4. The Size of the Formal Solution Space

Knowing what is big and what is small is more important than being able to solve partial differential equations.

Stan Ulam

The results in this chapter should be used with some care. We introduce measures for the size of the formal solution space of a differential equation. First of all, one cannot stress strongly enough that this only concerns *formal* solutions, i. e. formal power series satisfying the given equation. In contrast to Chapter 5, we do not bother here about the convergence of these series and thus the results do not imply any statement on the existence of strong solutions. So one might say that we perform in this chapter only combinatorial games. Nevertheless, there are situations where the results can be quite useful. This should be evident from the fact that scientists like Cartan and Einstein actively participated in these "games".

In the first section we briefly discuss the concept of a "general" solution. This word is often used but hardly anybody cares to give a rigorous definition. We do not give one either, as this turns out to be rather difficult and of doubtful value. So we only discuss some of the arising problems in order to motivate why we afterwards restrict to the formal solution space where the situation is much simpler.

Section 4.2 introduces the classical tools of the formal theory for measuring the size of the formal solution space: the Cartan characters and the Hilbert function. We exhibit the connection between them and show how they can be explicitly determined for an involutive equation. In applications to physics, one is interested in subtracting the effect of gauge symmetries; this is easily done using a pseudogroup approach. In the following section we analyse how the Cartan characters and the Hilbert function change under certain operations on the differential equation. This leads to the rather general concept of a differential relation.

The final section of this chapter is more of a "historical" nature. We show how Einstein's approach based on the so-called strength of a differential equation fits into our framework. It turns out that the machinery of formal theory is more powerful and flexible than this approach, as the strength corresponds to just one of the Cartan characters.

4.1 General Solutions

The notion of a "general solution" or a "general integral" is notoriously difficult. Although it is intuitively clear what it is supposed to mean, it is rather difficult to give a rigorous definition. From the point of view of the functional analytic approach that has dominated the theory of partial differential equations for many decades, this is not so surprising as obviously the answer decisively depends on the considered function spaces. The classical idea of a parameter dependent expression which yields all solutions through suitable specialisations of the parameters becomes rather meaningless, if one is interested in solutions living in distribution spaces.

Example 4.1.1. If one admits distributions as solutions, already for ordinary differential equations the classical counting rules fail. As a trivial example we may consider the following first-order equation $xu' = 0$. The classical point of view is that the solution space of a scalar linear first-order equation is a one-dimensional vector space. In our case a basis of this space is obviously given by $u(x) = 1$. However, because of the singularity of the coefficient at $x = 0$, the equation admits further solutions in a distributional sense, for example $u(x) = H(x)$ where $H(x)$ denotes the Heaviside or step function defined by $H(x) = 0$ for $x < 0$ and $H(x) = 1$ for $x \geq 0$ [149]. Thus we have an at least two-dimensional solution space! $\lhd$

Another problem with the concept of a general solution is the existence of so-called *singular solutions*. They appear only in non-linear equations. The word "singular" does not refer to singularities in the solution but to the fact that such solutions cannot be obtained from a general solution; typically they arise as envelopes of families of solutions.

Example 4.1.2. Again it suffices to consider ordinary differential equations. The general solution of the following first-order system in two unknown functions u and v [53, p. 159]

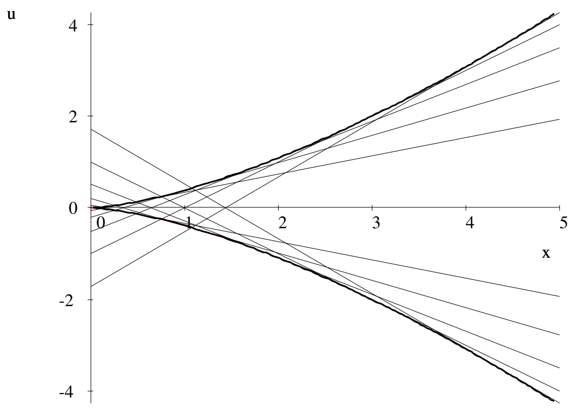

Fig. 4.1. Singular and general solution for (4.2)

$$(u')^2 + xu' + v' - u = 0 \; , \tag{4.1a}$$

$$u'v' + xv' - v = 0 \tag{4.1b}$$

is the two-parameter family $u(x) = \alpha x + \alpha^2 + \beta$ and $v(x) = \beta x + \alpha\beta$. However, there exists a further one-parameter family of solutions, namely $u(x) = -\frac{1}{4}(x + c)^2 + c^2$ and $v(x) = -\frac{1}{4}c(x - c)^2$. Obviously, it is impossible to choose the parameters α, β in such a way that these solutions are part of the general solution.

A similar example for a single equation is [52, p. 242]

$$(u')^3 - xu' + u = 0 \; . \tag{4.2}$$

Its general solution is the one-parameter family $u(x) = ax - a^3$. The two singular solutions are given by the quadratic equation $u^2 = \frac{4}{27}x^3$; they envelope the general solution (Figure 4.1). $\lhd$

In both examples above the origin of the problem is that the rank of the symbol matrix is not constant. In Example 4.1.1 the rank drops at $x = 0$, i. e. at a certain point of the base manifold $\mathcal{X}$. While this does not affect the classical solution space, it allows for additional distributional solutions. In Example 4.1.2, the determinant of the symbol matrix of the system (4.1) is $D = 2(u')^2 + 3xu' - v' + x^2$. Thus the rank of the symbol matrix drops whenever $D = 0$, i. e. on a submanifold of $J_1\mathcal{E}$, as this represents again a differential equation. The "general solution" consists of all those solutions for which $D \neq 0$. The "singular solutions" are the remaining ones; they arise as solutions of the overdetermined system obtained by adding $D = 0$ to (4.1).

Within our geometric approach to differential equations one may use a slightly different ansatz that avoids some of these problems, at least for smooth solutions. The downside is that we can only define *a* and not *the* general solution. Furthermore, the definition is not very constructive. The basic idea is very simple and rather old. Consider the inverse problem: we are given an expression containing some arbitrary constants and functions and we wish to determine differential equations for which this expression is a solution. For the "smallest" such differential equation our expression is a general solution. Following Hermann [76] we can make this idea more precise in our geometric framework.

Definition 4.1.3. *A* general solution *of the differential equation $\mathcal{R}_q \subseteq J_q\mathcal{E}$ is a subset $\Theta \subseteq \Gamma_{\mathrm{loc}}(\mathcal{E})$ such that the following two conditions are satisfied.*

(i) *For all $r \geq 0$ we have $j_{q+r}(\Theta) \subseteq \mathcal{R}_{q+r}$.*
(ii) *For no $r \geq 0$ a submanifold $\bar{\mathcal{R}}_{q+r} \subset \mathcal{R}_{q+r}$ of lower dimension exists such that $j_{q+r}(\Theta) \subseteq \bar{\mathcal{R}}_{q+r}$.*

Θ is a local general solution *at a point $x \in \mathcal{X}$, if all sections $\sigma \in \Theta$ are defined in neighbourhood of x.*

A natural question is how one may describe such subsets $\Theta \subseteq \Gamma_{\mathrm{loc}}(\mathcal{E})$. The intuitive idea is that it stems from an expression parametrised by some arbitrary constants and functions. Again Hermann [76] showed how this may be formulated in a geometric manner. Let $\mathcal{Y}_\lambda$ be manifolds and and $\mathcal{V}_\lambda$ vector spaces for $1 \leq \lambda \leq \ell$ and let $\mathcal{W}$ be a further manifolds. All manifolds and spaces are assumed to be finite-dimensional. Then a (local) *closed form representation* of a (local) general solution Θ is a smooth map

$$s : C^\infty(\mathcal{Y}_1, \mathcal{V}_1) \times \cdots \times C^\infty(\mathcal{Y}_\ell, \mathcal{V}_\ell) \times \mathcal{W} \rightarrow \Gamma_{\mathrm{loc}}(\mathcal{E}) \tag{4.3}$$

such that $\operatorname{im} s = \Theta$. $\dim \mathcal{W}$ corresponds to the number of arbitrary constants in the general solution. Each factor $C^\infty(\mathcal{Y}_\lambda, \mathcal{V}_\lambda)$ represents $\dim \mathcal{V}_\lambda$ functions depending on $\dim \mathcal{Y}_\lambda$ arguments.

