2) - (\nabla(Y)\,e_1, \nabla(\bar{X})\,e_2)$$

is linear over the ring of C^∞ functions in both variables, and it is skew symmetric. It therefore may be viewed as a differential 2-form, which I denote symbolically by $(\nabla e_1, \overline{\nabla} e_2)$. It has type $(1,1)$, because ∇ is a $(1,0)$-connection. The expression $(\Theta\, e_1, e_2)$ also defines a scalar valued $(1,1)$-form. The difference of these is the Levi form of the function (e_1, e_2); explicitly,

$$\partial\bar{\partial}(e_1, e_2) = (\nabla e_1, \overline{\nabla} e_2) - (\Theta\, e_1, e_2). \tag{7.13}$$

Since both sides of the equality are forms of type $(1,1)$, in order to prove (7.13), it suffices to check it on a pair of vector fields X and $\bar{Y}$, with both X and Y holomorphic. For such vector fields, $[X, \bar{Y}] = 0$; also, $\Theta(X, \bar{Y})\,e = -\nabla(\bar{Y})\,\nabla(X)\,e$, provided e is a holomorphic section. Thus

$$\begin{aligned}
\partial\bar{\partial}(e_1, e_2)(X, \bar{Y}) &= d\bar{\partial}(e_1, e_2)(X, \bar{Y}) \\
&= X(\bar{\partial}(e_1, e_2)(\bar{Y})) - \bar{Y}(\bar{\partial}(e_1, e_2)(X)) - \bar{\partial}(e_1, e_2)([X, \bar{Y}]) \\
&= X\bar{Y}(e_1, e_2) = \bar{Y}X(e_1, e_2) = \bar{Y}(\nabla(X)\,e_1, e_2) \\
&= (\nabla(X)\,e_1, \nabla(Y)\,e_2) + (\nabla(\bar{Y})\,\nabla(X)\,e_1, e_2) \\
&= (\nabla e_1, \overline{\nabla} e_2)(X, \bar{Y}) - (\Theta\, e_1, e_2)(X, \bar{Y}),
\end{aligned}$$

which verifies (7.13).

Again, I look at an exact sequence of holomorphic vector bundles

$$0 \longrightarrow \mathbf{E}' \xrightarrow{j} \mathbf{E} \xrightarrow{q} \mathbf{E}'' \longrightarrow 0,$$

in which $\mathbf{E}$ carries a Hermitian pseudometric with a nondegenerate restriction to the subbundle $\mathbf{E}'$. Both $\mathbf{E}'$ and $\mathbf{E}''$ then inherit a pseudometric from that of $\mathbf{E}$. I denote the metric connections on the three bundles by $\nabla, \nabla', \nabla''$, and the corresponding curvature forms by Θ, Θ', and Θ''. For C^∞ sections e of $\mathbf{E}'$ and vector fields X, the expression

$$\sigma(X)\, e = q\, \nabla(X)\, j\, e \in C^\infty(\mathbf{E}'')$$

depends linearly on both variables, linearly over the ring of C^∞ functions. Hence σ may be regarded as a $(1,0)$-form with values in $\operatorname{Hom}(\mathbf{E}', \mathbf{E}'')$; it is the *second fundamental form* of the subbundle $\mathbf{E}'$. I define $\sigma^*(X)$ as the adjoint of $\sigma(\bar{X})$. This makes σ^* C^∞-linear in X, so that σ^* becomes a $(0,1)$-form with values in $\operatorname{Hom}(\mathbf{E}'', \mathbf{E}')$. The wedge products $\sigma^* \wedge \sigma$ and $\sigma \wedge \sigma^*$ can therefore be viewed as $(1,1)$-forms with values in, respectively, $\operatorname{Hom}(\mathbf{E}', \mathbf{E}')$ and $\operatorname{Hom}(\mathbf{E}'', \mathbf{E}'')$. It should be remarked that if the Hermitian pseudometric of $\mathbf{E}$ happens to be positive definite, for any two vector fields X, Y of type $(1,0)$, the bundle map $\sigma^* \wedge \sigma(X, \bar{Y})$ of $\mathbf{E}'$ is pointwise negative semidefinite, whereas $\sigma \wedge \sigma^*(X, \bar{Y})$ is pointwise positive semidefinite.

(7.14) **Lemma.** *Let* $s\colon \mathbf{E}'' \to \mathbf{E}$ *be the* C^∞ *inclusion defined by the pseudometric. Then*

$$\Theta' = j^{-1}(1 - s\,q)\,\Theta\, j + \sigma^* \wedge \sigma$$

$$\Theta'' = q\,\Theta\, s + \sigma \wedge \sigma^*.$$

Proof. In order to verify the first identity, I consider two holomorphic sections e_1, e_2 of $\mathbf{E}'$ and two holomorphic vector fields X, Y. In particular, $\Theta'(X, \bar{Y})\, e_1 = -\nabla'(\bar{Y})\, \nabla(X)\, e_1$, and $\Theta(X, \bar{Y})\, j\, e_1 = -\nabla(\bar{Y})\, \nabla(X)\, j\, e_1$. Using (7.8), as well as the orthogonality of $j\mathbf{E}'$ and $s\mathbf{E}''$, one finds

$$\begin{aligned}
(\Theta'(X, \bar{Y})\, e_1, e_2) &= -(\nabla'(\bar{Y})\, \nabla'(X)\, e_1, e_2) \\
&= -(j^{-1}(1 - s\,q)\, \nabla(\bar{Y})(1 - s\,q)\, \nabla(X)\, j\, e_1, e_2) \\
&= -(\nabla(\bar{Y})(1 - s\,q)\, \nabla(X)\, j\, e_1, j\, e_2) \\
&= -(\nabla(\bar{Y})\, \nabla(X)\, j\, e_1, j\, e_2) + (\nabla(\bar{Y})\, s\, \sigma(X)\, e_1, j\, e_2) \\
&= (\Theta(X, \bar{Y})\, j\, e_1, j\, e_2) - (s\, \sigma(X)\, e_1, \nabla(Y)\, j\, e_2) \\
&= (\Theta(X, \bar{Y})\, j\, e_1, j\, e_2) - (\sigma(X)\, e_1, \sigma(Y)\, e_2) \\
&= (j^{-1}(1 - s\,q)\, \Theta(X, \bar{Y})\, j\, e_1, e_2) - (\sigma^*(\bar{Y})\, \sigma(X)\, e_1, e_2) \\
&= (j^{-1}(1 - s\,q)\, \Theta(X, \bar{Y})\, j\, e_1, e_2) + (\sigma^* \wedge \sigma(X, \bar{Y})\, e_1, e_2).
\end{aligned}$$

This proves the first identity, because both sides of it are (1, 1)-forms with values in $\operatorname{Hom}(\mathbf{E}', \mathbf{E}')$. By a straightforward computation, one checks that the first fundamental form of the bundle $\mathbf{E}''^*$ in the dual exact sequence

(7.15) $$0 \longrightarrow \mathbf{E}''^* \xrightarrow{{}^t q} \mathbf{E}^* \xrightarrow{{}^t j} \mathbf{E}'^* \longrightarrow 0$$

is $-{}^t\sigma$ (${}^t\sigma(X)$ = dual of $\sigma(X)$). I now use the first formula in the dual exact sequence, taking into account (7.10):

$$-{}^t\Theta'' = -{}^t s\, {}^t\Theta\, {}^t q + {}^t\sigma^* \wedge {}^t\sigma.$$

Since ${}^t\sigma^* \wedge {}^t\sigma$ is the transpose of $-\sigma \wedge \sigma^*$ (the wedge product is skew symmetric!), by dualizing the last identity, I obtain the second assertion of the lemma.

These basic facts about the curvature of Hermitian vector bundles will now be applied in the context of a variation of Hodge structure $\{M, \mathbf{H}_{\mathbb{C}}, \mathbf{F}^p\}$, of weight k. For each p, I let $\mathbf{E}^p$ denote the quotient bundle $\mathbf{F}^p/\mathbf{F}^{p+1}$. As was remarked just after the definition of a variation of Hodge structure in §2, one has a natural isomorphism of C^∞ vector bundles

(7.16) $$\mathbf{E}^p \cong \mathbf{H}^{p,k-p} \qquad (= \mathbf{F}^p \cap \overline{\mathbf{F}}^{k-p}).$$

The bundle $\mathbf{H}_{\mathbb{C}}$, it should be recalled, carries a flat, nondegenerate bilinear form $\mathbf{S}$, of parity $(-1)^k$, which polarizes the Hodge structures on the fibres of $\mathbf{H}_{\mathbb{C}}$. In particular

$$(e_1, e_2) = i^{2p-k}\, \mathbf{S}(e_1, \bar{e}_2), \qquad e_i \in C^\infty(\mathbf{H}^{p,k-p}),$$

defines a positive definite Hermitian metric on $\mathbf{H}^{p,k-p}$. Transferring it to $\mathbf{E}^p$ via the isomorphism (7.16), one obtains a positive definite Hermitian metric for the holomorphic vector bundle $\mathbf{E}^p$. The second fundamental form of the subbundle $\mathbf{F}^p \subset \mathbf{H}_{\mathbb{C}}$, with respect to the flat connection on $\mathbf{H}_{\mathbb{C}}$, will be denoted by σ^p; it is a (1, 0)-form with values in $\operatorname{Hom}(\mathbf{F}^p, \mathbf{H}_{\mathbb{C}}/\mathbf{F}^p)$. In view of the requirement ii) in the definition of a variation of Hodge structure, for every field X and every $e \in C^\infty(\mathbf{F}^p)$, $\sigma^p(X)\, e$ is really a section of the subbundle

$$\mathbf{E}^{p-1} = \mathbf{F}^{p-1}/\mathbf{F}^p \subset \mathbf{H}_{\mathbb{C}}/\mathbf{F}^p.$$

Also, $\sigma^p(X)\, e = 0$ if $e \in C^\infty(\mathbf{F}^{p+1})$. Hence σ^p defines a (1, 0)-form τ^p, with values in $\operatorname{Hom}(\mathbf{E}^p, \mathbf{E}^{p-1})$. I want to record an immediate consequence of the definition:

(7.17) if $e \in C^\infty(\mathbf{H}_{\mathbb{C}})$ takes values in the subbundle $\mathbf{F}^p \subset \mathbf{H}_{\mathbb{C}}$, and if $e' \in C^\infty(\mathbf{E}^p)$ denotes its projection to $\mathbf{E}^p$, then $\tau^p e' = 0$ if and only if $\nabla e \in C^\infty(\mathbf{T}^*_{\mathbb{C}} \otimes \mathbf{F}^p)$.

From τ^p, one can construct the (0, 1)-form $(\tau^p)^*$, with values in $\mathrm{Hom}(\mathbf{E}^{p-1}, \mathbf{E}^p)$, by requiring that $(\tau^p)^*(X)$ shall be the adjoint of $\tau^p(\overline{X})$, relative to the positive definite Hermitian metrics of $\mathbf{E}^p$ and $\mathbf{E}^{p-1}$.

(7.18) **Lemma** (cf. Theorem (6.2) of [11]). *The curvature form of the metric connection on* $\mathbf{E}^p$ *is*

$$\Theta = -(\tau^p)^* \wedge \tau^p - \tau^{p+1} \wedge (\tau^{p+1})^*.$$

Proof. On $\mathbf{H}_{\mathbb{C}}$, I consider the flat Hermitian form

$$(e_1, e_2) = (-i)^k \mathbf{S}(e_1, \bar{e}_2), \qquad e_i \in C^\infty(\mathbf{H}_{\mathbb{C}}).$$

As a consequence of the polarization condition, it restricts non-degenerately to each subbundle $\mathbf{F}^p$. Hence the bundles $\mathbf{F}^p$, as well as the quotients $\mathbf{E}^p = \mathbf{F}^p/\mathbf{F}^{p+1}$, inherit Hermitian pseudometrics from $\mathbf{H}_{\mathbb{C}}$. With respect to the pseudometric on $\mathbf{H}_{\mathbb{C}}$, $\oplus\, \mathbf{H}^{p,k-p}$ becomes an orthogonal direct sum. Also, the induced metric on each $\mathbf{E}^p$ agrees with the previously defined, positive definite metric, but only up to sign: I shall refer to the former as $(\ ,\ )_i$; then, on $\mathbf{E}^p$,

$$(\ ,\ )_i = (-1)^p (\ ,\ ).$$

This alternating change of sign will turn out to be crucial. However, the change of sign affects neither the metric connection, nor its curvature form.

For the rest of the proof, X and Y will denote vector fields of type (1, 0). Let Θ^p be the curvature form of $\mathbf{F}^p$. Since $\mathbf{H}_{\mathbb{C}}$ carries a flat pseudometric, its curvature vanishes. Hence, and according to (7.14), if $e_1, e_2 \in C^\infty(\mathbf{F}^p)$,

$$\begin{aligned}(\Theta^p(X, \overline{Y})\, e_1, e_2) &= ((\sigma^p)^* \wedge \sigma^p(X, \overline{Y})\, e_1, e_2)\\ &= -(\sigma^p(X)\, e_1, \sigma^p(Y)\, e_2);\end{aligned}$$

the inner product on the right is computed with respect to the induced pseudometric of $\mathbf{H}_{\mathbb{C}}/\mathbf{F}^p$. Under the inclusion $\mathbf{E}^{p-1} \hookrightarrow \mathbf{H}_{\mathbb{C}}/\mathbf{F}^p$, $\sigma^p(X)\, e_1$ coincides with $\tau^p(X)\, q^p e_1$ (q^p = projection of $\mathbf{F}^p$ onto $\mathbf{E}^p$), and similarly for e_2. Thus

$$(\Theta^p(X, \overline{Y})\, e_1, e_2) = -(\tau^p(X)\, q^p e_1, \tau^p(Y)\, q^p e_2)_i,$$

with the right hand side computed in the induced metric on $\mathbf{E}^p$. Next, I consider the exact sequence

$$0 \longrightarrow \mathbf{F}^{p+1} \xrightarrow{j^p} \mathbf{F}^p \xrightarrow{q^p} \mathbf{E}^p \longrightarrow 0.$$

The definition of τ^{p+1} shows that $\tau^{p+1} \circ q^{p+1}$ = second fundamental form of $\mathbf{F}^{p+1}$. The pseudometric determines a C^∞ splitting s^p: $\mathbf{E}^p \to \mathbf{F}^p$, which is adjoint to q^p, and which is an isometry, relative to $(\ ,\ )_i$. I use the second identity of (7.14) to compute the curvature Θ of $\mathbf{E}^p$: for $e_i \in C^\infty(\mathbf{E}^p)$,

$$\begin{aligned}(\Theta(X, \bar{Y})\, e_1, e_2)_i &= (q^p\, \Theta^p(X, \bar{Y})\, s^p\, r_1, e_2)_i \\ &\quad + ((\tau^{p+1} \circ q^{p+1}) \wedge (\tau^{p+1} \circ q^{p+1})^*(X, \bar{Y})\, e_1, e_2)_i \\ &= -(\tau^p(X)\, e_1, \tau^p(Y)\, e_2)_i \\ &\quad + ((\tau^{p+1} \circ q^{p+1})^*(\bar{Y})\, e_1, (\tau^{p+1} \circ q^{p+1})^*(\bar{X})\, e_2)_i \\ &= -(\tau^p(X)\, e_1, \tau^p(Y)\, e_2)_i \\ &\quad + (\tau^{p+1}(Y)^*\, e_1, \tau^{p+1}(X)^*\, e_2)_i\end{aligned}$$

($(q^{p+1})^*$: $\mathbf{E}^{p+1} \to \mathbf{F}^{p+1}$ is an isometry!). In this computation, the superscript * always designates the adjoint, relative to the induced pseudometrics on the bundles $\mathbf{E}^p$. Since these metrics agree with the positive definite metrics at least up to sign, so do the resulting adjoints. In the expression $(\tau^{p+1}(Y)^*\, e_1, \tau^{p+1}(X)^*\, e_2)_i$, the same adjoint occurs twice, so that the potential sign changes cancel. In other words,

$$\begin{aligned}(\Theta(X, \bar{Y})\, e_1, e_2)_i &= -(\tau^p(X)\, e_1, \tau^p(Y)\, e_2)_i \\ &\quad + ((\tau^{p+1})^*(\bar{Y})\, e_1, (\tau^{p+1})^*(\bar{X})\, e_2)_i,\end{aligned}$$

where now $(\tau^{p+1})^*$ is given the same meaning as in the statement of the lemma. The inner products on the right are computed in $\mathbf{E}^{p-1}$ and $\mathbf{E}^{p+1}$, the one on the left in $\mathbf{E}^p$. Because of the alternating changes in sign, in terms of the positive definite metrics $(\ ,\ )$, the identity becomes

$$\begin{aligned}(\Theta(X, \bar{Y})\, e_1, e_2) &= (\tau^p(X)\, e_1, \tau^p(Y)\, e_2) \\ &\quad - ((\tau^{p+1})^*(\bar{Y})\, e_1, (\tau^{p+1})^*(\bar{X})\, e_2) \\ &= -((\tau^p)^* \wedge \tau^p(X, \bar{Y})\, e_1, e_2) \\ &\quad - (\tau^{p+1} \wedge (\tau^{p+1})^*(X, Y)\, e_1, e_2),\end{aligned}$$

so that $\Theta = -(\tau^p)^* \wedge \tau^p - \tau^{p+1} \wedge (\tau^{p+1})^*$, as asserted.

(7.19) **Lemma** (cf. (5.5) and (5.8) of [11]). *Let e be a holomorphic section of $\mathbf{H}_{\mathbb{C}}$, with values in the subbundle $\mathbf{F}^p$, and such that ∇e takes values in $\mathbf{T}_{\mathbb{C}}^* \otimes \mathbf{F}^p$. Let e' be the induced section of $\mathbf{E}^p$. Then $\varphi = (e', e')$ is a plurisubharmonic function. Furthermore, if φ happens to be constant, there exists a flat section e_1 with values in the C^∞ subbundle $\mathbf{H}^{p,k-p} \subset \mathbf{H}_{\mathbb{C}}$, and a holomorphic section e_2 of $\mathbf{H}_{\mathbb{C}}$, with values in $\mathbf{F}^{p+1}$, so that $e = e_1 + e_2$.*

Proof. According to (7.13) and (7.18), for any (1, 0)-vector field X,

$$\begin{aligned}\partial\bar{\partial}\varphi(X,\bar{X}) &= (\nabla_{\mathbf{E}^p}(X)\,e', \nabla_{\mathbf{E}^p}(X)\,e') \\ &\quad + ((\tau^p)^* \wedge \tau^p(X,\bar{X})\,e', e') + (\tau^{p+1} \wedge (\tau^{p+1})^*(X,\bar{X})\,e', e') \\ &= (\nabla_{\mathbf{E}^p}(X)\,e', \nabla_{\mathbf{E}^p}(X)\,e') \\ &\quad - (\tau^p(X)\,e', \tau^p(X)\,e') + ((\tau^{p+1})^*(\bar{X})\,e', (\tau^{p+1})^*(\bar{X})\,e')\end{aligned}$$

($\nabla_{\mathbf{E}^p}$ = metric connection on $\mathbf{E}^p$). Because of (7.17) and the hypotheses, $\tau^p(X)\,e'=0$. Thus

$$\begin{aligned}\partial\bar{\partial}\varphi(X,\bar{X}) &= (\nabla_{\mathbf{E}^p}(X)\,e', \nabla_{\mathbf{E}^p}(X)\,e') \\ &\quad + ((\tau^{p+1})^*(\bar{X})\,e', (\tau^{p+1})^*(\bar{X})\,e') \geqq 0,\end{aligned}$$

which means that φ is plurisubharmonic. If φ is a constant function, the Levi form vanishes identically. In this case, the two nonnegative terms on the right must vanish, so that

$$\nabla_{\mathbf{E}^p}\,e'=0 \quad \text{and} \quad (\tau^{p+1})^*\,e'=0$$

(e' is a holomorphic section; hence $\nabla_{\mathbf{E}^p}(\bar{X})\,e'=0$ for any (1, 0)-vector field X!). With respect to the Hermitian pseudometric on $\mathbf{H}_{\mathbb{C}}$ which was defined in the proof of Lemma (7.18), and which will be used throughout the remainder of this proof, one has an orthogonal decomposition

$$\mathbf{F}^p = \mathbf{F}^{p+1} \oplus \mathbf{H}^{p,k-p}.$$

In particular, there exist C^∞ sections e_1, e_2 of $\mathbf{H}_{\mathbb{C}}$, with values in the subbundles $\mathbf{H}^{p,k-p}$ and $\mathbf{F}^{p+1}$, respectively, which add up to e. In view of (7.9), $\nabla_{\mathbf{E}^p}\,e'$ coincides with the image of $\nabla_{\mathbf{F}^p}\,e_1$ under the projection $\mathbf{F}^p \to \mathbf{E}^p$. On the other hand, because of (7.8), $\nabla_{\mathbf{F}^p}\,e_1$ can be computed by orthogonally projecting ∇e_1 onto $\mathbf{F}^p$ (∇ = flat connection of $\mathbf{H}_{\mathbb{C}}$). Therefore, the vanishing of $\nabla_{\mathbf{E}^p}\,e'$ becomes equivalent to:

(7.20) under the orthogonal projection of $\mathbf{H}_{\mathbb{C}}$ onto its subbundle $\mathbf{H}^{p,k-p}$, ∇e_1 goes to zero.

