ABZÄHLENDE KOMBINATORIK
VORLESUNG IM WINTERSEMESTER 1999/2000

MICHAEL STOLL

1. Einführung und Grundlagen

1.1. **Einführung.** Die erste Frage, die sich stellt, ist natürlich

Was ist Abzählende Kombinatorik?

Die Antwort lautet (recht allgemein und ein bißchen vage)

> *Abzählende Kombinatorik beschäftigt sich mit der (möglichst expliziten) Bestimmung der Mächtigkeit endlicher Mengen und mit Beziehungen zwischen solchen Mächtigkeiten.*

Die fraglichen endlichen Mengen sind dabei natürlich nicht durch Aufzählung ihrer Elemente gegeben (dann wäre die Anzahlbestimmung ja trivial), sondern durch Beschreibung. Ein Beispiel:

> *Wie viele unterscheidbare Möglichkeiten gibt es, die Seiten eines Würfels mit sechs (gegebenen) verschiedenen Farben zu bemalen?*

Meistens hat man es jedoch nicht mit einer einzelnen Menge zu tun, sondern mit einer ganzen Familie, die von einem oder mehreren Parametern abhängt. Gesucht ist dann eine Formel für die Anzahl der Elemente in Abhängigkeit von den Parametern, wobei die Formel möglichst „einfach“ oder in „geschlossener Form“ sein soll. (Was darunter zu verstehen ist, variiert je nach Schwierigkeit des Problems.) Zum Beispiel:

Wie viele Teilmengen hat eine n-elementige Menge?

Die Antwort — 2^n — kennt natürlich jeder. Die Frage ist, wie ein kombinatorischer Beweis dafür aussieht. Überhaupt sind für spätere Anwendungen weniger die *Ergebnisse* der Abzählenden Kombinatorik interessant, als vielmehr die *Methoden*. Wenn Sie einmal (z.B. im Rahmen einer Diplomarbeit) ein Abzählpeoblem zu lösen haben, werden Ihnen Resultate über andere Abzählprobleme wenig nützen (es sein denn, Ihr spezielles Problem wurde auch behandelt). Die Methoden aber, mit denen die Ergebnisse gewonnen wurden, lassen sich auch auf viele andere Probleme anwenden.

WIr wollen nun zwei Beweise geben, an denen man schon sehr schön zwei Grundvarianten eines kombinatorischen Beweises sehen kann. Zuerst noch ganz kurz etwas Notation: Die Mächtigkeit einer (endlichen) Menge A werden wir mit $\#A$ notieren, und die Potenzmenge (also die Menge aller Teilmengen) von A wird mit $\mathcal{P}(A)$ bezeichnet.

1. Beweis (Bijektion): Wir konstruieren eine bijektive Abbildung von $\mathcal{P}(A)$ auf eine Menge, von der wir schon wissen oder leicht sehen können, daß sie 2^n

Elemente hat. Eine geeignete Menge ist die Menge $\{0,1\}^A$ aller Abbildungen $f : A \to \{0,1\}$ (oder 0–1–Folgen $(f_a)_{a\in A}$ mit der Indexmenge A). Wir bilden eine Teilmenge $T \in \mathcal{P}(A)$ ab auf die Abbildung f mit $f(a) = 1$ für $a \in T$ und $f(a) = 0$ für $a \in A\backslash T$. In der anderen Richtung bilden wir f ab auf die Teilmenge $T = f^{-1}(1) = \{a \in A \mid f(a) = 1\}$. Offenbar sind diese Zuordnungen invers zueinander, so daß wir tatsächlich eine Bijektion $\mathcal{P}(A) \to \{0,1\}^A$ konstruiert haben. Dann folgt aber

$$\#\mathcal{P}(A) = \#(\{0,1\}^A) = 2^{\#A}\,.$$

2. Beweis (Rekursion/Induktion): Sei $a_n = \#\mathcal{P}(A)$, wenn $\#A = n$ ist. Wir versuchen, eine Rekursionsformel für die Folge (a_n) zu finden. Offenbar ist $a_0 = 1$ (denn $\mathcal{P}(\emptyset) = \{\emptyset\}$). Sei nun A eine n-elementige Menge und $A' = A \cup \{a\}$ eine $(n+1)$-elementige Menge. Dann können wir die Teilmengen von A' aufteilen in diejenigen, die a nicht enthalten, und diejenigen, die a enthalten. Die der ersten Sorte sind gerade die Elemente von $\mathcal{P}(A)$. Die anderen entstehen aus den Teilmengen von A, indem man das Element a hinzunimmt. Folglich gilt

$$a_0 = 1\,, \qquad a_{n+1} = a_n + a_n = 2\,a_n\,,$$

und damit (durch einen trivialen Induktionsbeweis) $a_n = 2^n$.

Zum Abschluß dieser Einführung noch ein weniger triviales, dafür aber vielleicht interessanteres Beispiel.

> *In wie viele Gebiete wird die Ebene durch n Geraden in allgemeiner Lage (d.h. keine zwei parallel, und keine drei durch einen Punkt) zerlegt?*

Es ist normalerweise eine gute Idee, zunächst einmal die ersten paar Anzahlen in der Folge zu bestimmen. Wenn a_n die Anzahl der Gebiete bei n Geraden bezeichnet, dann haben wir offenbar (Skizze machen!)

$$a_0 = 1, \quad a_1 = 2, \quad a_2 = 4, \quad a_3 = 7, \quad a_4 = 11, \quad \ldots$$

Wie bei manchen Intelligenztests (Wie geht diese Folge weiter?) kann es sich lohnen, nach einem einfachen Bildungsgesetz Ausschau zu halten. Hier werden wir schnell fündig: Es scheint $a_{n+1} = a_n + n + 1$ zu gelten. Wenn wir das beweisen können, sind wir im wesentlichen fertig. Stellen wir uns also vor, wir hätten bereits n Geraden gezeichnet, die die Ebene in a_n Gebiete zerlegen. Wie viele Gebiete kommen hinzu, wenn wir eine weitere Gerade einzeichnen? Offenbar kommt die Zunahme dadurch zustande, daß einige der vorhandenen Gebiete durch die neue Gerade in zwei Teile geteilt werden. Wie viele Gebiete trifft nun die neue Gerade? Offenbar wechselt sie genau dann von einem Gebiet in ein anderes, wenn sie eine der alten Geraden trifft. Die durchschnittenen Gebiete entsprechen also genau den Abschnitten, in die die neue Gerade durch die Schnittpunkte mit den alten Geraden geteilt wird. Da es n solcher Schnittpunkte gibt, wird die neue Gerade in $n+1$ Teile geteilt, also gibt es auch $n+1$ zerschnittene Gebiete, q.e.d. Aus der Rekursionsformel und dem Anfangswert $a_0 = 1$ bekommen wir schließlich

$$a_n = 1 + \sum_{k=1}^{n} k = 1 + \frac{n(n+1)}{2}\,.$$

Ein anderer Beweis, der ohne Induktion auskommt, geht so: Wir wählen eine Richtung in der Ebene (die nicht sekrecht auf einer der Geraden steht) als „nach oben“. Die Gebiete lassen sich dann einteilen in solche, die nach unten beschränkt sind, und solche, die nach unten unbeschränkt sind. Jedes nach unten beschränkte Gebiet hat einen eindeutig bestimmten tiefsten Punkt; dieser ist ein Schnittpunkt zweier Geraden. Umgekehrt ist jeder solche Schnittpunkt auch tiefster Punkt eines (dann natürlich nach unten beschränkten) Gebiets. Somit gibt es genau $\binom{n}{2}$ nach unten beschränkte Gebiete. Um die übrigen Gebite abzuzählen, betrachten wir eine waagerechte Gerade g, die unterhalb aller Geradenschnittpunkte liegt. Sie trifft genau die nach unten unbeschränkten Gebiete, und (wie im anderen Beweis) diese Gebiete entsprechen genau den Abschnitten, in die g durch die n gegebenen Geraden zerlegt wird. Also gibt es genau $n+1$ nach unten unbeschränkte Gebiete. Insgesamt haben wir also

$$\binom{n}{2} + n + 1 = \binom{n}{2} + \binom{n}{1} + \binom{n}{0}$$

Gebiete. (Die etwas künstlich erscheinende Schreibweise rechts wird klar, wenn man Ebenen im Raum oder noch höherdimensionalere Analoga betrachtet.)

1.2. **Grundlagen.** Nach diesen einführenden Beispielen wollen wir uns nun den Grundlagen Nach diesen einführenden Beispielen wollen wir uns nun den Grundlagen zuwenden. Die folgenden drei Grundregeln sind völlig elementar und im Grunde trivial; dennoch bilden sie das Fundament allen Abzählens. Alle vorkommenden Mengen sind endlich; die disjunkte Vereinigung von Mengen sei mit $A \uplus B$ etc. bezeichnet (im Handschriftlichen erscheint statt des Pluszeichens ein Punkt).

(1) **Bijektion.** Die fundamentalste aller Regeln:
Ist $f : A \to B$ bijektiv, so gilt $\#A = \#B$.

(2) **Summe.**

$$\#(A \uplus B) = \#A + \#B\,, \qquad \#\biguplus_{i\in I} A_i = \sum_{i\in I} \#A_i\,.$$

Und die Version für eine nicht-disjunkte Vereinigung:

$$\#(A \cup B) = \#A + \#B - \#(A \cap B)\,.$$

(3) **Produkt.**

$$\#(A \times B) = \#A \cdot \#B\,, \quad \#\prod_{i\in I} A_i = \prod_{i\in I} \#A_i\,,$$
$$\#(A^n) = (\#A)^n\,, \quad \#(A^B) = (\#A)^{\#B}\,.$$

Eine einfache Folgerung aus diesen Grundregeln ist das *Primzip des zweifachen Abzählens*: Sei $S \subset A \times B$. Wir setzen $A_b = \{a \in A \mid (a,b) \in S\}$ und $B_a = \{b \in B \mid (a,b) \in S\}$. Dann gilt

$$\#S = \sum_{a\in A} \#B_a = \sum_{b\in B} \#A_b\,.$$

Wir werden bald eine Anwendung davon sehen.

2. Binomialkoeffizienten

Binomialkoeffizienten treten in vielen Zusammenhängen immer wieder auf, deswegen wollen wir hier ein paar wichtige Eigenschaften zusammentragen.

Definition 2.1. Mit $\binom{n}{k}$ bezeichnen wir die Anzahl der k-elementigen Teilmengen einer n-elementigen Menge. Dabei sei n eine nichtnegative ganze Zahl und k eine ganze Zahl, mit der Konvention, daß $\binom{n}{k} = 0$ ist für $k < 0$.

Eine erste wichtige Eigenschaft ist die *Symmetrie* der Binomialkoeffizienten.

Proposition 2.2. *Es gilt* $\displaystyle\binom{n}{k} = \binom{n}{n-k}.$

Beweis: Die Sache ist klar für $k < 0$ oder $k > n$, da dann beide Seiten null sind. Wenn $0 \le k \le n$ ist, dann liefert $T \mapsto A \setminus T$ (mit $\#A = n$) offenbar eine Bijektion zwischen den k-elementigen und den $(n-k)$-elementigen Teilmengen. □

Die zweite wichtige Eigenschaft ist diejenige, die den Binomialkoeffizienten ihren Namen gegeben hat.

Proposition 2.3. *Es gilt* $\displaystyle(1+x)^n = \sum_{k=0}^{n} \binom{n}{k} x^k.$

Beweis: Wir wollen die x in den n Klammern zunächst als verschieden betrachten. Dann erhalten wir durch Ausmultiplizieren (in jedem Faktor können wir unabhängig entscheiden, ob wir 1 oder x_j wählen)

$$(1+x_1)(1+x_2)\dots(1+x_n) = \sum_{T\subset\{1,\dots,n\}} \prod_{j\in T} x_j\,.$$

Wenn wir jetzt jedes x_j wieder durch x ersetzen, haben wir

$$(1+x)^n = \sum_{T\subset\{1,\dots,n\}} x^{\#T} = \sum_{k=0}^{n} \binom{n}{k} x^k\,.$$

□

Dann gibt es die Formel für die Summe.

Proposition 2.4. *Es gilt* $\displaystyle\sum_{k=0}^{n} \binom{n}{k} = 2^n.$

Beweis: (1) Rechts steht die Anzahl aller Teilmengen einer n-elementigen Menge; links steht dieselbe Anzahl nach Größe der Teilmenge sortiert.
(2) Alternativ mit Prop. 2.3:

$$\sum_{k=0}^{n} \binom{n}{k} = \sum_{k=0}^{n} \binom{n}{k} 1^k = (1+1)^n = 2^n\,.$$

□

Folgende Identität ist unter dem Namen *Vandermondesche Konvolution* bekannt.

Proposition 2.5. *Es gilt* $\displaystyle\sum_{j=0}^{k}\binom{n}{j}\binom{m}{k-j}=\binom{n+m}{k}$.

BEWEIS: (1) Seien A und B disjunkte Mengen mit $\#A=n$ und $\#B=m$. Der Ausdruck rechts zählt die k-Teilmengen von $A\cup B$. Im linken Ausdruck werden dieselben Teilmengen T gezählt, aber sortiert nach der Größe j von $T\cap A$ (bzw. $\#(T\cap B)=k-j$).
(2) Alternativ mit Prop. 2.3 folgt die Behauptung durch Koeffizientenvergleich in $(1+x)^n(1+x)^m=(1+x)^{n+m}$. □

Korollar 2.6. *Es gilt* $\displaystyle\sum_{k=0}^{n}\binom{n}{k}^2=\binom{2n}{n}$.

BEWEIS: Das folgt unter Beachtung von Prop. 2.2 aus Prop. 2.5, wenn man dort $m=n$ setzt. □

Bemerkung 2.7. Es gibt die Formeln

$$\sum_{k=0}^{n}\binom{n}{k}^0=n+1\,,\quad \sum_{k=0}^{n}\binom{n}{k}^1=2^n\,,\quad \sum_{k=0}^{n}\binom{n}{k}^2=\binom{2n}{n}\,.$$

Man kann jedoch beweisen, daß es für $\sum_{k=0}^{n}\binom{n}{k}^3$ keine „einfache“ Formel gibt (mit einer vernünftigen und präzisen Definition von „einfach“).

Schließlich das angekündigte Beispiel für eine Anwendung des Prinzips des zweifachen Abzählens.

Proposition 2.8. *Es gilt* $(n+1)\binom{n}{k}=(k+1)\binom{n+1}{k+1}$.

BEWEIS: (1) Sei A eine Menge mit $\#A=n+1$. Wir zählen die Paare $(a,T)\in A\times\mathcal{P}(A)$ mit $a\in T$ und $\#T=k+1$ auf zwei Arten ab. Einmal können wir $a\in A$ beliebig wählen und dann k Elemente aus $A\setminus\{a\}$ hinzufügen; das ergibt den Term auf der linken Seite. Oder wir wählen zunächst die Teilmenge T aus und anschließend ein Element $a\in T$; das ergibt den Term auf der rechten Seite.
(2) Alternativ mit Prop. 2.3 folgt die Behauptung durch Koeffizientenvergleich in $(n+1)(1+x)^n=\frac{d}{dx}(1+x)^{n+1}$. □

Durch Induktion erhält man daraus folgende bekannte Formel.

Korollar 2.9.

$$\binom{n}{k}=\frac{n(n-1)\dots(n-k+1)}{k!}=\frac{n!}{k!(n-k)!}\,.$$

Ein einfacher Spezialfall der Vandermondeschen Kovolution Prop. 2.5 ist folgende Rekursionsformel für die Binomialkoeffizienten.

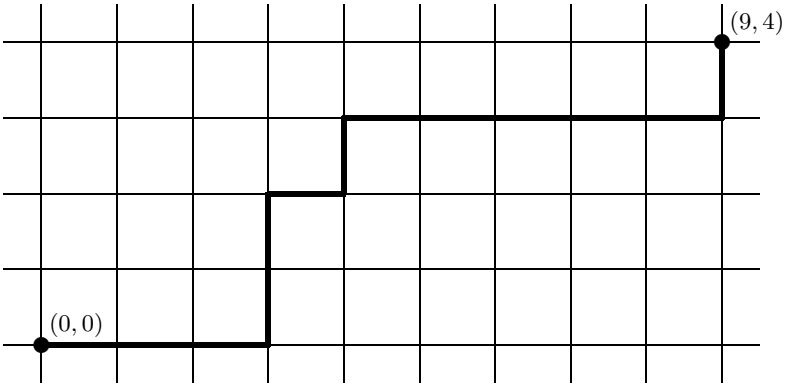
Korollar 2.10. *Es gilt* $\displaystyle\binom{n+1}{k+1}=\binom{n}{k}+\binom{n}{k+1}$.

BEWEIS: Setze in Prop. 2.5 $(n, m, k) := (n, 1, k+1)$. □

Mit dieser Formel und den Anfangswerten $\binom{0}{k} = 1$ für $k = 0$ und $= 0$ für $k \neq 0$ gewinnt man die Binomialkoeffizienten in Form des *Pascalschen Dreiecks*:

					1							
				1		1						
			1		2		1					
		1		3		3		1				
	1		4		6		4		1			
1		5		10		10		5		1		
	1		6		15		20		15		6		1
		1		7		21		35		35		21		7		1
	1		8		28		56		70		56		28		8		1
1		9		36		84		126		126		84		36		9		1
	1		10		45		120		210		252		210		120		45		10		1

Wir wollen jetzt noch ein paar weitere Situationen kennenlernen, in denen Binomialkoeffizienten auftauchen. Im einen Fall geht es um Wege in einem Gitter. Dazu betrachten wir das Quadratgitter in der Ebene, das durch $(x, y) \in \mathbb{R}^2 \mid x \in \mathbb{Z}$ oder $y \in \mathbb{Z}\}$ gegeben ist. Die Schnittpunkte der *Gitterlinien* heißen *Gitterpunkte*; das sind genau die Punkte mit ganzzahligen Koordinaten. Ein *Gitterweg* von $(a, b) \in \mathbb{Z}^2$ nach $(c, d) \in \mathbb{Z}^2$ soll für uns ein Weg vom Punkt (a, b) zum Punkt (c, d) sein, der stets längs der Gitterlinien verläuft und immer nach rechts oder nach oben geht (insbesondere muß also $c \geq a$ und $d \geq b$ sein). Das Bild zeigt ein Beispiel.



Wie viele Gitterwege gibt es nun von $(0, 0)$ zum Punkt (n, m), wenn n und m nichtnegative ganze Zahlen sind?

Proposition 2.11. *Es gibt genau* $\binom{n+m}{n} = \binom{n+m}{m} = \frac{(n+m)!}{n!m!}$ *Gitterwege von* $(0, 0)$ *nach* (n, m).

BEWEIS: Die Gitterwege entsprechen genau den Möglichkeiten, die n Schritte nach rechts (bzw. die m Schritte nach oben) auf die insgesamt $n + m$ Schritte zu verteilen. □

Als Übung sei empfohlen, die verschiedenen Eigenschaften der Binomialkoeffizienten anhand des Gitterweg-Modells zu beweisen.

Gitterwege werden uns im nächsten Kapitel noch einmal begegnen.

Nun zu einer ganz anderen Frage.