Example 4.1.4. We apply these ideas to the wave equation $u_{tt} - u_{xx} = 0$. As it is a linear equation, we do not have to bother about singular solutions. Furthermore, it is simple to derive closed form expressions for its general solution. Using the method of characteristics we obtain the classical superposition formula

$$s_1 : \begin{cases} C^\infty(\mathbb{R}, \mathbb{R}^2) & \longrightarrow & \Gamma_{\mathrm{loc}}(\mathbb{R}^2 \times \mathbb{R}) \\ \binom{f}{g} & \longmapsto & f(x-t) + g(x+t) \end{cases} . \tag{4.4}$$

Alternatively, we take the D'Alembert solution of the initial value problem

$$s_2 : \begin{cases} C^\infty(\mathbb{R}, \mathbb{R}^2) & \longrightarrow & \Gamma_{loc}(\mathbb{R}^2 \times \mathbb{R}) \\ \binom{f}{g} & \longmapsto & \frac{1}{2}\big[f(x-t) + f(x+t)\big] + \frac{1}{2}\int_{x-t}^{x+t} g(\tau)d\tau \end{cases} . \tag{4.5}$$

In both cases, the right hand side is to be interpreted as a section of the trivial fibre bundle $\pi : \mathbb{R}^2 \times \mathbb{R} \to \mathbb{R}^2$. ◁

4.2 Cartan Characters and Hilbert Function

We are interested in measuring the size of the formal solution space. In particular, we would like to be able to compare two differential equations. A simple possibility would be to count the parametric derivatives. However, for most partial differential equations the result would be infinity which does not help us much in comparing different equations. Intuitively it is clear that although most partial differential equations have an infinite-dimensional solution space, there are differences in the size. Let us for example assume that we could find a closed form representation like (4.3). Then for some equations more arbitrary functions or manifolds $\mathcal{Y}_\lambda$ of higher dimensions are needed than for others.

The basic trick to capture these ideas in a rigorous manner consists of exploiting the natural grading of derivatives. Instead of simply counting all parametric derivatives, we count them order by order. Obviously, this yields at each order a finite number. Just comparing these numbers for one fixed order does not give us much information, but studying the asymptotic growth turns out to be very useful.

In Section 1.5 we introduced the indices $\beta_q^{(k)}$ of the symbol $\mathcal{M}_q$ of a differential equation $\mathcal{R}_q$. If the equation $\mathcal{R}_q$ is involutive, $\beta_q^{(k)}$ gives the number of principal derivatives of class k and order q. Now we introduce similar numbers $\alpha_q^{(k)}$ for counting the remaining derivatives of class k and order q which are of course parametric.

Definition 4.2.1. *Let $\mathcal{R}_q$ be an involutive differential equation in n independent and m dependent variables with indices $\beta_q^{(k)}$. The* Cartan characters *$\alpha_q^{(k)}$ of $\mathcal{R}_q$ are*

$$\alpha_q^{(k)} = m\binom{q+n-k-1}{q-1} - \beta_q^{(k)}, \qquad 1 \le k \le n . \tag{4.6}$$

By (A.4b), the binomial coefficient on the right hand side of (4.6) counts the total number of derivatives of order q and class k. The Cartan characters of an involutive equation form a descending sequence. Recall that the corresponding property for the indices, Corollary 3.1.8, holds only for first-order equations.

Proposition 4.2.2. *The Cartan characters $\alpha_q^{(k)}$ of an involutive differential equation $\mathcal{R}_q$ in m dependent variables satisfy*

$$\alpha_q^{(1)} \geq \alpha_q^{(2)} \geq \cdots \geq \alpha_q^{(n)} \geq 0 \,. \tag{4.7}$$

Proof. For $q = 1$ this follows immediately from Corollary 3.1.8, as in this case $\alpha_1^{(k)} = m - \beta_1^{(k)}$. For a higher-order equation we exploit the fact (shown in Appendix A.3), that it can be transformed into an equivalent involutive first-order equation with exactly the same Cartan characters. □

One of the key properties of an involutive symbol is that it is possible to compute the indices and the Cartan characters of its prolongations without explicitly determining their equations. This is a simple consequence of our combinatorial interpretation of involutive bases.

Proposition 4.2.3. *Let $\mathcal{M}_q$ be an involutive symbol with Cartan characters $\alpha_q^{(k)}$ and indices $\beta_q^{(k)}$. Then the Cartan characters and indices of its prolongations $\mathcal{M}_{q+r}$ are given by*

$$\alpha_{q+r}^{(k)} = \sum_{i=k}^{n} \binom{r+i-k-1}{r-1} \alpha_q^{(i)} \,, \tag{4.8a}$$

$$\beta_{q+r}^{(k)} = \sum_{i=k}^{n} \binom{r+i-k-1}{r-1} \beta_q^{(i)} \,. \tag{4.8b}$$

Proof. It suffices to prove (4.8b), as (4.8a) is then an immediate consequence of (4.6). We use an induction over r. Prolonging an equation of class k or higher with respect to x^k yields an equation of class k. By Proposition 3.1.4 these are the only independent equations of order $q+1$ in the prolongation, as $\mathcal{M}_q$ is involutive. Thus $\beta_{q+1}^{(k)} = \beta_q^{(k)} + \cdots + \beta_q^{(n)}$ and we have proven (4.8b) for $r = 1$.

Let us assume that (4.8b) holds for $r-1$. Since according to Proposition 3.1.10 the prolongation of an involutive symbol is again involutive, we can apply the formula for $r = 1$ to get $\beta_{q+r}^{(k)} = \beta_{q+r-1}^{(k)} + \cdots + \beta_{q+r-1}^{(n)}$. The inductive hypothesis yields

$$\beta_{q+r}^{(k)} = \sum_{i=k}^{n} \sum_{l=k}^{i} \binom{r+l-k-2}{r-2} \beta_q^{(i)} = \sum_{i=k}^{n} \binom{r+i-k-1}{r-1} \beta_q^{(i)} \,. \tag{4.9}$$

The last equality follows from an identity for binomial coefficients. □

Definition 4.2.4. *The* Hilbert function $h(r)$ *of a differential equation* $\mathcal{R}_q$ *gives the number of parametric derivatives of order* r.