Now let X be a (1, 0)-vector field, f a C^∞ section of $\mathbf{F}^{p+1}$, f' its image in $C^\infty(\mathbf{E}^{p+1})$. Then $\sigma^{p+1}(X)f$ projects to $\tau^{p+1}(X)f'$ under $\mathbf{F}^p \to \mathbf{E}^p$. Since e_1 is an orthogonal lifting of the section e' of $\mathbf{E}^p$ back to $\mathbf{H}_{\mathbb{C}}$, and since $e_1 \perp \mathbf{F}^{p+1}$,

$$\begin{aligned}(\nabla(\bar{X})\,e_1, f) &= -(e_1, \nabla(X)\,f) + \bar{X}(e_1, f) \\ &= -(e_1, \nabla(X)\,f) = -(e, \sigma^{p+1}(X)\,f) \\ &= \pm(e', \tau^{p+1}(X)\,f) = \pm((\tau^{p+1})^*(\bar{X})\,e', f) = 0.\end{aligned}$$

In the final two expressions, the inner product is computed with respect to the positive definite metric of $\mathbf{E}^p$, and this accounts for the possible change in sign. Consequently, $\nabla(\bar{X})\,e_1 \perp \mathbf{F}^{p+1}$. Next, I consider a holomorphic section f of $\mathbf{F}^{p+1}$, and again a (1, 0)-vector field X. Then $e_1 \perp f$, and $\nabla(\bar{X})\,f=0$, so that

$$\big(\nabla(X)\,e_1, f\big) = -\big(e_1, \nabla(\bar{X})\,f\big) + X(e_1, f) = 0.$$

At each point of M, the local holomorphic sections of $\mathbf{F}^{p+1}$ span the fibre of $\mathbf{F}^{p+1}$. I may therefore conclude that $\nabla(X)\,e_1 \perp \mathbf{F}^{p+1}$. Together with the previous orthogonality statement and (7.20), this gives

$$\text{(7.21)} \qquad \nabla e_1 \perp \mathbf{F}^p.$$

Because of the condition ii) in the definition of a variation of Hodge structure, ∇e_2 takes values in $\mathbf{T}^*_{\mathbb{C}} \otimes \mathbf{F}^p$. According to the hypotheses, the same is true of ∇e, and hence also of $\nabla e_1 = \nabla e - \nabla e_2$. Since the Hermitian pseudometric of $\mathbf{H}_{\mathbb{C}}$ restricts nondegenerately to $\mathbf{F}^p$, (7.21) is now only possible if $\nabla e_1 = 0$, i.e. if e_1 is a flat section. As the difference of a holomorphic and a flat section, e_2 must also be holomorphic.

I can now prove the central technical result of this section. As before, I consider a variation of Hodge structure $\{M, \mathbf{H}_{\mathbb{C}}, \mathbf{F}^p\}$. The following theorem was originally proven by Griffiths, for the case of a compact base space M (Theorem 7.1 of [11]). Deligne then proved it in the geometric setting, for algebraic families with quasi-projective parameter space ((4.1.2) of [6]). It was remarked already that the proof which I shall give below is essentially that of Griffiths, coupled with the results of §6.

(7.22) **Theorem.** *If M can be embedded as a Zariski open subset in a compact analytic space, for any flat, global section e of $\mathbf{H}_{\mathbb{C}}$, the Hodge (p, q)-components of e are also flat.*

Proof. One can certainly express e as a sum $e = \sum_p e_p$, with $e_p \in C^\infty(\mathbf{H}^{p,k-p})$. Let l be the least integer such that $e_p = 0$ for $p > l$. I shall argue by induction on l. For the lowest possible value of l, there is nothing to prove. If l is arbitrary, I consider the function

$$\varphi = (e', e') = i^{2l-k}\,\mathbf{S}(e_l, e_l)$$

(e' = image of e under the projection $\mathbf{F}^l \to \mathbf{E}^l$). According to the previous lemma, φ is plurisubharmonic; also, if φ were known to be constant, e_l would have to be flat, hence also $e - e_l$, and this would complete the induction step. A Zariski open subset of a compact analytic space does not admit any nonconstant, bounded, plurisubharmonic functions [22]. It therefore suffices to prove the boundedness of φ. Although it is not strictly necessary to appeal to Hironaka's desingularization theorem, it will be convenient to do so. I assume that M lies as a Zariski open subset in a compact manifold $\bar{M}$, with complement $\bar{M}-M$, which is a divisor with no singularities other than normal crossings. One can choose an open neighborhood $\mathscr{U}$ of $\bar{M}-M$ in $\bar{M}$, having the following property: for every point of M, there exists an embedded copy of the unit disc $\Delta\subset\bar{M}$, passing through the point in question, intersecting $\bar{M}-M$ only in the origin or not at all, and such that $\partial\Delta\cap\mathscr{U}=\emptyset$. As will be shown presently, the restriction of φ to such a disc Δ remains bounded from above near the origin. Hence the restriction extends as a plurisubharmonic function to all of Δ [22]. By the maximum principle, $\varphi|_\Delta$ is bounded from above by its maximum on $\partial\Delta$, which in turn is bounded by the maximum of φ on the compact subset $\bar{M}-\mathscr{U}$ of M. Let then $\Delta\subset\bar{M}$ be an embedded disc, with $\Delta\cap(\bar{M}-M)=\{0\}$, and let $\Delta^*=\Delta-\{0\}$. On this copy of the punctured disc, e restricts to a well defined (i.e. single valued), flat section of $\mathbf{H}_{\mathbb{C}}$. Thus one can apply (6.7′), which asserts that the sum of the squares of the lengths of the Hodge components e_p remains bounded mear the puncture. In particular, the squared length of e_l, namely φ, stays bounded. This completes the proof.

(7.23) **Corollary.** *Under the hypotheses of the theorem, if a flat global section e of $\mathbf{H}_{\mathbb{C}}$ is of pure Hodge type (p,q) at some point, then it has Hodge type (p,q) everywhere.*

As was remarked in § 2, the operations of tensor products, symmetric and exterior products, Hom, and duality can be performed on polarized Hodge structures. Also, the category of flat vector bundles over a given space is closed under these operations. As a result, one can do the operations on variations of Hodge structure with a given base M: the only point which has to be checked is the property ii) in the definition of a variation of Hodge structure, and this presents no major problem.

By a morphism between two variations of Hodge structure $\{M,\mathbf{H}_{i,\mathbb{C}},\mathbf{F}_i^p\}$ with the same base space M $(i=1,2)$, I shall mean a morphism of the underlying flat bundles $\mathbf{H}_{i,\mathbb{C}}$, which induces a morphism of the Hodge structures corresponding to every $t\in M$. In other words, a morphism is a global, flat section of $\mathrm{Hom}(\mathbf{H}_{1,\mathbb{C}},\mathbf{H}_{2,\mathbb{C}})$, everywhere of Hodge type $(0,0)$, and rational, relative to the flat lattice bundles $\mathbf{H}_{i,\mathbb{Z}}$. According to (7.23), if M lies as a Zariski open set in a compact analytic space, the condition on the Hodge type needs to be checked only at one point. Without any condition on M, the rationality of a flat section at a single point implies the rationality everywhere. Hence, under the hypothesis of (7.22), for any $t\in M$, the morphisms between the two variations of Hodge structures correspond in a 1:1 manner to the $\pi_1(M,t)$-invariant morphisms between the two Hodge structures on the fibres over t. In particular, this proves the

(7.24) **Rigidity Theorem** (cf. (7.4) of [11]). *Let* $\{M, \mathbf{H}_{i,\mathbb{C}}, \mathbf{F}_i^p\}$ *be two variations of Hodge structure, with base M, which is Zariski open in a compact analytic space. If for some point* $t \in M$ *the two Hodge structures at t are isomorphic, and if the isomorphism preserves the action of* $\pi_1(M)$, *then the isomorphism extends to an isomorphism of the variations of Hodge structure.*

Loosely speaking, the period mapping of a variation of Hodge structure is completely determined by its value at a single point, plus the action of the fundamental group – provided the base can be compactified. For families of Abelian varieties, the rigidity theorem is due to Grothendieck [15]. Borel and Narasimham proved (7.24), in effect, whenever the classifying space D is Hermitian symmetric [2].

Deligne has deduced a remarkable fact from the flatness Theorem (7.22): under the usual hypothesis on the base, the monodromy group of a variation of Hodge structure acts semisimply:

(7.25) **Theorem** (Deligne, cf. 4.2.6 of [6]). *Let* $\{M, \mathbf{H}_{\mathbb{C}}, \mathbf{F}^p\}$ *be a variation of Hodge structure, such that the base M can be embedded into a compact analytic space as a Zariski open set. Then the representation* (3.23) *is completely reducible.*

I shall outline the proof. As usual, $\tilde{M}$ will refer to the universal covering of M, and $H_{\mathbb{C}}$ to the fibre of the pullback of $\mathbf{H}_{\mathbb{C}}$ to $\tilde{M}$. I keep fixed a point $t \in M$, and I choose some point lying above t in $\tilde{M}$. This choice determines an isomorphism between $H_{\mathbb{C}}$ and the fibre of $\mathbf{H}_{\mathbb{C}}$ at t, with which one can transfer the Hodge structure at t to $H_{\mathbb{C}}$. I shall denote the Weil operator of the Hodge structure on $H_{\mathbb{C}}$ by C. According to (7.22),

(7.26) the space of $\pi_1(M)$-invariants in $H_{\mathbb{C}}$ is C-stable.

Now let $L \subset H_{\mathbb{C}}$ be a one-dimensional, $\pi_1(M)$-invariant subspace, such that the action of $\pi_1(M)$ on L factors through a finite group. In a suitable symmetric product of $H_{\mathbb{C}}$, the line generated by L becomes trivial under the action of $\pi_1(M)$. If one applies (7.26) to the corresponding symmetric product of the variation of Hodge structure, it follows that the line generated by CL in the symmetric product of $H_{\mathbb{C}}$ must be $\pi_1(M)$-trivial. Hence, under the assumption at the beginning of the paragraph,

(7.27) CL is invariant under $\pi_1(M)$.

Because $\pi_1(M)$ preserves the lattice $H_{\mathbb{Z}} \subset H_{\mathbb{C}}$, if a line $L \subset H_{\mathbb{C}}$ is rationally defined and $\pi_1(M)$-invariant, $\pi_1(M)$ can only act on it by ± 1. In particular, (7.27) holds in this situation.

(7.28) **Lemma.** *Let V_1, V_2 be $\pi_1(M)$-invariant subspaces of $H_{\mathbb{C}}$, which do not intersect, and whose sum is defined over $\mathbb{Q}$. Then $\pi_1(M)$ leaves CV_i invariant, $i=1,2$.*

Proof. I set $n_i = \dim V_i$, $V = V_1 \oplus V_2$, $n = n_1 + n_2$. Since

$$\Lambda^n V \cong \Lambda^{n_1} V_1 \otimes \Lambda^{n_2} V_2$$

is an isomorphism of $\pi_1(M)$-modules, the remark below (7.27) shows that $\pi_1(M)$ can act only by ± 1 on the one-dimensional $\pi_1(M)$-submodule $\Lambda^{n_1} V_1 \otimes \Lambda^{n_2} V_2$ of $\Lambda^{n_1} H_{\mathbb{C}} \otimes \Lambda^{n_2} H_{\mathbb{C}}$. Applying (7.27) to the induced variation of Hodge structure on $\Lambda^{n_1} \mathbf{H}_{\mathbb{C}} \otimes \Lambda^{n_2} \mathbf{H}_{\mathbb{C}}$, one finds: $\Lambda^{n_i} CV_i$, and hence also CV_i, are $\pi_1(M)$-invariant, for $i=1,2$.

Deligne's argument for the complete reducibility of the $\pi_1(M)$-action on $H_{\mathbb{C}}$ now proceeds as follows. Let l be the least dimension of all $\pi_1(M)$-invariant subspaces, and W the span of all l-dimensional $\pi_1(M)$-submodules of $H_{\mathbb{C}}$. As a sum of irreducible submodules, W must be semi-simple. Thus each l-dimensional, $\pi_1(M)$-irreducible subspace V of $H_{\mathbb{C}}$ has a $\pi_1(M)$-invariant complement in W. Also, since $\pi_1(M)$ acts rationally, W is defined over $\mathbb{Q}$. Hence (7.28) gives the containment $CV \subset W$, for any such V. These subspaces V span W, so that $CW = W$. Consequently, the polarization form S restricts nondegenerately to W. Both $\pi_1(M)$ and C preserve the orthogonal complement of W. The same argument can therefore be applied again to $W^{\perp}$, and one can continue by induction.

Instead of the Weil operator C, Deligne works with an action of S^1, defined by $\rho(e^{i\theta})\, v = e^{i(p-q)\theta}\, v$, for $e^{i\theta} \in S^1$, $v \in H^{p,q}$. This does not really affect the argument, and gives the following additional information: the subspace W, and the other subspaces constructed inductively, correspond to a sub-variation of Hodge structure.

§ 8. Proof of the Nilpotent Orbit Theorem

I shall begin with some preliminary remarks and constructions. The choice of a base point $o \in D$ determines a reference Hodge structure on $H_{\mathbb{C}}$,

$$H_{\mathbb{C}} = \oplus H_0^{p,q}, \qquad p+q=k,$$

which in turn leads to a Hodge structure

$$\mathfrak{g} = \oplus \mathfrak{g}^{p,-p}$$

on $\mathfrak{g}$, of weight zero, as described in § 3. Let $\theta: \mathfrak{g} \to \mathfrak{g}$ be the Weil operator of this Hodge structure; explicitly

$$\theta X = (-1)^p X \qquad \text{for } X \in \mathfrak{g}^{p,-p}.$$

Then θ is an automorphism of $\mathfrak{g}$, as follows from (3.12), it is involutive, and is defined over $\mathbb{R}$. The $+1$ and -1 eigenspaces of θ will be denoted by, respectively, $\mathfrak{k}$ and $\mathfrak{p}$, and I set

$$\mathfrak{k}_0=\mathfrak{k}\cap\mathfrak{g}_0, \qquad \mathfrak{p}_0=\mathfrak{p}\cap\mathfrak{g}_0.$$

The properties of θ which were just mentioned imply

$$\begin{gathered}\mathfrak{g}=\mathfrak{k}\oplus\mathfrak{p}, \qquad \mathfrak{g}_0=\mathfrak{k}_0\oplus\mathfrak{p}_0,\\ [\mathfrak{k},\mathfrak{k}]\subset\mathfrak{k}, \quad [\mathfrak{k},\mathfrak{p}]\subset\mathfrak{p}, \qquad [\mathfrak{p},\mathfrak{p}]\subset\mathfrak{k}.\end{gathered} \tag{8.1}$$

Composing θ with complex conjugation (relative to $\mathfrak{g}_0$), one obtains a conjugate linear, involutive automorphism of $\mathfrak{g}$, whose fixed point set is

$$\mathfrak{m}_0=\mathfrak{k}_0\oplus i\mathfrak{p}_0. \tag{8.2}$$

It follows that $\mathfrak{m}_0$ is a real form for $\mathfrak{g}$, i.e. an $\mathbb{R}$-subalgebra such that

$$\mathfrak{g}=\mathfrak{m}_0\oplus i\,\mathfrak{m}_0.$$

Now let C be the Weil operator on $H_{\mathbb{C}}$ which corresponds to the reference Hodge filtration. Then

$$(u,v)=S(Cu,\bar{v}), \qquad u,v\in H_{\mathbb{C}}, \tag{8.3}$$

defines a Hermitian inner product on $H_{\mathbb{C}}$. As can be checked directly, C is an element of the group $G_{\mathbb{R}}$, whose adjoint action on $\mathfrak{g}$ coincides with θ. Thus $\mathfrak{m}_0$ can be described as the intersection of $\mathfrak{g}$ with the Lie algebra of all skew Hermitian transformations, relative to the inner product (8.3). One may conclude that the connected subgroup $M\subset G_{\mathbb{C}}$, which corresponds to the subalgebra $\mathfrak{m}_0\subset\mathfrak{g}$, coincides with the connected component of the identity in the intersection of $G_{\mathbb{C}}$ with the unitary group. In particular, this forces M to be compact. Also, a compact real form in a connected, complex semisimple Lie group is always connected and is its own normalizer; hence M equals the full intersection of $G_{\mathbb{C}}$ and the unitary group.

The intersection

$$K=M\cap G_{\mathbb{R}}$$

is a compact subgroup of $G_{\mathbb{R}}$, whose Lie algebra $\mathfrak{k}_0=\mathfrak{m}_0\cap\mathfrak{g}_0$ consists of skew Hermitian transformations, whereas the complement $\mathfrak{p}_0$ of $\mathfrak{k}_0$ in $\mathfrak{g}_0$ consists of Hermitian transformations. Now one can appeal to standard arguments on semisimple Lie groups (e.g. [16], Chapter III), to conclude: the connected component of the identity in K forms a maximal compact subgroup of the connected component of the identity in $G_{\mathbb{R}}$. The maximal compactness of K in $G_{\mathbb{R}}$ would follow, if it were known that K meets every component of $G_{\mathbb{R}}$. In order to establish this latter fact, I shall make use of Lemma 4.3 in Chapter IX of [16]. According to the description of M, K is equal to the intersection of $G_{\mathbb{R}}$ and the unitary group of the inner product (8.3). Also, the adjoint of $g \in G_{\mathbb{C}}$, relative to the inner product, is the complex conjugate of $C g C^{-1}$; since C lies in $G_{\mathbb{R}}$, $G_{\mathbb{R}}$ must thus be closed under adjoints. Hence the lemma applies, which asserts the connectedness of $G_{\mathbb{R}}/K$. I may deduce:

(8.4) K is a maximal compact subgroup of $G_{\mathbb{R}}$ and it meets every connected component of $G_{\mathbb{R}}$.

At this point, one should recall the identifications $\check{D} \cong G_{\mathbb{C}}/B$, $D \cong G_{\mathbb{R}}/V$.

(8.5) **Lemma.** *The groups M and B have V as their intersection, and K contains V.*

Proof. Since $K = G_{\mathbb{R}} \cap M$ and $V \subset G_{\mathbb{R}}$, the latter assertion follows from the former. The elements of $M \cap B$ leave the subspaces $F_0^p = \oplus_{s \geqq p} H_0^{s,k-s}$ invariant, and they preserve the inner product. With respect to the inner product, the reference Hodge decomposition of $H_{\mathbb{C}}$ is orthogonal, so that the elements of $M \cap B$ must also leave the subspaces $H_0^{p,q}$ invariant. On the other hand, every $g \in M$ is self-adjoint, i.e.

$$C g C^{-1} \bar{v} = \overline{g v} \qquad \text{for } v \in H_{\mathbb{C}}.$$

If $g \in M$ also happens to preserve the subspaces $H_0^{p,q}$, it commutes with C. It must then respect the real structure of $H_{\mathbb{C}}$, and thus belongs to $G_{\mathbb{R}}$. Hence $M \cap B \subset G_{\mathbb{R}}$. Conversely, any $g \in V = G_{\mathbb{R}} \cap B$ leaves the subspaces $H_0^{p,q}$ invariant (cf. § 3), therefore commutes with C and preserves the inner product (8.3). This gives the containment $V = G_{\mathbb{R}} \cap B \subset M$. Finally, then, $M \cap B = G_{\mathbb{R}} \cap B = V$.

In view of the lemma, the M-orbit of the identity coset in $G_{\mathbb{C}}/B \cong \check{D}$ becomes naturally isomorphic to M/V. Since $G_{\mathbb{R}}$ and M have the same dimension (both are real forms in $G_{\mathbb{C}}$), the dimensions of M/V and $D \cong G_{\mathbb{R}}/V$ also agree, so M/V must be an open orbit. Because of the compactness of M, the orbit is also closed. Hence M operates transitively on $\check{D}$, and this gives the identification

(8.6) $$D \cong M/V.$$

As usual, I regard the elements of $\mathfrak{g}$ as linear transformations on the vector space $H_{\mathbb{C}}$. The bilinear form

(8.7) $$B(X, Y) = \operatorname{trace} XY, \qquad X, Y \in \mathfrak{g},$$

is clearly symmetric, invariant under the adjoint action of $G_{\mathbb{C}}$ on $\mathfrak{g}$, and defined over $\mathbb{R}$, relative to the real structure $\mathfrak{g}_0 \subset \mathfrak{g}$.

(8.8) **Lemma.** *The bilinear form* $-B(\ ,\ )$ *polarizes the Hodge structure* $\mathfrak{g}=\oplus\,\mathfrak{g}^{p,-p}$.

Proof. For $X\in\mathfrak{g}^{p,-p}$, $Y\in\mathfrak{g}^{q,-q}$, with $p\neq -q$, the linear transformation XY shifts the subspaces $H_0^{p,q}$ of $H_{\mathbb{C}}$ nontrivially. It follows that XY has zero trace, and that $B(X,Y)=0$. With respect to the inner product (8.3), the elements of $\mathfrak{m}_0$ are skew Hermitian. This makes the eigenvalues of any $X\in\mathfrak{m}_0$ all purely imaginary, and if $X\neq 0$, not all eigenvalues can vanish. Thus $B(\ ,\ )$ is negative definite on $\mathfrak{m}_0$. By complex extension, since complex conjugation composed with θ is the conjugation operator relative to the real form $\mathfrak{m}_0\subset\mathfrak{g}$, one finds $-B(\theta X,\bar{X})>0$ for $X\in\mathfrak{g}$, $X\neq 0$. This concludes the proof.

As a consequence of (8.8), the inner product

$$(X,Y)=-B(\theta X,\bar{Y}),\qquad X,Y\in\mathfrak{g}, \tag{8.9}$$

turns $\mathfrak{g}$ into a complex Hilbert space and $\mathfrak{g}_0$ into a real Hilbert subspace. Under the adjoint action, the Lie algebra $\mathfrak{m}_0$ operates in a skew Hermitian manner, which makes the adjoint action of M on $\mathfrak{g}$ unitary. For emphasis:

(8.10) the adjoint actions of V, K, and M leave the inner product on $\mathfrak{g}$ invariant.

The norm which results from the inner product will be denoted by double bars.

At this point, it is useful to insert an observation for later reference:

(8.11) **Lemma.** *If* $T\in\mathfrak{g}_0$ *is nilpotent, and if* T *is expressed as* $T=Y+Z$, *with* $Y\in\mathfrak{k}_0$ *and* $Z\in\mathfrak{p}_0$, *then* $\|T\|=\sqrt{2}\,\|Y\|=\sqrt{2}\,\|Z\|$.

Proof. According to (8.8), $\mathfrak{k}$ and $\mathfrak{p}$ are orthogonal, so that $\|T\|^2=\|Y\|^2+\|Z\|^2$. On the other hand, the nilpotency of T forces T^2 to have zero trace, which leads to

$$\begin{aligned}0&=B(T,T)=B(T,\bar{T})=B(Y+Z,\bar{Y}+\bar{Z})\\&=B(Y,\bar{Y})+B(Z,\bar{Z})=B(\theta Y,\bar{Y})-B(\theta Z,\bar{Z})\\&=-\|Y\|^2+\|Z\|^2.\end{aligned}$$

The statement follows.