> *Wie viele Möglichkeiten gibt es, eine natürliche Zahl n als Summe von m positiven ganzen Zahlen (unter Beachtung der Reihenfolge) zu schreiben?*

Anders gesagt, handelt es sich um die Anzahl der Lösungen in positiven ganzen Zahlen der Gleichung

$$x_1 + x_2 + \cdots + x_m = n\,. \tag{2.1}$$

Zum Beispiel haben wir für $m = 3$ und $n = 5$ folgende Lösungen:

$$5 = 3+1+1 = 2+2+1 = 2+1+2 = 1+3+1 = 1+2+2 = 1+1+3\,.$$

Zur Lösung des Problems stellen wir uns n nebeneinander gemalte Punkte vor. Eine Lösung $(x_1, \ldots, x_m)$ obiger Gleichung (2.1) kann dann dargestellt werden, indem wir der Reihe nach von links x_1, x_2, ... Punkte abzählen und durch einen Strich abteilen. Wir müssen also $m-1$ Striche auf die $n-1$ Zwischenräume zwischen den Punkten verteilen. Es folgt:

Proposition 2.12. *Gleichung (2.1) hat genau* $\binom{n-1}{m-1}$ *Lösungen in positiven ganzen Zahlen.*

Mit Prop. 2.4 (oder direkt mit obigem Argument) sieht man dann:

Korollar 2.13. *Eine natürliche Zahl $n > 0$ läßt sich auf genau 2^{n-1} Weisen als (geordnete) Summe positiver ganzer Zahlen schreiben.*

Was passiert, wenn wir zulassen, daß $x_j = 0$ ist? Aus einer Lösung von (2.1) in nichtnegativen ganzen Zahlen erhalten wir eine Lösung von

$$x_1 + x_2 + \cdots + x_m = n + m$$

in positiven ganzen Zahlen, indem wir zu jedem x_j eins addieren.

Proposition 2.14. *Gleichung (2.1) hat genau* $\binom{n+m-1}{m-1}$ *Lösungen in nichtnegativen ganzen Zahlen.*

(Ein weiterer Beweis ergibt sich, indem man die Lösungen danach sortiert, welche der Variablen den Wert null annehmen. Man erhält

$$\sum_{T\subset\{1,\ldots,m\}} \binom{n-1}{m-\#T-1} = \sum_{k=0}^{m} \binom{m}{k}\binom{n-1}{m-k-1} = \binom{m+n-1}{m-1}$$

nach Prop. 2.5.)

Betrachten wir nun eine Zerlegung von n in $m+k$ positive ganze Zahlen:

$$n = x + y = (x_1 + x_2 + \cdots + x_m) + (y_1 + y_2 + \cdots + y_k)\,.$$

Zu jeder solchen Zerlegung gehört eine Zerlegung $n = x+y$, eine Zerlegung von x in m Teile und eine Zerlegung von y in k Teile. Wir erhalten:

Proposition 2.15. *Es gilt für $m, k > 0$:*

$$\sum_{j=1}^{n-1} \binom{j-1}{m-1}\binom{n-j-1}{k-1} = \binom{n-1}{m+k-1}.$$

Folgende äquivalente Form dürfte etwas einfacher anzuwenden sein. Sie gilt für $n, m, k \geq 0$.

$$\sum_{j=0}^{n} \binom{j}{m}\binom{n-j}{k} = \binom{n+1}{m+k+1}.$$

Einen wichtigen Spezialfall erhalten wir, wenn $m = 0$ ist.

Korollar 2.16. *Es gilt für $k \geq 0$:* $\displaystyle\sum_{j=0}^{n} \binom{j}{k} = \binom{n+1}{k+1}.$

Übungsaufgaben 2.17.

1. Sei A eine n-elementige Menge. Wie viele Teilmengen von A haben eine ungerade Anzahl von Elementen?
2. Wie viele Möglichkeiten gibt es, aus einer n-elementigen Menge zwei disjunkte Teilmengen T_1, T_2 auszuwählen?
3. Finden Sie zwei verschiedene Beweise für die Identität
$$(n-k)\binom{n}{k} = (k+1)\binom{n}{k+1}.$$

3. Kinowarteschlangen und Catalan-Zahlen

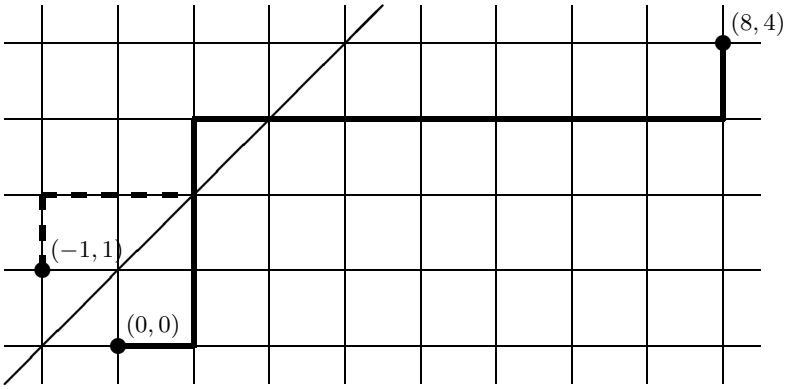
Wenden wir uns nun etwas schwierigeren Fragen zu.

> *Vor einem Kino (Eintritt DM 10) warten m Leute, die genau einen 10-DM-Schein dabei haben und n Leute, die genau einen 20-DM-Schein dabei haben. Die Kasse ist leer. Wie viele Möglichkeiten gibt es, die 10-DM-Leute und die 20-DM-Leute so auf die Schlange zu verteilen, daß immer genügend Wechselgeld vorhanden ist?*

Offenbar ist die Antwort 0, wenn $n > m$ ist. Wir wollen daher $n \leq m$ voraussetzen. Wenn wir jemanden mit einem 10-DM-Schein durch eine 1 und jemanden mit einem 20-DM-Schein durch eine 2 symbolisieren, dann läuft obige Aufgabe auf die Frage hinaus, wie viele Folgen aus m Einsen und n Zweien es gibt, so daß in jedem Anfangsstück stets mindestens so viele Einsen wie Zweien vorkommen. Wir können diese Folgen auch als Gitterwege betrachten, indem wir eine 1 als einen Schritt nach rechts und eine 2 als einen Schritt nach oben übersetzen. Dann läßt sich die Frage interpretieren als die nach der Zahl der Gitterwege von $(0, 0)$ nach (m, n), die die Winkelhalbierende des 1. Quadranten nicht überqueren. Eine äquivalente Formulierung der Bedingung ist, daß der Weg die Gerade $y = x + 1$ nicht berühren darf.

In vielen Fällen, in denen nach der Anzahl von Elementen (hier Gitterwegen) mit einer bestimmten Eigenschaft gefragt ist, ist es einfacher, die Elemente abzuzählen, die diese Eigenschaft nicht besitzen. Das gilt auch hier. Jeder schlechte

Gitterweg von $(0,0)$ nach (m,n) muß die Gerade $y = x+1$ irgendwann zum ersten Mal berühren, sagen wir im Punkt (r,s). Wenn wir den Weg von $(0,0)$ nach (r,s) an der Geraden $y = x+1$ spiegeln, erhalten wir einen Gitterweg von $(-1,1)$ nach (r,s), der zusammen mit dem ursprünglichen Weg von (r,s) nach (m,n) einen Gitterweg von $(-1,1)$ nach (m,n) bildet. (Diese Idee ist über 100 Jahre alt: D. André, 1887.)



Umgekehrt erhalten wir aus jedem Gitterweg von $(-1,1)$ nach (m,n) einen schlechten Weg von $(0,0)$ nach (m,n), indem wir den Abschnitt bis zum ersten Berühren der Geraden $y = x+1$ an dieser Geraden spiegeln. Wenn $n \le m$ ist, muß ja jeder Weg von $(-1,1)$ nach (m,n) diese Gerade wenigstens einmal berühren. Wir haben also gezeigt:

Satz 3.1. *Die Anzahl der unterhalb der Winkelhalbierenden des ersten Quadranten verlaufenden Gitterwege von* $(0,0)$ *nach* (m,n) *(mit* $n \le m$*) ist*

$$\binom{n+m}{m} - \binom{n+m}{m+1} = \frac{m+1-n}{m+1}\binom{n+m}{m} = \frac{m+1-n}{n}\binom{n+m}{m+1}.$$

Im Nachhinein läßt sich das auch noch anders einsehen. Sei $a_{m,n}$ die Anzahl der eingeschränkten (also „guten“) Gitterwege von $(0,0)$ nach (m,n). Dann haben wir offensichtlich immer noch die Rekursion

$$a_{m+1,n+1} = a_{m,n+1} + a_{m+1,n}\,,$$

gültig für $m,n \ge 0$ und $m \ge n$, aber mit den Randbedingungen

$$a_{m,0} = 1 \qquad \text{und} \qquad a_{m,m+1} = 0$$

für $m \ge 0$. Denn jeder Weg endet nach wie vor entweder mit einem Schritt nach rechts oder einem Schritt nach oben; zu einem Punkt $(m,0)$ gibt es stets genau einen Gitterweg, und die Gerade $y = x+1$ ist verboten. Beide Summanden in Satz 3.1 erfüllen die Rekursionsgleichung, und die Randwerte stimmen auch. Auf diesem Wege läßt sich der Satz auch verallgemeinern auf den Fall, daß die 20-DM-Scheine durch $(k\cdot 10)$-DM-Scheine ersetzt werden (siehe Übungsaufgabe).

Interessant ist der Spezialfall $m = n$; die dann auftretenden Zahlen

$$c_n = \frac{1}{n+1}\binom{2n}{n}$$

heißen *Catalan-Zahlen*. Sie treten in vielen Zusammenhängen auf. (Ganz nebenbei erhalten wir die Aussage, daß $\binom{2n}{n}$ durch $n+1$ teilbar ist.) Hier sind die

ersten Glieder dieser Folge:

$$c_0 = 1, \quad c_1 = 1, \quad c_2 = 2, \quad c_3 = 5, \quad c_4 = 14, \quad c_5 = 42, \quad \dots$$

Einige Abzählprobleme, die durch die Catalan-Zahlen gelöst werden, sind im folgenden Satz zusammengestellt.

Satz 3.2. *Folgende Anzahlen sind durch die Catalan-Zahl c_n gegeben.*

(1) *Die Folgen aus je n Nullen und Einsen, so daß jedes Anfangsstück mindestens so viele Nullen wie Einsen enthält.*
(2) *Die Folgen aus je $n+1$ Nullen und Einsen, so daß jedes Anfangsstück der Länge ungleich 0 oder $2n+2$ mehr Nullen als Einsen enthält.*
(3) *Die vollständigen Klammerungen eines Produkts aus $n+1$ Faktoren.*
(4) *Die ebenen Setzbäume mit $n+2$ Knoten (und daher $n+1$ Kanten).*
(5) *Die ebenen binären Setzbäume mit $n+1$ Blättern (ungleich der Wurzel).*
(6) *Die Zerlegungen eines konvexen $(n+2)$-Ecks in Dreiecke durch sich nicht schneidende Diagonalen.*
(7) *Die Möglichkeiten, $2n$ auf einem Kreis gelegene Punkte paarweise durch n sich nicht überschneidende Strecken zu verbinden.*

Hier sind zunächst einige Erklärungen nötig. Ein *Baum* ist ein zusammenhängender (einfacher, schlingenloser, ungerichteter) *Graph* ohne Kreise. Äquivalent zur letzten Bedingung ist, daß die Anzahl der Kanten eins weniger als die Anzahl der Ecken ist. Eine andere Charakterisierung von Bäumen ist, daß es zwischen je zwei Ecken einen eindeutigen Weg (ohne mehrfach durchlaufene Kanten) gibt. Ein *Wurzelbaum* ist ein Baum, in dem ein Knoten als Wurzel ausgezeichnet ist. Ein *Setzbaum* ist ein Wurzelbaum, dessen Wurzel Grad 1 hat (d.h. es geht nur eine Kante von ihr aus). Ein Wurzelbaum ist *eben* (oder *planar*), wenn die von der Wurzel weg weisenden Kanten in jedem Knoten geordnet sind (z.B. von links nach rechts). Ein Baum ist *binär*, wenn alle Knoten entweder Grad 1 (dann heißen sie *Blätter*) oder Grad 3 haben.

Als Beispiel sind in Abb. 1 die verschiedenen Figuren für $n = 3$ gezeigt.

Im folgenden sei $\mathcal{A}_m$ die Menge der kombinatorischen Objekte in der Behauptung (m) und $\mathcal{A}_m(n)$ die Teilmenge der Objekte, die zum Index n gehören. Die Behauptung des Satzes ist also $\#\mathcal{A}_m(n) = c_n$ für $m = 1, 2, 3, 4, 5, 6, 7$.

Zum Beweis des Satzes überlegen wir uns zunächst, daß (1) und (2) dieselbe Anzahl definieren. Da wir schon wissen, daß die zu (1) gehörende Anzahl c_n ist, wird dadurch Teil (2) des Satzes bewiesen. Dazu bilden wir die Objekte in $\mathcal{A}_1(n)$ und in $\mathcal{A}_2(n)$ bijektiv aufeinander ab. Wenn X in $\mathcal{A}_1(n)$ ist, dann ist offensichtlich $0X1$ in $\mathcal{A}_2(n)$. Ist umgekehrt Y in $\mathcal{A}_2(n)$, dann muß Y mit 0 beginnen (denn das Anfangsstück der Länge 1 enthält mehr Nullen als Einsen) und mit 1 enden (selbes Argument mit dem Anfangsstück der Länge $2n+1$). Also ist $Y = 0X1$, und X muß in $\mathcal{A}_1(n)$ sein. Damit ist die Bijektion hergestellt.

Nun gibt es mehrere Möglichkeiten, Folgen aus $\mathcal{A}_1$ rekursiv zu konstruieren beziehungsweise in kleinere Teile zu zerlegen. Zum einen hat eine solche Folge X eine eindeutige Zerlegung $X = X_1X_2\dots X_k$ in Teilfolgen $X_j \in \mathcal{A}_2$. Zum Beispiel zerlegt sich 001101 als 0011 | 01. Umgekehrt ist jede durch Aneinanderhängen von Folgen aus $\mathcal{A}_2$ entstehende Folge wieder in $\mathcal{A}_1$. Es ergibt sich also folgende *erste Rekursionsformel* für die Catalan-Zahlen.

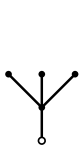
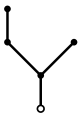
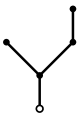
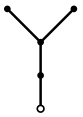
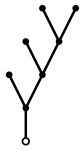
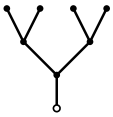
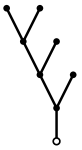
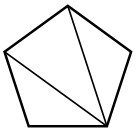
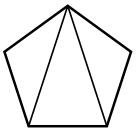
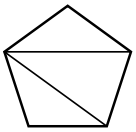
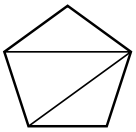
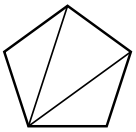
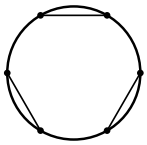
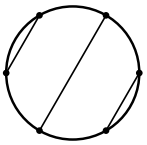
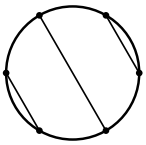
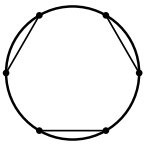
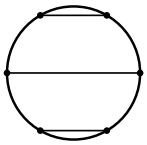
(1)	010101	001101	010011	001011	000111
(2)	00101011	00011011	00100111	00010111	00001111
(3)	$(x(x(xx)))$	$((xx)(xx))$	$(x((xx)x))$	$((x(xx))x)$	$(((xx)x)x)$
(4)					
(5)					
(6)					
(7)					

ABBILDUNG 1. Die kombinatorischen Figuren zu c_3

Korollar 3.3. *Es gilt*

$$c_n = \sum_{k=0}^{\infty} \sum_{\substack{m_1,\dots,m_k \geq 0 \\ m_1+\cdots+m_k=n-k}} c_{m_1} \dots c_{m_k} .$$

(Diese Formel enthält den Spezialfall $c_0 = 1$.)

Ausgeschrieben bedeutet das

$$c_0 = 1\,, \quad c_1 = c_0 = 1\,, \quad c_2 = c_1 + c_0^2 = 2\,, \quad c_3 = c_2 + c_0c_1 + c_1c_0 + c_0^3 = 5\,, \quad \dots$$

Wir können aber auch die Zerlegung $X = X_1X'$ betrachten (mit $X' = X_2 \dots X_k$ in der früheren Zerlegung), d.h. $X_1 \in \mathcal{A}_2$ und $X' \in \mathcal{A}_1$. Diese Zerlegung ist ebenfalls eindeutig, und umgekehrt ist $X_1X' \in \mathcal{A}_1$, wenn $X_1 \in \mathcal{A}_2$ und $X' \in \mathcal{A}_1$. Das ergibt die (praktischere) *zweite Rekursionsformel* für die Catalan-Zahlen.

Korollar 3.4. *Es gilt*

$$c_0 = 1\,, \qquad c_n = \sum_{k=0}^{n-1} c_k c_{n-1-k} \quad \textit{für } n \geq 1.$$

Ausgeschrieben bedeutet das

$$c_0 = 1\,, \quad c_1 = c_0^2 = 1\,, \quad c_2 = c_0c_1 + c_1c_0 = 2\,, \quad c_3 = c_0c_2 + c_1^2 + c_2c_0 = 5\,, \quad \dots$$

Um nun den Satz fertig zu beweisen, überzeugen wir uns davon, daß die Objekte in den verschiedenen Mengen $\mathcal{A}_m$ sich analog zerlegen lassen.

Zu (3): Jedes vollständig geklammerte Produkt aus $n+1$ Faktoren (mit $n \geq 1$) zerlegt sich eindeutig in zwei vollständig geklammerte Produkte (die beiden Faktoren in der äußersten Klammer), eines mit $k+1$ Faktoren, das andere mit $(n-1-k)+1$ Faktoren. Wir haben also eine Bijektion zwischen

$$\mathcal{A}_3(n) \quad \text{und} \quad \biguplus_{k=0}^{n-1} \mathcal{A}_3(k) \times \mathcal{A}_3(n-1-k)\,.$$

Da außerdem $\#\mathcal{A}_3(0) = 1$ ist, erfüllen die Zahlen $\#\mathcal{A}_3(n)$ die zweite Rekursionsformel, also ist $\#\mathcal{A}_3(n) = c_n$.

Zu (4): Ein ebener Setzbaum entsteht aus einer Folge ebener Setzbäume, indem man alle diese Setzbäume (in der durch die Folge gegebenen Ordnung) an ihren Wurzeln identifiziert und dann an diesen Knoten eine neue Wurzel durch eine Kante anhängt. Auf diese Weise wird aus Setzbäumen mit m_1+2, m_2+2, … m_k+2 Knoten ein Setzbaum mit $m_1+\cdots+m_k+k+2$ Knoten. Das zeigt, das die erste Rekursion erfüllt ist.

Zu (5): Dieselbe Konstruktion wie für die ebenen Setzbäume zeigt hier, daß die zweite Rekursion erfüllt ist.

Zu (6): Wir fixieren eine Kante des $(n+2)$-Ecks. Dann zerlegt sich das $(n+2)$-Eck in das Dreieck, zu dem die fixierte Kante gehört, ein trianguliertes $(k+2)$-Eck links von diesem Dreieck und ein trianguliertes $(n-1-k+2)$-Eck. Wir sehen, daß die zweite Rekursion gilt.