The term Hilbert function stems from commutative algebra where such a function is used to measure the size of graded modules (see Appendix B.1); in differential algebra sometimes the term *differential dimension polynomial* is preferred [94]. Here it measures the size of the formal solution space which is parametrised by the parametric coefficients, as we have seen in Section 1.4. For an *involutive* differential equation $\mathcal{R}_q$ it is straightforward to determine the Hilbert function. We have

$$h(r) = \begin{cases} \dim \mathcal{R}_0^{(q)} & \text{for } r = 0 \ , \\ \dim \mathcal{R}_r^{(q-r)} - \dim \mathcal{R}_{r-1}^{(q-r+1)} & \text{for } 0 < r \leq q \ , \\ \dim \mathcal{R}_r - \dim \mathcal{R}_{r-1} & \text{for } r > q \ . \end{cases} \tag{4.10}$$

However, this form is not very satisfying. In fact, as in commutative algebra, the Hilbert function $h(r)$ becomes a polynomial for sufficiently large values of r. More precisely, it is polynomial at least for $r \geq q$ and it is possible to derive a closed form expression for this polynomial.

Proposition 4.2.5. *The Hilbert polynomial* $H(r)$ *of the involutive differential equation* $\mathcal{R}_q$ *in* n *independent variables with Cartan characters* $\alpha_q^{(k)}$ *can be written in the form*

$$H(q + r) = \sum_{i=0}^{n-1} \left(\sum_{k=i}^{n-1} \frac{\alpha_q^{(k+1)}}{k!} s_{k-i}^{(k)}(0) \right) r^i \tag{4.11}$$

where the coefficients $s_k^{(i)}(0)$ *are the modified Stirling numbers introduced in Appendix A.4.*

Conversely, if the Hilbert polynomial of the differential equation $\mathcal{R}_q$ *has the form* $H(q + r) = \sum_{i=0}^{n-1} h_i r^i$, *then its Cartan characters can be computed recursively starting with* $k = n$ *by*

$$\alpha_q^{(k)} = (k-1)!\, h_{k-1} - \sum_{i=k+1}^{n} \frac{(k-1)!}{(i-1)!} s_{i-k}^{(i-1)}(0) \alpha_q^{(i)} \ . \tag{4.12}$$

Proof. By Theorem 1.5.4, $H(r) = \dim \mathcal{M}_r$ for $r \geq q$. It follows trivially from the definition of the indices and the Cartan characters, respectively, that $\dim \mathcal{M}_{q+r} = \sum_{i=1}^{n} \alpha_{q+r}^{(i)}$. Together with (4.8a) this yields

$$\begin{aligned} H(q + r) &= \sum_{i=1}^{n} \sum_{k=i}^{n} \binom{r + k - i - 1}{r - 1} \alpha_q^{(k)} \\ &= \sum_{k=1}^{n} \sum_{i=1}^{k} \binom{r + k - i - 1}{r - 1} \alpha_q^{(k)} = \sum_{k=1}^{n} \binom{r + k - 1}{r} \alpha_q^{(k)} \ . \end{aligned} \tag{4.13}$$

Here we used in the last step the identity (A.36). Now we can enter the defining relation (A.31) of the modified Stirling numbers and obtain after a simple index shift (4.11). It defines obviously a one-to-one correspondence between the coefficients h_i of the Hilbert polynomial and the Cartan characters $\alpha_q^{(k)}$. A straightforward inversion yields (4.12). □

Of course, this result represents nothing but a reformulation of Proposition 2.5.2 for the special case of a Rees decomposition.

The Cartan characters (or the Hilbert function) measure the size of the formal solution space in a rather abstract manner. But in some sense they are the best measures available, as they do not require any assumptions about the form of the general solution. The main caveat is that the Cartan characters are local quantities defined only at a point $\rho \in \mathcal{R}_q$. In particular, if singular solutions exist, the Cartan characters will vary over $\mathcal{R}_q$. As the question of global solutions is beyond the scope of this book anyway, we can live with this restriction.

For the remainder of this section we want to relate the Cartan characters to numbers that can be more easily interpreted, namely to the number of arbitrary functions appearing in a closed form solution like (4.3). However, such an interpretation can be given only under some restrictive assumptions that are often *not* satisfied in applications. The main assumption is that the map s is purely algebraic, i. e. does not involve any differentiations or integrations. Thus we exclude cases as the D'Alembert formula (4.5). More precisely, we require that our solutions can be written in the form

$$\mathbf{u}(\mathbf{x}) = \boldsymbol{\Psi}\Big(\dots, F_\lambda\big(\phi_1(\mathbf{x}), \dots, \phi_k(\mathbf{x})\big), \dots\Big) . \tag{4.14}$$

Here F_λ represents the arbitrary functions. The one shown depends on k arguments which are defined by the fixed functions ϕ_i (these functions define the manifold $\mathcal{Y}_\lambda$). Note that no explicit dependency on the independent variables $\mathbf{x}$ is permitted; they appear only via the functions ϕ_i.

Example 4.2.6. The nonlinear equation $uu_{xy} - u_x u_y = 0$ admits such an algebraic representation of its general solution. It contains two arbitrary functions F_1 and F_2 each depending on one argument and is given by $u(x,y) = \Psi\big(F_1(\phi_1(x,y)), F_2(\phi_2(x,y))\big)$ with $\Psi(a,b) = ab$, $\phi_1(x,y) = x$ and $\phi_2(x,y) = y$. Thus our differential equation describes all functions of x and y that can be separated multiplicatively: $u(x,y) = F_1(x)F_2(y)$. ◁

Our goal is to determine the number f_k of arbitrary functions in (4.14) that depend on k arguments for $1 \leq k \leq n$. It turns out that if such a representation exists, these numbers are unique. Following [114], we introduce the following short hand

$$\begin{bmatrix} k \\ r \end{bmatrix} = \binom{k+r-1}{k-1} . \tag{4.15}$$

Then a smooth function of k arguments possesses $\left[\begin{smallmatrix} k \\ r \end{smallmatrix}\right]$ Taylor coefficients of order r (cf. (1.24)).

We expand (4.14) in a Taylor series. In order to avoid degenerate cases, we choose the expansion point $x_0 \in \mathcal{X}$ such that if $\partial\phi_j/\partial x^i \not\equiv 0$ for some $1 \leq i \leq n$, then $\big(\partial\phi_j/\partial x^i\big)(x_0) \neq 0$. The right hand side of (4.14) depends thus on

$$T(r) = \sum_{k=1}^{n} f_k \begin{bmatrix} k \\ r \end{bmatrix} \tag{4.16}$$

arbitrary Taylor coefficients of order r. If (4.14) is indeed the general solution of the differential equation $\mathcal{R}_q$, these coefficients must be in a one-to-one correspondence with the parametric derivatives of $\mathcal{R}_q$. At least for orders $r \geq q$, it is straightforward to give a necessary condition for the existence of such a correspondence. We write both the Hilbert polynomial $H(r)$ of $\mathcal{R}_q$ and $T(r)$ as explicit polynomials in r and equate the coefficients. This yields a triangular system for the numbers f_k which is easily solved by back-substitution. We phrase the result in form of a proposition.