The holomorphic tangent space to $\check{D}\cong G_{\mathbb{C}}/B$ at the identity coset is naturally isomorphic to the quotient $\mathfrak{g}/\mathfrak{b}$; the quotient, in turn, is isomorphic as a V-module to the orthogonal complement of $\mathfrak{b}$ in $\mathfrak{g}$, namely

$$\mathfrak{c}=\oplus_{p>0}\,\mathfrak{g}^{p,-p}.$$

From $\mathfrak{g}$, $\mathfrak{c} \cong \mathfrak{g}/\mathfrak{b}$ inherits a V-invariant inner product. When this inner product is translated via $G_{\mathbb{R}}$ and M, one obtains, respectively, a $G_{\mathbb{R}}$-invariant Hermitian structure $h(\ ,\ )$ on $D \cong G_{\mathbb{R}}/V$, and an M-invariant Hermitian structure $h_M(\ ,\ )$ on $\check{D} \cong M/V$. The corresponding distance functions shall be denoted by $d(\ ,\ )$ and $d_M(\ ,\ )$; the value ∞ is allowed for $d(\ ,\ )$, since D need not be connected.

For $g \in G_{\mathbb{C}}$, I let $l(g)$: $\check{D} \to \check{D}$ denote left translation by g, and $l(g)_*$ the differential of this mapping. It will be necessary to estimate the operator norms of the linear transformations $l(g)_*$, with respect to the Hermitian metric h_M. Any given $g \in G_{\mathbb{C}}$ can be expressed as $g = mb$, with $m \in M$ and $b \in B$, because M operates transitively on $G_{\mathbb{C}}/B$. Then

$$l(g)_* = l(m)_* \circ l(b)_* .$$

At the identity coset, the action of $l(b)_*$ on the holomorphic tangent space corresponds to the action of $\operatorname{Ad} b$ on $\mathfrak{g}/\mathfrak{b}$. The operator norm of $\operatorname{Ad} b$ on this quotient is bounded by that of $\operatorname{Ad} b$ acting on $\mathfrak{g}$, and since $b = m^{-1} g$, with m operating unitarily on $\mathfrak{g}$, it is also bounded by the operator norm of $\operatorname{Ad} g$ on $\mathfrak{g}$. With respect to h_M, $l(m)_*$ becomes an isometry; hence: relative to the Hermitian metric h_M, the operator norm of the linear transformation $l(g)_*$ from the holomorphic tangent space at the identity coset to that at the g-coset is bounded by the operator norm of $\operatorname{Ad} g$, measured with respect to the inner product (8.9) on $\mathfrak{g}$. The same estimate must then hold at any point of $\check{D}$, for M operates unitarily on $\mathfrak{g}$ and transitively and isometrically on $\check{D}$. This proves:

(8.12) **Lemma.** *At each point of $\check{D}$, the operator norm of the linear transformation $l(g)_*$, measured relative to the Hermitian structure h_M, is bounded by the operator norm of* $\operatorname{Ad} g$ *acting on* $\mathfrak{g}$.

The statement (8.12) implies a relation between the two metrics h and h_M on D. Let X be a holomorphic tangent vector at the point $gV \in G_{\mathbb{R}}/V \cong D$, with $g \in G_{\mathbb{R}}$. Then there exists a holomorphic tangent vector Z at the identity coset, such that $l(g)_* Z = X$, or equivalently $Z = l(g^{-1})_* X$. By construction, the two metrics h and h_M agree at the identity coset. With respect to h, both $l(g)_*$ and $l(g^{-1})_*$ are isometries; with respect to h_M, they are not, but (8.12) limits their dilation. Hence:

(8.13) **Corollary.** *If the point $a \in D$ is the g-translate of the base point, with $g \in G_{\mathbb{R}}$, and if X is a holomorphic tangent vector at a, then*

$$h(X, X)^{\frac{1}{2}} \leqq \|\operatorname{Ad} g^{-1}\| \, h_M(X, X)^{\frac{1}{2}},$$
$$h_M(X, X)^{\frac{1}{2}} \leqq \|\operatorname{Ad} g\| \, h(X, X)^{\frac{1}{2}};$$

in this statement, double bars denote the operator norm, relative to the inner product (8.9) *on* $\mathfrak{g}$.

In order to prove (4.9), I consider a particular holomorphic, locally liftable map

$$\Phi\colon \Delta^* \to \Gamma \backslash D,$$

which has horizontal local liftings. Let τ be the universal covering map (4.2) of U onto Δ^*, $\tilde{\Phi}\colon U \to D$ a lifting of $\Phi \circ \tau$ to D, and $\gamma \in \Gamma$ an element with the transformation property (4.4). The map $\tilde{\Psi}\colon U \to \check{D}$ is defined by (4.7), and pushing $\tilde{\Psi}$ down to Δ^* via τ, one obtains the map $\Psi\colon \Delta^* \to \check{D}$. According to (3.17), $\tilde{\Phi}$ is uniformly bounded relative to the Poincaré metric on U and the $G_{\mathbb{R}}$-invariant metric on D, both globally and infinitesimally; the particular value of the bound depends only on the normalization of the metrics. This boundedness property of $\tilde{\Phi}$ is the most important ingredient of the proof of (4.9). In rough outline, the argument proceeds as follows: Under the translation $z \mapsto z+m$, $\tilde{\Phi}$ gets transformed by the element $\gamma_u^m = \exp(mN)$ of Γ. Because of the boundedness of $\tilde{\Phi}$, and because N is nilpotent, it can be deduced that $\tilde{\Phi}$ becomes asymptotically tangent to the holomorphic vector field determined by N, as the imaginary part of z tends to infinity. In view of its definition, $\tilde{\Psi}$ can then vary only little for large values of $\operatorname{Im} z$, and this forces Ψ to have a removable singularity. The lemmas below will be stated in slightly greater generality than would be necessary for (4.9) alone, in preparation for the proof of (4.12).

For each $z \in U$, I choose an element $g(z) \in G_{\mathbb{R}}$ whose V-coset represents the point $\tilde{\Phi}(z) \in D \cong G_{\mathbb{R}}/V$. Although $g(z)$ is determined only up to right multiplication by an element of V, which operates unitarily on $\mathfrak{g}$, the following statement is meaningful:

(8.14) **Lemma.** *There exist positive constants α, β, which depend only on the choice of base point in D and on the integer m, such that* $\operatorname{Im} z > \alpha$ *implies*

$$\|\operatorname{Ad} g(z)^{-1} N\| \leqq \beta (\operatorname{Im} z)^{-1}.$$

Proof. Two preliminary remarks are necessary. The tangent space to $G_{\mathbb{R}}/K$ at the identity coset is isomorphic as K-module to $\mathfrak{g}_0/\mathfrak{k}_0$. This quotient inherits an inner product from $\mathfrak{g}_0$, which is invariant under K and can thus be translated into a $G_{\mathbb{R}}$-invariant Riemannian structure; the Riemannian structure then defines a $G_{\mathbb{R}}$-invariant distance function $d_{G_{\mathbb{R}}/K}(\ ,\ )$ on $G_{\mathbb{R}}/K$. Because of the $G_{\mathbb{R}}$-invariance of the metrics of D and $G_{\mathbb{R}}/K$, the quotient map $D \cong G_{\mathbb{R}}/V \to G_{\mathbb{R}}/K$ has a uniformly bounded differential; by renormalizing the metric on $G_{\mathbb{R}}/K$, the bound can be arranged to have the value one. Now let $G_{\mathbb{R}} = UAK$ be an Iwasawa decomposition (cf. [16], Chapter VI, for example; if $G_{\mathbb{R}}$ happens not to be connected in the Lie topology, K still meets every connected component of $G_{\mathbb{R}}$ according to (8.4), so that this case does not create any additional difficulty). Here A stands for a vector subgroup of $G_{\mathbb{R}}$, and U for a suitable maximal unipotent subgroup. The Lie algebra $\mathfrak{u}_0$ of U is then known to contain a conjugate, relative to $\operatorname{Ad} G_{\mathbb{R}}$, of any given nilpotent element of $\mathfrak{g}_0$; since A and U normalize $\mathfrak{u}_0$, the conjugation can even be performed in K. In particular, for each $z \in U$, I can choose $k(z) \in K$ such that

$$\operatorname{Ad} g(z)^{-1} N \in \operatorname{Ad} k(z)^{-1} \mathfrak{u}_0.$$

Because of the uniform boundedness property of $\tilde{\Phi}$, for a suitable positive constant C, one finds

$$\begin{aligned} mC(\operatorname{Im} z)^{-1} &\geqq d(\tilde{\Phi}(z+m), \tilde{\Phi}(z)) \\ &= d(\gamma^m \tilde{\Phi}(z), \tilde{\Phi}(z)) = d(g(z)^{-1} \gamma_u^m g(z) V, eV) \\ &\geqq d_{G_{\mathbb{R}}/K}(g(z)^{-1} \gamma_u^m g(z) K, eK) \\ &= d_{G_{\mathbb{R}}/K}(k(z) g(z)^{-1} \gamma_u^m g(z) K, eK) \\ &= d_{G_{\mathbb{R}}/K}(\exp(m \operatorname{Ad} k(z) g(z)^{-1}(N)) K, eK). \end{aligned}$$

According to the choice of $k(z)$, $\operatorname{Ad} k(z) g(z)^{-1}(N)$ lies in $\mathfrak{u}_0$. The mapping from $\mathfrak{u}_0 \times A$ to $G_{\mathbb{R}}/K$, which is given by

$$(X, a) \mapsto (\exp X) aK, \qquad X \in \mathfrak{u}_0, \ a \in A,$$

is a diffeomorphism and maps $(0, e)$ to the identity coset. With respect to any given metrics, a diffeomorphism is locally bounded. Applying this remark to the Euclidean metric on $\mathfrak{u}_0$, an arbitrary metric on A, and the Riemannian metric on $G_{\mathbb{R}}/K$, one finds that $\|\operatorname{Ad} k(z) g(z)^{-1}(N)\|$ can be bounded by a multiple of $(\operatorname{Im} z)^{-1}$, if only $\operatorname{Im} z$ is sufficiently large. Since K operates unitarily on $\mathfrak{g}$, the lemma follows.

The mapping $(X, b) \mapsto \exp X \circ b$ of $\mathfrak{g} \times \check{D}$ into $\check{D}$ is infinitely differentiable and sends a sufficiently small neighborhood of $(0, eB)$ into D. Like any C^1 map, this one must be locally bounded. Hence:

(8.15) **Lemma.** *There exists a neighborhood $\mathscr{U}$ of the base point in $\check{D}$, and positive constants η, C, such that $X \in \mathfrak{g}$, $\|X\| < \eta$, $a \in \mathscr{U}$, together imply $\exp X \circ a \in D$, and $d(\exp X \circ a, a) < C \|X\|$.*

As before, for each $z \in U$, I let $g(z)$ be some element of $G_{\mathbb{R}}$ such that

$$\tilde{\Phi}(z) \cong g(z) V \in G_{\mathbb{R}}/V \cong D.$$

I now define a mapping $F_z: U - z \to \check{D}$ by

$$\begin{aligned} F_z(u) &= g(z)^{-1} \exp(-muN) \circ \tilde{\Phi}(z+mu) \\ &= \exp(-mu \operatorname{Ad} g(z)^{-1}(N)) g(z)^{-1} \circ \tilde{\Phi}(z+mu); \end{aligned}$$

it is holomorphic and periodic of period one. Let $\mathscr{P}$ be a polycylindrical coordinate neighborhood of the base point in D; after shrinking $\mathscr{P}$, if necessary, I may assume that the $G_{\mathbb{R}}$-invariant metric of D and the Euclidean metric of $\mathscr{P}$ are mutually uniformly bounded on $\mathscr{P}$.

(8.16) **Lemma.** *There exist positive constants* α, ζ, *which depend only on the choice of base point, the choice of* $\mathscr{P}$, *and on the integer* m, *such that* $F_z(u)\in\mathscr{P}$ *whenever* $\operatorname{Im} z>\alpha$, $|\operatorname{Im} u|<\zeta \operatorname{Im} z$.

Proof. To begin with, after shrinking $\mathscr{U}$ and η in (8.15), one can arrange that $\exp X\circ a$ lies in $\mathscr{P}$, provided $\|X\|<\eta$ and $a\in\mathscr{U}$. Until further notice, α shall have the same meaning as in (8.14). As a consequence of the uniform boundedness of $\tilde{\Phi}$, for a suitable positive constant ζ_1 (which depends on α), $\operatorname{Im} z\geqq\alpha$ and $|u|<\zeta_1(\operatorname{Im} z)^{-1}$ together imply

$$d(\tilde{\Phi}(z+mu), \tilde{\Phi}(z))\leqq \text{diameter of } \mathscr{U},$$

and hence also

$$g(z)^{-1}\circ\tilde{\Phi}(z+mu)\in\mathscr{U}.$$

According to the initial assumptions on $\mathscr{U}$ and η, if β has the same meaning as in (8.14), and if

$$\operatorname{Im} z>\alpha,\qquad |u|<\zeta_1(\operatorname{Im} z)^{-1},\qquad m\,|u|\,\beta(\operatorname{Im} z)^{-1}<\eta,$$

$F_z(u)$ must now lie in $\mathscr{P}$. Let ζ be the smaller of the two numbers $2^{-\frac{1}{2}}\zeta_1$, $2^{-\frac{1}{2}}\eta\,\beta^{-1}m^{-1}$; for $\operatorname{Im} z>\alpha$, F_z maps the u-disc of radius $\sqrt{2}\,\zeta\operatorname{Im} z$, centered at $u=0$, entirely onto $\mathscr{P}$. This disc contains the rectangle

$$|\operatorname{Re} u|\leqq\alpha\zeta,\qquad |\operatorname{Im} u|<\zeta\operatorname{Im} z.$$

Enlarging α does not destroy any of the previous properties, which allows me to assume that $\alpha\zeta\geqq 1$. The periodicity of F_z then gives the conclusion of the lemma.

The following lemma, which replaces a clumsier one in the original version of the proof of (4.9), was suggested by Deligne:

(8.17) **Lemma.** *For each* $\eta>0$, *let* $\mathscr{F}_\eta$ *be the collection of all bounded holomorphic functions on the strip* $|\operatorname{Im} u|<\eta$ *which are periodic of period one. Each* $f\in\mathscr{F}_\eta$ *then satisfies*

$$\left|\frac{\partial f}{\partial u}(0)\right|\leqq\pi(\sinh\pi\eta)^{-2}\sup|f(u)|.$$

Proof. Let $f\in\mathscr{F}_\eta$ be given, and let $M=\sup|f(u)|$. Since f is periodic of period one, there exists a holomorphic function $\varphi(t)$, defined on the anulus

$$e^{-2\pi\eta}<|t|<e^{2\pi\eta},$$

such that $\varphi(t)=f(u)$ if $t=\exp(2\pi i u)$; φ is then also bounded by M. Evidently,

$$\frac{\partial f}{\partial u}(0)=2\pi i\frac{\partial\varphi}{\partial t}(1).$$

For the purpose of estimating the derivative of φ at 1, let r be any real number between 1 and $e^{2\pi\eta}$. Then

$$\frac{\partial\varphi}{\partial t}(1)=(2\pi i)^{-1}\int_{|t|=r}(t-1)^{-2}\,\varphi(t)\,dt \\ -(2\pi i)^{-1}\int_{r|t|=1}(t-1)^{-2}\,\varphi(t)\,dt,$$

and therefore

$$\left|\frac{\partial\varphi}{\partial t}(1)\right|\leqq(2\pi)^{-1}M\int_0^{2\pi}\{r\,|r\,e^{i\theta}-1|^{-2}+r^{-1}\,|r^{-1}\,e^{i\theta}-1|^{-2}\}\,d\theta \\ \leqq M\{r(r-1)^{-2}+r^{-1}(r^{-1}-1)^{-2}\}=2Mr(r-1)^{-2}.$$

Letting r tend to $e^{2\pi\eta}$, one obtains the bound for $\frac{\partial\varphi}{\partial t}(1)$ which gives the desired inequality.

According to (8.16), for $\operatorname{Im} z>\alpha$, the coordinate functions of F_z, relative to the coordinate Polycylinder $\mathscr{P}$, belong to $\mathscr{F}_\eta$, with $\eta=\zeta\operatorname{Im} z$. Consequently, again for $\operatorname{Im} z>\alpha$, the image of the tangent vector $\frac{\partial}{\partial u}$ under the differential of the mapping F_z at $u=0$ is bounded by a multiple of $\exp(-2\pi\zeta\operatorname{Im} z)$; through increasing α and shrinking ζ, the value of the bound can be arranged to be one. By infinitesimal left translation, N determines a holomorphic tangent vector field on $\check{D}$; its value at a point $a\in\check{D}$ will be denoted by $N(a)$. In the following, stars as subscripts designate differentials of mappings, and vertical bars will be used, when necessary, to describe the point in the domain at which the differential of the mapping in question is being considered. For $g\in G_{\mathbb{C}}$ let $l(g)\colon\check{D}\to\check{D}$ be left translation. Then

$$F_{z*}\left(\frac{\partial}{\partial u}\right)\bigg|_{u=0}=l(g(z)^{-1})_*\left(m\tilde{\Phi}_*\left(\frac{\partial}{\partial z}\right)\bigg|_z-mN(\tilde{\Phi}(z))\right).$$

Since $g(z)^{-1}\in G_{\mathbb{R}}$, $l(g(z)^{-1})_*$ is an isometry with respect to the Hermitian structure h, and one finds:

(8.18) **Corollary.** *There exist positive constants α and ε, which depend only on the choice of base point and on the integer m, such that* $\operatorname{Im} z>\alpha$ *implies*

$$\left\|\tilde{\Phi}_*\left(\frac{\partial}{\partial z}\right)\bigg|_z-N(\tilde{\Phi}(z))\right\|\leqq\exp(-\varepsilon\operatorname{Im} z);$$

in this statement, length is measured with respect to h.

(8.19) **Lemma.** *Let $\alpha>0$ be given. For suitable positive constants C, β, the operator norms of both* $\operatorname{Ad} g(z)$ *and* $\operatorname{Ad} g(z)^{-1}$ *are bounded by* $C(\operatorname{Im} z)^\beta$, *provided that* $\operatorname{Im} z > \alpha$ *and* $|\operatorname{Re} z| \leqq m$. *The constant* β *can be chosen independently of* $\tilde{\Phi}$ *(but not independently of m), and as long as* $\tilde{\Phi}(i)$ *is known to lie in a particular compact subset of D, C can also be chosen uniformly.*

Proof. Let $G_{\mathbb{R}} = UAK$ be an Iwasawa decomposition, as in the proof of (8.14), and let $\mathfrak{a}_0$ be the Lie algebra of A. Then every element of $G_{\mathbb{R}}$ can be expressed as $k_1 \exp Y k_2$, with $Y \in \mathfrak{a}_0$, $k_1, k_2 \in K$ (cf. [16], again this holds even if $G_{\mathbb{R}}$ has more than one component in the Lie topology, since K meets all of them). In particular, I shall write $g(z)$ in this form: $g(z) = k_1 \exp Y k_2$. With respect to the Riemannian structure on $G_{\mathbb{R}}/K$, which was defined in the proof of (8.14), and the Euclidean metric on $\mathfrak{a}_0$, which comes from the inner product (8.9), the mapping $X \mapsto \exp XK$ embeds $\mathfrak{a}_0$ globally isometrically in $G_{\mathbb{R}}/K$, except possibly for the normalization of the metrics [16]. Thus, for a suitable positive constant η,

$$
\begin{aligned}
\|Y\| &= \eta\, d_{G_{\mathbb{R}}/K}(\exp YK, eK) = \eta\, d_{G_{\mathbb{R}}/K}(g(z)\,K, eK) \\
&\leqq \eta\, d_{G_{\mathbb{R}}/K}(g(z)\,K, g(i)\,K) + \eta\, d_{G_{\mathbb{R}}/K}(g(i)\,K, eK) \qquad (8.20)\\
&\leqq \eta\, d(\tilde{\Phi}(z), \tilde{\Phi}(i)) + \eta\, d_{G_{\mathbb{R}}/K}(g(i)\,K, eK).
\end{aligned}
$$

If $\tilde{\Phi}(i)$ is restricted to lie in some fixed compact subset of D, $g(i)$ will have to lie in a compact subset of $G_{\mathbb{R}}$, and so the second summand on the right hand side of the inequality can be bounded by a constant. Because of the boundedness of $\tilde{\Phi}$, the first summand can be at most propositional to the Poincaré distance between z and i in the upper half plane. For $|\operatorname{Re} z| \leqq m$, $\operatorname{Im} z > \alpha$, this distance in turn is bounded by $m\alpha^{-1} + |\log \operatorname{Im} z|$; since $\operatorname{Im} z$ stays away from zero, the absolute value bars may be dropped, provided $m\alpha^{-1}$ is replaced by a possibly larger constant. Under the adjoint action, $\mathfrak{a}_0$ operates semisimply on $\mathfrak{g}$, with real eigenvalues. Hence the operator norm of $\operatorname{Ad} \exp Y$ cannot exceed a multiple of the exponential of the largest eigenvalue of $\operatorname{Ad} Y$; the eigenvalues of $\operatorname{Ad} Y$ can be estimated in terms of $\|Y\|$. Putting all of these inequalities together, one finds a bound for the operator norm of $\operatorname{Ad} \exp Y$,

$$\|\operatorname{Ad} \exp Y\| \leqq C(\operatorname{Im} z)^{\beta},$$

which is valid under the assumptions made on z. The dependence on $\tilde{\Phi}$ is confined to the second term on the right hand side of (8.20), which affects only C. Since $g(z) = k_1 \exp Y k_2$, with k_1, $k_2 \in K$, the same bound holds for the operator norm of $\operatorname{Ad} g(z)$. Replacing Y by $-Y$ in the above argument, one also gets the same inequality for $\operatorname{Ad} g(z)^{-1}$.

By means of (8.13) and (8.19), the inequality contained in (8.18) gives an analogous one in terms of the M-invariant metric, which is valid, however, only on a vertical strip. Since the imaginary part of z remains bounded away from 0 by α, the constant C of (8.19) can be increased, and the constant ε of (8.18) decreased, so that the exponential term absorbs the factor $(\operatorname{Im} z)^\beta$. Thus there exist positive constants C, α, ε, such that $\operatorname{Im} z > \alpha$, $|\operatorname{Re} z| \leqq m$ together imply

$$\left\| \Phi_* \left(\frac{\partial}{\partial z}\right)\Big|_z - N(\tilde{\Phi}(z)) \right\| \leqq C \exp(-\varepsilon \operatorname{Im} z), \tag{8.21}$$

where length is now measured with respect to h_M.