Zu (7): Wir fixieren einen der $2n$ Punkte und betrachten die Sehne, die diesen Punkt enthält. Sie teilt die gegebene Figur in eine Sehnenfigur auf $2k$ Punkten auf der linken Seite und eine Sehnenfigur auf $2(n-1-k)$ Punkten auf der rechten Seite. Das zeigt, daß ebenfalls die zweite Rekursion erfüllt ist.

Der Satz ist damit bewiesen. □

Es ist auch möglich, die Mengen $\mathcal{A}_m(n)$ bijektiv aufeinander abzubilden. Hier sind in Kurzfassung einige Möglichkeiten.

$\mathcal{A}_3 \leftrightarrow \mathcal{A}_5$: Das ist sehr natürlich. Aus dem binären Baum bekommen wir ein vollständig geklammertes Produkt, indem wir die Blätter gemäß der Baumstruktur sukzessive zusammenfassen.

$\mathcal{A}_5 \leftrightarrow \mathcal{A}_6$: Aus der Triangulierung erhalten wir einen Binärbaum, indem wir innere Knoten den Dreiecken und Blätter den Seiten des Polygons zuordnen; eine Seite ist dabei fixiert und entspricht der Wurzel. Zwei innere Knoten werden miteinander verbunden, wenn die zugehörigen Dreiecke aneinanderstoßen; ein innerer Knoten wird mit einem Blatt verbunden, wenn das entsprechende Dreieck die dem Blatt zugeordnete Seite als Seite hat.

$\mathcal{A}_4 \leftrightarrow \mathcal{A}_7$: Aus der Sehnenfigur wird ein ebener Wurzelbaum, indem wir jedem der Gebiete, in die der Kreis geteilt wird, einen Knoten zuordnen; aneinandergrenzenden Gebieten entsprechen aneinandergrenzende Knoten. Ein Punkt (verschieden von den $2n$ gegebenen Punkten) auf der Kreislinie sei fixiert. Dann entspricht dem Gebiet, auf dessen Rand dieser Punkt wird, die Wurzel im Baum.

Einen Setzbaum erhalten wir, indem wir an dieser Wurzel eine neue Kante ansetzen, die zur neuen Wurzel führt.

Etwas weniger einsichtig ist es, wie man (zum Beispiel) die Bäume durch 0-1-Folgen kodieren kann.

$\mathcal{A}_2 \leftrightarrow \mathcal{A}_4$: Aus dem Setzbaum konstruieren wir eine 0-1-Folge, indem wir von der Wurzel ausgehend im Uhrzeigersinn um den Setzbaum herumlaufen. Für jeden Schritt von der Wurzel weg notieren wir eine 0, für jeden Schritt auf die Wurzel zu eine 1.

$\mathcal{A}_2 \leftrightarrow \mathcal{A}_5$: Wir laufen ebenfalls im Uhrzeigersinn um den Baum herum. Für jeden Schritt nach „links oben“ notieren wir eine 0, für jeden Schritt nach „rechts unten“ eine 1. Die Kante an der Wurzel soll dabei in beiden Fällen mitzählen.

In der Skizze vor dem Beweis stehen einander zugeordete Figuren untereinander. (Die fixierte Seite bei (6) ist dabei unten, und der fixierte Punkt bei (7) ist ebenfalls unten.)

4. Die Siebformel

Die Siebformel (auch *Ein-/Ausschaltformel* oder *Prinip von Inklusion und Exklusion* genannt) ist im Grunde eine Verallgemeinerung der Summenregel für Mächtigkeiten auf den Fall nicht notwendig disjunkter Mengen. Seien A_1, A_2 usw. Teilmengen einer Menge A. Wie man sich leicht an einem Mengendiagramm klarmacht, gilt dann

$$\begin{aligned}
\#(A \setminus (A_1 \cup A_2)) &= \#A - \#A_1 - \#A_2 + \#(A_1 \cap A_2) \\
\#(A \setminus (A_1 \cup A_2 \cup A_3)) &= \#A - \#A_1 - \#A_2 - \#A_3 \\
&\quad + \#(A_1 \cap A_2) + \#(A_1 \cap A_3) + \#(A_2 \cap A_3) \\
&\quad - \#(A_1 \cap A_2 \cap A_3)
\end{aligned}$$

Der folgende Satz sagt unter anderem, daß die offensichtliche Verallgemeinerung dieser Formeln richtig ist.

Zunächst führen wir aber noch eine abkürzende Schreibweise ein. Wenn I eine Indexmenge ist und A_i für $i \in I$ Teilmengen einer fest gegebenen Menge A sind, dann sei für eine Teilmenge $T \subset I$

$$A_T = \bigcap_{i \in T} A_i$$

mit dem Spezialfall $A_\emptyset = A$.

Satz 4.1. *Sei A eine endliche Menge mit Teilmengen $A_1, \dots, A_n$. Für $0 \le m \le n$ sei $A^{(m)} \subset A$ die Teilmenge, deren Elemente in genau m der Mengen A_j liegen. Dann gilt*

$$\begin{aligned}
\#A^{(m)} &= \sum_{T \subset \{1,\dots,n\}} (-1)^{\#T-m} \binom{\#T}{m} \#A_T \\
&= \sum_{k=m}^{n} (-1)^{k-m} \binom{k}{m} \sum_{1 \le j_1 < \dots < j_k \le n} \#A_{\{j_1,\dots,j_k\}} \,.
\end{aligned}$$

Wird die zweite Summe nach einem positiven (negativen) Term abgebrochen, so erhalten wir eine obere (untere) Abschätzung für $\#A^{(m)}$.

Der wichtigste Fall ist $m = 0$. Die Formel lautet dann einfach

$$\#(A \setminus \bigcup_{j=1}^{n} A_j) = \sum_{T \subset \{1,\dots,n\}} (-1)^{\#T} \#A_T \,.$$

Zum Beweis des Satzes benötigen wir ein paar Identitäten für Binomialkoeffizienten.

Lemma 4.2.

(1) $\displaystyle\sum_{k=0}^{s} (-1)^k \binom{r}{k} = (-1)^s \binom{r-1}{s}$ *für* $r \geq 1$.

(2) $\displaystyle\binom{k}{m}\binom{r}{k} = \binom{r}{m}\binom{r-m}{k-m}$.

(3) *Für* $s \geq m$ *gilt*

$$\sum_{k=m}^{s} (-1)^{k-m} \binom{k}{m}\binom{r}{k} = \begin{cases} (-1)^{s-m}\binom{r}{m}\binom{r-m-1}{s-m} & \text{falls } r > m \\ 1 & \text{falls } r = m \\ 0 & \text{falls } r < m \end{cases}$$

Beweis: (1) Induktion nach s. Klar für $s = 0$. Der Induktionsschritt ist nichts anderes als die übliche Rekursionsformel für Binomialkoeffizienten.
(2) Beide Seiten sind gleich $r!/(m!(k-m)!(r-k)!)$.
(3) Das folgt aus (1) und (2). □

Beweis (von Satz 4.1): Wir betrachten die Summe für $k = m$ bis $k = s$ (mit $m \leq s \leq n$) und bestimmen den Anteil, den $A^{(r)}$ dazu beisteuert. Ein Element, das in genau r der Teilmengen liegt, kommt offenbar in genau $\binom{r}{k}$ Durchschnitten von k der Teilmengen vor (denn diese k Teilmengen müssen aus den r Teilmengen ausgewählt werden, die das gegebene Element enthalten). Also ist der Beitrag von $A^{(r)}$ gerade

$$\#A^{(r)} \sum_{k=m}^{s} (-1)^{k-m} \binom{k}{m}\binom{r}{k} \,,$$

und die gesamte Summe $S(s)$ ist unter Verwendung von Lemma 4.2

$$\begin{aligned} S(s) &= \sum_{r=0}^{n} \#A^{(r)} \sum_{k=m}^{s} (-1)^{k-m} \binom{k}{m}\binom{r}{k} \\ &= \#A^{(m)} + (-1)^{s-m} \sum_{r=s+1}^{n} \binom{r}{m}\binom{r-m-1}{s-m} \#A^{(r)} \,; \end{aligned}$$

und wir haben $\#A^{(m)} \leq S(s)$, wenn $s-m$ gerade ist, $\#A^{(m)} \geq S(s)$, wenn $s-m$ ungerade ist, und $\#A^{(m)} = S(n)$. □

Kommen wir nun zu einigen Beispielen. Die Siebformel läßt sich immer dann gut anwenden, wenn es darauf ankommt, Objekte zu zählen, die keine (oder wenigstens eine) von einer Anzahl von Eigenschaften haben.

4.1. **Primzahlen bis 100.** Wie viele Primzahlen ≤ 100 gibt es?

Jede zusammengesetzte Zahl ≤ 100 hat einen Primteiler ≤ 7. Wir zählen also die Zahlen bis 100, die durch keine der Zahlen 2,3,5,7 teilbar sind. Dann müssen wir noch 4 addieren (für die Primzahlen 2,3,5,7) und eins abziehen (für die Nichtprimzahl 1). Sei also $A = \{1, 2, \ldots, 100\}$, und sei $A_n \subset A$ die Teilmenge der durch n teilbaren Zahlen. Dann gilt $\#A_n = \lfloor \frac{100}{n} \rfloor$ und $A_m \cap A_n = A_{mn}$, wenn m und n teilerfremd sind. Es folgt

$$\begin{aligned}
&\#(A \setminus (A_2 \cup A_3 \cup A_5 \cup A_7)) \\
&\quad = \#A - \#A_2 - \#A_3 - \#A_5 - \#A_7 \\
&\qquad + \#A_6 + \#A_{10} + \#A_{14} + \#A_{15} + \#A_{21} + \#A_{35} \\
&\qquad - \#A_{30} - \#A_{42} - \#A_{70} - \#A_{105} + \#A_{210} \\
&\quad = 100 - 50 - 33 - 20 - 14 + 16 + 10 + 7 + 6 + 4 + 2 - 3 - 2 - 1 - 0 + 0 \\
&\quad = 22\,,
\end{aligned}$$

also gibt es genau $22 + 4 - 1 = 25$ Primzahlen bis 100.

4.2. **Die Eulersche φ-Funktion.**

Sie ist definiert als

$$\varphi(n) = \#\{m \in \{1, 2, \ldots, n\} \mid \mathrm{ggT}(m, n) = 1\}\,.$$

Die zu zählenden Zahlen sind gerade die, die mit n keinen Primteiler gemeinsam haben. Sei also P die Menge der Primteiler von n. Wie eben sei wieder $A = \{1, 2, \ldots, n\}$, und für $d|n$ sei $A_d = \{d, 2d, \ldots, n\}$. Es gilt dann $\#A_d = n/d$, und wie eben haben wir

$$\varphi(n) = \sum_{T \subset P} (-1)^{\#T} \frac{n}{\prod\limits_{p \in T} p} = n \sum_{T \subset P} \prod_{p \in T} \frac{-1}{p} = n \prod_{p \in P} \left(1 - \frac{1}{p}\right)\,.$$

4.3. **Fixpunktfreie Permutationen.** Dieses Problem ist auch unter dem klassischen Namen „Problème des rencontres“ bekannt.

Gefragt ist nach der Anzahl der fixpunktfreien Permutationen von n Objekten, also der bijektiven Abbildungen $f : \{1, \ldots, n\} \longrightarrow \{1, \ldots, n\}$ mit $f(k) \neq k$ für alle k. (Ein Element k, so daß $f(k) = k$, heißt *Fixpunkt* von f.) Etwas allgemeiner können wir nach der Anzahl der Permutationen mit genau m Fixpunkten fragen. Zur Anwendung der Siebformel setzen wir $N = \{1, 2, \ldots, n\}$ und

$$A = \{f : N \to N \mid f \text{ bijektiv}\}\,, \qquad A_k = \{f \in A \mid f(k) = k\}\,.$$

Dann ist $\#A^{(m)}$ die Anzahl der Permutationen mit m Fixpunkten. Für $T \subset N$ gilt offenbar $\#A_T = (n - \#T)!$, da die Elemente des Durchschnitts in Bijektion mit den Permutationen von $N \setminus T$ stehen. Daraus ergibt sich

$$\#A^{(m)} = \sum_{k=m}^{n} (-1)^{k-m} \binom{k}{m} \binom{n}{k} (n-k)! = \frac{n!}{m!} \sum_{k=0}^{n-m} \frac{(-1)^k}{k!}\,.$$

Speziell haben wir für die Zahl der fixpunktfreien Permutationen

$$\#A^{(0)} = n! \sum_{k=0}^{n} \frac{(-1)^k}{k!} = \left\lfloor \frac{n!}{e} \right\rceil ,$$

wenn $n \geq 1$ (dabei sei $\lfloor x \rceil$ eine zu x am nächsten gelegene ganze Zahl). Eine zufällige Permutation ist also ziemlich genau mit Wahrscheinlichkeit $1/e = 0.367879\ldots$ fixpunktfrei.

4.4. **Surjektive Abbildungen und Äquivalenzrelationen.**

Seien M und N Mengen mit $\#M = m$ und $\#N = n$. Wir wollen die Anzahl der surjektiven Abbildungen von N auf M bestimmen. Dazu setzen wir $A = \{f : N \to M\}$, und für $x \in M$ setzen wir $A_x = \{f \in A \mid x \notin f(N)\}$. Für $T \subset M$ gilt dann $\#A_T = (m - \#T)^n$, denn A_T besteht gerade aus den Abbildungen von N in $M \setminus T$. Für die Anzahl der surjektiven Abbildungen ergibt sich also

$$\#A^{(0)} = \sum_{k=0}^{m} (-1)^k \binom{m}{k} (m-k)^n .$$

Da wir die Bilder beliebig permutieren können, muß diese Zahl durch $m!$ teilbar sein. Die Zahlen

$$S(n,m) = \frac{1}{m!} \sum_{k=0}^{m} (-1)^k \binom{m}{k} (m-k)^n = \sum_{k=0}^{m} \frac{(-1)^k (m-k)^n}{k!(m-k)!} \tag{4.1}$$

heißen *Stirling-Zahlen zweiter Art.* Die Zahl $S(n,m)$ ist die Anzahl der Partitionen einer n-elementigen Menge in m nichtleere Teilmengen (ohne Beachtung der Reihenfolge der Teilmengen) oder äquivalent die Anzahl der Äquivalenzrelationen auf einen n-elementigen Menge mit genau m Äquivalenzklassen. Die Gesamtzahl aller Äquivalenzrelationen auf einer n-elementigen Menge heißt *Bellsche Exponentialzahl*

$$B(n) = \sum_{m=0}^{n} S(n,m) .$$

Übungsaufgaben 4.3.

(1) Zeigen Sie folgende Identitäten für die Stirling-Zahlen 2. Art und die Bell-Zahlen.
   (a) $S(0,0) = 1$, $S(0,m) = 0$ für $m > 0$, $S(n,0) = 0$ für $n > 0$.
   (b) $S(n+1,m+1) = S(n,m) + (m+1)S(n,m+1)$.
   (c) $S(n+1,m+1) = \sum_{k=0}^{n} \binom{n}{k} S(k,m)$.
   (d) $B(0) = 1$, $B(n+1) = \sum_{k=0}^{n} \binom{n}{k} B(k)$.

(2) Von 20 Leuten gehört jeder wenigstens einem von fünf verschiedenen Vereinen an. Jeder Verein hat 10 Mitglieder. Zeigen Sie, daß es zwei Vereine gibt, die wenigstens drei Mitglieder gemeinsam haben. Läßt sich diese Aussage noch verbessern?

(3) n Paare von identischen Karten werden in einer Reihe ausgelegt. Wie groß ist die Wahrscheinlichkeit, daß kein Paar nebeneinander zu liegen kommt?

4.5. **Lösung der Übungsaufgaben.**

AUFGABE (1): Teil (a) läßt sich entweder durch Rückgriff auf die Formel (4.1) lösen oder indem man auf die kombinatorische Interpretation zurückgreift: Auf der leeren Menge gibt es genau eine (nämlich die leere) Äquivalenzrelation; sie hat 0 Äquivalenzklassen. Also ist $S(0,0) = 1$ und $S(0,m) = 0$ für $m > 0$. Jede Äquivalenzrelation auf einer nichtleeren Menge muß wenigstens eine Äquivalenzklasse haben, also gilt $S(n,0) = 0$ für $n > 0$. Da eine n-elementige Menge nicht in mehr als n Klassen zerfallen kann, gilt übrigens auch

$$m!\,S(n,m) = \sum_{k=0}^{m} (-1)^k \binom{m}{k} (m-k)^n = 0$$

für $m > n$.

Zu Teil (b): Sei A eine n-elementige Menge und a ein weiteres Element. Wenn ich $A' = A \cup \{a\}$ in $m+1$ Klassen zerlegen will, dann kann ich entweder a alleine in eine Klasse stecken; das ergibt dann die $S(n,m)$ Möglichkeiten, A in m Klassen aufzuteilen. Oder ich teile A in $m+1$ Klassen auf; dann kann ich a zu einer dieser Klassen dazunehmen, also gibt es dafür insgesamt $(m+1)S(n,m+1)$ Möglichkeiten.

Zu Teil (c): Wir betrachten die Klasse, in der a landet. Sei $n+1-k$ ihre Größe. Es gibt $\binom{n}{n-k} = \binom{n}{k}$ Möglichkeiten, die übrigen $n-k$ Elemente aus A auszuwählen; dann muß noch die verbleibende k-elementige Menge in m Klassen eingeteilt werden.

Teil (d) folgt aus (a) und (c).

AUFGABE (2): Sei A die Menge der 20 Leute und seien A_j $(j = 1,\dots,5)$ die Vereine. Nach Voraussetzung gilt $\#A = 20$ und $\#A_j = 10$. Die Siebformel liefert

$$0 \le \#A^{(0)} \le \#A - \sum_j \#A_j + \sum_{j<k} \#(A_j \cap A_k)\,,$$

also

$$\sum_{j<k} \#(A_j \cap A_k) \ge -20 + 5 \cdot 10 = 30\,.$$

Da es 10 verschiedene Durchschnitte von je zwei Vereinen gibt, muß einer wenigstens drei Elemente enthalten.

Tatsächlich muß es sogar einen Durchschnitt mit vier Elementen geben. Dazu benutzen wir die allgemeine Version der Siebformel. Wir setzen

$$a_k = \sum_{\#T=k} \#A_T\,,$$

also $a_0 = \#A = 20$, $a_1 = \sum_j \#A_j = 50$. Wir wollen a_2 nach unten abschätzen. Nach der Siebformel gilt

$$\begin{array}{lcccccccccc}
0 \le \#A^{(0)} & = & a_0 & - & a_1 & + & a_2 & - & a_3 & + & a_4 & - & a_5 \\
0 \le \#A^{(1)} & = & & & a_1 & - & 2a_2 & + & 3a_3 & - & 4a_4 & + & 5a_5 \\
0 \le \#A^{(4)} & = & & & & & & & & & a_4 & - & 5a_5 \\
0 \le \#A^{(5)} & = & & & & & & & & & & & a_5
\end{array}$$

Wenn wir diese Ungleichungen mit 3, 1, 1, 3 multiplizieren und addieren, erhalten wir

$$3a_0 - 2a_1 + a_2 \geq 0\,, \qquad \text{also} \qquad a_2 \geq 40\,.$$

Diese Schranke ist optimal, denn man kann die Leute so auf die Vereine verteilen, daß jeder Durchschnitt zweier Vereine genau vier Elemente hat.