Proposition 4.2.7. *If the involutive differential equation $\mathcal{R}_q$ with Cartan characters $\alpha_q^{(k)}$ possesses an algebraic representation of its general solution in the form* (4.14)*, then this representation contains f_k arbitrary functions depending on k arguments where the numbers f_k are determined by the recursion relation*

$$f_n = \alpha_q^{(n)} , \tag{4.17a}$$

$$f_k = \alpha_q^{(k)} + \sum_{i=k+1}^{n} \frac{(k-1)!}{(n-1)!} \left(s_{i-k}^{(i-1)}(0)\alpha_q^{(i)} - s_{i-k}^{(i-1)}(q) f_i \right) . \tag{4.17b}$$

Such an algebraic representation can exist only, if the solution of this recursion relation contains only non-negative integers.[1]

Now we see why Definition 3.3.6 of an underdetermined equation makes sense. By Proposition 3.3.7, we get for the highest index of an underdetermined equation $\beta_q^{(n)} < m$ and thus $\alpha_q^{(n)} > 0$. Hence any algebraic representation of the general solution of such an equation depends on at least

[1] One can show by explicit computations that even for fairly large values of n, the recursion relation (4.17) always yields integer solutions. Thus it is tempting to conjecture that this holds for all n. However, one can also show that this fact is highly non-trivial, as the modified Stirling number $s_{i-k}^{(i-1)}(q)$ is in general not divisible by the denominator $(n-1)!$.

one arbitrary function of n arguments. Obviously, we may simply choose $\mathbf{x}$ for these arguments and at least one component of $\mathbf{u}$ remains completely unrestricted by the differential equation.

Example 4.2.8. We consider again Maxwell's equations (1.64). We know from Example 1.6.1 that they are formally integrable and from Example 3.1.9 that their symbol is involutive. We determined already that $\beta_1^{(4)} = 6$, $\beta_1^{(3)} = 2$ and all other indices vanish. This yields the Cartan characters

$$\alpha_1^{(4)} = 0 \,, \quad \alpha_1^{(3)} = 4 \,, \quad \alpha_1^{(2)} = \alpha_1^{(1)} = 6 \,. \tag{4.18}$$

The Hilbert polynomial is quadratic, $H(1+r) = 2r^2 + 12r + 16$, and for the numbers of arbitrary functions in an algebraic representation of the general solution we compute from (4.17)

$$f_4 = 0 \,, \quad f_3 = 4 \,, \quad f_2 = 2 \,, \quad f_1 = 0 \,. \tag{4.19}$$

In this particular case, these numbers are easy to interpret. Maxwell's equations are obviously not underdetermined. For each of the x- and y-components of the fields we have one evolution equation; thus we may prescribe in an initial value problem arbitrary values on a three-dimensional hypersurface. For each of the z-components we have two equations; as we will discuss in more detail in Chapter 5, this implies that, at least formally, we may prescribe initial data for them only on a two-dimensional surface. ◁

For first-order equations the situation becomes very simple and we can solve (4.17) in closed form. We can even show that the solution consists always of non-negative integers.

Corollary 4.2.9. *For an involutive first-order equation $\mathcal{R}_1$ the solution of* (4.17) *is given by*

$$f_n = \alpha_1^{(n)} = m - \beta_1^{(n)} \,, \tag{4.20a}$$

$$f_k = \alpha_1^{(k)} - \alpha_1^{(k+1)} = \beta_1^{(k+1)} - \beta_1^{(k)} \,, \quad 1 \le k < n \,. \tag{4.20b}$$

It contains only non-negative integers.

Proof. The proof is by a descending induction. The case $k = n$ is obvious. If we assume that (4.20b) holds for $k+1$, we get for k that

$$\begin{aligned} f_k = \alpha_1^{(k)} - \alpha_1^{(k+1)} &+ \\ (k-1)! \sum_{i=k+2}^{n-1} i!\, \alpha_1^{(i+1)} &\Bigl(i s_{i-k}^{(i-1)}(1) + s_{i-k+1}^{(i)}(0) - s_{i-k+1}^{(i)}(1) \Bigr) \,. \end{aligned} \tag{4.21}$$

By Lemma A.4.3 the expression in the bracket vanishes for all values of i and k; this yields (4.20b) for k. It is obvious that all f_k are integers and Proposition 4.2.2 implies that they are non-negative. □

One might be tempted to think that the general solutions of most partial differential equations should possess such algebraic representation. In fact, for higher-order equations it is a rather rare phenomenon. In particular, in the important case of normal equations no algebraic representations exist.

Example 4.2.10. Consider a normal differential equation $\mathcal{R}_q$ of order $q > 1$ in n independent and m dependent variables. By Proposition 3.3.7 the index $\beta_q^{(n-1)} = m$ and all other indices vanish. Thus its Cartan characters are

$$\alpha_q^{(n)} = 0 , \qquad \alpha_q^{(k)} = m\binom{q+k-1}{q-1} , \quad 1 \leq k < n . \tag{4.22}$$

Evaluating (4.17) yields $f_n = 0$, $f_{n-1} = mq$ and $f_{n-2} = mq(1-q)/2$. Thus for $q > 1$ and $n > 2$ we always get a negative value for f_{n-2}. $\lhd$

This simple example shows that the notion of an algebraic representation is only of limited use. One can consider more general representations involving differentiations and integrations of the arbitrary functions [160], but one obtains in general no unique values for the numbers f_k. Furthermore, a number of technical assumptions are required. So the value of such a formal counting of arbitrary functions is somewhat dubious. We have presented it here only because it helps to get a better understanding of how the Cartan characters and the Hilbert polynomial measure the size of the formal solution space. As already mentioned above, these abstract tools do not depend on any representation of the general solution and thus can always be applied without problems.

It is no coincidence that for first-order equation one always obtains reasonable values for the numbers f_k. As we will see in Chapter 5, these numbers are related to a concrete initial value problem. Solving this problem with a power series ansatz yields then indeed at each order a one-to-one correspondence between the Taylor coefficients of the initial data and the parametric coefficients of the solution.

For higher-order equations the situation becomes more difficult, as in general it is no longer so easy to relate the orders of the coefficients of the data and of the solution.[2] Only one thing is clear: the asymptotic growth of the number of parametric coefficients is dominated by the arbitrary functions with the highest number of arguments. The number k_0 of their arguments determines the degree of the Hilbert polynomial, $\deg H = k_0 - 1$ and is equal to the highest index k_0 such that $\alpha_q^{(k_0)} > 0$; it is sometimes called the *Cartan genus* of the differential equation and corresponds to the (Krull) dimension of an ideal in algebraic geometry.

[2] Of course, a simple solution is to rewrite the equation as an equivalent first-order equation as described in Appendix A.3 and then to analyse this equation.

The value $\alpha_q^{(k_0)}$ is known as *index of generality* [34, pp. 42–43]. It is not difficult to see that any representation of the general solution depends on precisely $\alpha_q^{(k_0)}$ arbitrary functions of k_0 arguments. In field theories $\alpha_q^{(k_0)}$ gives the number of degrees of freedom [164]; in algebraic geometry it corresponds to the multiplicity of an ideal.

4.3 Differential Relations and Gauge Symmetries

One can act with various kinds of transformations on a differential equation. In the simplest case, the transformation is just a change of variables in a local representation of the equation. Obviously, this does not affect the Hilbert function or the Cartan characters. However, more "brutal" transformations may well change these measures for the size of the formal solution space. A simple example is the reduction to a first-order equation as discussed in Appendix A.3; depending on the way the reduction is performed, the Cartan characters may change.