Since N is nilpotent, the operator norm of $\operatorname{Ad}\exp(zN)$ can be bounded by a polynomial in $|z|$. The definition of $\tilde{\Psi}$ gives

$$\tilde{\Psi}_* \left(\frac{\partial}{\partial z}\right)\Big|_z = l(\exp(-mzN))_* \left(m\tilde{\Phi}_* \left(\frac{\partial}{\partial z}\right)\Big|_{mz} - mN(\tilde{\Phi}(mz))\right).$$

In view of (8.12), one now obtains an estimate for the h_M-length of $\tilde{\Psi}_* \left(\frac{\partial}{\partial z}\right)$. For this purpose, I adjust the constants α, C, ε, to take into account the change of scale by the factor m. By further changing the constants, I can absorb the polynomial which estimates the norm of $\operatorname{Ad}\exp(-mzN)$ into the exponential factor. The estimate can thus be brought into the form

$$h_M\left(\tilde{\Psi}_* \left(\frac{\partial}{\partial z}\right)\Big|_z, \tilde{\Psi}_* \left(\frac{\partial}{\partial z}\right)\Big|_z\right)^{\frac{1}{2}} \leqq C \exp(-\varepsilon \operatorname{Im} z), \tag{8.22}$$

provided $\operatorname{Im} z > \alpha$. The restriction on the real part of z becomes unnecessary, since $\tilde{\Psi}$ is invariant under $z \mapsto z+1$. Now let β be a fixed positive number greater than α, and let z_1, z_2 be two points in U, with $\operatorname{Im} z_1, \operatorname{Im} z_2 \geqq \beta$. By integrating the inequality (8.22), taking into account the periodicity of $\tilde{\Psi}$, one finds that the M-invariant distance between $\tilde{\Psi}(z_1)$ and $\tilde{\Psi}(z_2)$ can amount to at most

$$C(1+\varepsilon^{-1}) \exp(-\varepsilon\beta).$$

When this estimate is rephrased in terms of the mapping Ψ and uncluttered by absorbing the factor $(1+\varepsilon^{-1})$ into C, one obtains:

(8.23) **Corollary.** *There exist positive constants ρ, λ, C, such that for any two complex numbers t_1 and t_2, with $0 < |t_1|, |t_2| \leqq r$, $r < \rho$, the inequality $d_M(\Psi(t_1), \Psi(t_2)) \leqq C r^\lambda$ holds.*

Clearly, then, Ψ extends continuously, and therefore holomorphicly to a mapping from Δ into $\check{D}$. Let a denote the point $\Psi(0)$. The transformation property $\tilde{\Psi}\left(z+\frac{1}{m}\right) = \gamma_s \circ \tilde{\Psi}(z)$ translates into $\Psi(e^{2\pi i m^{-1}} t) = \gamma_s \circ \tilde{\Psi}(t)$; letting t tend to zero in this identity, one finds that a is a fixed point for γ_s.

For sufficiently small values of $t \in \Delta$, the d_M-distance between a and $\Psi(t)$ grows at most linearly with $|t|$. Hence

$$d_M(a, \exp(-zN) \circ \tilde{\Phi}(z)) = d_M(a, \Psi(\exp 2\pi i m^{-1} z))$$

is bounded by a constant multiple of $\exp(-2\pi m^{-1} \operatorname{Im} z)$, for all large values of $\operatorname{Im} z$. According to (8.12), left translation by any $g \in G_{\mathbb{C}}$ distorts the metric h_M by no more than a factor equal to the operator norm of $\operatorname{Ad} g$; the global version of this statement clearly follows from the infinitesimal one. As was pointed out before, if z is restricted to a vertical strip and bounded away from the real axis, the operator norm of $\operatorname{Ad} \exp zN$ does not increase faster than a power of $\operatorname{Im} z$. On such a vertical strip, then, if $\operatorname{Im} z$ is large enough, one obtains an estimate

$$d_M(\exp(zN) \circ a, \tilde{\Phi}(z)) \leqq C (\operatorname{Im} z)^\beta \exp(-2\pi m^{-1} \operatorname{Im} z),$$

where C and β are suitable constants. I recall the choices of $g(z) \in G_{\mathbb{R}}$, for $z \in U$, which were made just before the statement (8.14). From the preceeding estimate, translating both points by $g(z)^{-1}$, again using the global version of (8.12), together with (8.19), one can deduce the same type of bound for the d_M-distance between $g(z)^{-1} \exp(zN) \circ a$ and the base point eB in $G_{\mathbb{C}}/B \cong \check{D}$. In particular, for any given neighborhood of the base point, if z is constrained to a vertical strip and has a sufficiently large imaginary part, the points $g(z)^{-1} \exp(zN) \circ a$ are forced to lie in that neighborhood. One can pick such a neighborhood, entirely contained in D, such that the two distance functions d and d_M are mutually bounded on it. Hence there exist positive constants α, β, and C with the following property: if $|\operatorname{Re} z| \leqq m$ and $\operatorname{Im} z > \alpha$, the point $g(z)^{-1} \exp(zN) \circ a$ lies in D, and its $G_{\mathbb{R}}$-invariant distance from the base point does not exceed $C (\operatorname{Im} z)^\beta \exp(-2\pi m^{-1} \operatorname{Im} z)$. By enlarging α and β, if necessary, the constant C can be absorbed. Translation by $g(z)$ leaves D invariant and preserves the distance function d. I have therefore proven that $\operatorname{Im} z > \alpha$ implies

$$\exp(zN) \circ a \in D, \quad \text{and} \quad d(\exp(zN) \circ a, \tilde{\Phi}(z)) < (\operatorname{Im} z)^\beta \exp(-2\pi m^{-1} \operatorname{Im} z);$$

the restriction $|\operatorname{Re} z| \leqq m$ becomes unnecessary, because the substitution $z \mapsto z + m$ has the effect of translating both $\tilde{\Phi}(z)$ and $\exp(zN) \circ a$ by $\gamma^m = \gamma_u^m$.

In order to complete the proof of (4.9), it remains to be shown that $z \mapsto \exp(zN) \circ a$ is a horizontal mapping. As usual, $N(b)$, for $b \in \check{D}$, will refer to the value at b of the holomorphic tangent vector field which is defined by infinitesimal left translation by N. The tangent vector to the holomorphic curve $z \mapsto \exp(zN) \circ a$, for any particular value of z, equals the $\exp(zN)$-translate of the vector $N(a)$ at a. Since $G_{\mathbb{C}}$ leaves the horizontal tangent subbundle invariant, it suffices to verify that $N(a)$ lies in the fibre of this subbundle at a. For reasons of continuity, this would follow if it were known that $N(\tilde{\Psi}(z))$ differs from a horizontal vector at $\tilde{\Psi}(z)$ by a vector whose h_M-length tends to zero, as z tends to ∞ in a vertical half strip. Since Ψ extends holomorphically over the origin, and since $\tilde{\Psi}(z)=\Psi(\exp 2\pi i z)$, the h_M-length of $\tilde{\Psi}_*\left(\frac{\partial}{\partial z}\right)\Big|_z$ does tend to zero as $\operatorname{Im} z \to +\infty$. On the other hand,

$$\frac{1}{m}\tilde{\Psi}_*\left(\frac{\partial}{\partial z}\right)\Big|_z = -N(\tilde{\Psi}(z)) + l(\exp(-mzN))_* \tilde{\Phi}_*\left(\frac{\partial}{\partial z}\right)\Big|_{mz}.$$

The vector $\tilde{\Phi}_*\left(\frac{\partial}{\partial z}\right)\Big|_{mz}$ is horizontal because $\tilde{\Phi}$ is a horizontal mapping, and thus the translate of this vector by $l(\exp(-mzN))$ must also be horizontal. This concludes the argument.

I now turn to the situation described above the statement (4.12). To begin with, I shall prove that Ψ continues holomorphically to Δ^k. For this purpose, two simplifications can be made. First, since the problem is a local one, I only need to continue Ψ to a neighborhood of the origin in Δ^k; for any other point of the subvariety on which Ψ is not yet determined, the same argument can be applied to a smaller polycylinder centered at the point in question. Secondly, nothing is lost by assuming $k=l$: I merely choose $\gamma_i=1$, $N_i=0$ for $l+1\leqq i\leqq k$. By applying (8.18) to each of the variables separately, one can find positive constants α and ε, such that the restrictions $\operatorname{Im} z_i>\alpha$, $1\leqq i\leqq k$, imply the inequalities

$$(8.24)\qquad \left\|\tilde{\Phi}_*\left(\frac{\partial}{\partial z_j}\right)\Big|_{(z)} - N_j(\tilde{\Phi}(z))\right\| \leqq \exp(-\varepsilon \operatorname{Im} z_j),$$

$1\leqq j\leqq k$. Here (z) is shorthand for the k-tuple $(z_1,\ldots,z_k)\in U^k$, and the metric h is used to measure length. Just as in the proof of (4.9), the next step is to replace the metric h by h_M in the inequality. To this end, for $(z)\in U^k$, I choose an element $g(z)\in G_{\mathbb{R}}$ whose V-coset represents the point $\tilde{\Phi}(z)\in D\cong G_{\mathbb{R}}/V$. In complete analogy to the statement and proof of (8.19), one may conclude:

(8.25) **Lemma.** *For any $\alpha>0$, there exist positive constants C, β, such that, subject to the constraints $|\operatorname{Re} z_i|\leqq m_i$, $\operatorname{Im} z_i>\alpha$, $1\leqq i\leqq k$, the operator norms of $\operatorname{Ad} g(z)$ and $\operatorname{Ad} g(z)^{-1}$ remain bounded by $C\prod_{i=1}^k (\operatorname{Im} z_i)^\beta$. The constant β can be chosen independently of $\tilde{\Phi}$ (though not independently of the m_i), and if $\tilde{\Phi}(i,\ldots,i)$ is known to lie in a particular compact subset of D, C can also be chosen independently of $\tilde{\Phi}$.*

In conjunction with (8.13), Lemma (8.25) allows me to rephrase the inequality (8.24) in terms of the metric h_M. For $\operatorname{Im} z_i>\alpha$, $|\operatorname{Re} z_i|\leqq m_i$,

$1 \leqq i \leqq k$, the h_M-length of the tangent vector

$$\tilde{\Phi}_* \left(\frac{\partial}{\partial z_j} \right) \bigg|_{(z)} - N_j(\tilde{\Phi}(z))$$

is bounded by $C(\prod_{i=1}^k \operatorname{Im} z_i)^\beta \exp(-\varepsilon \operatorname{Im} z_j)$; here j is any integer between 1 and k, and the various constants have the same meaning as before. Since the N_j are nilpotent, on each product of truncated vertical strips $\operatorname{Im} z_i > \alpha$, $|\operatorname{Re} z_i| \leqq m$, $1 \leqq i \leqq m$, one can bound the operator norm of $\operatorname{Ad} \exp(-\sum_{i=1}^k m_i z_i N_i)$ by a constant multiple of a suitable power of $\prod_{i=1}^l \operatorname{Im} z_i$. Let (mz) denote the k-tuple $(m_1 z_1, \ldots, m_k z_k)$; the definition of $\tilde{\Psi}$ gives the formula

$$\text{(8.26)} \quad \begin{aligned} &\tilde{\Psi}_* \left(\frac{\partial}{\partial z_j} \right) \bigg|_{(z)} \\ &= l(\exp(-\textstyle\sum_{i=1}^k m_i z_i N_i))_* \left(m_j \tilde{\Phi}_* \left(\frac{\partial}{\partial z_j} \right) \bigg|_{(mz)} - m_j N_j(\tilde{\Phi}(mz)) \right). \end{aligned}$$

By combining the preceeding estimate with (8.26), and using (8.12), one obtains an estimate for the h_M-length of the tangent vector $\tilde{\Psi}_* \left(\frac{\partial}{\partial z_j} \right) \Big|_{(z)}$. I adjust the constants C, α, β, ε to take into account the scale factors m_i, and to absorb the estimate for the operator norm of $\operatorname{Ad} \exp(-\sum_{i=1}^k m_i z_i N_i)$ into the term $C(\prod_{i=1}^k \operatorname{Im} z_i)^\beta$. Thus, for $\operatorname{Im} z_i > \alpha$, and $1 \leqq j \leqq k$,

$$\text{(8.27)} \quad \left\| \tilde{\Psi}_* \left(\frac{\partial}{\partial z_j} \right) \bigg|_{(z)} \right\| \leqq C(\textstyle\prod_{i=1}^k \operatorname{Im} z_i)^\beta \exp(-\varepsilon \operatorname{Im} z_j);$$

here length is measured with respect to h_M. The restriction on the real parts of the z_i can be dropped, because $\tilde{\Psi}$ is invariant under $z_i \mapsto z_i + 1$.

Since $\check{D}$ is acted on transitively and isometrically by the group M, there exists a constant $\delta > 0$, with the following property:

(8.28) any subset of $\check{D}$ of h_M-diameter less than δ can be enclosed in a polycylindrical coordinate neighborhood.

It is certainly permissible to enlarge the constant α, without destroying the estimate (8.27). I may therefore assume that

$$\text{(8.29)} \quad C\left\{ \left(\alpha + \frac{1}{2\pi} \log \varepsilon \right)^{(k-1)\beta} \int_\alpha^\infty y^\beta e^{-\varepsilon y} \, dy + \left(\alpha + \frac{1}{2\pi} \log 2 \right)^{k\beta} e^{-\varepsilon\alpha} \right\} < \frac{\delta}{3}.$$

Let (z), $(z') \in U^k$ be two points which satisfy, for some integer j between 1 and k,

$$\text{(8.30)} \quad \begin{aligned} &\operatorname{Im} z_j,\ \operatorname{Im} z'_j > \alpha; \quad z_i = z'_i && \text{for } i \neq j; \\ &\alpha < \operatorname{Im} z_i = \operatorname{Im} z'_i < \alpha + \frac{1}{2\pi} \log 2 && \text{for } i \neq j. \end{aligned}$$

By integrating the inequality (8.27), taking into account the periodicity of $\tilde{\Psi}$, one finds:

(8.31) under the hypotheses (8.30),

$$d_M(\tilde{\Psi}(z), \tilde{\Psi}(z')) < \frac{\delta}{3}.$$

If necessary, I increase α further, so that

$$(8.32) \qquad kC\left(1+\frac{1}{2\pi}\log 2\right)\left(\alpha+\frac{1}{2\pi}\log 2\right)^{k\beta} e^{-\varepsilon\alpha} < \frac{\delta}{3}.$$

With this new choice of α, as another consequence of (8.27) and the periodicity of $\tilde{\Psi}$, I conclude:

(8.33) the set $\tilde{\Psi}\left(\left\{(z)\in U^k \,\middle|\, \alpha < \operatorname{Im} z_i < \alpha + \frac{1}{2\pi}\log 2\right\}\right)$

has h_M-diameter less than $\frac{\delta}{3}$.

Let $\rho = e^{-2\pi\alpha}$; then $0<\rho<1$. The next statement is obtained by combining (8.28), (8.31), and (8.33), and rephrasing the result in terms of the mapping Ψ.

(8.34) **Lemma.** *There exists a polycylindrical coordinate neighborhood $\mathscr{P} \subset \check{D}$, such that Ψ maps the set*

$$\bigcup_{j=1}^k \{(t)\in\Delta^k \mid 0<|t_j|<\rho;\ \tfrac{1}{2}\rho<|t_i|<\rho \ \text{ for } i\neq j\}$$

into $\mathscr{P}$.

Since the coordinate functions of Ψ with respect to $\mathscr{P}$ are bounded, Ψ extends holomorphically to the set

$$\bigcup_{j=1}^k \{(t)\in\Delta^k \mid |t_j|<\rho;\ \tfrac{1}{2}\rho<|t_i|<\rho \ \text{ for } i\neq j\},$$

and maps the enlarged set also into $\mathscr{P}$. According to a classical theorem of Hartogs, a holomorphic function which is defined on a set of this form extends holomorphically to all of

$$\{(t)\in\Delta^k \mid |t_i|<\rho \ \text{ for } 1\leq i\leq k\}.$$

The mapping Ψ, therefore, extends at least to some neighborhood of the origin in Δ^k. As was pointed out before, this already implies that Ψ continues holomorphically to all of Δ^k. Thus:

(8.35) **Corollary.** *The mapping Ψ has a holomorphic extension to Δ^k.*

I now drop the assumption $l=k$, and for $(w)\in\Delta^{k-l}$, I set $a(w)=\Psi(0,w)$. Let η be given, with $0<\eta<1$. Corresponding to each $(z,w)\in U^l\times\Delta^{k-l}$, I choose an element $g(z, w) \in G_{\mathbb{R}}$, whose V-coset represents the point $\tilde{\Phi}(z, w) \in D \cong G_{\mathbb{R}}/V$. Under the restriction $|w_j| \leqq \eta$, for $l+1 \leqq j \leqq k$, the points $\tilde{\Phi}(i, \ldots, i, w_{l+1}, \ldots, w_k)$ are confined to a compact subset of D. Thus, applied to the first l variables, for any given $\alpha > 0$, Lemma (8.25) asserts the existence of positive constants β, C, so that:

(8.36) under the hypotheses $\operatorname{Im} z_i > \alpha$, $|\operatorname{Re} z_i| < m_i$ for $1 \leqq i \leqq l$, and $|w_j| \leqq \eta$ for $l+1 \leqq j \leqq k$, the operator norms of $\operatorname{Ad} g(z)$ and $\operatorname{Ad} g(z)^{-1}$ are bounded by $C(\prod_{i=1}^{l} \operatorname{Im} z_i)^{\beta}$.

Since Ψ is holomorphic on all of Δ^k, its restriction to any compact subset dilates distances at most linearly. Reinterpreted in terms of $\tilde{\Phi}$, this statement implies:

(8.37) **Lemma.** *Let $\alpha > 0$ be given. For a suitable positive constant ζ, if $\operatorname{Im} z_i > \alpha$, $1 \leqq i \leqq l$, and if $|w_j| \leqq \eta$, $l+1 \leqq j \leqq k$, then*

$$d_M\left(a(w), \exp\left(-\sum_{i=1}^{l} Z_i N_i\right) \circ \tilde{\Phi}(z, w)\right) \leqq \zeta \sum_{i=1}^{l} \exp(-2\pi m_i^{-1} \operatorname{Im} z_i).$$

At this point, with the help of (8.36) and (8.37) the argument can be continued just as in the proof of (4.9). As the conclusion of this argument, for a suitable constant $\alpha > 0$, if $\operatorname{Im} z_i > \alpha$ for $1 \leqq i \leqq l$, and if $|w_j| \leqq \eta$ for $l+1 \leqq j \leqq k$, the point $\exp\left(\sum_{i=1}^{l} z_i N_i\right) \circ a(w)$ lies in D, and its $G_{\mathbb{R}}$-invariant distance from the point $\tilde{\Phi}(z, w)$ does not exceed

$$\left(\prod_{i=1}^{l} \operatorname{Im} z_i\right)^{\beta} \sum_{i=1}^{l} \exp(-2\pi m_i^{-1} \operatorname{Im} z_i);$$

in this estimate, the constant β may have to be larger than in (8.36).

In order to finish the proof of (4.12), I have to show that the mapping

$$(z, w) \mapsto \exp\left(\sum_{i=1}^{l} z_i N_i\right) \circ a(w) \tag{8.38}$$

is horizontal. As usual, for $b \in \check{D}$, I let $N_i(b)$ denote the value at b of the holomorphic vector field generated by N_i. If one applies Theorem (4.9) to the j-th variable separately, for $1 \leqq j \leqq l$, with all other variables kept fixed, one finds that the tangent vector

$$N_j\left(\exp\left(\sum_{i \neq j} m_i^{-1} z_i N_i\right) \circ \lim_{\operatorname{Im} z_i \to \infty} \tilde{\Psi}(m^{-1} z, w)\right)$$

lies in the appropriate fibre of the horizontal tangent subbundle; here $(m^{-1} z)$ is shorthand for the l-tuple $(m_1^{-1} z_1, \ldots, m_l^{-1} z_l)$, and $z_1, \ldots, z_{j-1}, z_{j+1}, \ldots, z_k \in U$, as well as $(w) \in \Delta^{k-l}$, are arbitrary. Since the N_i commute, for any $b \in \check{D}$ one has

$$N_j\left(\exp\left(\sum_{i \neq j} m_i^{-1} z_i N_i\right) \circ b\right) = l\left(\exp\left(\sum_{i \neq j} m_i^{-1} z_i N_i\right)\right)_* N_j(b).$$

Moreover, translation by any element of $G_{\mathbb{C}}$ preserves the horizontal tangent subbundle. Thus under the restriction $t_j = 0$, and $t_i \neq 0$ if

$1 \leqq i \leqq j-1$ or $j+1 \leqq i \leqq l$, the tangent vector $N_j(\Psi(t))$ must lie in the fibre of the horizontal tangent subbundle of $\Psi(t)$. When the first l variables are made to tend to zero (except for the j-th one, which equals zero already) and the others are kept fixed, $\Psi(w)$ tends to $a(w_{l+1}, \ldots, w_k)$. For reasons of continuity, then

(8.39) $N_j(a(w))$ lies in the fibre of the horizontal tangent subbundle, for all $(w) \in \Delta^{k-l}$ and $1 \leqq j \leqq l$.

The mapping $(z, w) \mapsto \tilde{\Phi}(z, w)$ is known to be horizontal. Since translation by elements of $G_{\mathbb{C}}$ preserves the horizontal property of a map, in view of the definition of $\tilde{\Psi}$

$$(w) \mapsto \tilde{\Psi}(z, w)$$

is horizontal, for each $(z) \in U^l$. Thus, as long as each of the first l variables differs from zero, $\Psi(w)$ depends horizontally on the last $k-l$ variables. Since Ψ is holomorphic on all of Δ^k, I may let the first l variables tend to zero, and conclude:

(8.40) $(w) \to a(w)$ is a horizontal map of Δ^{k-l} into $\check{D}$.

Taken together, (8.39) and (8.40) imply the horizontality of the mapping (8.38).