Der gegebene Beweis funktioniert auch noch, wenn alle Zahlen mit einer Konstante multipliziert werden (2000 Leute, 5 Vereine mit je 1000 Mitgliedern, dann gibt es zwei Vereine mit mindestens 400 gemeinsamen Mitgliedern) oder im Grenzfall für Maße von Mengen (5 meßbare Teilmengen des Einheitsquadrats, jede mit der Fläche $\frac{1}{2}$; dann gibt es zwei, deren Durchschnitt mindestens Fläche $\frac{1}{5}$ hat). Für die gegebene Aufgabe gibt es noch eine einfachere Lösung wie folgt. Wir nehmen an, für alle Durchschnitte gelte $\#(A_j \cap A_k) \leq 3$. Dann gilt (ebenfalls nach der Siebformel)

$$\begin{aligned}\#(A_j \setminus (A_1 \cup A_2 \cup \cdots \cup A_{j-1})) &\geq \#A_j - \#(A_1 \cap A_j) - \cdots - \#(A_{j-1} \cap A_j) \\ &\geq 10 - 3(j-1)\,.\end{aligned}$$

Also haben wir den Widerspruch

$$20 = \#A \geq \#A_1 + \#(A_2 \setminus A_1) + \#(A_3 \setminus (A_1 \cup A_2)) \geq 10 + 7 + 4 = 21\,.$$

Aufgabe (3): Wir müssen uns zunächst überlegen, wie viele Möglichkeiten es ohne die Nebenbedingung gibt. Wenn wir nur darauf achten, welche Karten gleich und welche verschieden sind, dann hat unsere Grundmenge A die Mächtigkeit $\frac{(2n)!}{2^n n!}$. (Wenn wir uns alle $2n$ Karten als verschieden vorstellen, gibt es $(2n)!$ Möglichkeiten. Da die beiden Karten in jedem Paar ununterscheidbar sind, müssen wir durch 2^n teilen. Da es uns auf die Reihenfolge der Paare nicht ankommt, müssen wir noch durch $n!$ teilen.)

Als Ausschlußmengen nehmen wir A_j als die Teilmenge von Konfigurationen, bei denen auf den Plätzen j und $j+1$ zwei gleiche Karten liegen ($j = 1, 2, \ldots, 2n-1$). Für eine Teilmenge $T \subset \{1, 2, \ldots, 2n-1\}$ gilt offenbar $A_T = \emptyset$ genau dann, wenn T zwei aufeinanderfolgende Zahlen enthält. Im anderen Fall ist $\#A_T = \frac{(2n-2k)!}{2^{n-k}(n-k)!}$, wenn $\#T = k$. Wir müssen also noch die Anzahl der k-elementigen Teilmengen von $\{1, 2, \ldots, 2n-1\}$ bestimmen, die keine aufeinanderfolgenden Zahlen enthalten.

Dazu ordnen wir die Elemente von T aufsteigend an:

$$j_1 < j_2 < \cdots < j_k\,.$$

Die Bedingung ist dann $j_2 > j_1 + 1$, $j_3 > j_2 + 1$, … , $j_k > j_{k-1} + 1$. Wir können also eine neue Menge bilden:

$$T' = \{j_1, j_2 - 1, j_3 - 2, \ldots, j_k - (k-1)\} \subset \{1, 2, \ldots, 2n-k\}\,.$$

Da sich diese Konstruktion umkehren läßt, haben wir gezeigt, daß die gesuchte Zahl gleich der Zahl der k-Teilmengen einer $(2n-k)$-elementigen Menge ist, also gleich $\binom{2n-k}{k}$.

Jetzt können wir die Siebformel anwenden, um die gesuchte Wahrscheinlichkeit zu berechnen.

$$\begin{aligned} w_n &= \frac{\#A^{(0)}}{\#A} \\ &= \frac{2^n n!}{(2n)!} \sum_{k=0}^{n} (-1)^k \binom{2n-k}{k} \frac{(2n-2k)!}{2^{n-k}(n-k)!} \\ &= \sum_{k=0}^{n} \frac{(-1)^k}{k!} \frac{(2n-k)!}{(2n)!} \frac{2^k n!}{(n-k)!} \\ &= \sum_{k=0}^{n} \frac{(-1)^k}{k!} \frac{(2n)(2n-2)(2n-4)\dots(2n-2k+2)}{(2n)(2n-1)(2n-2)\dots(2n-k+1)} \end{aligned}$$

Die ersten paar Werte

$$w_0 = 1, \quad w_1 = 0, \quad w_2 = w_3 = \tfrac{1}{3}, \quad w_4 = \tfrac{12}{35}, \quad w_5 = \tfrac{47}{135}, \quad \dots$$

deuten darauf hin, daß w_n einen Grenzwert hat. Tatsächlich gilt

Proposition 4.4.

$$\lim_{n\to\infty} w_n = e^{-1} = 0.367879\dots \ .$$

Zum Beweis benutzen wir ein etwas allgemeineres Lemma, das wir zunächst formulieren und beweisen wollen.

Lemma 4.5. *In der Siebformel (Satz 4.1) seien die Mengen A und A_j von einem Parameter n abhängig: $A(n)$ usw. Wir bezeichnen mit $a(n,k)$ den k-ten Term*

$$a(n,k) = \sum_{\#T=k} \#A(n)_T$$

in der Siebformel.

Wenn für jedes k der Grenzwert

$$\lim_{n\to\infty} \frac{a(n,k)}{\#A(n)} =: a_k$$

existiert, dann gilt für alle $m \geq 0$, für die die Summe auf der rechten Seite konvergiert,

$$\lim_{n\to\infty} \frac{\#A(n)^{(m)}}{\#A(n)} = \sum_{k=m}^{\infty} (-1)^{k-m} \binom{k}{m} a_k \,. \tag{4.2}$$

Beweis: Nach der Siebformel gilt

$$\frac{\#A(n)^{(m)}}{\#A(n)} \begin{cases} \leq \displaystyle\sum_{k=m}^{m+2s} (-1)^{k-m} \binom{k}{m} \frac{a(n,k)}{\#A(n)} \\ \geq \displaystyle\sum_{k=m}^{m+2s+1} (-1)^{k-m} \binom{k}{m} \frac{a(n,k)}{\#A(n)} \end{cases}$$

Es folgt

$$\limsup_{n\to\infty} \frac{\#A(n)^{(m)}}{\#A(n)} \le \sum_{k=m}^{m+2s} (-1)^{k-m} \binom{k}{m} a_k$$
$$\liminf_{n\to\infty} \frac{\#A(n)^{(m)}}{\#A(n)} \ge \sum_{k=m}^{m+2s+1} (-1)^{k-m} \binom{k}{m} a_k \,.$$

Wenn die Summe auf der rechten Seite von (4.2) konvergiert, dann folgt die Behauptung, wenn wir $s \to \infty$ gehen lassen. □

Beweis (von Prop. 4.4): Hier haben wir (mit den Bezeichnungen des Lemmas)

$$\frac{a(n,k)}{\#A(n)} = \frac{1}{k!} \frac{(2n)(2n-2)(2n-4)\dots(2n-2k+2)}{(2n)(2n-1)(2n-2)\dots(2n-k+1)} \,,$$

also

$$\lim_{n\to\infty} \frac{a(n,k)}{\#A(n)} = \frac{1}{k!} \,.$$

Aus dem Lemma folgt dann

$$\lim_{n\to\infty} \frac{\#A(n)^{(m)}}{\#A(n)} = \sum_{k=m}^{\infty} \frac{(-1)^{k-m}}{k!} \binom{k}{m} = \frac{e^{-1}}{m!} \,;$$

der Fall $m = 0$ gibt die Behauptung. □

Man könnte auch an folgenden „Beweis“ denken:

Für ein gegebenes Paar ist die Wahrscheinlichkeit, daß es nebeneinander zu liegen kommt, gleich

$$\frac{1}{2n-1} \binom{2n}{2} = \frac{1}{n} \,.$$

Die Wahrscheinlichkeit, daß alle n Paare nicht zusammenkommen, ist also

$$\left(1 - \frac{1}{n}\right)^n \to e^{-1} \,.$$

Der Fehler in dieser Argumentation ist, daß implizit angenommen wird, daß die n Paare bezüglich des Zusammentreffens voneinander unabhängig sind. Das ist natürlich falsch (das sieht man auch daran, daß w_n für $n = 2, 3, \dots$ von $(1 - \frac{1}{n})^n$ verschieden ist). Das Ergebnis bleibt aber richtig, was daran liegt, daß sich die Paare doch annähernd unabhängig verhalten, wenn man nicht zu viele (relativ zu n) gleichzeitig betrachtet.

4.6. **Memory.**

Es ist Ihnen sicher auch schon aufgefallen, daß beim Memory-Spiel sehr häufig ein Paar gleicher Karten nebeneinander zu liegen kommt, auch wenn die Karten sehr gut gemischt waren. Wir wollen hier die Wahrscheinlichkeit (für viele Karten) dafür bestimmen, daß das nicht passiert. Diese Frage ist das zwiedimensionale Analogon zu Aufgabe (3). Der Einfachheit halber wollen wir annehmen, daß die Karten in einem Quadrat ausgelegt werden (das Ergebnis bleibt richtig, wenn man Rechtecke betrachtet, deren beide Seitenlängen beliebig groß werden). Wir betrachten also $N = 2n^2$ Paare von Karten, die in einem $(2n) \times (2n)$-Quadrat

ausgelegt werden. Es sei W_n die Wahrscheinlichkeit, daß dabei kein Paar von Karten zusammen kommt. Wir wollen den Grenzwert $\lim_{n\to\infty} W_n$ bestimmen.

Wir gehen so vor wie für Aufgabe (3). Die Menge $A(n)$ hat die Mächtigkeit $\#A(n) = \frac{(2N)!}{2^N N!}$. Die Möglichkeiten, einen Dominostein (oder ein aneinanderliegendes Kartenpaar) auf dem Quadrat zu plazieren, bilden eine Indexmenge $I(n)$, die die Ausnahmemengen $A_i(n) \subset A(n)$ indiziert. Mit den Bezeichnungen des Lemmas haben wir dann offenbar

$$a(n,k) = b(n,k)\frac{(2N-2k)!}{2^{N-k}(N-k)!}\,,$$

wo $b(n,k)$ die Möglichkeiten zählt, k Dominosteine auf dem Quadrat unterzubringen (ohne daß sie sich überschneiden). Die Anzahlen $b(n,k)$ zu bestimmen ist sicher sehr schwierig, aber das ist auch gar nicht nötig. Für unsere Zwecke genügt folgende Abschätzung.

Lemma 4.6. *Sei $m = 4n(2n-1) = 4(N-n)$. Für $0 \le k \le \frac{1}{7}m$ gilt*

$$\frac{m(m-7)(m-14)\dots(m-7(k-1))}{k!} = 7^k\binom{m/7}{k} \le b(n,k) \le \binom{m}{k}.$$

Insbesondere haben wir

$$\lim_{n\to\infty}\frac{b(n,k)}{(2N)^k} = \frac{2^k}{k!}$$

für alle $k \ge 0$.

Beweis: Die Zahl m ist die Größe von $I(n)$, also die Anzahl der Möglichkeiten, einen Dominostein unterzubringen. Wir stellen uns vor, die k Dominosteine der Reihe nach auf das Quadrat zu legen. Jedesmal, wenn wir einen Stein plazieren, verringern wir die Anzahl der Möglichkeiten für weitere Steine. Diese Anzahl verringert sich mindestens um 1 (die Position, die der gerade abgelegte Stein einnimmt) und höchstens um 7 (die Positionen, die „verboten“ werden, wenn der Stein nicht am Rand liegt und nicht an einen anderen Stein angrenzt). Daraus folgt die Behauptung. (Der Term $k!$ im Nenner kommt daher, daß die Reihenfolge der Dominosteine irrelevant ist.)

Die Limesformel folgt, wenn man $\lim_{n\to\infty}\frac{m}{2N} = 2$ beachtet. □

Nun ist es nicht mehr schwer. Wir haben

$$\lim_{n\to\infty}\frac{a(n,k)}{\#A(n)} = \lim_{n\to\infty}\frac{b(n,k)}{(2N)^k}\,\frac{2^k(2N)^k(2N-2k)!N!}{(2N)!(N-k)!} = \frac{2^k}{k!}\,.$$

Mit Lemma 4.5 folgt dann

$$\lim_{n\to\infty}\frac{\#A(n)^{(m)}}{\#A(n)} = \frac{2^m}{m!}\,e^{-2} \tag{4.3}$$

und speziell

$$\lim_{n\to\infty} W_n = e^{-2} = 0.135335\dots$$

Die Wahrscheinlichkeit für ein zusammenliegendes Paar nähert sich also

$$1 - e^{-2} = 0.8646647\dots\,,$$

was die zu Beginn angesprochene Beobachtung erklärt.

Formel (4.3) sagt, daß die Wahrscheinlichkeitsverteilung der Anzahl nebeneinanderliegender Paare sich einer *Poissonverteilung* mit Mittelwert 2 nähert. Man wird also durchschnittlich zwei Paare finden, die unmittelbar zusammen liegen.

5. Abzählen von Bahnen

Bei Abzählproblemen tritt oft der Fall ein, daß man gewisse Objekte als ununterscheidbar betrachtet und deshalb insgesamt nur einmal zählen möchte. Meistens läßt sich diese Situation durch die Operation einer (endlichen) Gruppe G auf der (endlichen) Menge X der Objekte beschreiben, wobei zwei Objekte nicht unterschieden werden, wenn sie durch diese Operation ineinander übergeführt werden (also in derselben Bahn liegen). Das Problem läßt sich also allgemein so formulieren.

> *Eine endliche Gruppe G operiere (von links) auf einer endlichen Menge X. Wie groß ist der Bahnenraum $G\backslash X$? (D.h., wie viele Bahnen hat diese Operation?)*

Wir wollen zunächst einen einfachen Spezialfall formulieren (den wir auch schon häufig implizit angewendet haben).

Lemma 5.1. *Die Gruppe G operiere auf X mit trivialen Stabilisatoren (d.h. für jedes $x \in X$ ist der Stabilisator $G_x = \{g \in G \mid g \cdot x = x\}$ trivial). Dann gilt*

$$\#(G\backslash X) = \frac{\#X}{\#G}\,.$$

Beweis: Nach der grundlegenden Formel $\#G = \#(G\cdot x)\#G_x$ gilt hier, daß jede Bahn $G \cdot x$ so groß ist wie G. □

Eine ähnliche Aussage gilt, wenn etwas allgemeiner alle Stabilisatoren G_x dieselbe Mächtigkeit haben.

Normalerweise kann man aber nicht erwarten, daß alle Stabilisatoren gleich groß sind. Den allgemeinen Fall behandelt das folgende Resultat, das unter dem Namen „Lemma von Burnside“ bekannt ist, jedoch zuerst von Cauchy und von Frobenius bewiesen wurde.

Proposition 5.2. *Die endliche Gruppe G operiere auf der endlichen Menge X. Dann gilt*

$$\#(G\backslash X) = \frac{1}{\#G}\sum_{g\in G} \#\{x \in X \mid g \cdot x = x\}\,.$$

In Worten: Die Größe des Bahnenraums ist gleich der durchschnittlichen Anzahl von Fixpunkten in X der Elemente von G.

Beweis: Für den Beweis zählen wir die Menge

$$M = \{(g, x) \in G \times X \mid g \cdot x = x\}$$

zweimal ab. Zum einen haben wir

$$\#M = \sum_{g\in G} \#\{x \in X \mid (g, x) \in M\} = \sum_{g\in G} \#\{x \in X \mid g \cdot x = x\}\,,$$

auf der anderen Seite gilt

$$\begin{aligned}\#M &= \sum_{x\in X} \#\{g\in G \mid (g,x)\in M\} = \sum_{x\in X} \#G_x = \sum_{x\in X} \frac{\#G}{\#(G\cdot x)} \\ &= \#G \sum_{B\in G\backslash X} \sum_{x\in B} \frac{1}{\#B} = \#G \sum_{B\in G\backslash X} 1 = \#G\#(G\backslash X)\,.\end{aligned}$$

Daraus folgt die Behauptung. □

Es folgen einige Anwendungsbeispiele.

5.1. **Der kleine Satz von Fermat.**

Dieser sagt bekanntlich, daß für alle Primzahlen p und alle ganzen Zahlen a die Zahl $a^p - a$ von p geteilt wird. Für den folgenden Beweis nehmen wir an, daß a positiv ist. (Da die Aussage des Satzes nur von der Restklasse von a modulo p abhängt, ist das keine Einschränkung.) Anschaulich stellen wir uns Halsketten (oder Armbänder) vor, die aus p Perlen von (bis zu) a verschiedenen Sorten bestehen. Auf diesen Objekten operiert die zyklische Gruppe C_p der Ordnung p durch Rotation. Etwas formaler können wir als Menge X die Menge der Abbildungen $C_p \to \{1,2,\ldots,a\}$ betrachten, auf der C_p in natürlicher Weise operiert. Wir müssen nun die Fixpunkte abzählen. Das Nullelement von C_p hat offensichtlich alle Elemente von X als Fixpunkte. Für alle anderen Elemente gilt, daß sie die Gruppe C_p erzeugen. Da C_p transitiv auf sich selbst operiert, bedeutet das, daß die Fixpunkte genau die konstanten Abbildungen sind (anschaulich: Halsketten aus nur einer Sorte Perlen). Also gilt

$$\mathbb{Z} \ni \#(C_p\backslash X) = \frac{a^p + (p-1)a}{p} = a + \frac{a^p - a}{p}\,,$$

womit der Satz bewiesen ist.

5.2. **Färbungen des Würfels.**

Wir wollen uns überlegen, auf wie viele Weisen man die Seitenflächen eines Würfels schwarz und weiß färben kann (jede Fläche soll natürlich einfarbig sein). Dabei sollen Färbungen, die durch eine Drehung des Würfels ineinander übergehen, nicht unterschieden werden.

Sei also G die Symmetriegruppe des Würfels. Wir müssen uns einen Überblick über die Elemente von G und ihre Operation auf der Menge S der Seiten des Würfels verschaffen. G hat insgesamt 24 Elemente (als abstrakte Gruppe ist G isomorph zur symmetrischen Gruppe S_4), die sich wie folgt einteilen lassen.

(a) Die Identität. Sie hat 6 Bahnen auf S.
(b) 6 Drehungen um 180°, deren Drehachse zwei Kantenmittelpunkte verbindet. Sie haben jeweils 3 Bahnen auf S.
(c) 8 Drehungen um ±120°, deren Drehachse zwei Ecken verbindet. Sie haben jeweils 2 Bahnen auf S.
(d) 6 Drehungen um ±90°, deren Drehachse zwei Flächenmittelpunkte verbindet. Sie haben jeweils 3 Bahnen auf S.
(e) 3 Drehungen um 180°, deren Drehachse zwei Flächenmittelpunkte verbindet. Sie haben jeweils 4 Bahnen auf S.

Die möglichen Färbungen lassen sich interpretieren als die Abbildungen $S \to \{\text{schwarz}, \text{weiß}\}$. Auf dieser Menge X operiert G durch die Operation auf S wie folgt. Für eine Abbildung $f \in X$ gilt

$$(g \cdot f)(s) = f(g^{-1} \cdot s)\,.$$

(Die auf den ersten Blick vielleicht etwas seltsam wirkende Bildung des Inversen ist notwendig, um wieder eine Operation von links zu bekommen. Das merkt man, wenn man versucht nachzurechnen, daß $g_1 \cdot (g_2 \cdot f) = (g_1 g_2) \cdot f$ gilt.)