In this section we provide a rather general framework for such transformations: differential relations. These are slightly more complicated objects than differential equations; in a certain sense they contain two equations in one. Using this concept we can derive an equation for the Hilbert functions of differentially related differential equations.

Definition 4.3.1. *Let $\pi : \mathcal{E} \to \mathcal{X}$ and $\bar{\pi} : \bar{\mathcal{E}} \to \mathcal{X}$ be two fibred manifolds over the same base space $\mathcal{X}$. A* differential relation *of order $r, \bar{r}$ is a submanifold $\mathcal{T}_{r,\bar{r}} \subseteq J_r\mathcal{E} \times J_{\bar{r}}\bar{\mathcal{E}}$ such that $\mathcal{T}_{r,\bar{r}}$ is a fibred submanifold with respect to the two fibrations induced by π and $\bar{\pi}$, namely $\pi^r \times \mathrm{id} : J_r\mathcal{E} \times J_{\bar{r}}\bar{\mathcal{E}} \to \mathcal{X} \times J_{\bar{r}}\bar{\mathcal{E}}$ and $\mathrm{id} \times \bar{\pi}^{\bar{r}} : J_r\mathcal{E} \times J_{\bar{r}}\bar{\mathcal{E}} \to J_r\mathcal{E} \times \mathcal{X}$.*

The defining properties of a differential relation allow us to consider for a given section $\sigma : \mathcal{X} \to \mathcal{E}$ the submanifold $\mathcal{T}_{\bar{r}}[\sigma] = \mathcal{T}_{r,\bar{r}} \cap \big(\mathrm{im}\, j_r\sigma \times J_{\bar{r}}\bar{\mathcal{E}}\big)$ as a differential equation in $J_{\bar{r}}\bar{\mathcal{E}}$ and similarly for any section $\bar{\sigma} : \mathcal{X} \to \bar{\mathcal{E}}$ the submanifold $\mathcal{T}_r[\bar{\sigma}] = \mathcal{T}_{r,\bar{r}} \cap \big(J_r\mathcal{E} \times \mathrm{im}\, j_{\bar{r}}\bar{\sigma}\big)$ as a differential equation in $J_r\mathcal{E}$.

Definition 4.3.2. *Let $\mathcal{R}_q \subseteq J_q\mathcal{E}$ and $\bar{\mathcal{R}}_{\bar{q}} \subseteq J_{\bar{q}}\bar{\mathcal{E}}$ be two differential equations. They are* differentially related *via $\mathcal{T}_{r,\bar{r}} \subseteq J_r\mathcal{E} \times J_{\bar{r}}\bar{\mathcal{E}}$, if the following condition is satisfied: if the section σ is a solution of $\mathcal{R}_q$, then every solution of $\mathcal{T}_{\bar{r}}[\sigma]$ solves $\bar{\mathcal{R}}_q$, and conversely if $\bar{\sigma}$ is a solution of $\bar{\mathcal{R}}_q$, then every solution of $\mathcal{T}_r[\bar{\sigma}]$ solves $\mathcal{R}_q$.*

Example 4.3.3. A classical instance of a differential relation are *Bäcklund transformations* [137, 153]. Two additional conditions are imposed on them: (i) $\mathcal{T}_s[\bar{\sigma}]$ and $\mathcal{T}_{\bar{s}}[\sigma]$ are solvable, if and only if σ and $\bar{\sigma}$ are solutions of $\mathcal{R}_q$ and

$\bar{\mathcal{R}}_{\bar{q}}$, respectively; (ii) both $\mathcal{T}_s[\bar{\sigma}]$ and $\mathcal{T}_{\bar{s}}[\sigma]$ are equations of finite type. These two conditions allow for a geometric interpretation of Bäcklund transformations as so-called zero-curvature representations based on Remark 1.3.2; we do not pursue here this point of view but refer to Brandt [24].

Bäcklund [11] introduced these transformations for the sine-Gordon equation $u_{xt} = \sin u$. Here we even find an auto-Bäcklund transformation mapping the equation into itself. It has the following form:

$$\left(\frac{u+v}{2}\right)_x = a \sin\left(\frac{u-v}{2}\right) , \qquad \left(\frac{u-v}{2}\right)_t = \frac{1}{a} \sin\left(\frac{u+v}{2}\right) \tag{4.23}$$

where $a \neq 0$ is an arbitrary real parameter. Obviously, given v this is an equation of finite type for u and vice versa. In order to check whether these finite type equations are formally integrable, we differentiate the first equation with respect to t and the second one with respect to x and enter (4.23) into the result. With the help of some elementary trigonometric identities we find that it is necessary and sufficient for the formal integrability that both u and v satisfy the sine-Gordon equation.

Bäcklund transformations play an important role in the theory of soliton equations. In particular, they are related to the inverse scattering transform and allow for the iterative construction of N-soliton solutions [1, Section 3.1]. For example, we consider the transformation (4.23) for $u(x,t) \equiv 0$ which leads to $v_x = -2a\sin(v/2)$ and $v_t = -(2/a)\sin(v/2)$. Combining these two equations yields the simple linear equation $v_x - a^2 v_t = 0$ whose general solution is trivially given by $v(x,t) = F(t + a^2 x)$ with a yet arbitrary function F. Entering it into the first equation of our system shows that F must satisfy the ordinary differential equation $aF' = -2\sin(F/2)$. This is a separable equation and an elementary integration gives $F(\tau) = 4\arctan\bigl(\exp(c - \tau/a)\bigr)$ where c is an arbitrary integration constant. Thus, starting from the trivial solution $u(x,t) \equiv 0$, we have constructed the single kink solution

$$v(x,t) = 4\arctan\Bigl(\exp(ax + \frac{t}{a} + c)\Bigr) \tag{4.24}$$

of the sine-Gordon equation. Iteration yields solutions with more kinks. $\lhd$

If two differential equations are differentially related, we can associate to each solution of one of them some solutions of the other one. But in general this is not a one-to-one correspondence. In the case of the Bäcklund transformation (4.23), $v(x,t)$ solves an equation of finite type; and each solution $u(x,t)$ of the sine-Gordon equation leads via the transformation to a one-parameter family of new solutions. Our goal is now to compare the Hilbert functions of two such equations. For simplicity we will assume that all differential equations involved are involutive.

Let $h(s)$ and $\bar{h}(s)$ be the Hilbert functions of the differential equations $\mathcal{R}_q$ and $\bar{\mathcal{R}}_{\bar{q}}$, respectively. We introduce two different Hilbert functions, $g(s)$ and $\bar{g}(s)$, for the differential relation $\mathcal{T}_{r,\bar{r}}$: for the former one we consider a differential equation $\mathcal{T}_r[\bar{\sigma}]$ and for the second one an equation $\mathcal{T}_{\bar{r}}[\sigma]$. We assume here that the Hilbert functions of these equations do not depend on the sections σ and $\bar{\sigma}$, respectively.