§ 9. Proof of the SL_2-Orbit Theorem

In this section, I shall freely use the notation established in the beginning of § 8. In particular, a base point $o \in D$ will be chosen, which corresponds to the identity coset under the identification $D \cong G_{\mathbb{R}}/V$. The base point determines an Ad V-invariant Hodge structure of weight zero on $\mathfrak{g}$,

$$\mathfrak{g} = \oplus_p \mathfrak{g}^{p,-p}; \tag{9.1}$$

it has the property that $\mathfrak{g}^{0,0} = \mathfrak{v} =$ complexified Lie algebra of V. I shall denote the Weil operator of this Hodge structure by θ. In view of (3.12),

$$J: \mathfrak{g} \to \mathfrak{g}, \quad \text{with} \quad JX = i^{-p} X \quad \text{for } X \in \mathfrak{g}^{p,-p}, \tag{9.2}$$

defines an automorphism of $\mathfrak{g}$. Then

$$\begin{aligned} &J \text{ preserves the real structure } \mathfrak{g}_0 \subset \mathfrak{g}, \\ &J^2 = \theta, \quad \text{and} \quad JX = X \quad \text{for } X \in \mathfrak{v}, \end{aligned} \tag{9.3}$$

as can be checked directly. It will be convenient to have the notation

$$\mathfrak{q} = \mathfrak{g}^{1,-1} \oplus \mathfrak{g}^{-1,1}, \quad \mathfrak{q}_0 = \mathfrak{q} \cap \mathfrak{g}_0. \tag{9.4}$$

For the remainder of this section, I shall keep fixed a point $a \in \check{D}$, a nilpotent element $N \in \mathfrak{g}_0$, and a positive constant α, such that the conditions (5.1) are satisfied. As was pointed out in § 5, (5.1 b) can be replaced by (5.2). Since V preserves the Hodge structure (9.1),

$$\mathfrak{g}_0 = \mathfrak{v}_0 \oplus \{\mathfrak{g}_0 \cap (\oplus_{p \neq 0} \mathfrak{g}^{p,-p})\} \tag{9.5}$$

defines an Ad V-invariant splitting. This splitting determines a $G_{\mathbb{R}}$-invariant connection on the principal bundle

$$V \to G_{\mathbb{R}} \to G_{\mathbb{R}}/V \cong D.$$

The mapping

$$y \mapsto \exp(i y N) \circ a \in D, \quad y \in \mathbb{R}, \ y > \alpha,$$

has a unique (except for the choice of some initial point) lifting from D to $G_{\mathbb{R}}$ which is tangent to the connection. I shall denote this lifting by

$$y \mapsto h(y) \in G_{\mathbb{R}}, \quad y \in \mathbb{R}, \ y > \alpha. \tag{9.6}$$

For the sake of notational simplicity, apostrophes will be used to designate derivatives. Because the map (9.6) takes values in a matrix group, $h(y)^{-1} h'(y)$ lies in its Lie algebra, $\mathfrak{g}_0$, for all $y > \alpha$. Thus

$$\begin{gathered} A(y) = -2 h(y)^{-1} h'(y), \\ F(y) = \operatorname{Ad} h(y)^{-1} N, \quad E(y) = -\theta F(y), \end{gathered} \tag{9.7}$$

defines three $\mathfrak{g}_0$-valued functions on the interval $\{y \in \mathbb{R} \mid y > \alpha\}$. I recall (9.4).

(9.8) **Lemma.** *For all $y \in \mathbb{R}$, with $y > \alpha$, $A(y) \in \mathfrak{q}_0$, $E(y) \in \mathfrak{v}_0 \oplus \mathfrak{q}_0$, $F(y) \in \mathfrak{v}_0 \oplus \mathfrak{q}_0$, $E(y) - F(y) \in \mathfrak{v}_0$, and $A(y) = -J(E(y) + F(y))$. Moreover, these three functions satisfy the differential equations*

$$\begin{aligned} &2E'(y) = -[A(y), E(y)], \quad 2F'(y) = [A(y), F(y)], \\ &A'(y) = -[E(y), F(y)]. \end{aligned}$$

Proof. The fact that the mapping (9.6) is tangent to the connection determined by the splitting (9.5) immediately implies

$$h(y)^{-1} h'(y) \in \mathfrak{g}_0 \cap (\oplus_{p \neq 0} \mathfrak{g}^{p,-p}). \tag{9.9}$$

I choose an element $g_0 \in G_{\mathbb{C}}$ whose B-coset represents the point $a \in \check{D} \cong G_{\mathbb{C}}/B$. The condition (5.1 a) is equivalent to

$$\operatorname{Ad} g_0^{-1} N \in \mathfrak{g}^{-1,1} \oplus \mathfrak{b}. \tag{9.10}$$

By construction, the coset $h(y)B$ represents the point $\exp(i y N) \circ a$. Hence there exists a B-valued function $y \mapsto b(y)$, such that $h(y) =$

$\exp(iyN)\,g_0\,b(y)$. Differentiation of this identity gives

$$(9.11)\qquad h(y)^{-1}\,h'(y) = i\,\mathrm{Ad}\,b(y)^{-1}\,\mathrm{Ad}\,g_0^{-1}\,N + b(y)^{-1}\,b'(y).$$

Since $b(y)$ takes values in B, $b(y)^{-1}\,b'(y)$ lies in the Lie algebra $\mathfrak{b}$. Moreover, the subspace $\mathfrak{g}^{-1,1}\oplus\mathfrak{b}$ of $\mathfrak{g}$ is $\mathrm{Ad}\,B$-invariant. Thus (9.10) implies

$$(9.12)\qquad h(y)^{-1}\,h'(y)\in\mathfrak{g}^{-1,1}\oplus\mathfrak{b}.$$

The values of $h(y)^{-1}\,h'(y)$ lie in $\mathfrak{g}_0$, hence

$$\begin{aligned} h(y)^{-1}\,h'(y) &\in \mathfrak{g}_0\cap(\mathfrak{g}^{-1,1}\oplus\mathfrak{b})\\ &= \mathfrak{g}_0\cap(\mathfrak{g}^{-1,1}\oplus\mathfrak{b})\cap(\mathfrak{g}^{1,-1}\oplus\bar{\mathfrak{b}}) = \mathfrak{v}_0\oplus\mathfrak{q}_0. \end{aligned}$$

In combination with (9.9), this gives $A(y)\in\mathfrak{q}_0$.

The formula (9.11) can be rewritten as $A(y) = -2i\,\mathrm{Ad}\,h(y)^{-1}N - 2b(y)^{-1}\,b'(y)$, so that

$$(9.13)\qquad A(y) = -2i\,F(y) - 2b(y)^{-1}\,b'(y).$$

In particular, $F(y)$ must lie in $\mathfrak{b}\oplus\mathfrak{g}^{-1,1}$. But $F(y)$ also lies in $\mathfrak{g}_0$, and I can conclude $F(y)\in\mathfrak{v}_0\oplus\mathfrak{q}_0$, just as in the case of $h(y)^{-1}\,h'(y)$. The subspace $\mathfrak{v}_0\oplus\mathfrak{q}_0$ is θ-invariant, so that it must also contain $E(y)$. According to the definition of $E(y)$, the components of $E(y)$ and $F(y)$ in $\mathfrak{q}_0$ coincide; hence $E(y)-F(y)$ lies in $\mathfrak{v}_0$. On $\mathfrak{g}^{-1,1}$, J acts as multiplication by i. Thus, and in view of (9.13), $A(y)+2JF(y)$ does not have a component in $\mathfrak{g}^{-1,1}$. As was shown above, $A(y)$ has a zero component in $\mathfrak{v}$. From the definitions (9.7), it follows that $J(E(y)+F(y))$ also has a zero component in $\mathfrak{v}$, and that its component in $\mathfrak{g}^{-1,1}$ is equal to twice that of $JF(y)$. Thus

$$A(y)+J\big(E(y)+F(y)\big)\in\mathfrak{g}^{1,-1}\cap\mathfrak{g}_0 = 0.$$

It remains to verify the differential equations. Differentiation of the identity $F(y)=\mathrm{Ad}\,h(y)^{-1}N$ yields $2F'(y)=[A(y),F(y)]$. Applying the automorphism θ to both sides, I obtain the first of the three equations. Combining these two, and taking into account that J operates as the identity on $E(y)-F(y)$, and that $J^2=\theta$, I find

$$\begin{aligned} A'(y) &= -J\big(E'(y)+F'(y)\big) = \tfrac{1}{2}J\,[A(y),E(y)-F(y)]\\ &= \tfrac{1}{2}[JA(y),E(y)-F(y)] = \tfrac{1}{2}[E(y)+F(y),E(y)-F(y)]\\ &= -[E(y),F(y)], \end{aligned}$$

as was to be shown.

The strategy of the proof of (5.13) is to show that $A(y)$, $E(y)$, $F(y)$ can be developed in a convergent series near $y=\infty$. The differential equations in (9.8) then give recursive relations on the coefficients of the series. If all possible information is squeezed out of these relations and applied to the function $h(y)$, Theorem (5.13) follows. The next few lemmas are aimed at obtaining the convergent series for $A(y)$, etc. As in §8, $\| \ \|$ will denote the $\operatorname{Ad} K$-invariant norm on $\mathfrak{g}$ which comes from the inner product (8.9).

(9.14) **Lemma.** *For all sufficiently large values of y, $\|A(y)\|$, $\|E(y)\|$, $\|F(y)\|$ are bounded by a constant multiple of y^{-1}.*

Proof. According to (5.1),

$$z \mapsto \exp((z+i\alpha)N) \circ a = \Xi(z)$$

defines a horizontal mapping Ξ of the upper half plane into D. In view of (3.17), the length of the tangent vector $\Xi_* \left(\frac{\partial}{\partial z}\right)\Big|_z$, measured with respect to h, is bounded by a multiple of $(\operatorname{Im} z)^{-1}$. For $z = i(y-\alpha)$, with $y > \alpha$, $\Xi(z)$ coincides with the $h(y)$-translate of the base point $o \in D$. Since translation by $h(y)$ preserves the $G_{\mathbb{R}}$-invariant metric, the length of

$$l(h(y)^{-1})_* \, \Xi_* \left(\frac{\partial}{\partial z}\right)\Big|_{i(y-\alpha)},$$

which is a tangent vector at o, does not exceed a multiple of $(y-\alpha)^{-1}$. For $X \in \mathfrak{g}$ and $b \in \check{D}$, the value of the holomorphic vector field "infinitesimal translation by X" will be referred to as $X(b)$. Then

$$\begin{aligned} &l(h(y)^{-1})_* \, \Xi_* \left(\frac{\partial}{\partial z}\right)\Big|_{i(y-\alpha)} \\ &\quad = l(h(y)^{-1})_* \, N(h(y) \circ o) = (\operatorname{Ad} h(y)^{-1} N)(o) = F(y)(o); \end{aligned}$$

under the natural identification between the holomorphic tangent bundle at o and $\mathfrak{g}/\mathfrak{b}$, this tangent vector corresponds to the image of $F(y)$ in $\mathfrak{g}/\mathfrak{b}$. It follows that the length of the projection of $F(y)$ into $\mathfrak{g}/\mathfrak{b}$ – equivalently, the length of the component of $F(y)$ in $\mathfrak{g}^{-1,1}$ – is bounded by a multiple of $(y-\alpha)^{-1}$, and hence also by a multiple of y^{-1}, as long as y is sufficiently large. Since $F(y)$ is real, and since $\mathfrak{g}^{1,-1}$ and $\mathfrak{g}^{-1,1}$ are mutually conjugate, also the length of the component of $F(y)$ in $\mathfrak{q} = \mathfrak{g}^{-1,1} \oplus \mathfrak{g}^{1,-1}$ remains bounded by a multiple of y^{-1}, as $y \to \infty$. Being a conjugate of N, $F(y)$ must be nilpotent; also, $F(y) \in \mathfrak{v}_0 \oplus \mathfrak{q}_0$, with $\mathfrak{v}_0 \subset \mathfrak{k}_0$, $\mathfrak{q}_0 \subset \mathfrak{p}_0$. Thus Lemma (8.11) gives the desired estimate for $F(y)$. The estimates for $A(y)$ and $E(y)$ follow, because $E(y) = -\theta F(y)$, $A(y) = -J(E(y)+F(y))$.

In view of the differential equations (9.8), the k-th derivative of $A(y)$ can be expressed as a Lie polynomial, homogeneous of degree $k+1$, in $A(y)$, $E(y)$, and $F(y)$. Thus:

(9.15) **Corollary.** *For* $k=0,1,2,\ldots,$

$$\|A^{(k)}(y)\|=O(y^{-k-1}) \quad as \quad y\to\infty.$$

I let $G_{\mathbb{R}}=UAK$ be an Iwasawa decomposition, as in the proof of Lemma (8.14). For $g\in G_{\mathbb{R}}$, the components of g in U, A, K depend real analyticly on g. Thus, if I express $h(y)$ as

$$(9.16) \qquad h(y)=u(y)\,a(y)\,k(y), \qquad \text{with } u(y)\in U, \text{ etc},$$

$u(y)$, $a(y)$, $k(y)$ will be real analytic functions on the interval $\{y\in\mathbb{R}\,|\,y>\alpha\}$, with values in U, A, K.

(9.17) **Lemma.** *The matrix entries of* $u(y)$ *and* $a(y)^2$ *are rational functions of* y*, which are regular on the interval* $y\in\mathbb{R}$, $y>\alpha$.

Proof. For $y>\alpha$, the point $\exp(iyN)\circ a$ defines a polarized Hodge structure on $H_{\mathbb{C}}$, which will be denoted as

$$(9.18) \qquad H_{\mathbb{C}}=\oplus H^{p,q}(y).$$

Let $H_{\mathbb{C}}=\oplus H_0^{p,q}$ be the reference Hodge filtration. Since $h(y)$ lies in $G_{\mathbb{R}}$, and since $\exp(iyN)\circ a$ coincides with the $h(y)$-translate of the reference point,

$$(9.19) \qquad H^{p,q}(y)=h(y)\,H_0^{p,q},$$

for all p,q. I choose an element $g_0\in G_{\mathbb{C}}$, whose B-coset represents the point $a\in\check{D}\cong G_{\mathbb{C}}/B$. Then for each integer p,

$$\exp(iyN)\,g_0(\oplus_{j\geqq p}H_0^{j,k-j})=\oplus_{j\geqq p}H^{j,k-j}(y).$$

Let $(w_1,w_2,\ldots,w_s)$ be a basis of $H_{\mathbb{C}}$, such that the first few basis vectors span the last nonzero subspace in the reference Hodge filtration

$$H_{\mathbb{C}}\supset\cdots\supset F_0^{p-1}\supset F_0^p\supset F_0^{p+1}\supset\cdots\supset 0,$$

the next few basis vectors span the next to last nonzero subspace, and so forth. Then, for $y>\alpha$,

$$(9.20) \qquad \{\exp(iyN)\,g_0\,w_j\,|\,1\leqq j\leqq s\}$$

forms a basis with the same property relative to the Hodge structure (9.18); also, because N is a nilpotent linear transformation, these basis vectors have polynomial dependence on y.

I shall have to digress briefly on the Gramm-Schmidt orthogonalization process. Let E be a finite dimensional complex vector space, $F\subset E$ a subspace, $(x,y)\mapsto h(x,y)$ a nondegenerate Hermitian form, such that h is positive definite on F, negative definite on the orthogonal complement of F in E. I suppose that a basis $\{v_1,\ldots,v_n\}$ for E is given, with $v_1,\ldots,v_l$ spanning F, for some $l\leqq n$. Inductively, I define vectors

$u_1, u_2, \ldots, u_n \in E$ by

$$u_1 = v_1, \quad u_r = v_r - \textstyle\sum_{j<r} h(v_r, u_j)\, h(u_j, u_j)^{-1}\, u_j.$$

Because of the hypotheses, the process is well-defined: each u_r lies either in F or $F^\perp$, and thus has nonzero "length". One verifies easily that the u_j are mutually orthogonal with respect to h, and that $\{u_1, \ldots, u_n\}$ forms a basis for E, $u_j \in F$ for $1 \leqq j \leqq l$, $u_j \in F^\perp$ for $j > l$. With respect to an arbitrary fixed basis of E, the coordinates of the u_j are rational functions of the coordinates of the v_j. This "Gramm Schmidt process without normalization" evidently works in the more general setting of a filtration $F_1 \subset \cdots \subset F_k \subset E$, as long as h is nondegenerate on F_j and of definite sign on the orthogonal complement of F_{j-1} in F_j, for $1 \leqq j \leqq k$.

Since S polarizes the Hodge structure (9.18), I can apply the Gramm-Schmidt process without normalization to the basis (9.20) and the Hermitian form $(u, v) \mapsto S(u, \bar{v})$. The result will be a basis $\{w_1(y), \ldots, w_s(y)\}$ for $H_{\mathbb{C}}$, with the following properties:

a) $S\big(w_j(y), \overline{w_l(y)}\big) = 0$ for $j \neq l$, $S\big(w_j(y), \overline{w_j(y)}\big) \neq 0$;

b) $w_j(y)$ is a rational function of y, for $1 \leqq j \leqq S$;

c) each subspace $H^{p,q}(y)$ has a basis $\{w_j(y) \mid l(p,q) \leqq j < l(p-1, q+1)\}$, for suitable integers $l(p,q)$.

The projection of $H_{\mathbb{C}}$ onto the subspace $H^{p,q}(y)$ in the decomposition (9.18) can therefore be represented explicitly as

$$v \mapsto \sum S\big(v, \overline{w_j(y)}\big)\, S\big(w_j(y), \overline{w_j(y)}\big)^{-1}\, w_j(y),$$

with j running from $l(p,q)$ to $l(p-1, q+1) - 1$. In particular, this projection becomes a rational function of y, regular for $y \in \mathbb{R}$, $y > \alpha$. The same assertion then holds also for the Weil operator $C(y)$ corresponding to the Hodge structure (9.18). Let C be the Weil operator of the reference Hodge structure. As was pointed out in §8, $C \in G_{\mathbb{R}}$ and $\theta = \operatorname{Ad} C$. In particular, θ lifts from the Lie algebra to an automorphism of $G_{\mathbb{R}}$, which shall also be denoted by θ. According to the definition of K, θ acts on K as the identity. The Lie algebra $\mathfrak{a}_0$ of the group A in the Iwasawa decomposition lies in the (-1)-eigenspace of θ, so that $\theta(a) = a^{-1}$ for $a \in A$ (cf. [16]). In view of (9.19), $C(y) = h(y)\, C h(y)^{-1}$; as was pointed out before, this linear transformation depends rationally on y, with no singularities on the interval $y \in \mathbb{R}$, $y > \alpha$. Hence the same is true of

$$\begin{aligned} C^{-1} C(y) &= C^{-1} h(y)\, C h(y)^{-1} = \big(\theta^{-1} h(y)\big)\, h(y)^{-1} \\ &= \big(\theta h(y)\big)\, h(y)^{-1} = \big\{\theta\big(u(y)\, a(y)\, k(y)\big)\big\}\, k(y)^{-1}\, a(y)^{-1}\, u(y)^{-1} \\ &= \big(\theta u(y)\big)\, a(y)^{-1}\, k(y)\, k(y)^{-1}\, a(y)^{-1}\, u(y)^{-1} \\ &= \theta u(y)\, a(y)^{-2}\, u(y)^{-1}, \end{aligned}$$

and thus also of its inverse

(9.21) $$u(y)\,a(y)^2\,\theta u(y)^{-1}.$$

Since the Weil operator C is defined over $\mathbb{R}$, the inner product (8.3) restricts to a real inner product on $H_{\mathbb{R}}$. With respect to this inner product, the adjoint of each $g \in G_{\mathbb{R}}$ equals $C^{-1}\,g^{-1}\,C = \theta g^{-1}$. Thus, relative to any orthonomal basis of $H_{\mathbb{R}}$, K is represented by orthogonal matrices, and $g \mapsto \theta g^{-1}$, for $g \in G_{\mathbb{R}}$, corresponds to transposition of matrices. As can be deduced either directly, or from Lemma 9.3 and the proof of Proposition 11.25 of [1], there exists some orthonormal basis of $H_{\mathbb{R}}$, relative to which

a) U is represented by lower triangular matrices, with 1's in all diagonal entries;

b) A is represented by diagonal matrices;

c) $g \mapsto \theta g^{-1}$ corresponds to matrix transposition.

If a nonsingular square matrix M can be expressed as $M = TD\,{}^tT$, where D is diagonal and T lower triangular, with 1's on the diagonal (${}^tT$ = transpose of T), then this expression is unique, and the matrix entries of T and D depend rationally on those of M. Since the matrix entries of (9.21) are rational functions of y, this concludes the proof of the lemma.

According to the definition of an Iwasawa decomposition, the subgroup $A \subset G_{\mathbb{R}}$ is Abelian, connected, simply connected, and consists wholly of semisimple elements with real eigenvalues. Thus, as a matrix group, A can be diagonalized over $\mathbb{R}$. From this, together with the preceeding lemma, it follows that the entries of $a(y)^{-1}\,a'(y)$ must be rational functions, with no singularities on the interval $y \in \mathbb{R}$, $y > \alpha$; also, the entries of $a(y)$ must be linear combinations of square roots of rational functions which assume positive values on $\{y \in \mathbb{R} \mid y > \alpha\}$. Rational functions of y can be expanded as Laurent series in y^{-1}, converging on some punctured neighborhood of $y = \infty$. Likewise, the square root of a rational function $r(y)$, with $r(y) > 0$ for $y \in \mathbb{R}$, $y > \alpha$, can be expressed as a Laurent series in the variable $y^{-\frac{1}{2}}$, such that the series converges and represents the function in question on some interval of the form $y \in \mathbb{R}$, $y > \beta$. Thus:

(9.22) **Corollary.** *The matrix entries of $u(y)$, $u(y)^{-1}\,u'(y)$, and $a(y)^{-1}\,a'(y)$ all have series expansions of the form*

$$a_{m_0}\,y^{-m_0} + a_{m_0+1}\,y^{-m_0-1} + a_{m_0+2}\,y^{-m_0-2} + \cdots, \qquad m_0 \in \mathbb{Z},$$

and the matrix entries of $a(y)$ can be expressed as a series in $y^{-\frac{1}{2}}$, of the type

$$b_{n_0}\,y^{-n_0/2} + b_{n_0+1}\,y^{-(n_0+1)/2} + b_{n_0+2}\,y^{-(n_0+2)/2} + \cdots, \qquad n_0 \in \mathbb{Z};$$

these series expansions are valid and converge on some interval $\{y \in \mathbb{R} \mid y > \beta\}$.