Sei nun $g \in G$ eine beliebige Drehung. Wie viele Fixpunkte hat g auf X? Wenn eine Färbung von g fixiert werden soll, dann müssen offenbar Seiten, die von g ineinander übergeführt werden, dieselbe Farbe haben. Anders ausgedrückt: Eine Färbung ist genau dann Fixpunkt von g, wenn sie auf den Bahnen von g auf S konstant ist. Das zeigt

$$\#\{x \in X \mid g \cdot x = x\} = 2^{\#(\langle g\rangle\backslash S)}\,.$$

(Im Exponenten steht die Anzahl der Bahnen von g (oder äquivalent: der von g erzeugten Untergruppe $\langle g\rangle \subset G$) auf S.)

Aus der Aufzählung der Elemente von G und der Anzahl ihrer Bahnen auf S ergibt sich daher für die Anzahl der Färbungen

$$\#(G\backslash X) = \frac{1}{24}(2^6 + 6\cdot 2^3 + 8\cdot 2^2 + 6\cdot 2^3 + 3\cdot 2^4) = 10\,.$$

Wie ist es nun, wenn wir den Würfel nicht schwarz-weiß, sondern mit einer beliebigen Anzahl r von Farben anmalen möchten?

Unsere Argumentation bleibt gültig; es gilt also immer noch, daß die von $g \in G$ fixierten Färbungen gerade die sind, die auf den Bahnen von g auf S konstant sind. Damit folgt (wir müssen einfach 2^e durch r^e ersetzen) für die Anzahl $a(r)$ der Färbungen mit r Farben:

$$a(r) = \frac{1}{24}(r^6 + 3\,r^4 + 12\,r^3 + 8\,r^2)\,.$$

5.3. **Verallgemeinerung.**

Die Argumentation, die wir beim Würfel benutzt haben, läßt sich ganz allgemein in folgender Situation anwenden. Wir haben eine Gruppe G, die auf einer Menge S operiert, und wir wollen die Bahnen von G auf der Menge X der Abbildungen von S nach F abzählen. Wie oben ergibt sich, daß die Fixpunkte von g gerade die Abbildungen sind, die auf den Bahnen von g auf S konstant sind. Das führt zu der Formel

$$\#(G\backslash X) = \frac{1}{\#G}\sum_{g\in G}(\#F)^{\#(\langle g\rangle\backslash S)}\,.$$

Bemerkung 5.3. Formeln wie diese stehen am Anfang einer relativ weit ausgebauten Methode (der sogenannten *Polyaschen Abzähltheorie*) zum Abzählen von Bahnenräumen unter verschiedenen Arten von Operationen. Um diese Methode zu verstehen, braucht man allerdings gute Kenntnisse über erzeugende Funktionen. Der zweite Teil dieser Vorlesung wird eine Einführung in erzeugende Funktionen bringen.

6. Graphen und Bäume. Der Satz von Cayley

In diesem Kapitel soll nun endlich (wie in der Vorlesungsankündigung versprochen) der Satz von Cayley über die Anzahl der markierten Bäume bewiesen werden. Wir werden gleich drei verschiedene Beweise besprechen. Zwei davon (einer mittels Rekursion/Induktion, der andere mittels Bijektion) werden dabei recht ausführlich behandelt; der dritte besteht darin, das Ergebnis auf einen anderen Satz zurückzuführen, den wir hier nicht beweisen wollen.

Zuerst brauchen wir aber einige Definitionen, damit wir überhaupt wissen, worüber wir reden.

Definition 6.1.

(1) Ein *(einfacher) ungerichteter (schlingenloser) Graph* Γ besteht aus einer (endlichen) Menge E von *Ecken* (oder *Knoten*, engl. vertex/vertices) und einer Teilmenge K der Menge der zweielementigen Teilmengen von E; die Elemente von K werden *Kanten* (engl. edge(s)) genannt. Wir schreiben $\Gamma = (E, K)$ und $E = E_\Gamma$, $K = K_\Gamma$.
Ist $k = \{a, b\} \in K$ eine Kante, so nennt man $a, b \in E$ die *Endpunkte* von k. Der *Grad* (auch die Valenz genannt) $\deg(e)$ einer Ecke $e \in E$ ist die Anzahl der Kanten, die e als Endpunkt haben.
(2) Ein *(einfacher) gerichteter Graph* Γ besteht aus einer (endlichen) Menge E von *Ecken* und einer Menge $K \subset E \times E$ (deren Elemente wieder *Kanten* heißen) von geordneten Paaren von Ecken. Wir schreiben wieder $\Gamma = (E, K)$ und $E = E_\Gamma$, $K = K_\Gamma$.
Ist $k = (a, b) \in K$ eine Kante, so nennt man $a \in E$ den *Anfangspunkt* und $b \in E$ den *Endpunkt* von k.
(3) Ein Graph $\Gamma_1 = (E_1, K_1)$ heiß *Teilgraph* von $\Gamma = (E, K)$, wenn $E_1 \subset E$ und $K_1 \subset K$ gilt. Der Teilgraph heißt *voll*, wenn K_1 aus allen Kanten in K besteht, deren Endpunkte (bzw. Anfangs- und Endpunkt) in E_1 liegen.
(4) Der Graph $K_n = (\{1, 2, \ldots, n\}, \{\{i, j\} \mid 1 \leq i < j \leq n\})$ heißt der *vollständige Graph* auf n Ecken.

Wenn im Folgenden einfach von einem Graphen die Rede ist, dann ist stets ein ungerichteter Graph gemeint. Man veranschaulicht einen Graphen meist durch eine Zeichnung, in der die Ecken durch Punkte und die Kanten durch Strecken oder Bögen dargestellt werden, die die Endpunkte der Kante miteinander verbinden. Bei orientierten Graphen gibt man die Richtung durch einen Pfeil an.

Definition 6.2.

(1) Sei $\Gamma = (E, K)$ ein (ungerichteter) Graph. Eine Folge
$$\gamma = e_0, k_1, e_1, k_2, e_2, \ldots, e_{n-1}, k_n, e_n$$
mit $e_0, e_1, \ldots, e_n \in E$ und $k_1, k_2, \ldots, k_n \in K$, so daß $k_m = \{e_{m-1}, e_m\}$ für alle $m = 1, 2, \ldots, n$ und $k_m \neq k_{m+1}$ für alle $m = 1, 2, \ldots, n-1$ gilt, heißt ein *Weg* in Γ von e_0 (dem *Anfangspunkt* von γ) nach e_n (dem *Endpunkt* von γ). n heißt die *Länge* von γ.
(2) Sei $\Gamma = (E, K)$ ein gerichteter Graph. Eine Folge
$$\gamma = e_0, k_1, e_1, k_2, e_2, \ldots, e_{n-1}, k_n, e_n$$

mit $e_0, e_1, \dots, e_n \in E$ und $k_1, k_2, \dots, k_n \in K$, so daß $k_m = (e_{m-1}, e_m)$ für alle $m = 1, 2, \dots, n$ und $k_m \neq k_{m+1}$ für alle $m = 1, 2, \dots, n-1$ gilt, heißt ein *Weg* in Γ von e_0 (dem *Anfangspunkt* von γ) nach e_n (dem *Endpunkt* von γ). n heißt die *Länge* von γ.

(3) Ein Weg positiver Länge, dessen Anfangs- und Endpunkt übereinstimmen, heißt ein *Kreis*.
(4) Ein (ungerichteter) Graph heißt *zusammenhängend*, wenn zwischen je zwei seiner Ecken ein Weg existiert.
(5) Ein zusammenhängender Graph mit mindestens einer Ecke, der keine Kreise enthält, heißt ein *Baum*.
(6) Sei Γ ein Graph. Ein Teilgraph B von Γ heißt *aufspannender Baum* von Γ, wenn $E_B = E_\Gamma$ gilt und B ein Baum ist.

Jetzt wissen wir also endlich, was ein Baum ist. Ein Topologe würde sagen, ein Baum sei ein zusammenhängender und einfach zusammenhängender Graph. Wir werden gleich sehen, daß Bäume gerade minimale zusammenhängende Graphen sind (bei vorgegebener Eckenmenge). Zuerst brauchen wir aber ein Lemma, das zeigt, wie man Bäume schrittweise auf- oder abbauen kann.

Lemma 6.3. *Sei $B = (E, K)$ ein Baum mit $\#E \geq 2$. Dann gilt:*

(a) *Es gibt eine Ecke $e \in E$ mit* $\deg(e) = 1$.
(b) *Sei $k \in K$ die (einzige) Kante mit e als Endpunkt. Dann ist der Graph $B' = (E \setminus \{e\}, K \setminus \{k\})$ ebenfalls ein Baum.*

Beweis: (a): Da B nach Voraussetzung zusammenhängend ist und es mindenstens zwei Ecken gibt, hat jede Ecke mindestens Grad 1. (Eine Ecke vom Grad 0 wäre isoliert — es gäbe keinen Weg von dieser Ecke zu einer anderen.) Wir nehmen also an, all Ecken hätten Grad mindestens 2. Dann können wir, ausgehend von einer beliebigen Ecke $e_0 \in E$, einen unendlich langen Weg

$$e_0, k_1, e_1, k_2, e_2, \dots$$

konstruieren — da es an jeder Ecke, an die wir kommen, stets noch wenigstens eine andere Kante gibt als die, längs der wir gekommen sind, läßt sich der Weg stets fortsetzen. Weil es aber nur endlich viele Ecken gibt, muß es Zahlen $0 \leq m < n$ geben, so daß $e_m = e_n$ ist. Dann ist aber das Teilstück von e_m bis e_n unseres Weges ein Kreis, im Widerspruch zur Voraussetzung, daß B ein Baum ist.
(b): Wenn B' einen Kreis enthielte, wäre das auch ein Kreis in B. Da B ein Baum ist, kann also auch B' keinen Kreis enthalten. Es bleibt zu zeigen, daß B' zusammenhängend ist. Dazu bemerken wir, daß jeder Weg in B, der die Ecke e enthält, dort beginnen oder enden muß (denn nach der Definition ist $\dots, k, e, k, \dots$ in einem Weg verboten). Seien nun $e_1, e_2 \in E \setminus \{e\}$ zwei Ecken von B'. Da B zusammenhängend ist, gibt es in B einen Weg von e_1 nach e_2. Da dieser Weg in e weder beginnt noch endet, kommt e und damit auch k in diesem Weg nicht vor. Der Weg ist also auch ein Weg in B'. □

Dieses Lemma ist deswegen wichtig, weil es uns Induktionsbeweise für Aussagen über Bäume ermöglicht. Das folgende Lemma gibt ein Beispiel.

Lemma 6.4. *Sei Γ ein Graph mit $n \geq 1$ Ecken.*

(a) *Ist* Γ *ein Baum, so hat* Γ *genau* $n-1$ *Kanten.*
(b) *Ist* Γ *zusammenhängend, so hat* Γ *mindestens* $n-1$ *Kanten.*
(c) *Ist* Γ *zusammenhängend mit* $n-1$ *Kanten, dann ist* Γ *ein Baum.*

BEWEIS: (a): Durch Induktion nach n. Der Fall $n = 1$ ist klar. Sei also jetzt $B = (E, K)$ ein Baum mit $n + 1 \geq 2$ Ecken, und sei $B' = (E', K')$ zu B gemäß Lemma 6.3 (b) gebildet. Dann ist B' ein Baum mit n Ecken; nach Induktionsvoraussetzung gilt dann also $\#K' = n - 1$. Es folgt $\#K = \#K' + 1 = n = (n + 1) - 1 = \#E - 1$.
(b): Sei Γ zusammenhängend. Dann ist Γ entweder ein Baum, und die Behauptung folgt aus Teil (a). Oder Γ enthält einen Kreis. Dann können wir eine Kante des Kreises aus Γ entfernen, und der resultierende Graph Γ' ist immer noch zusammenhängend (und hat dieselbe Eckenmenge wie Γ, aber eine Kante weniger). [Diese Aussage ist anschaulich klar. Der formale Beweis ist ein wenig umständlich, aber nicht schwierig.] Die Behauptung folgt durch Induktion über die Anzahl der Kanten (mit dem Fall „Γ ist Baum“ als Induktionsanfang).
(c): Wäre Γ kein Baum, dann hätten wir wieder einen Kreis, und das Argument im Beweis von Teil (b) würde einen zusammenhängenden Graphen mit weniger als $n - 1$ Kanten produzieren; Widerspruch. □

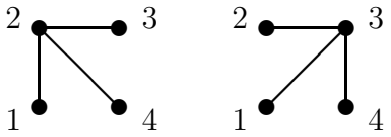
Es folgt noch eine weitere Definition. Dabei verwenden wir folgende Schreibweise. Sei $f : E \to M$ eine Abbildung auf der Eckenmenge eines Graphen $\Gamma = (E, K)$. Dann setzen wir

$$f(K) = \{\{f(a), f(b)\} \mid \{a, b\} \in K\} \subset \mathcal{P}(M)\,.$$

Definition 6.5.

(1) Seien $\Gamma_1 = (E_1, K_1)$ und $\Gamma_2 = (E_2, K_2)$ zwei Graphen. Eine Abbildung $f : E_1 \to E_2$ heißt *Isomorphismus* zwischen Γ_1 und Γ_2, wenn f bijektiv ist und $f(K_1) = K_2$ gilt. Zwei Graphen, zwischen denen es einen Isomorphismus gibt, heißen (wie üblich) *isomorph*. Diese Beziehung definiert eine Äquivalenzrelation; wir können also von *Isomorphieklassen* von Graphen sprechen.
(2) Sei $\Gamma = (E, K)$ ein Graph. Ein Isomorphismus zwischen Γ und sich selbst (also eine Permutation f von E mit $f(K) = K$) heißt *Automorphismus* von Γ. Die Menge aller Automorphismen von Γ bildet (wie üblich) eine Gruppe, die *Automorphismengruppe* $\mathrm{Aut}(\Gamma)$ von Γ. Sie ist eine Untergruppe der Permutationsgruppe von E.

Man spricht von *markierten* Graphen (oder Bäumen), wenn man die Eckenmenge E festhält. Man kann sich das so vorstellen, daß man die Ecken numeriert. Wenn es einem nur auf die „Gestalt“ des Graphen ankommt, spricht man von *freien* Graphen (oder Bäumen). Das heißt, man betrachtet Isomorphieklassen von Graphen. Zum Beispiel sind die folgenden beiden Graphen als markierte Graphen versschieden, aber als freie Graphen gleich.



Das folgende Lemma sagt uns, wie viele markierte Graphen (bei festem E) zu einem gegebenen freien Graphen gehören.

Lemma 6.6. *Sei Γ ein freier Graph mit n Ecken. Dann ist die Anzahl der markierten Graphen (E, K) (mit festem E, so daß $\#E = n$), die zu Γ isomorph sind, gegeben durch*

$$\frac{n!}{\#\operatorname{Aut}(\Gamma)}.$$

Beweis: Sei (E, K_0) ein beliebiger zu Γ isomorpher Graph. Dann sind alle zu Γ isomorphen Graphen mit Eckenmenge E gegeben durch $(E, f(K_0))$, wobei f alle Permutationen von E durchläuft. Wir müssen also die Länge der Bahn von K_0 unter der Permutationsgruppe $\mathcal{S}(E)$ bestimmen. Diese ist gegeben durch

$$\#(\mathcal{S}(E) \cdot K_0) = \frac{\#\mathcal{S}(E)}{\#\mathcal{S}(E)_{K_0}} = \frac{n!}{\#\operatorname{Aut}(\Gamma)},$$

denn der Stabilisator von K_0 ist nichts anderes als die Automorphismengruppe von $\Gamma = (E, K_0)$. □

Wir wollen jetzt die Anzahl $t(n)$ der markierten Bäume bestimmen für kleine Werte von n.

$n = 1$:

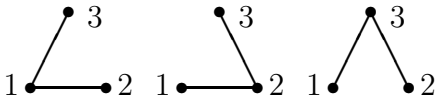
Hier gibt es nur einen Baum: •1 Also haben wir $\boxed{t(1) = 1.}$

$n = 2$:

Hier gibt es ebenfalls nur einen Baum: 1•——•2 Also haben wir $\boxed{t(2) = 1.}$

$n = 3$:

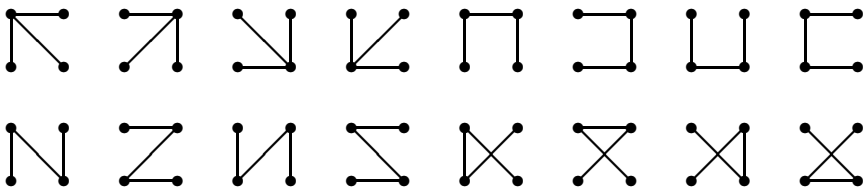
Es gibt die folgenden Bäume:



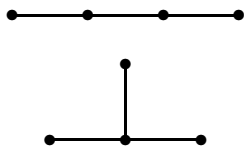
Alternativ kann man sich überlegen, daß die Automorphismengruppe des einzigen freien Baums mit drei Ecken: zwei Elemente hat. Beides führt zu $\boxed{t(3) = 3.}$

$n = 4$:

Für größere Werte von n ist es etwas mühsam, alle markierten Bäume einzeln aufzuzählen, zumal es schnell sehr viele werden. Für $n = 4$ ist das gerade noch machbar:



Es geht einfacher, wenn man das Lemma 6.6 benutzt. Es gibt zwei freie Bäume mit vier Ecken:



$\#\,\mathrm{Aut} = 2$, das ergibt $\frac{4!}{2} = 12$ Bäume

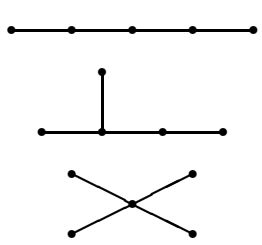
$\#\,\mathrm{Aut} = 6$, das ergibt $\frac{4!}{6} = 4$ Bäume

} 16 Bäume

Also ist $\boxed{t(4) = 16.}$

$n = 5$:

Hier gibt es drei freie Bäume.



$\#\,\mathrm{Aut} = 2$, das ergibt $\frac{5!}{2} = 60$ Bäume

$\#\,\mathrm{Aut} = 2$, das ergibt $\frac{5!}{2} = 60$ Bäume

$\#\,\mathrm{Aut} = 24$, das ergibt $\frac{5!}{24} = 5$ Bäume

} 125 Bäume

Also ist $\boxed{t(5) = 125.}$

Daß es jeweils keine weiteren freien Bäume gibt, läßt sich mit Lemma 6.3 nachprüfen. Es sei als kleine Übung enpfohlen, sich davon zu überzeugen, daß $t(6) = 1296$ ist. All diese Ergebnisse lassen den folgenden Satz vermuten.

Satz 6.7 (Cayley). *Es gilt $t(n) = n^{n-2}$ für alle $n \geq 1$.*

Cayley hat diesen Satz als erster formuliert, wenn auch nicht vollständig bewiesen. Inzwischen gibt es unzählige Beweise dafür. Drei (oder eigentlich zweieinhalb) davon will ich hier vorführen.