Proposition 4.3.4. *The above defined Hilbert functions satisfy for all arguments* $s \geq \max\{r, \bar{r}\}$

$$h(s) - g(s) = \bar{h}(s + \bar{r} - r) - \bar{g}(s + \bar{r} - r) . \tag{4.25}$$

Proof. The formal power series solution of $\mathcal{R}_q$ contains $h(s)$ parametric coefficients of order s. Via the differential relation $\mathcal{T}_{r,\bar{r}}$, a family of solutions of $\bar{\mathcal{R}}_{\bar{q}}$ depending on $\bar{g}(s)$ parametric coefficients of order s corresponds to each solution σ of $\mathcal{R}_q$. However, each of these solutions of $\bar{\mathcal{R}}_{\bar{q}}$ can be obtained not only by starting with σ but also by starting with any solution in a whole family parameterised at order s by $g(s)$ coefficients (obtained by applying $\mathcal{T}_{r,\bar{r}}$ "backwards"). Since for $s \geq \max\{r, \bar{r}\}$ a coefficient of σ of order s corresponds via $\mathcal{T}_{r,\bar{r}}$ to coefficients of order $s + \bar{r} - r$ of $\bar{\sigma}$, the formal power series solution of $\bar{\mathcal{R}}_{\bar{q}}$ contains $\bar{h}(s + \bar{r} - r) = h(s) - g(s) + \bar{g}(s + \bar{r} - r)$ parametric coefficients of order $s + \bar{r} - r$. □

For lower values of s, the equality (4.25) does not need to hold. This is due to the fact that in general we cannot say that the coefficients of σ of order $s \leq \max\{r, \bar{r}\}$ are related to coefficients of $\bar{\sigma}$ of order $s + \bar{r} - r$. We only know that the coefficients of σ *up to order* r correspond to the coefficients of $\bar{\sigma}$ *up to order* $\bar{r}$. Only when the relation $\mathcal{T}_{r,\bar{r}}$ has a special form, we may be able to make stronger statements. In particular, if both equations $\mathcal{T}_r[\bar{\sigma}]$ and $\mathcal{T}_{\bar{r}}[\sigma]$ are of Cauchy–Kovalevskaya form, the relation (4.25) holds for the full Hilbert functions.

Example 4.3.5. We exhibited the relation between Maxwell's equations and the $U(1)$ Yang–Mills equations in Example 1.6.3. Maxwell's equations use as dependent variables the electric and magnetic fields $\mathbf{E}$ and $\mathbf{B}$ which may be combined into the field strength tensor F given by (1.72); the Yang–Mills equations the vector potential A. As already mentioned, in the language of differential forms F may be considered as the exterior derivative of A. At the level of the components the two sets of fields are related by

$$F_{ij} = \frac{\partial A_i}{\partial x^j} - \frac{\partial A_j}{\partial x^i} . \tag{4.26}$$

This is nothing else but a differential relation $\mathcal{T}_{1,0}$. For a given potential A the field strength F is uniquely determined; hence the corresponding Hilbert function is simply $\bar{g}(s) = 0$. For given field strength F, (4.26) represents a first-order differential equation in the potential A; its Hilbert function is easily determined to

$$g(0) = 4\,, \quad g(1) = 10\,, \quad g(2+s) = 20 + \frac{37}{3}s + \frac{5}{2}s^2 + \frac{1}{6}s^3\,. \tag{4.27}$$

The Hilbert functions of the two differential equations are given by

$$h(0) = 4\,, \quad h(1) = 16\,, \quad h(2+s) = 36 + \frac{73}{3}s + \frac{9}{2}s^2 + \frac{1}{6}s^3 \tag{4.28}$$

for the second-order Yang–Mills equations and

$$\bar{h}(0) = 6\,, \quad \bar{h}(1+s) = 16 + 12s + 2s^2 \tag{4.29}$$

for the first-order Maxwell equations. As predicted by our theory, these Hilbert functions satisfy the equality (4.25) for all values $s \geq 0$. $\lhd$

A somewhat similar situation arises in the formal analysis of gauge theories. Such theories are of great importance in physics; today all fundamental interactions are modelled by field theories invariant under gauge symmetries. In fact, this is the guiding principle in modern elementary particle physics. In this section we use a very broad definition of the term "gauge symmetry" that is much more general than the one appearing in elementary particle physics. In particular, it turns out that when a symmetry is to be promoted to a gauge symmetry depends generally on the physical interpretation and cannot be decided on the basis of purely mathematical considerations.

We assume that we are given a differential equation $\mathcal{R}_q \subseteq J_q\mathcal{E}$, the "field equations". We do not distinguish whether these are ordinary or partial differential equations and we do not assume any special structure of $\mathcal{R}_q$ like that they are Lagrangian or Hamiltonian field equations. Our considerations are valid for any kind of formalism to set up the field equations.

For us symmetries are diffeomorphisms $\mathcal{E} \to \mathcal{E}$. Thus we may treat them as global sections of the trivial bundle $\mathrm{pr}_1 : \mathcal{E} \times \mathcal{E} \to \mathcal{E}$ and form the jet bundles $J_r(\mathcal{E} \times \mathcal{E})$ with $r \geq 0$ over it. However, not every section of this bundle corresponds to a diffeomorphism. We define the open subbundles $I_r(\mathcal{E} \times \mathcal{E})$ consisting only of the jets of those local sections that are invertible. Now (local) symmetry transformations correspond to solutions of differential equations in $I_r(\mathcal{E} \times \mathcal{E})$; this is the basic idea behind the theory of Lie pseudogroups. For simplicity of notation, we identify local sections $\mathcal{U} \subseteq \mathcal{E} \to \mathcal{E} \times \mathcal{E}$ with maps $\mathcal{U} \to \mathcal{E}$.

Definition 4.3.6. *A* Lie pseudogroup *is a differential equation $\mathcal{G}_r \subseteq I_r(\mathcal{E} \times \mathcal{E})$ (often called a* finite Lie equation*) whose solutions satisfy the following two conditions.*

(i) *If the two sections* $\gamma_1 : \mathcal{U}_1 \subseteq \mathcal{E} \to \mathcal{E}$ *and* $\gamma_2 : \mathcal{U}_2 \subseteq \mathcal{E} \to \mathcal{E}$ *are both local solutions of* $\mathcal{G}_r$ *with* $\gamma_1(\mathcal{U}_1) \cap \mathcal{U}_2 \neq \emptyset$*, then the section* $\gamma_2 \circ \gamma_1$ *defined on some subset* $\mathcal{U}_3 \subseteq \mathcal{U}_1$ *with* $\gamma_1(\mathcal{U}_3) \subseteq \mathcal{U}_2$ *is also a local solution of* $\mathcal{G}_r$*.*
(ii) *If the section* $\gamma : \mathcal{U} \subseteq \mathcal{E} \to \mathcal{E}$ *is a local solution of* $\mathcal{G}_r$*, then the section* $\gamma^{-1} : \gamma(\mathcal{U}) \to \mathcal{U}$ *is also a local solution of* $\mathcal{G}_r$*.*

Obviously, this definition just rephrases the usual axioms of a local group: composition and inversion of group elements yield (where defined) again group elements. The associativity follows trivially from the properties of the composition of maps.