(9.23) **Lemma.** *For* $y \in \mathbb{R}$, $y > \alpha$,

$$k'(y)\,k(y)^{-1} = -\tfrac{1}{2}\operatorname{Ad} a(y)^{-1}\big(u(y)^{-1}\,u'(u)\big) - \tfrac{1}{2}\operatorname{Ad} a(y)\big(\theta\big(u(y)^{-1}\,u'(y)\big)\big),$$

and

$$\operatorname{Ad} k(y)\big(A(y)\big) = -2a(y)^{-1}\,a'(y) - \operatorname{Ad} a(y)^{-1}\big(u(y)^{-1}\,u'(y)\big) + \operatorname{Ad} a(y)\big(\theta\big(u(y)^{-1}\,u'(y)\big)\big).$$

Proof. If the identity (9.16) is differentiated logarithmically and rearranged, using the Definition (9.7), one obtains

$$\begin{aligned}
&\tfrac{1}{2}\operatorname{Ad} k(y)\big(A(y)\big) + k'(y)\,k(y)^{-1}\\
&\quad = -\operatorname{Ad} k(y)\big(h(y)^{-1}\,h(y)\big) + k'(y)\,k(y)^{-1}\\
&\quad = -\operatorname{Ad} a(y)^{-1}\big(u(y)^{-1}\,u'(y)\big) - a(y)^{-1}\,a'(y)\\
&\quad = -a(y)^{-1}\,a'(y) - \tfrac{1}{2}\big\{\operatorname{Ad} a(y)^{-1}\big(u(y)^{-1}\,u'(y)\big)\\
&\qquad - \theta\operatorname{Ad} a(y)^{-1}\big(u(y)^{-1}\,u'(y)\big)\big\}\\
&\qquad - \tfrac{1}{2}\big\{\operatorname{Ad} a(y)^{-1}\big(u(y)^{-1}\,u'(y)\big) + \theta\operatorname{Ad} a(y)^{-1}\big(u(y)^{-1}\,u'(y)\big)\big\}.
\end{aligned}$$

According to (9.8), $A(y)$ takes values in $\mathfrak{q}_0$. Since $\mathfrak{q}_0 \subset \mathfrak{p}_0$, and since $\operatorname{Ad} K(\mathfrak{p}_0) \subset \mathfrak{p}_0$, the first summand on the left hand side of the identity above must lie in $\mathscr{P}_0$; the second summand clearly lies in the Lie algebra of K, namely $\mathfrak{k}_0$. The subspaces $\mathfrak{k}_0$ and $\mathfrak{p}_0$ of $\mathfrak{g}_0$ can be described as, respectively, the $(+1)$- and (-1)-eigenspace of θ; furthermore, by the very definition of an Iwasawa decomposition, the Lie algebra of A lies in $\mathfrak{p}_0$. Thus, equating the components in $\mathfrak{k}_0$, I find

$$k'(y)\,k(y)^{-1} = -\tfrac{1}{2}\big\{\operatorname{Ad} a(y)^{-1}\big(u(y)^{-1}\,u'(y)\big) + \theta\operatorname{Ad} a(y)^{-1}\big(u(y)\,u'(y)\big)\big\}.$$

Since θ acts as multiplication by -1 on the Lie algebra of the connected subgroup $A \subset G_{\mathbb{R}}$, the lifting of θ to an automorphism of $G_{\mathbb{R}}$ acts on A as inversion. Hence, for $a \in A$, $\theta \circ \operatorname{Ad} a = \operatorname{Ad} a^{-1} \circ \theta$, and this now proves the first of the two assertions; the second can be verified similarly.

In view of (9.22) and (9.23), the matrix valued function $k'(y)\,k(y)^{-1}$ can be expanded as a series

(9.24) $$k'(y)\,k(y)^{-1} = \textstyle\sum_{n \geq n_0} Z_n\, y^{-n/2}$$

which converges and represents the function on some interval $\{y \in \mathbb{R} \mid y > \beta\}$. The values of this function lie in $\mathfrak{k}_0$; hence $Z_n \in \mathfrak{k}_0$, for $n = n_0, n_0 + 1, \ldots$.

(9.25) **Lemma.** *For* $n < 2$, $Z_n = 0$.

Proof. I define $Z(y) = \sum_{n \geq n_0} Z_n\, y^{-n/2}$, so that

$$k'(y) = Z(y)\,k(y).$$

If $n_0 \geqq 2$, there is nothing to prove. I shall therefore assume that the leading coefficient Z_{n_0} does not vanish, and that $n_0 < 2$; from this I shall deduce a contradiction. By induction on l, one finds that the l-th derivative of $k(y)$ satisfies an identity

$$k^{(l)}(y) = P_l(Z(y), Z'(y), \ldots, Z^{(l-1)}(y))\, k(y),$$

where P_l is a noncommutative polynomial, homogeneous of weighted degree l when $Z^{(r)}(y)$ is assigned the weight $r+1$, for $r=0, 1, \ldots$; moreover, the monomial $(Z(y))^l$ occurs in P_l with coefficient 1. Since n_0 was assumed to be less than 2, this implies

$$P_l(Z(y), Z'(y), \ldots, Z^{(l-1)}(y)) \\ = (Z_{n_0})^l\, y^{-l n_0/2} + \text{terms involving higher powers of } y^{-\frac{1}{2}}.$$

As an element of the Lie algebra of a compact group of matrices, Z_{n_0} must be semisimple; in particular, $Z_{n_0} \neq 0$ implies $(Z_{n_0})^l \neq 0$, for every $l \geqq 0$. Since $k(y)^{-1}$ ranges over the compact group K, its matrix entries are uniformly bounded. Putting these statements together, I find that, under the hypotheses mentioned above,

$$\|k^{(l)}(y)\| = O(y^s) \quad \text{as } y \to \infty \quad \text{if and only if} \quad s \geqq -\tfrac{1}{2} l\, n_0, \\ \text{for } l = 0, 1, \ldots. \tag{9.26}$$

As a consequence of (9.22), there exists a constant ζ, such that, for $l \geqq 0$,

$$\left\| \frac{d^{(l)}}{d y^l} \left(u(y)^{-1}\right) \right\| = O(y^{\zeta - l}), \quad \text{and} \\ \left\| \frac{d^{(l)}}{d y^l} \left(a(y)^{-1}\right) \right\| = O(y^{\zeta - l}), \quad \text{as } y \to \infty. \tag{9.27}$$

Similarly, one obtains a bound for $h(y) = u(y)\, a(y)\, k(y)$: because the entries of $k(y)$ are uniformly bounded, after enlarging ζ, if necessary,

$$\|h(y)\| = O(y^\zeta) \quad \text{as } y \to \infty. \tag{9.28}$$

Since $2h'(y) = -h(y)\, A(y)$, there exists a noncommutative polynomial Q_l, such that

$$h^{(l)}(y) = h(y)\, Q_l(A(y), A'(y), \ldots, A^{(l-1)}(y));$$

Q_l is homogeneous of weighted degree l if $A^{(r)}(y)$ is assigned the weight $r+1$. Combined with (9.15) and (9.28), this gives the estimate

$$\|h^{(l)}(y)\| = O(y^{\zeta - l}) \quad \text{as } y \to \infty, \tag{9.29}$$

for $l=0, 1, \ldots$. In view of (9.27) and (9.29), when the identity $k(y)=a(y)^{-1}\,u(y)^{-1}\,h(y)$ is differentiated l times, each single differentiation of one of the factors pushes down the order of growth by a factor of y^{-1}; thus

$$(9.30) \qquad \|k^{(l)}(y)\| = O(y^{3\zeta-l}) \quad \text{as } y\to\infty,\ l=0,1,\ldots.$$

Taken together, (9.26) and (9.30) contradict the original hypotheses, namely that $n_0<2$ and $Z_{n_0}\neq 0$. Hence the lemma is proven.

(9.31) **Lemma.** *There exists an element L of the Lie algebra $\mathfrak{k}_0$, and a K-valued function $k_1(y)$, which is a real analytic function of the real variable $y^{-\frac{1}{2}}$ on some interval $|y^{-\frac{1}{2}}|<\eta$, such that $k(y)=k_1(y)\exp(\log y L)$.*

(9.32) *Remark.* It will turn out that $L=0$.

Proof. When the differential equation

$$k'(y) = \left(\textstyle\sum_{n\geq 2} Z_n\, Y^{-n/2}\right) k(y)$$

is rewritten in terms of the variable $t=y^{-\frac{1}{2}}$, it becomes

$$\frac{dk}{dt} = (-2Z_{-2}\,t^{-1} + \text{higher order terms})\,k(t).$$

This latter equation at worst has a regular singular point at $t=0$. Moreover, the leading coefficient $-2Z_{-2}$, which is an element of the Lie algebra of a compact matrix group, has purely imaginary eigenvalues. In particular, no two distinct eigenvalues of the leading coefficient differ by an integer. Hence, from the theory of linear differential equations with regular singular points (e.g. Chapter IV, §4 of [4]) it follows that any solution, and hence also the definite solution $k(y)$ occuring in (9.16), is of the form

$$k(y) = m(y)\exp(\log y\, Z_{-2})\, m_0$$

(because $\log y = -2\log t$); here $m(y)$ denotes a matrix valued function whose entries have power series expansions in $y^{-\frac{1}{2}}$ near $y=\infty$, and such that $m(\infty)=1$; m_0 stands for a constant matrix. Since $k(y)$, when defined, is invertible, the matrix m_0 must also have an inverse. The matrix entries of $m(y)^{-1}$ also have convergent power series expansions in $y^{-\frac{1}{2}}$, because $m(\infty)=1$. Thus

$$\|m(y)^{-1}\, m'(y)\| = O(y^{-\frac{3}{2}}) \quad \text{as } y\to\infty.$$

Since $\exp(\log y\, Z_{-2})\in K$ for $y\in\mathbb{R}$, the entries of this matrix, as well as those of its inverse, remain bounded near $y=\infty$. I deduce:

$$\begin{aligned}
&\lim_{y\to\infty}(y\,k(y)^{-1}\,k'(y)) \\
&\quad = m_0^{-1}\, Z_{-2}\, m_0 + \lim_{y\to\infty}\{\mathrm{Ad}(m_0^{-1}\exp(-\log y\, Z_{-2}))(y\,m(y)^{-1}\,m'(y))\} \\
&\quad = m_0^{-1}\, Z_{-2}\, m_0 .
\end{aligned}$$

On the other hand, for those real values of y for which it is defined, $k(y)^{-1}\,k'(y)$ lies in $\mathfrak{k}_0$, so that $m_0^{-1}\,Z_{-2}\,m_0 \in \mathfrak{k}_0$. I now define $L = m_0^{-1}\,Z_{-2}\,m_0$, $k_1(y) = m(y)\,m_0$. Then $k(y) = k_1(y)\exp(\log y\,L)$, as desired. This last identity also forces $k_1(y)$ to take values in K, since $L \in \mathfrak{k}_0$. All the other stated properties have already been verified.

The next step of the proof consists of showing that $L=0$. For this purpose, a certain algebraic statement is needed. In my original version of the proof of (5.13), this statement was implicit, though rather well hidden; Deligne extracted it, cleaned up its proof, and recast it in the following neat form:

(9.33) **Lemma.** *Let $\mathfrak{a}_0$, $\mathfrak{b}_0$ be Lie algebras of compact Lie groups, $\rho\colon \mathfrak{a}_0 \to \mathfrak{b}_0$ a linear map, and $L \in \mathfrak{b}_0$ some element of $\mathfrak{b}_0$, such that*

$$[\rho X, \rho Y] = (1 + \operatorname{ad} L) \circ \rho\,[X, Y]$$

for all X, $Y \in \mathfrak{a}_0$. Then ρ is a Lie algebra homomorphism.

Proof. It must be shown that L commutes with $\rho\,[\mathfrak{a}_0, \mathfrak{a}_0]$. The derived algebra of a compact Lie algebra is semisimple and therefore equal to its derived algebra; moreover, it is again the Lie algebra of a compact group [16]. One may therefore assume that $\mathfrak{a}_0$ is semisimple. A compact semisimple Lie algebra is the linear span of some number of subalgebras which are isomorphic to $\mathfrak{su}(2)$ [16]; this reduces the problem to the case when $\mathfrak{a}_0 = \mathfrak{su}(2)$. I let X, Y, Z, denote the images under ρ of, respectively,

$$\tfrac{1}{2}\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \quad \tfrac{1}{2}\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad \tfrac{1}{2}\begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix};$$

then, according to the hypotheses,

$$[X, Y] = Z + [L, Z],$$
$$[Y, Z] = X + [L, X],$$
$$[Z, X] = Y + [L, Y].$$

Using this, together with the Jacobi identity, one finds

$$\begin{aligned}&[X, [X, L]] + [Y, [Y, L]] + [Z, [Z, L]] \\ &\quad = -[X, [Y, Z]] - [Y, [Z, X]] - [Z, [X, Y]] = 0.\end{aligned}$$

Since $\mathfrak{b}_0$ is the Lie algebra of a compact Lie group, one can choose a negative definite, ad-invariant bilinear form $B(\ ,\ )$ on $\mathfrak{b}_0$ (for example, the trace form of a faithful finite dimensional representation). Then

$$\begin{aligned}&B([X, L], [X, L]) + B([Y, L], [Y, L]) + B([Z, L], [Z, L]) \\ &\quad = -B([X, [X, L]], L) - B([Y, [Y, L]], L) - B([Z, [Z, L]], L) \\ &\quad = 0.\end{aligned}$$

Because of the negative definiteness of B, L must commute with $\rho[\mathfrak{a}_0, \mathfrak{a}_0] = \rho\, \mathfrak{a}_0$, as was to be shown.

(9.34) **Corollary.** *Let $\mathfrak{g}_1$, $\mathfrak{g}_2$ be real semisimple Lie algebras, with Cartan involutions θ_1, θ_2, and $\rho: \mathfrak{g}_1 \to \mathfrak{g}_2$ a linear map which is compatible with the involutions* (i.e. $\theta_2 \circ \rho = \rho \circ \theta_1$). *If there exists some $L \in \mathfrak{g}_2$, with $\theta_2 L = L$, such that $[\rho X, \rho Y] = (1 + \operatorname{ad} L) \circ \rho[X, Y]$ for all $X, Y \in \mathfrak{g}_1$, then ρ is a homomorphism of Lie algebras.*

Proof. I set $\mathfrak{m}_i = \{X \in \mathfrak{g}_i \otimes_{\mathbb{R}} \mathbb{C} \mid \theta_i X = \bar{X}\}$; then $\mathfrak{m}_1$, $\mathfrak{m}_2$ are Lie algebras of compact Lie groups [16]. Because of the hypotheses on ρ, the complex linear extension maps $\mathfrak{m}_1$ to $\mathfrak{m}_2$, and ρ, L, $\mathfrak{m}_1$, $\mathfrak{m}_2$ satisfy the assumptions of the lemma.

(9.35) **Lemma.** *In the notation of Lemma* (9.31), $L = 0$.

Proof. According to (9.16), (9.22), and (9.31), one can express the function $h(y)$ as

$$h(y) = h_1(y) \exp(\log y\, L),$$

where $h_1(y)$ is a $G_{\mathbb{R}}$-valued function whose matrix entries have convergent Laurant series expansions in $y^{-\frac{1}{2}}$ near $y = \infty$, and $L \in \mathfrak{k}_0$. I define

$$\tilde{A}(y) = \operatorname{Ad} \exp(\log y\, L)(A(y)),$$
$$\tilde{E}(y) = \operatorname{Ad} \exp(\log y L)(E(y)),$$
$$\tilde{F}(y) = \operatorname{Ad} \exp(\log y L)(F(y)).$$

As one can compute easily,

$$\tilde{A}(y) = -2 h_1(h)^{-1} h_1'(y) - 2 y^{-1} L, \tag{9.36}$$

which implies that $\tilde{A}(y)$ has a Laurant series expansion in $y^{-\frac{1}{2}}$ near $y = \infty$. Similarly,

$$\tilde{F}(y) = \operatorname{Ad} h_1(y)^{-1} N$$

and $\tilde{E}(y) = -\theta \tilde{F}(y)$ (θ commutes with $\operatorname{Ad} K$!) have such Laurant series expansions. Since K operates unitarily on $\mathfrak{g}$ under the adjoint action, and in view of (9.14), the lowest power of y^{-1} which can occur in these expansions is the first power, so that

$$\tilde{A}(y) = \tilde{A}_2\, y^{-1} + \tilde{A}_3\, y^{-\frac{3}{2}} + \cdots,$$
$$\tilde{E}(y) = \tilde{E}_2\, y^{-1} + \tilde{E}_3\, y^{-\frac{3}{2}} + \cdots,$$
$$\tilde{F}(y) = \tilde{F}_2\, y^{-1} + \tilde{F}_3\, y^{-\frac{3}{2}} + \cdots.$$

Using the differential equation for $A(y)$ in (9.8), I find

$$-\tilde{A}_2\, y^{-2} - \tfrac{3}{2} A_3\, y^{-\frac{5}{2}} - \cdots = \tilde{A}'(y)$$
$$= \operatorname{Ad}\exp(\log y L)(A'(y)) + y^{-1}[L, \tilde{A}(y)]$$
$$= -[\tilde{E}(y), \tilde{F}(y)] + y^{-1}[L, \tilde{A}(y)]$$
$$= (-[\tilde{E}_2, \tilde{F}_2] + [L, \tilde{A}_2])\, y^{-2} + \cdots,$$

and hence $[\tilde{E}_2, \tilde{F}_2] = (1 + \operatorname{ad} L)(\tilde{A}_2)$. Similarly,

$$[\tilde{A}_2, \tilde{E}_2] = (1 + \operatorname{ad} L)(2\tilde{E}_2), \qquad [\tilde{A}_2, \tilde{F}_2] = (1 + \operatorname{ad} L)(-2\tilde{F}_2).$$

Since $\operatorname{Ad} K$ commutes with θ, $\theta \tilde{A}_2 = -\tilde{A}_2$, and $\theta \tilde{E}_2 = \tilde{F}_2$. There exists a unique linear map $\rho\colon \mathfrak{sl}(2, \mathbb{R}) \to \mathfrak{g}_0$ which sends

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$$

to, respectively, $\tilde{A}_2, \tilde{E}_2, \tilde{F}_2$. Under this map, the usual Cartan involution on $\mathfrak{sl}(2, \mathbb{R})$, namely $X \to -{}^tX$, corresponds to the Cartan involution θ of $\mathfrak{g}_0$. In view of the identities above, and because $\theta L = L$, (9.34) can be applied to this situation; in particular, $[L, \tilde{A}_2] = 0$. As a consequence of (9.36),

$$h_1(y)^{-1}\, h_1'(y) = (\tfrac{1}{2}\tilde{A}_2 - L)\, y^{-1} + \cdots;$$

when this differential equation is rewritten in terms of the variable $t = y^{-\frac{1}{2}}$, it becomes

$$\frac{dh_1}{dt} = \{(2L - \tilde{A}_2)\, t^{-1} + \text{higher order terms}\}\, h_1(t).$$

According to the elementary theory of ordinary differential equations with regular singular points (e.g. Chapter IV, §4 in [4]), the general solution, near $t = 0$, cannot have a Laurent series expansion in the variable t, unless the leading coefficient $(2L - \tilde{A}_2)$ has only integral eigenvalues. On the other hand, $h_1(t)$ does have a Laurent expansion in t near $t = 0$; hence the eigenvalues of $(2L - \tilde{A}_2)$ are all integral. Since $L \in \mathfrak{k}_0$, L is semisimple and has only purely imaginary eigenvalues, whereas the eigenvalues of $\tilde{A}_2 \in \mathfrak{p}_0$ must be real ($\mathfrak{k}_0 \oplus i\,\mathfrak{p}_0$ is the Lie algebra of a compact Lie group; cf. §8). Also, L and $\tilde{A}_2$ are known to commute. But all this is only possible if $L = 0$.

Combined with (9.22) and (9.31), the lemma implies

(9.37) **Corollary.** *The function $k(y)$ extends to a real analytic K-valued function of the real variable $y^{-\frac{1}{2}}$ near $y = \infty$. In particular, $k(\infty)$ lies in K. The matrix entries of $h(y), h(y)^{-1}, A(y), E(y), F(y)$ all have convergent Laurent series expansions in $y^{-\frac{1}{2}}$ near $y = \infty$.*

Because of the statement (9.14), the lowest power of $y^{-\frac{1}{2}}$ which can occur in the series expansions of $A(y), E(y), F(y)$ is y^{-1}. Thus I can write

$$(9.38)\qquad \begin{aligned} &A(y)=\textstyle\sum_{n\geqq 0} A_n\, y^{-(n+2)/2}, \qquad E(y)=\sum_{n\geqq 0} E_n\, y^{-(n+2)/2}\\ &F(y)=\textstyle\sum_{n\geqq 0} F_n\, y^{-(n+2)/2}.\end{aligned}$$

The coefficients A_n, E_n, F_n lie in $\mathfrak{g}_0$, because $A(y)$, $E(y)$, $F(y)$ are $\mathfrak{g}_0$-valued functions. The labelling of the indices in (9.38) does not agree with my previous practice, but it will turn out to be more convenient. In terms of the coefficients of the series, the differential equations (9.8) become

$$(9.39)\qquad \begin{aligned} (n+2)\, E_n &= \textstyle\sum_{k=0}^{n} [A_k, E_{n-k}],\\ (n+2)\, F_n &= \textstyle\sum_{k=0}^{n} [A_k, F_{n-k}],\\ (n+2)\, A_n &= \textstyle\sum_{k=0}^{n} [E_k, F_{n-k}].\end{aligned}$$

By setting $n=0$, one finds

$$(9.40)\qquad [A_0, E_0]=2E_0, \qquad [A_0, F_0]=-2F_0, \qquad [E_0, F_0]=A_0.$$

Thus A_0, E_0, F_0 span a subalgebra of $\mathfrak{g}_0$, which I shall denote by $\mathfrak{s}_0$; $\mathfrak{s}$ will stand for the complexification of $\mathfrak{s}_0$ in $\mathfrak{g}$. The commutation relations (9.40) are those of the standard generators of $\mathfrak{sl}(2,\mathbb{R})$. Hence $\mathfrak{s}_0\cong\mathfrak{sl}(2,\mathbb{R})$, unless A_0, E_0, F_0 all vanish. In this latter case, from (9.39) one deduces by induction on n that all the coefficients A_n, E_n, F_n must be zero. But then $A(y)\equiv 0$, $F(y)\equiv 0$, so that $h(y)$ must remain constant, and

$$N=\operatorname{Ad} h(y)(F(y))=0.$$

For emphasis:

$$(9.41)\qquad \mathfrak{s}_0\cong\mathfrak{sl}(2,\mathbb{R}), \qquad \mathfrak{s}\cong\mathfrak{sl}(2,\mathbb{C}), \qquad \text{unless } N=0.$$

From now on, I specifically exclude the trivial situation, when $N=0$. Via ad, $\mathfrak{s}$ is represented on $\mathfrak{g}$. I recall the basic facts about representations of $\mathfrak{sl}_2$, which were stated in §6. In particular, $\mathfrak{g}$ breaks up into a direct sum of $\mathfrak{s}$-invariant, $\mathfrak{s}$-irreducible subspaces. For each such irreducible subspace, the collection of eigenvalues of $\operatorname{ad} A_0$ is of the form

$$\{r, r-2, r-4, \ldots, -r\}, \qquad \text{with } r\in\mathbb{Z},\ r\geqq 0;$$

the eigenvalues all have multiplicity one. Corresponding to every non-negative integer r and every integer s, I define subspaces

(9.42) $\mathfrak{g}(r)=$ linear span of all $\mathfrak{s}$-invariant, irreducible subspaces of $\mathfrak{g}$ on which $\operatorname{ad} A_0$ has eigenvalues $r, r-2, \ldots, -r$; $\mathfrak{g}(r,s)=s$-eigenspace of $\operatorname{ad} A_0$ in $\mathfrak{g}(r)$.