Erster Beweis: Dieser Beweis verwendet Rekursion und Induktion. Da es keine vernünftige Rekursionsformel für die Folge $t(n) = n^{n-2}$ gibt (jedenfalls keine offensichtliche), muß man die Fragestellung erst etwas verfeinern. Wir denken uns die Ecken irgendwie total geordnet und definieren für nichtnegative ganze Zahlen $d_1, \ldots, d_n$

$$T(d_1, d_2, \ldots, d_n) = \#\{B \mid B \text{ Baum mit } E_B = \{e_1, \ldots, e_n\} \text{ und } \deg(e_j) = d_j \text{ für alle } j\}\,.$$

Dann gilt folgendes.

(a) $T(d_1, \ldots, d_n) = 0$, falls $d_1 + \cdots + d_n \neq 2n - 2$.
(b) $T(0) = 1$.
(c) $T(d_1, \ldots, d_n) = 0$, falls $n \geq 2$ und ein $d_j = 0$ ist.
(d) Wenn $\sigma \in \mathcal{S}_n$ eine Permutation von $\{1, 2, \ldots, n\}$ ist, dann gilt
$$T(d_{\sigma(1)}, \ldots, d_{\sigma(n)}) = T(d_1, \ldots, d_n).$$

Zum Beweis von Aussage (a) beachten wir, daß für jeden Graphen $\Gamma = (E, K)$ gilt: $\sum_{e \in E} \deg(e) = 2\,\#K$. Das sieht man durch zweifaches Abzählen von

$$\{(e, k) \in E \times K \mid e \text{ Endpunkt von } k\}\,.$$

Da ein Baum mit n Ecken gerade $n - 1$ Kanten hat, folgt die Behauptung.

Aussage (b) ist trivial. Aussage (c) ist auch klar, denn eine Ecke vom Grad 0 wäre isoliert, und der Graph nicht zusammenhängend (da $n \geq 2$, gibt es weitere Ecken). Aussage (d) ist klar.

Sei nun $B = (E, K)$ ein Baum mit $n \geq 2$ Ecken, die die Grade $d_1, \dots, d_n$ haben. Nach Lemma 6.3 gibt es eine Ecke vom Grad 1. Durch Umnumerieren können wir annehmen, daß $d_n = \deg(e_n) = 1$ gilt. Die einzige Kante k mit e_n als Endpunkt verbindet e_n mit einer anderen Ecke e_m, und wenn wir e_n zusammen mit k entfernen, verbleibt ein Baum, dessen Ecken die Grade $d_1, \dots, d_{m-1}, d_m - 1, d_{m+1}, \dots, d_{n-1}$ haben. Umgekehrt entsteht aus solch einem Baum in eindeutiger Weise ein Baum mit Eckengraden $d_1, \dots, d_{n-1}, d_n = 1$, in dem die Ecke e_n mit der Ecke e_m verbunden ist. Das zeigt, daß folgende Aussage gilt.

(e) Für $n \geq 2$ gilt

$$T(d_1, \dots, d_{n-1}, 1) = \sum_{m=1}^{n-1} T(d_1, \dots, d_{m-1}, d_m - 1, d_{m+1}, \dots, d_{n-1}) \,.$$

Zum Beispiel haben wir

$$T(d_1, d_2, 1) = T(d_1 - 1, d_2) + T(d_1, d_2 - 1) \,.$$

Das erinnert an die bekannte Rekursion für die Binomialkoeffizienten. Tatsächlich gilt für $n \geq 2$ und $d_1 + d_2 + \cdots + d_n = 2n - 2$, daß

$$T(d_1, d_2, \dots, d_n) = \frac{(n-2)!}{(d_1 - 1)!(d_2 - 1)! \dots (d_n - 1)!} \,; \tag{6.1}$$

das ist ein sogenannter *Multinomialkoeffizient*, eine natürliche Verallgemeinerung der Binomialkoeffizienten.

Daß die Behauptung für $n = 2$ richtig ist, überprüft man (unter Beachtung von $\frac{1}{(-1)!} = 0$) anhand von $T(2, 0) = T(0, 2) = 0$ und $T(1, 1) = 1$. Der Beweis für $n > 2$ geschieht induktiv. Da beide Seiten symmetrisch in den d_j sind, können wir ohne Einschränkung $d_n = 1$ annehmen. Dann haben wir mit Eigenschaft (e) und der Induktionsvoraussetzung

$$\begin{aligned} T(d_1, \dots, d_n) &= \sum_{m=1}^{n-1} T(d_1, \dots, d_{m-1}, d_m - 1, d_{m+1}, \dots, d_{n-1}) \\ &= \sum_{m=1}^{n-1} \frac{(n-3)!}{(d_1 - 1)! \dots (d_{m-1} - 1)!(d_m - 2)!(d_{m+1} - 1)! \dots (d_{n-1} - 1)!} \\ &= \frac{(n-3)! \sum_{m=1}^{n-1} (d_m - 1)}{(d_1 - 1)! \dots (d_{n-1} - 1)!(1-1)!} \\ &= \frac{(n-2)!}{(d_1 - 1)! \dots (d_n - 1)!} \,. \end{aligned}$$

Um den Beweis abzuschließen, brauchen wir noch folgende Eigenschaft der Multinomialkoeffizienten (die auch ihren Namen erklärt).

Lemma 6.8.

$$(x_1 + x_2 + \cdots + x_k)^n = \sum_{j_1+j_2+\cdots+j_k=n} \frac{n!}{j_1!j_2!\ldots j_k!} x_1^{j_1} x_2^{j_2} \ldots x_k^{j_k} .$$

BEWEIS: Durch Induktion nach k. Der Fall $k = 2$ ist bekannt, die Fälle $k = 0$ und $k = 1$ trivial. Für $k > 2$ gilt

$$\begin{aligned}(x_1 + x_2 + \cdots + x_k)^n &= ((x_1 + \cdots + x_{k-1}) + x_k)^n \\ &= \sum_m \frac{n!}{m!(n-m)!} (x_1 + \cdots + x_{k-1})^{n-m} x_k^m \\ &= \sum_m \frac{n!}{m!(n-m)!} \sum_{j_1+\cdots+j_{k-1}=n-m} \frac{(n-m)!}{j_1!\ldots j_{k-1}!} x_1^{j_1} \ldots x_{k-1}^{j_{k-1}} x_k^m \\ &= \sum_{j_1+\cdots+j_{k-1}+m=n} \frac{n!}{j_1!\ldots j_{k-1}!m!} x_1^{j_1} \ldots x_{k-1}^{j_{k-1}} x_k^m .\end{aligned}$$

□

Damit ergibt sich (für $n \geq 2$)

$$\begin{aligned}t(n) &= \sum_{d_1+d_2+\cdots+d_n=2n-2} T(d_1, d_2, \ldots, d_n) \\ &= \sum_{d_1+d_2+\cdots+d_n=2n-2} \frac{(n-2)!}{(d_1-1)!(d_2-1)!\ldots(d_n-1)!} \\ &= \sum_{j_1+j_2+\cdots+j_n=n-2} \frac{(n-2)!}{j_1!j_2!\ldots j_n!} \\ &= (\underbrace{1+1+\cdots+1}_{n \text{ Einsen}})^{n-2} = n^{n-2} .\end{aligned}$$

□

Für den nun folgenden zweiten Beweis (der eigentlich nur ein halber ist, da wir uns auf ein hier nicht bewiesenes Resultat stützen werden) formulieren wir den Satz von Cayley wie folgt.

Der vollständige Graph K_n hat genau n^{n-2} aufspannende Bäume.

Diese Formulierung legt nahe, nach einem allgemeinen Ergebnis über aufspannende Bäume in Graphen zu suchen. Tatsächlich gibt es so etwas.

Satz 6.9 („Matrix-Tree-Theorem“)**.** *Sei $\Gamma = (E, K)$ ein zusammenhängender ungerichteter (schlingenloser) Graph mit n Ecken. Wir definieren eine $E \times E$-Matrix M durch*

$$M_{e,e'} = \begin{cases} \deg(e) & \text{falls } e = e', \\ -1 & \text{falls } \{e, e'\} \in K, \\ 0 & \text{sonst.} \end{cases}$$

Dann gilt: Die Anzahl der aufspannenden Bäume von Γ ist gegeben durch einen beliebigen $(n-1) \times (n-1)$-Minor von M (d.h. der Determinante einer Untermatrix, die aus M durch Streichen der e-ten Zeile und Spalte entsteht, für ein beliebiges $e \in E$).

Beachte, daß sich die Zeilen (und Spalten) von M zu null addieren; die Determinante von M selbst verschwindet also.

In gewisser Weise macht dieser Satz die Aussage, daß das Zählen der aufspannenden Bäume in einem Graphen (im Gegensatz zu vielen anderen Problemen, die beweisbar schwierig sind) „einfach“ ist. Wir werden uns das gleich zu Nutze machen.

Zweiter Beweis: Nach Satz 6.9 ist die gesuchte Zahl $t(n)$ gleich einem beliebigen $(n-1)\times(n-1)$-Minor der zum vollständigen Graphen K_n gehörigen Matrix M_n. Diese Matrix hat die Form

$$M_n = n\cdot I_n - E_n\,,$$

wobei I_n die $n\times n$-Einheitsmatrix und E_n die $n\times n$-Matrix bestehend aus lauter Einsen ist. Für die Berechnung von $t(n)$ können wir zum Beispiel die letzte Zeile und Spalte entfernen und bekommen dann

$$t(n) = \det(M_n') \quad \text{mit} \quad M_n' = n\cdot I_{n-1} - E_{n-1}\,.$$

Nun beachten wir, daß die Matrix E_{n-1} eine Basis von Eigenvektoren hat, bestehend aus

$$(1,1,\dots,1)^\top, (1,-1,0,\dots,0)^\top, (0,1,-1,0,\dots,0)^\top, \dots, (0,\dots,0,1,-1)^\top$$

mit den Eigenwerten $n-1,0,0,\dots,0$. Folglich hat M_n' dieselben Eigenvektoren, aber mit den Eigenwerten $1,n,n,\dots,n$, und es ergibt sich

$$t(n) = \det(M_n') = n^{n-2}\,.$$

□

Schließlich folgt hier noch ein dritter Beweis, diesmal durch Bijektion.

Dritter Beweis: Wir wollen eine Bijektion herstellen zwischen markierten Bäumen und Abbildungen $\{1,2,\dots,n\}\to\{1,2,\dots,n\}$. Da es n^n solche Abbildungen gibt, aber nur n^{n-2} markierte Bäume, müssen wir dafür sorgen, daß jeweils n^2 Abbildungen zu jedem Baum gehören. Wir tun das dadurch, daß wir die Bäume mit Zusatzstruktur versehen, und zwar zeichnen wir zwei Ecken (die auch zusammenfallen dürfen) als *Wurzel* w und als *Spitze* s aus. Wir betrachten also Tripel (B,w,s), wobei $B=(\{1,2,\dots,n\},K)$ ein Baum ist und $w,s\in\{1,2,\dots,n\}$ die ausgewählten Ecken sind. Es ist dann klar, daß zu jedem Baum B gerade n^2 solcher Tripel gehören.

Im Hinblick auf die Zurodnung zu den Abbildungen machen wir aus dem Baum B im Tripel (B,w,s) einen gerichteten Graphen $tilde B$, indem wir alle Kante in Richtung der Wuruel w orientieren. Wir können uns $(\tilde B,w,s)$ dann vorstellen als einen (zur Wurzel hin orientierten) Baum mit *Baumstamm* von der Spitze s zur Wurzel w, an dem seitlich noch verschiedene Äste hängen, die ihrerseits Bäume sind, die zum Baumstamm hin orientiert sind.

Auf der anderen Seite gehört zu jeder Abbildung $f:\{1,2,\dots,n\}\to\{1,2,\dots,n\}$ in natürlicher Weise ein gerichteter Graph

$$\Gamma_f = (\{1,2,\dots,n\}, \{(k,f(k)) \mid k\in\{1,2,\dots,n\}\})\,.$$

(D.h., von jeder Ecke k geht genau eine Kante aus; sie endet in $f(k)$.) Da die Folge $k, f(k), f(f(k)),\dots$ für jedes k schließlich periodisch werden muß, sieht man, daß der Graph Γ_f in Komponenten zerfällt, die aus einem (orientierten)

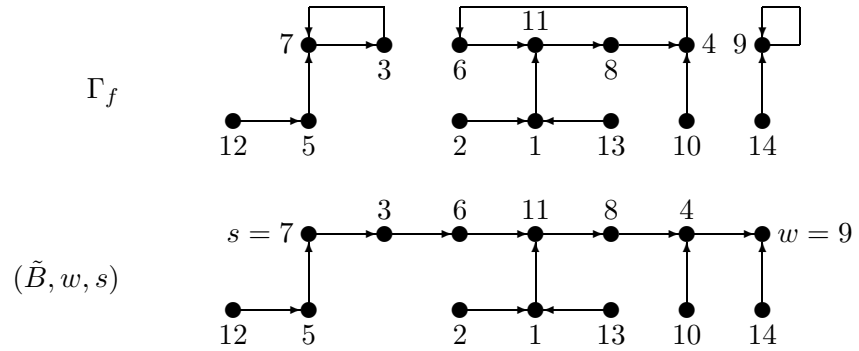
Kreis eventuell zusammen mit daran angehängten zum Kreis hin orientierten Bäumen bestehen.

Für die Bijektion zwischen den Abbildungen f und den Tripeln (B, w, s) wollen wir nun die Kreise in Γ_f mit dem Baumstamm von $(\tilde{B}, w, s)$ identifizieren. Die angehängten Teilbäume werden dabei einfach mitgenommen.

Aus einer Menge von Kreisen machen wir einen Baumstamm wie folgt. Wir suchen in jedem Kreis die kleinste Ecke; dann ordnen wir die Kreise nach aufsteigender kleinster Ecke an: $k_1, k_2, \ldots, k_r$. Die kleinste Ecke im Kreis k sei $m(k)$; weiter sei $n(k) = f(m(k))$, also die auf die kleinste Ecke im Kreis folgende Ecke. Aus der Menge von Kreisen wird nun ein Weg, indem wir die Kanten $(m(k_j), n(k_j))$ für alle $j = 1, 2, \ldots, r$ entfernen und dafür neue Kanten $(m(k_j), n(k_{j+1}))$ für $j = 1, 2, \ldots, r-1$ einfügen. Dieser Weg wird dann der Baumstamm (mit $s = n(k_1)$ und $w = m(k_r)$); die angehängten Seitenbäume der Kreise werden zu den Ästen.

Diese Konstruktion läßt sich umkehren. Dazu nehmen wir ein Tripel $(\tilde{B}, w, s)$ her und betrachten seinen Baumstamm. Wir setzen $j = 1$ und $n_1 = s$. Dann wiederholen wir folgende Schritte: Wir suchen die kleinste Ecke m_j auf dem Weg von n_j nach w. Aus dem Teil des Baumstammes von n_j bis m_j machen wir einen Kreis, indem wir die von m_j ausgehende Kante (falls vorhanden) entfernen und eine neue Kante (m_j, n_j) einfäugen. Falls $m_j = w$, dann sind wir fertig. Sonst sei n_{j+1} die im ursprünglichen Baumstamm auf m_j folgende Ecke. Wir erhöhen j um 1 und setzen das Verfahren fort.

Es ist klar, daß diese Konstruktion eine Bijektion liefert zwischen (orientierten) Wegen mit Eckenmenge E und disjunkten Vereinigungen von Kreisen, die zusammen ebenfalls die Eckenmenge E haben. Wenn wir die angehängten Bäume mitnehmen, erhalten wir also wie gewünscht eine Bijektion zwischen den Abbildungen $\{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$ und den Tripeln (B, w, s). Die folgende Skizze verdeutlicht dies an einem Beispiel.



Eine Variante dieses Beweises ergibt sich, wenn man nur Abbildungen f betrachtet, die $f(1) = 1$ und $f(n) = n$ erfüllen. Dem entspricht, daß man für jeden Baum die Wurzel $w = n$ und die Spitze $s = 1$ wählt. Auf diese Weise erhält man dann direkt eine Bijektion zwischen den Bäumen und einer Menge der Größe n^{n-2}. □

Erzeugende Funktionen

7. Formale Potenzreihen

Die Grundidee bei der Methode der Erzeugenden Funktionen ist, die Zahlenfolge, die einen interessiert (zum Beispiel die Anzahlen a_n von kombinatorischen Objekten, die von einem Paramter n abhängen), als Koeffizienten in eine Potenzreihe zu verpacken. Beziehungen zwischen kombinatorischen Objekten übersetzen sich dabei in Beziehungen zwischen den zugehörigen Potenzreihen. Dabei werden diese Potenzreihen zunächst ganz formal als algebraische Objekte behandelt, das heißt, ohne irgendwelche Konvergenzvoraussetzungen zu machen. (Insofern führt der Begriff „Erzeugende *Funktion*“ etwas in die Irre, weil man die erzeugende Potenzreihe zunächst ja gerade nicht als Funktion (also etwas, das man an irgendwelchen Stellen auswerten kann) betrachtet.) Es ist jedoch oft der Fall, daß die sich ergebende Potenzreihe tatsächlich positiven Konvergenzradius hat, also als (in einer Umgebung der Null holomorphe) Funktion interpretiert werden kann. Dann lassen sich mit analytischen Methoden zusätzliche Informationen über die Koeffizienten gewinnen, zum Beispiel über ihr asymptotisches Verhalten.

Zunächst einmal müssen wir aber die zugrundeliegenden rein algebraischen Objekte verstshen. Dazu sei R ein kommutativer Integritätsring der Charakteristik 0, also zum Beispiel $R = \mathbb{Z}, \mathbb{Q}, \mathbb{C}$ oder auch $R = \mathbb{Q}[Y]$.

Definition 7.1. Der *Ring der formalen Potenzreihen (in einer Variablen)* über R, geschrieben $R[\![X]\!]$, bestheht aus Elementen, die in der Form $\sum_{n=0}^{\infty} a_n X^n$ mit $a_n \in R$ geschrieben werden. (Ganz formal kann man die Elemente mit Folgen $(a_n)_{n\ge 0}$ von Elementen aus R identifizieren.)

Zwei Elemente $f = \sum_{n=0}^{\infty} a_n X^n$ und $g = \sum_{n=0}^{\infty} b_n X^n$ werden wie folgt addiert bzw. subtrahiert:

$$f \pm g = \sum_{n=0}^{\infty} (a_n \pm b_n) X^n \,.$$

Das Produkt von f und g ist

$$f \cdot g = \sum_{n=0}^{\infty} c_n X^n \quad \text{mit} \quad c_n = \sum_{k=0}^{n} a_k b_{n-k} \,.$$

Man kann nachrechnen, daß $R[\![X]\!]$ auf diese Weise zu einem kommutativen Integritätsring wird.