In our case, $\mathcal{E}$ is not an arbitrary manifold but the total space of a fibration $\pi : \mathcal{E} \to \mathcal{X}$. A *gauge transformation* γ preserves the fibration; in other words, it is a fibred morphism and induces a map $\bar{\gamma} : \mathcal{X} \to \mathcal{X}$ such that $\pi \circ \gamma = \bar{\gamma} \circ \pi$. For simplicity, we assume that $\bar{\gamma}$ is the identity map on $\mathcal{X}$. Such a gauge transformation is a *symmetry* of the differential equation $\mathcal{R}_q$, if it maps solutions into solutions. This means, if $\sigma : \mathcal{X} \to \mathcal{E}$ is a solution of $\mathcal{R}_q$, then $\gamma \circ \sigma$ is again a solution.

This is the above mentioned very general mathematical definition of a gauge symmetry; it is just a fibre preserving symmetry. In physical applications, gauge symmetries have a more narrow meaning which, however, incorporates already physical interpretations. Thus no mathematical criterion exists when a fibre preserving symmetry should be called a gauge symmetry in the physical sense. In physics, gauge symmetries represent a certain redundancy in the mathematical description of a theory. Solutions related by a gauge symmetry are identified, as one says that they correspond to the same physical state and hence cannot be distinguished by measurements.

These considerations imply that a physicist is not interested in the size of the full solution space but only in the size modulo the action of the gauge group, i. e. in the size of the "state space". We determine now this "gauge corrected" size from the Hilbert functions of the field equation $\mathcal{R}_q$ and the Lie pseudogroup $\mathcal{G}_r$.

Proposition 4.3.7. *Let* $h(s)$ *be the Hilbert function of the involutive differential equation* $\mathcal{R}_q$ *and* $g(s)$ *the Hilbert function of the involutive Lie pseudogroup* $\mathcal{G}_r$*. Then the size of the state space is measured by the gauge correct Hilbert function*

$$\bar{h}(s) = h(s) - g(s) \; . \tag{4.30}$$

Proof. We proceed as in the proof of Proposition 4.3.4. By definition of the Hilbert function, the formal power series solution of $\mathcal{R}_q$ depends on $h(s)$ parametric coefficients of order s. Using the gauge invariance $g(s)$ of these coefficients can be given arbitrary values. Thus $\bar{h}(s)$ parametric coefficients remain to distinguish physical states. □

Example 4.3.8. Let us consider again the $U(1)$ Yang–Mills equations. For them Maxwell's equations represent a gauge-invariant form, so that we can explicitly compare the result of a gauge correction with the values obtained directly from Maxwell's equations.

As already mentioned in Example 3.3.5, the $U(1)$ Yang–Mills equations are invariant under the gauge transformation $\bar{A}_i = A_i + \partial_{x^i}\Lambda$ (or $\bar{A} = A + d\Lambda$ in differential forms language) with an arbitrary function Λ. These transformations build the solution space of the following Lie equation

$$\mathcal{G}_1 : \left\{ \quad \frac{\partial \bar{A}_i}{\partial x^j} - \frac{\partial \bar{A}_j}{\partial x^i} = 0 \,, \quad \frac{\partial \bar{A}_i}{\partial A_j} = \delta_i^j \,. \right. \tag{4.31}$$

$\mathcal{G}_1$ is involutive and in a four-dimensional space-time its Hilbert function is ($\dim I_1(\mathcal{E} \times \mathcal{E}) = 36$ here and (4.31) consists of 22 first-order equations and of no algebraic equations)

$$g(0) = 4 \,, \quad g(1) = 10 \,, \quad g(2+s) = 20 + \frac{37}{3}s + \frac{5}{2}s^2 + \frac{1}{6}s^3 \,. \tag{4.32}$$

According to Proposition 4.3.7 we must subtract it from the Hilbert function (4.28) of the $U(1)$ Yang–Mills equations. Thus the gauge corrected Hilbert function is

$$\bar{h}(0) = 0 \quad \bar{h}(1) = 6 \,, \quad \bar{h}(2+s) = 16 + 12s + 2s^2 \,. \tag{4.33}$$

Up to a shift of the argument by one which is necessary because of the lower order of Maxwell's equations, we find indeed the Hilbert function (4.29) of Maxwell's equations. It is no coincidence that (4.32) is identical with (4.27): the gauge symmetry is of course sitting in the differential relation $\mathcal{T}_{1,0}$; if A and $\bar{A}$ are connected by a gauge transformation, then they yields the same field strength F.

The gauge corrected Hilbert function corresponds to the following gauge corrected Cartan characters for the Yang–Mills equations:

$$\bar{\alpha}_2^{(4)} = 0 \,, \quad \bar{\alpha}_2^{(3)} = 4 \,, \quad \bar{\alpha}_2^{(2)} = \bar{\alpha}_2^{(1)} = 6 \,. \tag{4.34}$$

Of course, they are identical with the characters (4.18) of Maxwell's equations. Since the highest character vanishes after the gauge correction, the underdeterminacy of the Yang–Mills equations is solely due to the gauge freedom. In fact, we expect this for any reasonable physical theory.

The Hilbert function of an arbitrary Yang–Mills theory over a four-dimensional space-time is simply d times (4.28), respectively (4.33) after the gauge correction, where d is the dimension of the gauge group. This is not surprising, since it is determined by the symbol where the structure constants of the Lie algebra do not appear. The principal part of an arbitrary Yang–Mills theory consists of d copies of the principal part of the $U(1)$ Yang–Mills theory; there is no coupling between terms belonging to different generators of the Lie algebra. $\lhd$

Gauge symmetries can be "fixed" by adding additional equations such that the augmented differential equation is no longer invariant. Obviously, this reduces the solution space. However, the gauge fixing conditions much be chosen so that we still have at least one solution in each orbit of the gauge group. Otherwise we have not removed gauge freedom but parts of the state space. If we are able to completely fix the gauge, then we obtain a differential equation such that each of these solutions lies on a different orbit. Thus its solution space can be bijectively mapped to the state space. The Hilbert function of this equation is then $\bar{h}(s)$. Using (4.12) we can also determine in advance its Cartan characters.

Example 4.3.9. The Lorentz gauge of Example 3.3.9 can be extended to general Yang–Mills theories. In does not eliminate all arbitrariness due to the gauge group, but it manages at least to render the originally underdetermined equation overdetermined. The gauge fixing condition is $\eta^{\mu\nu}\partial_\mu A^a_\nu = 0$. As it is of lower order, the augmented equation is not formally integrable but it becomes involutive after one prolongation with subsequent projection. It is straightforward to determine its Cartan characters:

$$\tilde{\alpha}_2^{(4)} = 0\,, \quad \tilde{\alpha}_2^{(3)} = 4d\,, \quad \tilde{\alpha}_2^{(2)} = 12d\,, \quad \tilde{\alpha}_2^{(1)} = 16d\,. \tag{4.35}$$

Comparing with (4.34) we find that while the two higher characters have the same value as the gauge corrected ones the two lower ones are larger. Thus the gauge fixed system still contains a residual gauge freedom. $\lhd$

4.4 Einstein's Strength

Einstein [46] introduced an alternative measure for the size of the formal solution space of a differential equation $\mathcal{R}_q$: the *strength*. We show now how his approach fits into our framework. Let $\mathcal{R}_q$ be an involutive differential equation in n independent variables. The starting point is the function

$$Z(r) = h(r)/\begin{bmatrix} n \\ r \end{bmatrix} \tag{4.36}$$

where h is the Hilbert function of $\mathcal{R}_q$. For values $r \geq q$ the function Z is rational in r: its denominator is of degree $n-1$; the numerator is then the Hilbert polynomial (4.11) whose degree is at most $n-1$. The function Z describes the growth of the number of parametric coefficients relative to the growth of number of Taylor coefficients of an analytic function of n variables. We expand it in powers of $1/r$:

$$Z(q+r) = Z_q^{(0)} + Z_q^{(1)}/r + O(1/r^2)\,. \tag{4.37}$$

Definition 4.4.1. $Z_q^{(0)}$ *is the* compatibility coefficient, $Z_q^{(1)}$ *the* strength[3] *of the differential equation* $\mathcal{R}_q$. *An equation with* $Z_q^{(0)} = 0$ *is called* absolutely compatible.