Then $\mathfrak{g}$ becomes the direct sum of these subspaces:

(9.43) $$\mathfrak{g}=\oplus_{r\geqq 0}\mathfrak{g}(r), \quad \mathfrak{g}(r)=\oplus_{0\leqq i\leqq r}\mathfrak{g}(r, r-2i).$$

The mapping $[\ ,\ ]: \mathfrak{g}\oplus\mathfrak{g}\to\mathfrak{g}$ commutes with the action of $\mathfrak{s}$. Hence, for $r_1, r_2\geqq 0$, $[\mathfrak{g}(r_1), \mathfrak{g}(r_2)]$ is an $\mathfrak{s}$-invariant subspace of $\mathfrak{g}$, on which $\operatorname{ad} A_0$ has no eigenvalues exceeding r_1+r_2, and all of the same parity as r_1+r_2. This implies

(9.44) $$[\mathfrak{g}(r_1, s_1), \mathfrak{g}(r_2, s_2)]\subset\oplus_{0\leqq i}\mathfrak{g}(r_1+r_2-2i, s_1+s_2).$$

For future reference, I record the identities

(9.45) $$[E_0, \mathfrak{g}(r, s)]\subset\mathfrak{g}(r, s+2), \quad [F_0, \mathfrak{g}(r, s)]\subset\mathfrak{g}(r, s-2),$$ $$[A_0, \mathfrak{g}(r, s)]\subset\mathfrak{g}(r, s),$$

which are a simple consequence of the commutation relations (9.40) and the $\mathfrak{s}$-invariance of the subspaces $\mathfrak{g}(r)$.

I define an element Ω of the universal enveloping algebra of $\mathfrak{s}$ by the formula

(9.46) $$\Omega=2E_0F_0+2F_0E_0+A_0A_0.$$

One can verify that Ω lies in the center of the universal enveloping algebra of $\mathfrak{s}$, either directly or by identifying Ω with eight times the Casimir operator of $\mathfrak{s}$. It follows that Ω acts as a constant mutiple of the identity under every irreducible representation of $\mathfrak{s}$. According to the Definition (9.42) (if one takes into account the first identity in (9.45)), each irreducible subspace of $\mathfrak{g}(r)$ contains a nonzero element X, with the properties $[A_0, X]=rX$, $[E_0, X]=0$. One computes easily

$$\begin{aligned}\operatorname{ad}\Omega(X)&=(2\operatorname{ad}E_0\operatorname{ad}F_0+2\operatorname{ad}F_0\operatorname{ad}E_0+(\operatorname{ad}A_0)^2)(X)\\&=(4\operatorname{ad}F_0\operatorname{ad}E_0+2\operatorname{ad}[E_0, F_0]+(\operatorname{ad}A_0)^2)(X)\\&=(r^2+2r)X.\end{aligned}$$

Hence Ω acts on all of $\mathfrak{g}(r)$ as multiplication by (r^2+2r). For distinct nonnegative r's, these numbers are also distinct, and this gives another characterization of $\mathfrak{g}(r)$:

(9.47) $$\mathfrak{g}(r)=\{X\in\mathfrak{g}\,|\operatorname{ad}\Omega(X)=(r^2+2r)X\}.$$

I now decompose the coefficients of the series (9.38) into their components in the various subspaces $\mathfrak{g}(r, s)$:

$$A_n=\textstyle\sum_{r,s}A_n^{r,s}, \quad E_n=\sum_{r,s}E_n^{r,s}, \quad F_n=\sum_{r,s}F_n^{r,s},$$ $$\text{with } A_n^{r,s}, E_n^{r,s}, F_n^{r,s}\in\mathfrak{g}(r, s), \quad \text{for } n=1, 2, \ldots$$

(9.48) **Lemma.** *The following statements holds for* $n>0$:

a) *If* $n<r$, *or if* $n-r$ *is not an even integer,* $A_n^{r,s}=E_n^{r,s}=F_n^{r,s}=0$;

b) $A_n^{n,n}=A_n^{n,-n}=E_n^{n,-n}=E_n^{n,2-n}=F_n^{n,n}=F_n^{n,n-2}=0$;

c) $A_{n-2}^{n-2,n-2}$, $A_{n-2}^{n-2,2-n}$, $E_{n-2}^{n-2,2-n}$, $E_{n-2}^{n-2,4-n}$, $F_{n-2}^{n-2,n-2}$, $F_{n-2}^{n-2,n-4}$ *all vanish.*

Proof. By induction on n. I let $n>0$ be given, and I assume that the statements in the lemma have been verified for all coefficients A_k, E_k, F_k with $0<k<n$. This induction hypothesis is vacuously satisfied for $n=1$. I define

$$X_n=2\sum_{0<k<n}[E_k, F_{n-k}], \qquad X_n^{r,s}=\text{component of } X_n \text{ in } \mathfrak{g}(r,s);$$
$$Y_n=\sum_{0<k<n}[A_k, E_{n-k}], \qquad Y_n^{r,s}=\text{component of } Y_n \text{ in } \mathfrak{g}(r,s);$$
$$Z_n=-\sum_{0<k<n}[A_k, F_{n-k}], \qquad Z_n^{r,s}=\text{component of } Z_n \text{ in } \mathfrak{g}(r,s).$$

The induction hypotheses, together with (9.44), imply

(9.49)

a) if $n<r$, or if $n-r$ is not an even integer, $X_n^{r,s}=Y_n^{r,s}=Z_n^{r,s}=0$;

b) $X_n^{n,s}=0$ if $s=\pm n$, $\pm(n-2)$; $Y_n^{n,s}=0$ if $s=-n, 2-n, 4-n$, or $s=n$; $Z_n^{n,s}=0$ if $s=n, n-2, n-4$, or $s=-n$;

c) $X_n^{n-2,n-2}=X_n^{n-2,2-n}=Y_n^{n-2,2-n}=Y_n^{n-2,4-n}=Z_n^{n-2,n-2}=Z_n^{n-2,n-4}=0$.

From (9.39), by equating the components in the subspaces $\mathfrak{g}(r,s)$, taking into account (9.45), one finds

$$\begin{aligned} (n+2)\,E_n^{r,s}&=[A_0, E_n^{r,s}]-[E_0, A_n^{r,s-2}]+Y_n^{r,s},\\ (n+2)\,F_n^{r,s}&=-[A_0, F_n^{r,s}]+[F_0, A_n^{r,s+2}]+Z_n^{r,s},\\ (n+2)\,A_n^{r,s}&=2\,[E_0, F_n^{r,s-2}]-2\,[F_0, E_n^{r,s+2}]+X_n^{r,s}. \end{aligned} \tag{9.50}$$

In the first and second of these equations, I replace s by, respectively, $s+2$ and $s-2$; since $\operatorname{ad} A_0$ operates on $\mathfrak{g}(r,s)$ as multiplication by s, this gives

$$\begin{aligned} (n-s)\,E_n^{r,s+2}&=-[E_0, A_n^{r,s}]+Y_n^{r,s+2},\\ (n+s)\,F_n^{r,s-2}&=[F_0, A_n^{r,s}]+Z_n^{r,s-2}. \end{aligned} \tag{9.51}$$

Next, I multiply the third equation in (9.50) by n^2-s^2 and substitute the equations (9.51):

$$
\begin{aligned}
&(n+2)(n^2-s^2)\,A_n^{r,s}\\
&\quad=(n^2-s^2)\,X_n^{r,s}+2(n-s)\big[E_0,[F_0,A_n^{r,s}]\big]+2(n-s)[E_0,A_n^{r,s-2}]\\
&\qquad+2(n+s)\big[F_0,[E_0,A_n^{r,s}]\big]-2(n+s)[F_0,Y_n^{r,s+2}]\\
&\quad=2n(\operatorname{ad}E_0\operatorname{ad}F_0+\operatorname{ad}F_0\operatorname{ad}E_0)(A_n^{r,s})-2s[A_0,A_n^{r,s}]\\
&\qquad+2(n-s)[E_0,Z_n^{r,s-2}]-2(n+s)[F_0,Y_n^{r,s+2}]+(n^2-s^2)\,X_n^{r,s}\\
&\quad=n(\Omega-(\operatorname{ad}A_0)^2)(A_n^{r,s})-2s^2A_n^{r,s}+2(n-s)[E_0,Z_n^{r,s-2}]\\
&\qquad-2(n+s)[F_0,Y_n^{r,s+2}]+(n^2-s^2)\,X_n^{r,s}\\
&\quad=n\big((r^2+2r)-(n+2)s^2\big)A_n^{r,s}+2(n-s)[E_0,Z_n^{r,s-2}]\\
&\qquad-2(n+s)[F_0,Y_n^{r,s+2}]+(n^2-s^2)\,X_n^{r,s}
\end{aligned}
$$

(cf. (9.46) and (9.47)). Thus

$$
\begin{aligned}
&n\big((n+1)^2-(r+1)^2\big)A_n^{r,s}\\
&\quad=2(n-s)[E_0,Z_n^{r,s-2}]-2(n-s)[F_0,Y_n^{r,s+2}]+(n^2-s^2)\,X_n^{r,s}.
\end{aligned}\tag{9.52}
$$

If $n<r$, or if $n-r$ is not an even integer, the right hand side of (9.52) vanishes because of the induction hypotheses. In this case, since the coefficient of $A_n^{r,s}$ in (9.53) is nonzero, $A_n^{r,s}$ must also vanish. But then (9.52) and (9.49 a) give

$$E_n^{r,s}=0,\qquad \text{except possibly for } s=n+2;$$

$$F_n^{r,s}=0,\qquad \text{except possibly for } s=-n-2.$$

Still under the hypothesis that $n<r$ or $n-r\notin 2\mathbb{Z}$, according to the previous conclusions and the third equation in (9.51),

$$[F_0,E_n^{r,n+2}]=0,\qquad [E_0,F_n^{r,-n-2}]=0.$$

For $s\neq -r$, $\operatorname{ad}F_0$ maps $\mathfrak{g}(r,s)$ injectively to $\mathfrak{g}(r,s-2)$ (cf. (6.3)); also, r is known to be nonnegative, and $n>0$. Hence $E_n^{r,n+2}=0$, and similarly one finds that $F_n^{r,-n-2}=0$. This concludes the induction step for (9.48 a).

As a consequence of the definition, $\mathfrak{g}(r,s)=0$ unless s is an integer between r and $-r$, of the same parity as r. Together with (9.49 b, c), (9.50), and (9.51), the preceeding statement gives the following information:

(9.53) if $r=n$ or $r=n-2$, and if $s=\pm n,\ \pm(n-2)$,

$$(n+2)\,A_n^{r,s}=2[E_0,F_n^{r,s-2}]-2[F_0,E_n^{r,s+2}],$$

$$(n-s)\,E_n^{r,s+2}=-[E_0,A_n^{r,s}],\ \text{ and}$$

$$(n+s)\,F_n^{r,s-2}=[F_0,A_n^{r,s}].$$

Since $E_n^{n,n+2}=0$ and $[E_0, A_n^{n,n}]=0$, I deduce

$$\begin{aligned} n(n+2)\, A_n^{n,n} &= 2n[E_0, F_n^{n,n-2}] = [E_0, [F_0, A_n^{n,n}]] \\ &= [E_0, [F_0, A_n^{n,n}]] - [F_0, [E_0, A_n^{n,n}]] \\ &= [A_0, A_n^{n,n}] = n A_n^{n,n}, \end{aligned}$$

so that $A_n^{n,n}=0$. Similarly,

$$\begin{aligned} (n-1)(n+2)\, A_n^{n-2,n-2} &= (2n-2)[E_0, F_n^{n-2,n-4}] \\ &= [E_0, [F_0, A_n^{n-2,n-2}]] \\ &= [A_0, A_n^{n-2,n-2}] = (n-2)\, A_n^{n-2,n-2}, \end{aligned}$$

and hence $A_n^{n-2,n-2}=0$. For trivial reasons, $A_n^{n,n+2}=A_n^{n-2,n}=0$. If these conclusions are fed back into the last equation in (9.53), one finds

$$F_n^{n,n} = F_n^{n,n-2} = F_n^{n-2,n-2} = F_n^{n-2,n-4} = 0.$$

(For $n=1$, the coefficient of $F_n^{n-2,n-4}$ in (9.54) vanishes, but $F_1^{-1,-3}$ is zero because of trivial reasons.) By completely analogous arguments,

$$A_n^{n,-n} = A_n^{n-2,2-n} = E_n^{n,-n} = E_n^{n,2-n} = E_n^{n-2,2-n} = E_n^{n-2,4-n} = 0.$$

This concludes the proof by induction.

Since A_0 lies in the Lie algebra of $\mathfrak{g}_0$, the function $y \mapsto \exp(-\frac{1}{2}\log y\, A_0)$ takes values in $G_{\mathbf{R}}$. Hence the relation

$$h(y) = g(y)\exp(-\tfrac{1}{2}\log y\, A_0) \tag{9.54}$$

defines a $G_{\mathbf{R}}$-valued function $g(y)$.

(9.55) **Lemma.** *The function $g(y)$ is a real analytic, $G_{\mathbf{R}}$-valued function of the real variable y^{-1} near $y=\infty$. Moreover, $F_0 = \operatorname{Ad} g(\infty)^{-1}(N)$.*

Proof. Logarithmic differentiation of (9.54) gives

$$A(y) = A_0\, y^{-1} - 2\operatorname{Ad}\exp(\tfrac{1}{2}\log y\, A_0)(g(y)^{-1} g'(y)),$$

or equivalently,

$$\begin{aligned} g(y)^{-1} g'(y) &= -\tfrac{1}{2}\operatorname{Ad}\exp(-\tfrac{1}{2}\log y\, A_0)(A(y)) + \tfrac{1}{2} A_0\, y^{-1} \\ &= -\tfrac{1}{2}\operatorname{Ad}\exp(-\tfrac{1}{2}\log y\, A_0) \\ &\quad \cdot (A_0\, y^{-1} + \textstyle\sum_{n>0}\sum_{r,s} A_n^{r,s}\, y^{-(n+2)/2}) + \tfrac{1}{2} A_0\, y^{-1} \\ &= -\tfrac{1}{2}\textstyle\sum_{n>0}\sum_{r,s} A_n^{r,s}\, y^{-(n+s+2)/2}. \end{aligned} \tag{9.56}$$

According to (9.48), $A_n^{r,s}=0$ unless n and r have the same parity. Also, $\mathfrak{g}(r,s)=0$ unless $r-s$ is even. Hence only integral powers of y occur with nonzero coefficients on the right hand side of (9.56). Again by (9.48), if

$n>0$ and $A_n^{r,s}\neq 0$, $n+s$ must be strictly positive. Thus $g(y)^{-1}g'(y)$ has a series expansion of the form

(9.57) $$g(y)^{-1}g'(y)=\sum_{k\geqq 2}B_k\,y^{-k},$$

which converges near $y=\infty$. One can regard (9.57) as a differential equation for $g(y)$; as such, it has a regular point at $y=\infty$, and this implies the first conclusion of the lemma. In view of the definition of the function $F(y)$,

$$\begin{aligned}\operatorname{Ad}g(y)^{-1}(N)&=\operatorname{Ad}\exp(-\tfrac{1}{2}\log y\,A_0)(F(y))\\&=\operatorname{Ad}\exp(-\tfrac{1}{2}\log y\,A_0)\big(F_0\,y^{-1}+\textstyle\sum_{n>0}\sum_{r,s}F_n^{r,s}\,y^{-(n+2)/2}\big)\\&=F_0+\textstyle\sum_{n>0}\sum_{r,s}F_n^{r,s}\,y^{-(n+s+2)/2}.\end{aligned}$$

In this last series, only positive powers of y^{-1} can occur, since $F_n^{r,s}=0$ for $s<-n$. By letting y tend to ∞, one obtains the second statement.

According to the lemma, $g(y)$ and $g(y)^{-1}$ can be expanded in series of the form

(9.58) $$\begin{aligned}g(y)&=g(\infty)(1+g_1\,y^{-1}+g_2\,y^{-2}+\cdots),\\ g(y)^{-1}&=(1+f_1\,y^{-1}+f_2\,y^{-2}+\cdots)\,g(\infty)^{-1},\end{aligned}$$

which converge and are valid for all sufficiently large values of y. The coefficients g_k, f_k lie in $\operatorname{Hom}(H_{\mathbb{R}},H_{\mathbb{R}})$. In view of the facts about representations of $\mathfrak{sl}_2$ which were mentioned in §6, A_0 operates on $\operatorname{Hom}(H_{\mathbb{R}},H_{\mathbb{R}})$ semisimply, with integral eigenvalues. Thus I can decompose the coefficients g_k and f_k as follows:

(9.59) $$\begin{aligned}&g_k=\textstyle\sum_{s\in\mathbb{Z}}g_{k,s},\quad f_k=\textstyle\sum_{s\in\mathbb{Z}}f_{k,s},\quad \text{with } g_{k,s},f_{k,s}\in\operatorname{Hom}(H_{\mathbb{R}},H_{\mathbb{R}}),\ \text{and}\\ &[A_0,g_{k,s}]=s\,g_{k,s},\quad [A_0,f_{k,s}]=s\,f_{k,s}.\end{aligned}$$

(9.60) **Lemma.** *Unless* $s\leqq k-1$, $g_{k,s}$ *and* $f_{k,s}$ *vanish.*

Proof. Slightly more detailed information will be needed below. I shall therefore record certain statements which are not important for the proof of (9.60) itself. Also, with a shift in notation, I shall replace $g(y)$ by $g(\infty)^{-1}g(y)$; this has the effect of setting $g(\infty)$ equal to 1, which simplifies the formulas below, but does not affect the general validity of the argument. According to (9.56),

$$g'(y)=g(y)\{-\tfrac{1}{2}\textstyle\sum_{n\geqq 1}\sum_{r,s}A_n^{r,s}\,y^{-(n+s+2)/2}\},$$

and $A_n^{r,s}=0$ unless $n+s$ is even, $r\leqq n$, and $|s|\leqq r-2$. Thus

(9.61) $$\begin{aligned}&g'(y)=g(y)\{\textstyle\sum_{k\geqq 2}B_k\,y^{-k}\},\quad\text{with}\\ &B_{k+1}=-\tfrac{1}{2}\textstyle\sum_{s\leqq k-1}\sum_{r\leqq 2k-s}A_{2k-s}^{r,s}.\end{aligned}$$

When the series (9.58) is substituted in (9.61), one finds

$$g_k = -\frac{1}{k} \sum_{2 \leqq l \leqq k+1} g_{k+1-l} B_l.$$

Hence, by induction on k:

(9.62) g_k is a noncommutative polynomial in $B_2, B_3, \ldots, B_{k+1}$, homogeneous of weighted degree k when B_{l+1} is assigned the weight l; B_{k+1} occurs with coefficient $-\frac{1}{k}$, and B_2^k with coefficient $(-1)^k \frac{1}{k!}$.

Multiplying the two series in (9.58) gives the identities

$$f_k = -\sum_{1 \leqq l \leqq k} f_{k-l} g_l, \qquad \text{for } k \geqq 1.$$

Hence, by induction on k, and using (9.62), it can be shown that

(9.63) f_k is a noncommutative polynomial in $B_2, B_3, \ldots, B_{k+1}$, homogeneous of weighted degree k when B_{l+1} is assigned the weight l; B_{k+1} occurs with coefficient $\frac{1}{k}$, and B_2^k with coefficient $\frac{1}{k!}$.

As is asserted by (9.61), B_{l+1} has a nonzero component in the s-eigenspace of A_0 on $\mathrm{Hom}(H_{\mathbb{R}}, H_{\mathbb{R}})$ only if $s \leqq l-1$. Under composition in $\mathrm{Hom}(H_{\mathbb{R}}, H_{\mathbb{R}})$, the s_1- and the s_2-eigenspace of A_0 get mapped into the (s_1+s_2)-eigenspace. Thus the lemma follows from (9.62) and (9.63).

Instead of considering the action of $G_{\mathbb{R}}$ on $H_{\mathbb{R}}$, one may compose $g(y)$ with any finite dimensional representation of $G_{\mathbb{C}}$. The proof of (9.60) carries over to this more general situation without difficulty.