Es gibt eine kanonische Einbettung

$$R \longrightarrow R[\![X]\!] \,, \qquad r \longmapsto r + 0 \cdot X + 0 \cdot X^2 + \ldots$$

und einen kanonischen Epimorphismus

$$R[\![X]\!] \longrightarrow R \,, \qquad f = \sum_{n=0}^{\infty} a_n X^n \longmapsto a_0 = f(0) \,.$$

Proposition 7.2. *Ein Element $f \in R[\![X]\!]$ besitzt genau dann einen Kehrwert (also ein multiplikatives Inverses), wenn $f(0) \in R^\times$ ist.*

BEWEIS: Notwendigkeit ist klar, denn aus $f \cdot g = 1$ in $R[\![X]\!]$ folgt $f(0) \cdot g(0) = 1$ in R. Zum Beweis, daß die Bedingung auch hinreichend ist, schreiben wir $f = \sum_n a_n X^n$, $g = \sum_n b_n X^n$. Die Gleichung $fg = 1$ bedeutet dann

$$a_0 b_0 = 1 \quad \text{und} \quad a_0 b_n + a_1 b_{n-1} + \cdots + a_n b_0 = 0 \text{ für } n \geq 1.$$

Wenn $a_0 \in R^\times$ ist, dann lassen sich also die Koeffizienten von g rekursiv berechnen als

$$b_0 = a_0^{-1}, \quad b_n = -a_0^{-1}(a_n b_0 + \cdots + a_1 b_{n-1}) \text{ für } n \geq 1.$$

□

Eine sehr wichtige Eigenschaft des Rings der formalen Potenzreihen ist, daß er eine natürliche Topologie trägt, die ihn zu einem vollständigen metrischen Raum macht.

Definition 7.3. Für eine Potenzreihe $f = \sum_n a_n X^n \in R[\![X]\!]$ sei

$$v(f) = \begin{cases} +\infty & \text{falls } f = 0 \\ \min\{n \mid a_n \neq 0\} & \text{falls } f \neq 0 \end{cases}$$

und $|f| = (\frac{1}{2})^{v(f)}$ ($= 0$ falls $f = 0$). $v(f)$ heißt die *Bewertung* von f, und $|f|$ heißt der *Betrag* von f.

Proposition 7.4. *Bewertung und Betrag haben folgende Eigenschaften.*

(a) $v(f \pm g) \geq \min\{v(f), v(g)\}$ *bzw.* $|f \pm g| \leq \max\{|f|, |g|\} \leq |f| + |g|$.
(b) $v(fg) = v(f) + v(g)$ *bzw.* $|fg| = |f|\,|g|$.
(c) *Der Ring* $R[\![X]\!]$ *wird durch die Metrik* $d(f, g) = |f - g|$ *zu einem* vollständigen metrischen Raum.

BEWEIS: (a) Das ist klar, wenn $f = 0$ oder $g = 0$. Der allgemeine Fall folgt aus der trivialen Tatsache $a_n = b_n = 0 \Longrightarrow a_n \pm b_n = 0$. Die erste Ungleichung für den Betrag ist lediglich eine Übersetzung der Aussage über die Bewertung; die zweite ist trivial (denn $|f|, |g| \geq 0$).
(b) $v(fg) \geq v(f) + v(g)$ folgt ähnlich wie in Teil (a). Die Gleichheit folgt daraus, daß R Integritätsring ist (d.h. $a, b \neq 0 \Longrightarrow ab \neq 0$). Die Aussage über die Beträge ist wiederum nur die Übersetzung der Aussage über die Bewertungen.
(c) Daß d eine Metrik ist, ist klar ($d(f, f) = |f - f| = 0$, $d(f, g) = |f - g| = |g - f| = d(g, f)$, die Dreiecksungleichung wurde in Teil (a) bewiesen). Es bleibt zu zeigen, daß diese Metrik *vollständig* ist, d.h. daß jede Cauchy-Folge in $R[\![X]\!]$ konvergiert. Hier ist eine Beweisskizze. Sei (f_j) eine Cauchy-Folge in $R[\![X]\!]$ mit $f_j = \sum_n a_{jn} X^n$. Der erste Schritt ist zu zeigen, daß für jedes n ein N existiert, so daß a_{jn} konstant ($= a_n$) ist für $j \geq N$. Das folgt leicht aus der Cauchy-Eigenschaft. Dann zeigt man, daß f_j gegen $\sum_n a_n X^n$ konvergiert. □

Die verschärfte (sogenannte *nichtarchimedische*) Version der Dreiecksungleichung im Teil (a) der Proposition führt zu einem besonders einfachen Kriterium für die Konvergenz von unendlichen Reihen und Produkten.

Proposition 7.5. *Sei* (f_j) *eine Folge von Potenzreihen in* $R[\![X]\!]$.

(a) *Die Reihe* $\sum_{j=0}^{\infty} f_j$ *konvergiert genau dann in* $R[\![X]\!]$, *wenn* $|f_j| \to 0$ *(oder äquivalent* $v(f_j) \to \infty$*) für* $j \to \infty$.

(b) *Das Produkt $\prod_{j=0}^{\infty} f_j$ konvergiert genau dann in $R[\![X]\!]$, wenn $|f_j - 1| \to 0$ (oder äquivalent $v(f_j - 1) \to \infty$) für $j \to \infty$.*

Beweis: (a) Die Notwendingkeit der Bedingung ist klar (sonst wäre die Folge der Partialsummen keine Cauchy-Folge). Die Bedingung ist aber auch hinreichend: Sei $\varepsilon > 0$ gegeben. Dann gibt es nach Voraussetzung ein N, so daß $|f_j| < \varepsilon$ ist für alle $j \geq N$. Nach der verschärften Dreiecksungleichung gilt dann

$$|\sum_{j=m}^{m+r} f_j| \leq \max\{|f_j| \mid m \leq j \leq m+r\} < \varepsilon$$

für alle $m \geq N$ und $r \geq 0$. Also ist die Folge der Partialsummen eine Cauchy-Folge und damit konvergent.
(b) Übung. □

Diese Proposition zeigt zum Beispiel, daß für eine beliebige Folge (a_n) in R die Reihe $\sum_{n=0}^{\infty} a_n X^n$ konvergiert, was die Potenzreihenschreibweise für die Elemente von $R[\![X]\!]$ rechtfertigt.

Ein Beispiel für ein konvergentes Produkt ist

$$\prod_{n=1}^{\infty}(1 - X^n) = 1 - X - X^2 + X^5 + X^7 \mp \cdots = \sum_{m\in\mathbb{Z}} (-1)^m X^{(3m^2-m)/2}.$$

Diese Gleichung ist der berühmte *Pentagonalzahlensatz* von Euler. Er läßt sich wie folgt kombinatorisch interpretieren (und auch mittels dieser Interpretation beweisen): Für jede natürliche Zahl n gibt es gleich viele Partitionen von n in ungerade viele verschiedene Teile wie in gerade viele verschiedene Teile, außer n ist von der Form $n = \frac{1}{2}(3m^2 - m)$ für ein $m \in \mathbb{Z}$. In diesem Fall gibt es von der einen Sorte eine Partition mehr als von der anderen Sorte, je nach Parität von m. (Eine *Partition* von n ist eine Zerlegung von n in eine Summe positiver ganzer Zahlen ohne Beachtung der Reihenfolge.)

Ein wichtiger Spzeialfall einer konvergenten Reihe von Potenzreihen tritt auf bei der *Verknüpfung* von Potenzreihen. Dabei möchte man eine Potenzreihe in eine andere einsetzen.

Proposition 7.6. *Seien $f = \sum_{n=0}^{\infty} a_n X^n$ und g Elemente von $R[\![X]\!]$ mit $g(0) = 0$. Dann konvergiert die Reihe*

$$f(g) := \sum_{n=0}^{\infty} a_n g^n$$

in $R[\![X]\!]$.

Beweis: Die Bedingung $g(0) = 0$ bedeutet $v(g) \geq 1$. Es folgt $v(a_n g^n) \geq n$, und damit konvergiert die Reihe nach Prop. 7.5. □

Bemerkung 7.7. Ist $g(0) \neq 0$, dann gilt $v(a_n g^n) = 0$, falls $a_n \neq 0$. In diesem Fall konvergiert die Reihe genau dann, wenn f ein Polynom ist.

Es ist nun einfach zu sehen, daß die Potenzreihen f mit $f(0) = 0$ bezüglich dieser Verknüpfung eine (nicht-kommutative) Halbgruppe (d.h. die Verknüpfung ist assoziativ) mit dem neutralen Element X bilden. Dies wirft die Frage auf,

wann eine Potenzreihe ein Inverses hat, d.h. für welche f gibt es ein g mit $f(g) = g(f) = X$?

Proposition 7.8. *Ist $f = \sum_{n=0}^{\infty} a_n X^n \in R[\![X]\!]$ mit $a_0 = 0$ und $a_1 \in R^\times$, dann hat f eine inverse Potenzreihe g, d.h. eine Potenzreihe $g \in R[\![X]\!]$ mit $g(0) = 0$, so daß $f(g) = g(f) = X$ gilt.*

Hat umgekehrt $f \in R[\![X]\!]$ mit $f(0) = 0$ eine Links- oder Rechtsinverse g (d.h. $f(g) = X$ oder $g(f) = X$), so gilt $a_1 \in R^\times$.

Beweis: Wir beginnen mit dem zweiten Teil. Sei $g = \sum_n b_n X^n$ mit $b_0 = 0$. Dann folgt aus $f(g) = X$ und auch aus $g(f) = X$, daß $a_1 b_1 = 1$ ist, also sind $a_1, b_1 \in R^\times$.

Für den ersten Teil genügt es zu zeigen, daß f sowohl eine linksinverse als auch eine rechtsinverse Potenzreihe hat, denn dann folgt automatisch, daß beide übereinstimmen ($g_1(f) = X = f(g_2)$ impliziert $g_1 = g_1(f(g_2)) = g_2$).

Zur Existenz der linksinversen Reihe: Wir schreiben wieder $g = \sum_n b_n X^n$. Die Gleichung $g(f) = X$ bedeutet $\sum_{n=1}^{\infty} b_n f^n = X$. Daraus folgt schon einmal $b_1 = a_1^{-1}$ (durch Vergleich des Koeffizienten von X^1). Jetzt betrachten wir den Koeffizienten von X^n für $n \geq 2$. Da $v(f) = 1$, tragen die Terme f^m mit $m > n$ nichts dazu bei, und der Beitrag von f^n ist a_1^n. Also erhalten wir eine Gleichung

$$b_n = a_1^{-n}(\text{Ausdruck in } a_1, \ldots, a_n \text{ und } b_1, \ldots, b_{n-1}),$$

aus der sich die Koeffizienten von g eindeutig rekursiv bestimmen lassen.

Nun zur Existenz der rechtsinversen Reihe. Diesmal haben wir die Gleichung $f(g) = \sum_{n=1}^{\infty} a_n g^n = X$ zu lösen. Wir haben wieder $b_1 = a_1^{-1}$. Wenn wir jetzt den Koeffizienten von X^n ($n \geq 2$) betrachten, dann sehen wir, daß der einzige Term, in dem b_n vorkommt, durch $a_1 b_n$ gegeben ist. Alle anderen Terme enthalten (außer den Koeffizienten von f) nur Koeffizienten b_j mit $j < n$. Daraus ergibt sich wieder eine Rekursionsgleichung für die Koeffizienten von g, durch die diese eindeutig bestimmt sind. □

Bemerkung 7.9. Es gibt eine Formel für die Koeffizienten der inversen Reihe, die sich manchmal nutzbringend anwenden läßt. Wir werden diese sogenannte *Lagrangesche Inversionsformel* später beweisen.

Wie Funktionen in der Analysis, so kann man auch Potenzreihen ableiten, indem man sie formal gliedweise differenziert.

Definition 7.10. Für eine Potenzreihe $f = \sum_{n=0}^{\infty} a_n X^n \in R[\![X]\!]$ ist die *Ableitung* definiert als

$$f' = Df = \frac{d}{dX} f = \sum_{n=0}^{\infty} (n+1) a_{n+1} X^n .$$

Insbesondere gilt

$$Xf' = \sum_{n=0}^{\infty} n a_n X^n .$$

Wie man leicht sieht, kann man Differentiation mit (konvergenten) Reihen vertauschen, woraus folgt, daß alle die üblichen Rechenregeln auch für Ableitungen von formalen Potenzreihen gelten. Zum Beispiel gilt $f' = 0 \iff f$ konstant (dabei wird benutzt, daß der Ring R Charakteristik 0 hat). Außerdem gilt ganz trivial die Taylorformel

$$f = \sum_{n=0}^{\infty} \frac{D^n f(0)}{n!} X^n \,.$$

Nach all der Theorie soll es auch ein paar Beispiele geben. Zunächst ein paar wichtige Potenzreihen.

Beispiele 7.11.

(1) Die geometrische Reihe: $\quad (1-X)^{-1} = \dfrac{1}{1-X} = \displaystyle\sum_{n=0}^{\infty} X^n.$

(2) Die Exponentialreihe: $\quad \exp(X) = \displaystyle\sum_{n=0}^{\infty} \frac{1}{n!} X^n.$

(3) Die Logarithmusreihe: $\quad \log(1+X) = \displaystyle\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} X^n.$

Variante: $\quad \log\Big(\dfrac{1}{1-X}\Big) = \displaystyle\sum_{n=1}^{\infty} \frac{1}{n} X^n.$

(4) Die Binomialreihe: $\quad (1+X)^r = \exp(r\log(1+X)) = \displaystyle\sum_{n=0}^{\infty} \binom{r}{n} X^n.$

Diese Gleichung gilt in $\mathbb{Q}[r][\![X]\!]$. Dabei ist der Binomialkoeffizient $\binom{r}{n}$ definiert als

$$\binom{r}{n} = \frac{r(r-1)\dots(r-n+1)}{n!}$$

(das ist $= 1$, wenn $n = 0$ und $= 0$, wenn $n < 0$).

Es gelten dann die üblichen Beziehungen wie zum Beispiel

$$\begin{aligned} \exp(\log(1+X)) &= 1+X \\ \log(1+(\exp(X)-1)) &= X \\ (1+X)^r(1+X)^s &= (1+X)^{r+s} \quad (\text{in } \mathbb{Q}[r,s][\![X]\!]) \end{aligned}$$

Als einen kleinen Appetithappen, der die Methode der Erzeugenden Funktionen schmackhaft machen soll, wollen wir sehen, wie sich die Fibonacci-Zahlen damit behandeln lassen.

Beispiel 7.12. Die Fibonacci-Zahlen F_n sind definiert durch

$$F_0 = 0\,, \quad F_1 = 1\,, \quad F_{n+2} = F_{n+1} + F_n \ \text{ für } n \geq 0\,.$$

Wir multiplizieren die Rekursionsgleichung mit X^{n+2} und summieren über alle $n \geq 0$. Das ergibt mit $f(X) = \sum_{n=0}^{\infty} F_n X^n = X + X^2 + 2X^3 + \dots$

$$\begin{aligned}\sum_{n=0}^{\infty} F_{n+2} X^{n+2} &= \sum_{n=0}^{\infty} F_{n+1} X^{n+2} + \sum_{n=0}^{\infty} F_n X^{n+2} \\ f(X) - X &= X f(X) + X^2 f(X) \\ (1 - X - X^2) f(X) &= X \\ f(X) &= X(1 - X - X^2)^{-1} = \frac{X}{1 - X - X^2}\,.\end{aligned}$$

Damit haben wir die Erzeugende Funktion für die Fibonacci-Zahlen gefunden; sie ist eine *rationale* Potenzreihe. Jede rationale Funktion hat eine *Partialbruchzerlegung*. Hier erhalten wir aus

$$1 - X - X^2 = \Big(1 - \frac{1+\sqrt{5}}{2} X\Big)\Big(1 - \frac{1-\sqrt{5}}{2} X\Big)$$

die Gleichung (in $\mathbb{Q}(\sqrt{5})[\![X]\!]$ oder in $\mathbb{C}[\![X]\!]$)

$$f(X) = \frac{1}{\sqrt{5}} \left(\frac{1}{1 - \frac{1+\sqrt{5}}{2} X} - \frac{1}{1 - \frac{1-\sqrt{5}}{2} X} \right)$$

und daraus durch Entwickeln der geometrischen Reihe die Formel

$$F_n = \frac{1}{\sqrt{5}} \left(\Big(\frac{1+\sqrt{5}}{2}\Big)^n - \Big(\frac{1-\sqrt{5}}{2}\Big)^n \right).$$

8. Rationale Erzeugende Funktionen

Hier kommt nun endlich die formale Definition, was eine Erzeugende Funktion ist.

Definition 8.1. Sei $a = (a_n)_{n \geq 0}$ eine Folge von Elementen des Rings R, von dem wir ab jetzt voraussetzen wollen, daß er $\mathbb{Q}$ enthält. Dann heißt

$$F_a(X) = \sum_{n=0}^{\infty} a_n X^n \in R[\![X]\!]$$

die *gewöhnliche Erzeugende Funktion* (kurz GEF) von a und

$$E_a(X) = \sum_{n=0}^{\infty} a_n \frac{X^n}{n!} \in R[\![X]\!]$$

die *exponentielle Erzeugende Funktion* (kurz EEF) von a.

Es gibt neben diesen beiden auch noch eine Reihe weiterer Arten von Erzeugenden Funktionen der Bauart $\sum_n a_n X^n / f(n)$, die besser an die Struktur gewisser kombinatorischer Probleme angepaßt sind. Siehe zum Beispiel Stanley, Enumerative Combinatorics, Vol. 2. Die gewöhnliche und die exponentielle Erzeugende Funktion sind aber die wichtigsten und häufigsten Varianten.

Wir brauchen noch eine Schreibweise, um aus einer Potenzreihe die Koeffizienten zurückzuerhalten. Folgende Notation hat sich dafür eingebürgert.

Definition 8.2. Sei $f(X) = \sum_{n=0}^{\infty} a_n X^n \in R[\![X]\!]$. Wir schreiben

$$[X^n]f(X) = a_n$$

und etwas allgemeiner für $a \in R^\times$

$$\Big[\frac{X^n}{a}\Big]f(X) = a \cdot a_n\,.$$

Bemerkung 8.3. Es gilt offensichtlich $[X^n]X^k f(X) = [X^{n-k}]f(X)$.

Wenn wir mit Erzeugenden Funktionen sinnvoll arbeiten wollen, dann brauchen wir eine Art Lexikon, das Operationen mit Folgen in Operationen mit ihren Erzeugenden Funktionen übersetzt (und umgekehrt).

Lemma 8.4. *Seien $A = (a_n)$ und $b = (b_n)$ Folgen in R.*

(1) *Die GEF von $(a_{n+1})_{n\ge0}$ ist $\dfrac{F_a(X) - a_0}{X} = \dfrac{F_a(X) - F_a(0)}{X}$.*
Die GEF von $(a_{n-1})_{n\ge0}$ (mit $a_{-1} = 0$) ist $XF_a(X)$.
Die GEF von $(na_n)_{n\ge0}$ ist $XF_a'(X)$.
Die GEF von $(\sum_{k=0}^{n} a_k b_{n-k})_{n\ge0}$ ist $F_a(X)F_b(X)$.
(2) *Die EEF von $(a_{n+1})_{n\ge0}$ ist $E_a'(X)$.*
Die EEF von $(na_n)_{n\ge0}$ ist $XE_a'(X)$.
Die EEF von $(\sum_{k=0}^{n} \binom{n}{k} a_k b_{n-k})_{n\ge0}$ ist $E_a(X)E_b(X)$.

Beweis: Leicht. □

Die Eignung einer Sorte von Erzeugenden Funktionen für die Behandlung einer Klasse kombinatorischer Probleme entscheidet sich hauptsächlich an der Art der Verknüpfung zweier Folgen, die dem Produkt der Erzeugenden Funktionen entspricht. Dabei stellt sich heraus, daß GEF gut an Probleme angepaßt sind, die mit natürlichen Zahlen zu tun haben, während EEF sich gut dafür eignen, Probleme über endliche Mengen zu behandeln (denn der Faktor $\binom{n}{k}$, der in der Produktformel auftaucht, entspricht der Auswahl einer k-Teilmenge).