As the next proposition shows (but this is also clear from the interpretation of the function Z), an equation which is not absolutely compatible is necessarily underdetermined.

Proposition 4.4.2. *The compatibility coefficient and the strength of an involutive differential equation* $\mathcal{R}_q$ *in* n *independent variables with Cartan characters* $\alpha_q^{(k)}$ *are*

$$Z_q^{(0)} = \alpha_q^{(n)}\,, \tag{4.38a}$$

$$Z_q^{(1)} = (n-1)\Big(\frac{1}{2}n\,\alpha_q^{(n)} + \alpha_q^{(n-1)}\Big)\,. \tag{4.38b}$$

Proof. Proposition 4.2.5 provides the explicit form (4.11) of the Hilbert polynomial in terms of the Cartan characters. Together with (A.31) this yields readily the compatibility coefficient $Z_q^{(0)}$ by taking the limit $r \to \infty$. Subtracting the result from (4.37), multiplying by r and taking again the limit $r \to \infty$ gives the formula for the strength $Z_q^{(1)}$. □

In the case of gauge theories, Einstein used the gauge corrected Hilbert function for the definition of Z. Then the gauge corrected Cartan characters appear in Proposition 4.4.2. As we mentioned in the last section, we always expect for a reasonable physical theory that all underdeterminacy stems from the gauge freedom. After the gauge correction, the field equations of such a theory are always absolutely compatible and we get $Z_q^{(1)} = (n-1)\bar{\alpha}_q^{(n-1)}$. Thus up to a constant factor $n-1$ depending solely on the dimension of space-time, the strength coincides with Cartan's index of generality (we assume here that $\bar{\alpha}_q^{(n-1)} > 0$ which is the case for all classical examples of field theories).

[3] Actually, Einstein called $Z_q^{(1)}$ the *coefficient of freedom*; he only titled his whole discussion of the subject with the word strength. Later, other authors started to call $Z_q^{(1)}$ strength, although this is quite contradictory: the stronger the differential equation is in the sense of Einstein, the smaller is its strength!

In any case, we may conclude that Einstein's approach provides us with less information than a full involution analysis of the symbol. The compatibility coefficient and the strength correspond only to the two highest Cartan characters; the lower characters are ignored. One could probably recover them by determining further coefficients in the expansion (4.37). But it is clear that analysing the symbol yields the characters much faster and easier.

4.5 Notes

The notions of "general" and "singular" integrals have disappeared from modern textbooks on differential equations. Older books discuss them often at considerable length; see e. g. [52, Chapters VIII and XIII]. Our approach is based on ideas of Ampère [6] who was apparently the first to propose the definition that a general solution uniquely determine the differential equation and follows a formulation given by Hermann [76]. Within differential algebra, i. e. for differential equations with at most polynomial non-linearities, a rigorous definition of general and singular solutions exists. Apparently Ritt [151] was the first to introduce it; more recent work in this direction is due to Hubert [80, 81].

The here used notion of a differential relation was introduced recently in [164]. It generalises the classical concept of a Bäcklund transformation and allows for rather rough transformations of a differential equation. The gauge correction of Cartan characters is also fairly new. [160] presented a first approach based on an explicit representation of the gauge transformation; the approach via Lie pseudogroups is again from [164]. This article also discusses the use of the gauge corrected Cartan characters for counting the degrees of freedom of a gauge theory. The formal analysis of gauge fixings conditions was first proposed in [162]. More on symmetries of differential equations can be found in the textbooks [20, 129]

The Notes of Chapter 3 discussed the definition of involution via the Spencer cohomology. In this approach the Cartan characters arise as follows: assume that we are given a quasi-regular basis of $T^*\mathcal{X}$ for the symbol $\mathcal{M}_q$ and consider the subspaces $\mathcal{M}_{q,k}$ defined by (3.53); the Cartan characters are then the differences $\alpha_q^{(k)} = \dim \mathcal{M}_{q,k-1} - \dim \mathcal{M}_{q,k}$.

Einstein showed considerable interest in formal measures for the size of the solution space. His reason [46, p. 133] was the following:

> Given certain field variables and a system of field equations for them, the latter will not in general determine the field completely. There still remain certain free data for a solution of the field equation. The smaller the number of free data consistent with the system of field equations, the "stronger" is the system. It is clear that in the absence of any other viewpoint from which to select the equations, one will prefer the "stronger" system to a less strong one.

Around 1930 he discussed this question in his correspondence with Cartan [34]. Cartan tried to explain Einstein his theory of involutive systems; in fact, one letter is one of very few places where Cartan explained his theory in terms of partial differential equations and not of exterior differential systems. But for some reasons Einstein apparently never used this approach.

Many years later, he developed his own notion, the strength, in an appendix of his book on the meaning of relativity [46, pp. 133–139]. This appendix on a generalised theory of gravitation was introduced in the third edition and completely rewritten for the fifth edition under the title "Relativistic Theory of the Non-Symmetric Field". The second version starts with discussing the strength, as

> This discussion is of intrinsic interest quite apart from the particular theory presented here. For a deeper understanding of our problem, however, it is indispensable.

Of course, Einstein did not use the Hilbert function for defining the strength. He gave a much more cumbersome recipe for determining it. In order to obtain the numerator of (4.36), he counted the number of field equations, the number of identities satisfied by them, the number of identities of these identities and so on. Obviously, this process is prone to errors, in particular when applied to an equation which is not involutive. Indeed, one can find several examples in the literature where wrong results for the strength were computed. In contrast, the determination of the Cartan characters via the symbol is trivial.

The strength did not spark much interest. Only few authors applied it to various field theories [29, 78, 114–116, 127, 158, 167]. Penney [135] used it for some metaphysical speculations on the dimensionality of the world. He conjectured that the world is four-dimensional, because only in four dimensions the Maxwell and the Einstein equations possess the same strength. In fact, already Einstein [46, p. 139] remarked:

> It is surprising that the gravitational equations for empty space determine their field just as strongly as do Maxwell's equations in the case of the electromagnetic field.

However, if one computes the full set of (gauge corrected) Cartan characters for both theories, one sees that even in four dimensions only the highest non-vanishing one is identical; the other ones and thus the size of the formal solution spaces of the two theories are different [160]. Furthermore, there is no compelling physical reason why the two theories should have formal solution spaces of the same size.

Sué [174] was apparently the first who tried to relate Einstein's and Cartan's approach, especially for gauge theories. But he compared a strength which was corrected for the gauge freedom with Cartan characters which were not corrected, so some of his conclusions are misleading. The relation (4.38) between strength and compatibility index on one side and the two highest Cartan characters on the other side was first derived in [160].