(9.64) **Corollary.** *Let π be a representation of $G_{\mathbb{C}}$ on a finite dimensional complex vector space W. The coefficients of y^{-k} in the power series expansions of $\pi(g(\infty)^{-1} g(y))$ and of $\pi(g(y)^{-1} g(\infty))$ have zero components in the s-eigenspace of A_0 on* $\mathrm{Hom}(W, W)$, *whenever $s \geqq k$.*

So far, the choice of the base point $o \in D$ has remained arbitrary. If o is replaced by $g_0 o$, for some $g_0 \in G_{\mathbb{R}}$, V will be replaced by $\mathrm{Ad}\, g_0 V$, $\mathfrak{v}$ by $\mathrm{Ad}\, g_0 \mathfrak{v}$, $\mathfrak{q}$ by $\mathrm{Ad}\, g_0 \mathfrak{q}$, $h(y)$ by $h(y) g_0^{-1}$, $g(y)$ by $g(y) g_0^{-1}$, and $F(y)$ by $\mathrm{Ad}\, g_0(F(y))$. In particular, setting $g_0 = g(\infty)$, one finds

(9.65) **Lemma.** *By choosing the base point $o \in D$ appropriately, it can be arranged that $g(\infty) = 1$ and that $F_0 = N$.*

At this point, the proof of (5.13) is more or less complete. The generators

$$\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

of $\mathfrak{sl}(2, \mathbb{R})$ satisfy the same commutation relations as the triple A_0, E_0, F_0. Hence there exists a unique homomorphism $\psi_*: \mathfrak{sl}(2, \mathbb{R}) \to \mathfrak{g}_0$, which maps these generators, in the given order, to A_0, E_0, and F_0. Clearly ψ_* extends complex linearly to the complexifications, and since $SL(2, \mathbb{C})$ is simply connected, it lifts to a homomorphism

$$\psi: SL(2, \mathbb{C}) \to G_{\mathbb{C}}.$$

As in (5.8), ψ determines the embedding $\tilde{\psi}$. The $G_{\mathbb{R}}$-valued function $g(y)$ has a convergent power series expansion in y^{-1} near $y = \infty$. Replacing y by the complex variable z, one obtains a holomorphic function $g(z)$, defined near $z = \infty$, with values in $\mathrm{Hom}(H_{\mathbb{C}}, H_{\mathbb{C}})$. For $z \in \mathbb{R}$, $g(z)$ lies in the complex submanifold $G_{\mathbb{C}}$ of $\mathrm{Hom}(H_{\mathbb{C}}, H_{\mathbb{C}})$; hence $g(z) \in G_{\mathbb{C}}$, also for complex z. According to (9.54), the definition of $\tilde{\psi}$, and (5.5),

$$\begin{aligned} \exp(iyN) \circ a &= g(y) \exp(-\tfrac{1}{2} \log y\, A_0) \circ o \\ &= g(y) \exp(-\tfrac{1}{2} \log y\, \psi_*(Y)) \circ o = g(y) \circ \tilde{\psi}(iy), \end{aligned}$$

which is the assertion a) for $z \in U \cap i\mathbb{R}$. By the identity theorem for holomorphic functions, a) follows. The statement b) is a consequence of the containment $\psi_*(\mathfrak{sl}(2, \mathbb{R})) \subset \mathfrak{g}_0$. As for c), Lemma (9.8) gives

$$\psi_*(Z) = i(E_0 - F_0) \in \mathfrak{v};$$

$$\psi_*(X_+) = \tfrac{1}{2}(iA_0 + E_0 + F_0) \in \mathfrak{v} \oplus \mathfrak{q},$$

and

$$J\psi_*(X_+) = i\psi_*(X_+),$$

hence $\psi_*(X_+) \in \mathfrak{g}^{-1,1}$ (cf. (9.2)). Similarly, $\psi_*(X_-) \in \mathfrak{g}^{1,-1}$. Next, d) follows directly from the construction of $g(y)$, and e) is asserted by (9.55). The definition of $h(y)$ in f) agrees with (9.54), and so the differential equation is just the one which was used to characterize $h(y)$ in the first place. Lemma (9.60) leads to g), and the first part of h) is embodied in (9.65). I assume then, that $g(\infty) = 1$ and $N = F_0$ (cf. (9.55)). In the notation of the proof of Lemma (9.60), for $l \geqq 1$,

$$(\mathrm{Ad}\, N)^{l+1} B_{l+1} = (\mathrm{Ad}\, F_0)^{l+1} B_{l+1} = \textstyle\sum_{s \leqq l-1} \sum_{r \leqq 2l-s} (\mathrm{Ad}\, F_0)^{l+1} A_{2l-s}^{r,s}.$$

Under the indicated restrictions, $s - 2l - 2 < s - 2l \leqq -r$, so that $\mathfrak{g}(r, s-2l-2) = 0$. On the other hand, $(\mathrm{Ad}\, F_0)^{l+1} A_{2l-s}^{r,s}$ must lie in $\mathfrak{g}(r, s-2l-2)$; hence $(\mathrm{Ad}\, N)^{l+1} B_{l+1} = 0$. Combined with (9.62) and (9.63), this implies the vanishing of $(\mathrm{Ad}\, N)^{k+1} g_k$ and $(\mathrm{Ad}\, N)^{k+1} f_k$, which completes the proof of the theorem.

I now turn to the proof of Lemma (5.25). The strategy will be to investigate closely the functions $u(y)$, $a(y)$, $k(y)$ in (9.16), with an appropriate choice of Iwasawa decomposition.

Since I do not insist on the conclusions of (9.65), the choice of base point can be made as in (5.19). It will be necessary to know that the group K of the Iwasawa decomposition, which enters the statement (9.17), agrees with the choice of K in §5. For this purpose, I consider a subalgebra $\mathfrak{f}_0 \subset \mathfrak{g}_0$, such that $\mathfrak{f}_0$ belongs to a compact subgroup of $G_{\mathbb{R}}$, and such that $\mathfrak{f}_0$ contains $\mathfrak{v}_0$, the Lie algebra of V. Since $\theta = \operatorname{Ad} C$, with $C \in V$ (C is the Weil operator of the reference Hodge structure), $\mathfrak{f}_0$ splits up as

$$\mathfrak{f}_0 = (\mathfrak{f}_0 \cap \mathfrak{k}_0) \oplus (\mathfrak{f}_0 \cap \mathfrak{p}_0)$$

(cf. (8.1)). Any $X \in \mathfrak{f}_0$ has purely imaginary eigenvalues, which makes the trace form (8.7) negative definite on $\mathfrak{f}_0$. On the other hand, the trace form is positive definite on $\mathfrak{p}_0$. I conclude $\mathfrak{f}_0 \subset \mathfrak{k}_0$. As a result, the group K of the Iwasawa decomposition, which was defined in §8, contains the identity component of every compact subgroup of $G_{\mathbb{R}}$ which contains V. In a real, semisimple matrix group (which may have more than a single component in the ordinary topology), a maximal compact subgroup is the normalizer of its own identity component. It follows that K can be described as the unique maximal compact subgroup of $G_{\mathbb{R}}$ which contains V; hence the choices of K in §5 and §8 are consistent.

One of the identities which was used to define ψ_* is $\psi_*(Y) = A_0$. The change of base point, which may have been necessary to accomplish (5.19), does not destroy this identity. Similarly, F_0 is the image under ψ_* of the element (5.18) of $\mathfrak{sl}(2, \mathbb{R})$; combined with (5.19), this gives $F_0 = N$. For emphasis;

(9.66) $$A_0 = \psi_*(Y), \qquad F_0 = N.$$

Let $\mathfrak{a}_0$ be the Lie algebra of the group A in the Iwasawa decomposition. Aside from the datum of K, no special demands were made on the decomposition. Thus $\mathfrak{a}_0$ may be any maximal Abelian subspace of the (-1)-eigenspace of θ on $\mathfrak{g}_0$. According to (5.13c),

$$A_0 = \psi_*(Y) = i\psi_*(X_- - X_+)$$

belongs to this (-1)-eigenspace of θ. Hence, without loss of generality, I may assume that $A_0 \in \mathfrak{a}_0$.

Once A and K have been determined, the choice of U amounts to the choice of a system of positive roots for $(\mathfrak{g}_0, \mathfrak{a}_0)$, such a system can be selected so that it contains every root which assumes a strictly negative value on A_0. In this case, the Lie algebra $\mathfrak{u}_0$ of U has the following property:

(9.67) for every strictly negative number l, $\mathfrak{u}_0$ contains the l-eigenspace of $\operatorname{ad} A_0 : \mathfrak{g}_0 \to \mathfrak{g}_0$.

Again, I can make this hypothesis without loss of generality. I recall the definition of $\mathfrak{c}_0$ in (5.16). According to the representation theory of $\mathfrak{sl}_2$, as it was reviewed in §6, $\mathfrak{c}_0$ lies inside the direct sum of the eigenspaces of $\operatorname{ad} A_0$ corresponding to strictly negative, integral eigenvalues. Hence $\mathfrak{u}_0$ contains $\mathfrak{c}_0$, and because of (5.19), $g(\infty)\in\exp\mathfrak{c}_0\subset U$. Also, the relation $[A_0, F_0] = -2F_0$ gives $N = F_0 \in \mathfrak{u}_0$.

For the statement of the next lemma, one should remember the decomposition (9.16) of the function $h(y)$.

(9.68) **Lemma.** *Suppose that the base point* o *and the Iwasawa decomposition* $G_{\mathbb{R}} = UAK$ *have been chosen subject to the following conditions:* $K\supset V$, $A_0\in\mathfrak{a}_0$, $g(\infty)\in U$, *and* (9.67). *Then* $u(y)$, $k(y)$, $\exp(\frac{1}{2}\log y\, A_0)\, a(y)$ *are regular functions of the variable* $y^{-\frac{1}{2}}$ *near* $y=\infty$, *and* $u(\infty)=g(\infty)$, $k(\infty)=1$, $\lim_{y\to\infty}\exp(\frac{1}{2}\log y\, A_0)\, a(y)=1$.

Proof. I define auxiliary functions

$$u_1(y) = \operatorname{Ad}\exp(\tfrac{1}{2}\log y\, A_0)\big(g(\infty)^{-1}\, u(y)\big),$$

$$a_1(y) = \exp(\tfrac{1}{2}\log y\, A_0)\, a(y).$$

The linear transformation A_0 acts semisimply, with integral eigenvalues. Because of this, and because of (9.22), (9.31), and (9.35), all of the functions $u(y)$, $u_1(y)$, $a(y)$, $a_1(y)$, $k(y)$ have Laurant series expansions in terms of the variable $y^{-\frac{1}{2}}$, which converge and are valid around $y=\infty$. The group A normalizes U, and $A_0\in\mathfrak{a}_0$; hence $u_1(y)$ takes values in U, and $a_1(y)$ in A. From the definition (9.54) of $g(y)$, one obtains

$$\operatorname{Ad}\exp(\tfrac{1}{2}\log y\, A_0)\big(g(\infty)^{-1}\, g(y)\big) = u_1(y)\, a_1(y)\, k(y).$$

In the notation of (9.60), the left hand side can be expressed as

$$1 + \textstyle\sum_{k\geqq 1}\sum_{s\leqq k-1} g_{k,s}\, y^{-(2k-s)/2}$$

which tends to the identity as $y\to\infty$. The Iwasawa decomposition establishes a diffeomorphism between $G_{\mathbb{R}}$ and $U\times A\times K$. Thus each of the factors on the right tends to the identity individually: $k(y)$, $a_1(y)$, $u_1(y)$ are regular functions of $y^{-\frac{1}{2}}$ near $y=\infty$, and $k(\infty)=1$, $a_1(\infty)=1$, $u_1(\infty)=1$.

The statement about $u(y)$ is more difficult; it will come out of a careful consideration of the differential equations in (9.23). I recall that $A(y)=\sum_{n\geqq 0} A_n\, y^{-(n+2)/2}$, and I define

$$L(y) = \operatorname{Ad} k(y)\big(A(y)\big) - A_0\, y^{-1} = \textstyle\sum_{n\geqq 3} L_n\, y^{-n/2},$$

$$k'(y)\, k(y)^{-1} = \textstyle\sum_{n\geqq 3} B_n\, y^{-n/2} \quad \text{with } B_n\in\mathfrak{k}_0.$$

By induction on n, one shows:

(9.69) the coefficient of $y^{-n/2}$ in the series expansion of $k(y)$ (respectively, of $k(y)^{-1}$) is a noncommutative polynomial in the B_k's, homogeneous of weighted degree n when B_{k+2} is assigned the weight k; B_{n+2} occurs with coefficient $-\frac{2}{n}$ $\left(\text{respectively, with coefficient } \frac{2}{n}\right)$.

As a consequence,

(9.70) L_n is a noncommutative polynomial in the B_k's and the A_k's, homogeneous of weighted degree $n-2$ when B_{k+2} and A_k are assigned the weight k; the only term involving B_n and A_0 is $\frac{2}{n-2}[A_0, B_n]$.

I let $A_{n,s}$, $B_{n,s}$, $L_{n,s}$ denote the components of A_n, B_n, L_n in the s-eigenspace of $\operatorname{ad} A_0$ on $\mathfrak{g}_0$. Lemma (9.48) asserts that

(9.71) $$A_{n,s}=0 \quad \text{if } n>0 \quad \text{and} \quad |s|>n-2.$$

The Cartan involution θ is the identity on $\mathfrak{k}_0$, and $\theta A_0 = -A_0$. Hence

(9.72) $$\theta B_{n,s} = B_{n,-s}.$$

The Lie algebra $\mathfrak{u}_0$ contains all eigenspaces of $\operatorname{ad} A_0$ corresponding to strictly negative eigenvalues, and $\theta \mathfrak{u}_0$ contains all eigenspaces corresponding to strictly positive eigenvalues. Also, $\operatorname{ad} A_0$ acts trivially on $\mathfrak{a}_0$. These facts, together with (9.23), give

(9.73) $$L_{n,s} = 2B_{n,s} \quad \text{for } s<0.$$

By induction on n, I shall show that

(9.74) $$B_{n,s}=0 \quad \text{unless} \quad |s| \leqq n-3.$$

For $n<3$, this is vacuously satisfied. Let me assume, then, that (9.74) has been verified for all integers less than a given $n \geqq 3$. According to the induction hypothesis and (9.70),

$$L_{n,s} = \frac{2}{n-2}[A_0, B_{n,s}] = \frac{2s}{n-2} B_{n,s}, \quad \text{provided } s<3-n.$$

If $s<3-n$, s is negative, and (9.73) now implies the relation

$$(n-2)\, B_{n,s} = s\, B_{n,s} \quad \text{if } s<3-n,\ n \geqq 3.$$

This is possible only if $B_{n,s}=0$. By invoking (9.72), I complete the induction step.

In the power series expansion of $k(y)^{-1}$, the coefficient of $y^{-n/2}$ has a zero component in the s-eigenspace of A_0 on $\operatorname{Hom}(H_{\mathbb{C}}, H_{\mathbb{C}})$, unless $n=0$ or $|s| \leqq n-1$, as follows from (9.69) and (9.74). I may conclude that

$$\lim_{y\to\infty} \operatorname{Ad} \exp(-\tfrac{1}{2} \log y\, A_0)(k(y)^{-1}) = 1.$$

Finally, then,

$$u(y)=h(y)\,k(y)^{-1}\,a(y)^{-1}=g(y)\exp(-\tfrac{1}{2}\log y\,A_0)\,k(y)^{-1}\,a(y)^{-1}$$
$$=g(y)\,\mathrm{Ad}\exp(-\tfrac{1}{2}\log y\,A_0)(k(y)^{-1})\,a_1(y)\to g(\infty)$$

as $y\to\infty$, and this proves the lemma.

Let $\mathfrak{r}$ be the subalgebra of $\mathfrak{g}$ which was defined in (5.21), and $\mathfrak{r}_0=\mathfrak{r}\cap\mathfrak{g}_0$. From the construction of $\mathfrak{r}$ and (9.67), it follows that $\mathfrak{r}_0\subset\mathfrak{u}_0$; in fact, $\mathfrak{r}_0$ is an ideal in $\mathfrak{u}_0$. The centralizer of $A_0=\psi_*(Y)$ in $\mathfrak{u}_0$, which will be denoted by $\mathfrak{r}_0'$, is a complementary subalgebra for $\mathfrak{r}_0$ in $\mathfrak{u}_0$, so that

(9.75) $$\mathfrak{u}_0=\mathfrak{r}_0\oplus\mathfrak{r}_0' \quad \text{(semidirect product).}$$

The same argument which gave the containment $\mathfrak{c}_0\subset\mathfrak{u}_0$ also implies $\mathfrak{c}_0\subset\mathfrak{r}_0$. Hence, and because of (9.68),

$$u(\infty)=g(\infty)\in\exp\mathfrak{c}_0\subset\exp\mathfrak{r}_0.$$

Near the identity, the group U is real analytically diffeomorphic to the product of its two subgroups $\exp\mathfrak{r}_0=R\cap G_{\mathbb{R}}$ and $\exp\mathfrak{r}_0'$; the global statement is also true, but not necessary in this context. Hence the U-valued function $u(\infty)^{-1}\,u(y)$ can be expressed as a product

$$u(\infty)^{-1}\,u(y)=r_1(y)\,r_2(y);$$

here $r_1(y)$ takes values in $R\cap G_{\mathbb{R}}$, $r_2(y)$ in $\exp\mathfrak{r}_0'$, both are regular functions of the variables $y^{-\frac{1}{2}}$ near $y=\infty$, and $r_1(\infty)=1$, $r_2(\infty)=1$. As in the proof of (9.68), I set

$$a_1(y)=\exp(\tfrac{1}{2}\log y\,A_0)\,a(y);$$

then $a_1(\infty)=1$. Since $\mathfrak{r}_0'$ centralizes A_0,

$$h(y)=u(\infty)\,r_1(y)\,r_2(y)\,a(y)\,k(y)$$
$$=u(\infty)\,r_1(y)\exp(-\tfrac{1}{2}\log y\,A_0)\,z(y)\,k(y),$$

with $z(y)=r_2(y)\,a_1(y)$. The function $z(y)$ takes values in the centralizer of A_0 in $G_{\mathbb{R}}$, it is a regular function of $y^{-\frac{1}{2}}$ near $y=\infty$, and $z(\infty)=1$.

I shall denote by $Z(A_0)$ the centralizer of A_0 in $G_{\mathbb{C}}$, or, equivalently, the centralizer of the image under ψ of the diagonal subgroup of $SL(2,\mathbb{C})$. Then $Z(A_0)$ is a reductive $\mathbb{Q}$-subgroup of $G_{\mathbb{C}}$; it contains the maximal $\mathbb{Q}$-split torus $T\subset G_{\mathbb{C}}$ which was defined in §5. Consequently, $P\cap Z(A_0)$ must be a minimal $\mathbb{Q}$-parabolic subgroup of $Z(A_0)$. Since $\theta A_0=-A_0$, the Lie algebra of $Z(A_0)$ is θ-stable. It follows that $K\cap Z(A)$ is maximal compact in $Z(A_0)\cap G_{\mathbb{R}}$, certainly if one considers only the connected components of the identity, in the Lie topology. Hence the product

$$(R\cap Z(A_0)_{\mathbb{R}})\,T_{\mathbb{R}}\,M_{\mathbb{R}}(Z(A_0)\cap K)$$

contains at least the identity component of the centralizer of A_0 in $G_{\mathbb{R}}$ (the subscripts $\mathbb{R}$ refer to the groups of real points). Now I can express the function $z(y)$ as

$$z(y)=r_3(y)\,t_1(y)\,m(y)\,k_1(y);$$

the factors are regular functions of the variable $y^{-\frac{1}{2}}$ near $y=\infty$, with values in, respectively, $R\cap Z(A_0)_{\mathbb{R}}$, $T_{\mathbb{R}}$, $M_{\mathbb{R}}$, and $Z(A_0)\cap K$; all assume the identity as value at $y=\infty$. Since $r_3(y)$ commutes with A_0,

$$h(y)=u(\infty)\,r_1(y)\,r_3(y)\exp(-\tfrac{1}{2}\log y\,A_0)\,t_1(y)\,m(y)\,k_1(y)\,k(y).$$

I define $r(y)=u(\infty)\,r_1(y)\,r_3(y)$, $t(y)=\exp(-\frac{1}{2}\log y\,A_0)\,t_1(y)$, and I change notation, replacing $k_1(y)\,k(y)$ by $k(y)$. This proves Lemma (5.25).

References

1. Borel, A.: Introduction aux Groupes Arithmétiques. Paris: Hermann 1969
2. Borel, A., Narasimhan, E.: Uniqueness conditions for certain holomorphic mappings. Inventiones math. **2**, 247–255 (1967)
3. Clemens, H.: Picard-Lefschetz theorem for families of nonsingular algebraic varieties acquiring ordinary singularities. Trans. Amer. Math. Soc. **136**, 93–108 (1969)
4. Coddington, E., Levinson, N.: Theory of Ordinary Differential Equations. New York: McGraw-Hill 1955
5. Deligne, P.: Travaux de Griffiths. Séminaire Bourbaki Exp. 376
6. Deligne, P.: Théorie de Hodge II. Publ. Math. I.H.E.S. **40**, 5–57 (1972)
7. Deligne, P.: Théorie de Hodge III. To appear in Publ. Math. I.H.E.S.
8. Deligne, P.: Equations différentielles à points singuliers réguliers. Lecture Notes in Mathematics **163**. Berlin-Heidelberg-New York: Springer 1970
9. Grauert, H.: Ein Theorem der analytischen Garbentheorie und die Modulräume komplexe Strukturen. Publ. Math. I.H.E.S., 5 (1960)
10. Griffiths, P.: Periods of integrals on algebraic manifolds I, II. Amer. J. Math. **90**, 568–626, 805–865 (1968)
11. Griffiths, P.: Periods of integrals on algebraic manifolds III. Publ. Math. I.H.E.S. **38**, 125–180 (1970)
12. Griffiths, P.: Periods of integrals on algebraic manifolds: summary of main results and discussion of open problems. Bull. Amer. Math. Soc. **76**, 228–296 (1970)
13. Griffiths, P., Schmid, W.: Locally homogeneous complex manifolds. Acta Math. **123**, 253–302 (1969)
14. Griffiths, P., Schmid, W.: Recent developments in Hodge theory: a discussion of techniques and results. To appear in the Proceedings of the International Colloquium on Discrete Subgroups of Lie Groups. Bombay, 1973
15. Grothendieck, A.: Un théorème sur les homomorphismes des schémas abéliens. Inventiones math. **2**, 59–78 (1966)
16. Helgason, S.: Differential Geometry and Symmetric Spaces. New York: Academic Press 1962
17. Katz, N.: Nilpotent connections and the monodromy theorem; applications of a result of Turritin. Publ. Math. I.H.E.S. **39**, 175–232 (1971)
18. Kobayashi, S.: Hyperbolic Manifolds and Holomorphic Mappings. New York: Marcel Dekker 1970

19. Kodaira, K., Spencer, D.C.: On deformations of complex analytic structures III. Ann. of Math. **71**, 43–76 (1960)
20. Kostant, B.: The principal three-dimensional subgroup and the Betti numbers of a complex simple Lie group. Amer. J. Math. **81**, 973–1032 (1969)
21. Landman, A.: On the Picard-Lefschetz transformations. To appear in Trans. Amer. Math. Soc.
22. Lelong, P.: Fonctions Plurisousharmoniques et Formes différentielles positives. Paris-London-New York: Gordon and Breach 1968
23. Serre, J.P.: Géométrie algébrique et géométrique analytique. Ann. Inst. Fourier, Grenoble **6**, 1–42 (1955)
24. Serre, J.-P.: Algèbres de Lie Semi-simples Complexes. New York: Benjamin 1966
25. Weil, A.: Variétés Kählériennes. Publ. Inst. Math. Univ. Nancago, VI, Paris: Hermann 1958
26. Wu, H.: Normal families of holomorphic mappings. Acta Math. **119**, 193–233 (1967)

Wilfried Schmid
Mathematisches Institut der Universität
D-5300 Bonn
Federal Republic of Germany

(Received July 3, 1973)