Beispiele 8.5.

(1) Es gilt

$$\sum_{n=0}^{\infty} nX^n = X\frac{d}{dX}\,\frac{1}{1-X} = \frac{X}{(1-X)^2}$$

und

$$\sum_{n=0}^{\infty} n^2X^n = X\frac{d}{dX}\,\frac{X}{(1-X)^2} = \frac{X+X^2}{(1-X)^3}\,.$$

(2) Wir haben für $d \ge 1$

$$\frac{1}{(1-X)^d} = \sum_{n=0}^{\infty} \binom{n+d-1}{d-1} X^n\,.$$

Das beweist man durch Induktion nach d; der Induktionsschritt verwendet die übliche Rekursionsformel für die Binomialkoeffizienten.

(3) Etwas allgemeiner gilt für $0 \le k < d$

$$\frac{X^k}{(1-X)^d} = \sum_{n=0}^{\infty} \binom{n-k+d-1}{d-1} X^n .$$

Dafür brauchen wir, daß der Binomialkoeffizient für $0 \le n < k$ verschwindet. Man beachte, daß der Koeffizient in dieser Potenzreihe ein *Polynom* in n vom Grad $d-1$ ist. Dies werden wir gleich noch verwenden.

Bevor wir den ersten Satz über (spezielle) rationale Erzeugende Funktionen formulieren können, brauchen wir noch eine kleine Definition.

Definition 8.6. Sei $a = (a_n)$ eine Folge in R. Dann setzen wir

$$\Delta a = (a_{n+1} - a_n)_{n \ge 0} .$$

Sei $P \in R[X]$ ein Polynom. Dann setzen wir analog

$$\Delta P = P(X+1) - P(X) .$$

Δ heißt der *Differenzoperator*.

Satz 8.7. *Sei $a = (a_n)_{n\ge 0}$ eine Folge in $\mathbb{C}$ und sei $d \in \mathbb{N}$. Folgende Aussagen sind äquivalent.*

(i) *Für alle $n \ge 0$ ist $a_n = P(n)$ mit einem Polynom $P \in \mathbb{C}[X]$ vom Grad $\deg P < d$.*
(ii) $\Delta^d a = 0$.
(iii) $F_a(X) = \dfrac{\tilde P(X)}{(1-X)^d}$ *mit einem Polynom $\tilde P \in \mathbb{C}[X]$ vom Grad $\deg \tilde P < d$.*

Beweis: (i) $\Rightarrow$ (ii): Für jedes Polynom P gilt $\deg \Delta P = \deg P - 1$, falls $\Delta P \neq 0$. Also muß $\Delta^d P = 0$ sein. Dann ist aber auch $(\Delta^d a)_n = \Delta^d P(n) = 0$, also $\Delta^d a = 0$.
(ii) $\Rightarrow$ (iii): Durch Induktion nach d. Der Fall $d = 0$ ist klar. Sei die Implikation für d gezeigt; wir wollen sie für $d+1$ zeigen. Nach unserem „Lexikon“ gilt

$$F_{\Delta a}(X) = \frac{(1-X)F_a(X) - a_0}{X} . \tag{8.1}$$

Da $\Delta^d(\Delta a) = 0$, können wir die Induktionsvoraussetzung auf Δa anwenden und erhalten

$$F_{\Delta a}(X) = \frac{Q(X)}{(1-X)^d}$$

mit einem Polynom Q von Grad $< d$. Wenn wir das in (8.1) einsetzen, erhalten wir

$$F_a(X) = \frac{XF_{\Delta a}(X) + a_0}{1-X} = \frac{XQ(X) + a_0(1-X)^d}{(1-X)^{d+1}} = \frac{\tilde P(X)}{(1-X)^{d+1}} ,$$

wobei $\tilde P(X) = XQ(X) + a_0(1-X)^d$ ein Polynom vom Grad $\deg \tilde P < d+1$ ist.
(iii) $\Rightarrow$ (i): Sei $\tilde P(X) = b_0 + b_1 X + \cdots + b_{d-1}X^{d-1}$. Dann folgt mit Beispiel 8.5, (3), daß

$$\frac{\tilde P(X)}{(1-X)^d} = \sum_{k=0}^{d-1} b_k \frac{X^k}{(1-X)^d} = \sum_{n=0}^{\infty}\sum_{k=0}^{d-1} b_k \binom{n-k+d-1}{d-1} X^n ,$$

also sind die Koeffizienten a_n gegeben durch ein Polynom in n vom Grad höchstens $d-1$. □

Ein wichtiger Spezialfall von GEF sind *rationale GEF*, also GEF der Form $F_a(X) = P(X)/Q(X)$, wo P und Q Polynome sind mit $Q(0) = 1$. Der folgende Satz zeigt, daß zu solchen GEF Folgen gehören, die einer linearen Rekursion mit konstanten Koeffizienten genügen.

Satz 8.8. *Sei $a = (a_n)$ eine Folge in $\mathbb{C}$ und sei $Q(X) = 1-\alpha_1 X-\cdots-\alpha_m X^M \in \mathbb{C}[X]$ mit $\alpha_m \neq 0$. Dann ist $Q(X) = \prod_{j=1}^{k}(1-\gamma_j)^{d_j}$ mit paarweise verschiedenen komplexen Zahlen γ_j und $d_j \geq 1$.*

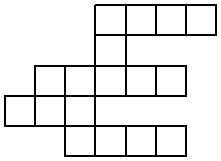
Folgende Aussagen sind äquivalent.

(i) $F_a(X) = \dfrac{P(X)}{Q(X)}$ *mit* $P \in \mathbb{C}[X]$, $\deg P < m$.
(ii) *Für alle* $n \geq 0$ *gilt* $a_{n+m} = \alpha_1 a_{n+m-1} + \alpha_2 a_{n+m-2} + \cdots + \alpha_m a_n$.
(iii) *Für alle* $n \geq 0$ *gilt* $a_n = P_1(n)\gamma_1^n + \cdots + P_k(n)\gamma_k^n$ *mit Polynomen* $P_j \in \mathbb{C}[X]$, $\deg P_j < d_j$.
(iv) $F_a(X) = \dfrac{R_1(X)}{(1-\gamma_1 X)^{d_1}} + \cdots + \dfrac{R_k(X)}{(1-\gamma_k X)^{d_k}}$ *mit Polynomen* $R_j \in \mathbb{C}[X]$, $\deg R_j < d_j$.

Beweis: (i) ⇔ (iv): Das ist genau die Partialbruchzerlegung rationaler Funktionen.
(iii) ⇔ (iv): Das folgt für jeden Summanden einzeln aus Satz 8.7, wenn man dort X durch $\gamma_j X$ ersetzt.
(i) ⇒ (ii): Wir betrachten den Koeffizienten von X^{n+m} in der Gleichung $Q(X)F_a(X) = P(X)$. Da $\deg P < m$, ist dieser Koeffizient auf der rechten Seite null. Auf der linken Seite erhalten wir gerade $a_{n+m} - \alpha_1 a_{n+m-1} - \cdots - \alpha_m a_n$.
(ii) ⇒ (i): Die Rekursionsgleichung bedeutet gerade, daß alle Koeffizienten von X^n mit $n \geq m$ in $Q(X)F_a(X)$ verschwinden, also ist $P(X) = Q(X)F_a(X)$ ein Polynom mit $\deg P < m$. □

Beispiel 8.9. Als Beispiel für die Verwendung rationaler GEF möchte ich hier *horizontal konvexe Polyominos* abzählen. Das Beispiel ist deswegen instruktiv, weil sich am Ende eine rationale GEF ergibt, die zu einer ganz und gar nicht offensichtlichen linearen Rekursion für diese Anzahlen führt.

Ein horizontal konvexes Polyomino ist eine Anordnung von Kästchen in zusammenhängenden waagerechten Reihen, so daß übereinander liegende Reihen wenigstens längs eines Kästchens aneinanderstoßen.



Die Anzahl solcher Objekte mit insgesamt n Kästchen sei mit a_n bezeichnet. Die ersten paar Werte sind

$$a_0 = 1\,, \quad a_1 = 1\,, \quad a_2 = 2\,, \quad a_3 = 6\,, \ldots$$

Da es keine offensichtliche Beziehung zwischen den a_n zu geben scheint, verfeinern wir das Zählproblem. Wir definieren $a(n,k)$ als die Anzahl der horizontal konvexen Polyominos mit insgesamt n Kästchen und davon k in der untersten

Reihe. Dann sieht man leicht, daß folgende Beziehungen gelten.

$$a(n,0) = \begin{cases} 1 & n = 0 \\ 0 & n > 0 \end{cases}$$

$$a(n,k) = \begin{cases} 0 & n < k \\ 1 & n = k \\ \sum_{l=1}^{\infty}(k+l-1)a(n-k,l) & n > k \geq 1 \end{cases}$$

Wir führen folgende GEF ein:

$$F_k(X) = \sum_{n=0}^{\infty} a(n,k)X^n = X^k + \dots$$

$$F(X) = \sum_{n=0}^{\infty} a_n X^n = \sum_{k=0}^{\infty} F_k(X)$$

$$G(X) = \sum_{k=0}^{\infty} kF_k(X)$$

(Der Grund für die Einführung von G wird gleich klar werden.) Die oben aufgestellten Rekursionsgleichungen für die $a(n,k)$ übersetzen sich dann in folgende Gleichungen für die GEF.

$$\begin{aligned} F_0(X) &= 1 \\ F_k(X) &= X^k + \sum_{n=k+1}^{\infty}\sum_{l=1}^{\infty}(k+l-1)a(n-k,l)X^n \\ &= X^k\Big(1 + \sum_{m=1}^{\infty}\sum_{l=1}^{\infty}(k+l-1)a(m,l)X^m\Big) \\ &= X^k\big(1 + (k-1)(F(X)-1) + G(X)\big) \end{aligned}$$

Die letzte Gleichung gilt für $k \geq 1$.

Aus diesen Gleichungen gewinnen wir ein System von linearen Gleichungen für $F(X)$ und $G(X)$, indem wir einmal über k summieren und einmal mit k multiplizieren und dann über k summieren. Das ergibt zum einen

$$\begin{aligned} F(X) &= 1 + \sum_{k=1}^{\infty}(2-k)X^k + \sum_{k=1}^{\infty}(k-1)X^kF(X) + \sum_{k=1}^{\infty}X^kG(X) \\ &= \frac{1-X-X^2}{(1-X)^2} + \frac{X^2}{(1-X)^2}F(X) + \frac{X}{1-X}G(X) \end{aligned}$$

und zum anderen

$$\begin{aligned} G(X) &= \sum_{k=1}^{\infty}k(2-k)X^k + \sum_{k=1}^{\infty}k(k-1)X^kF(X) + \sum_{k=1}^{\infty}kX^kG(X) \\ &= \frac{X(1-3X)}{(1-X)^3} + \frac{2X^2}{(1-X)^3}F(X) + \frac{X}{(1-X)^2}G(X) \end{aligned}$$

(dabei haben wir die Formeln

$$\sum_k X^k = \frac{1}{1-X}\,, \qquad \sum_k kX^k = \frac{X}{(1-X)^2}\,, \qquad \sum_k k^2X^k = \frac{X(1+X)}{(1-X)^3}$$

benutzt). Beides zusammen ergibt das lineare Gleichungssystem

$$\begin{aligned} (1-2X)\,F(X) \qquad\qquad - X(1-X)\,G(X) &= 1-X-X^2 \\ -2X^2\,F(X) + (1-X)(1-3X+X^2)\,G(X) &= X(1-3X) \end{aligned}$$

mit der Lösung

$$F(X) = \frac{1-4X+4X^2-X^3-X^4}{1-5X+7X^2-4X^3} = 1 + \frac{X(1-X)^3}{1-5X+7X^2-4X^3}\,.$$

Nach Satz 8.8 bekommen wir daraus (wenn wir berücksichtigen, daß der Zähler größeren Grad als der Nenner hat) folgendes Ergebnis.

Proposition 8.10. *Für die Zahlen a_n gilt*

$$a_0 = 1\,, \quad a_1 = 1\,, \quad a_2 = 2\,, \quad a_3 = 6\,, \quad a_4 = 19\,,$$

und für alle $n \geq 2$

$$a_{n+3} = 5\,a_{n+2} - 7\,a_{n+1} + 4\,a_n\,.$$

9. Formale Laurent-Reihen und die Lagrangesche Inversionsformel

Ds Hauptziel in diesem Kapitel ist es, die Lagrangesche Inversionsformel zu beweisen, die es erlaubt, die Koeffizienten der zu einer Potenzreihe $F \in R[\![X]\!]$ inversen Reihe G (d.h. mit $F(G(X)) = G(F(X)) = X$) durch Koeffizienten von Potenzen von F auszudrücken. Für diesen Beweis müssen wir aber zunächst die *formalen Laurent-Reihen* einführen.

Es ist weiterhin R ein kommutativer (Integritäts-)Ring mit $\mathbb{Q} \subset R$.

Definition 9.1. Der *Ring der formalen Laurent-Reihen* über R ist definiert als

$$R(\!(X)\!) = R[\![X]\!][X^{-1}] = \{X^{-n}F(X) \mid n \in \mathbb{N}, F(X) \in R[\![X]\!]\}\,.$$

Seine von 0 verschiedenen Elemente können also geschrieben werden als

$$F(X) = \sum_{n=m}^{\infty} a_n X^n = X^m F_0(X)\,,$$

wobei $m \in \mathbb{Z}$, $a_m \neq 0$ und $F_0(X) \in R[\![X]\!]$ mit $F_0(0) \neq 0$. Die Bewertung $v : R[\![X]\!] \longrightarrow \mathbb{N} \cup \{+\infty\}$ setzt sich kanonisch fort zu $v : R(\!(X)\!) \longrightarrow \mathbb{Z} \cup \{+\infty\}$; für die Laurent-Reihe F wie oben ist $v(F) = m$. Die wesentlichen Eigenschaften der Bewertung bleiben erhalten.

Bemerkung 9.2. Die Differentiation von formalen Laurent-Reihen ist genauso (nämlich gliedweise mittels $\frac{d}{dX}X^n = nX^{n-1}$) definiert wie für formale Potenzreihen. Die üblichen Rechenregeln bleiben dabei gültig.

Eine Reihe $F(X) = \sum_{n=m}^{\infty} a_n X^n \in R(\!(X)\!)$ mit $a_m \neq 0$ hat genau dann einen Kehrwert in $R(\!(X)\!)$, wenn $a_m \in R^\times$. (Das folgt aus der entsprechenden Aussage für Potenzreihen.) Ist R speziell ein Körper, so folgt, daß auch $R(\!(X)\!)$ ein Körper ist.

Laurent-Reihen (sogar etwas allgemeinerer Art als hier) sind sicher schon aus der Funktionentheorie geläufig. In der Funktionentheorie lernt man auch, daß das *Residuum* einer Funktion wichtig ist. Das gilt auch für unsere formalen Laurent-Reihen.

Definition 9.3. Das *Residuum* ist folgende R-lineare Abbildung.

$$\operatorname{Res} : R(\!(X)\!) \longrightarrow R\,, \qquad F(X) \longmapsto [X^{-1}]F(X)\,.$$

(Die Schreibweise $[X^n]F(X)$ verwenden wir in offensichtlicher Weise auch für Laurent-Reihen.)

Das Residuum hat einige bemerkenswerte Eigenschaften im Zusammenhang mit der Differentiation.

Lemma 9.4.

(1) *Für* $F(X) \in R(\!(X)\!)$ *gilt* $\operatorname{Res}(F'(X)) = 0$.

(2) *Für* $F(X) \in R(\!(X)\!)^\times$ *gilt* $\mathit{Res}\left(\dfrac{F'(X)}{F(X)}\right) = v(F(X)) \in \mathbb{Z} \subset R$.

(3) *Für* $F(X) \in R(\!(X)\!)^\times$ *und* $k \in \mathbb{Z}$ *gilt*

$$\operatorname{Res}(F(X)^k F'(X)) = \begin{cases} v(F(X)), & k = -1 \\ 0, & \textit{sonst.} \end{cases}$$

Beweis: (1) Der Koeffizient von X^{-1} in $\frac{d}{dX}X^n = nX^{n-1}$ ist stets null.
(2) Sei zunächst $v(F) = 0$. Dann ist $F'(X)/F(X) \in R[\![X]\!]$ eine Potenzreihe, hat also keinen X^{-1}-Term. Im allgemeinen Fall schreiben wir $F(X) = X^m G(X)$ mit $v(G) = 0$, $m = v(F)$. Dann gilt $F'(X)/F(X) = m/X + G'(X)/G(X)$, also $\operatorname{Res}(F'(X)/F(X)) = \operatorname{Res}(m/X) + \operatorname{Res}(G'(X)/G(X)) = m + 0 = v(F)$.
(3) Der Fall $k = -1$ ist gerade Teil (2). Falls $k \neq -1$, dann ist $F(X)^k F'(X) = \frac{d}{dX}\frac{1}{k+1}F(X)^{k+1}$, und die Behauptung folgt aus Teil (1). □

Nach diesen Vorbereitungen können wir eine erste Version der Inversionsformel formulieren und beweisen.

Satz 9.5. *Sei* $F(X) \in R[\![X]\!]$ *mit* $v(F) = 1$ *und* $[X^1]F(X) \in R^\times$. *Sei* $G(X) \in R[\![X]\!]$ *die dann existierende und eindeutig bestimmte Reihe mit* $F(G(X)) = X$. *Dann gilt für* $n \geq 1$

$$[X^n]G(X) = \frac{1}{n}\operatorname{Res} F(X)^{-n}\,.$$

(Beachte, daß $F(X)^{-n}$ *hier eine formale Laurent-Reihe ist mit* $v(F^{-n}) = -n$.*)*

Beweis: Sei $G(X) = \sum_{n=1}^\infty a_n X^n$. Aus $F(G(X)) = X$ folgt, daß auch $G(F(X)) = X$ sein muß (siehe Prop. 7.8). Wenn wir die Gleichung $X = \sum_{k=1}^\infty a_k F(X)^k$ differenzieren, erhalten wir

$$1 = \sum_{k=1}^\infty k a_k F(X)^{k-1} F'(X)\,,$$

also

$$\operatorname{Res}(F(X)^{-n}) = \sum_{k=1}^\infty k a_k \operatorname{Res}(F(X)^{k-n-1}F'(X)) = n a_n v(F) = n a_n$$

nach Lemma 9.4, Teil (3). □

Üblicherweise wird diese Formel in einer etwas anderen Form angegeben, in der keine Laurent-Reihen vorkommen und die für die praktische Anwendung meistens günstiger ist.

Korollar 9.6. *Sei $F(X) \in R[\![X]\!]$ mit $F(0) \in R^\times$, und sei $G(X) \in R[\![X]\!]$ mit $G(X) = XF(G(X))$ (G existiert und ist eindeutig bestimmt). Dann gilt für $n \geq 1$*

$$[X^n]G(X) = \frac{1}{n}[X^{n-1}]F(X)^n = \frac{1}{n!}\frac{d^{n-1}}{dX^{n-1}}F(X)^n\Big|_{X=0}$$

oder auch (besser angepaßt an EEF)

$$\left[\frac{X^n}{n!}\right]G(X) = \left[\frac{X^{n-1}}{(n-1)!}\right]F(X)^n\,.$$

Beweis: Sei $H(X) = X/F(X)$. Dann erfüllt H die Voraussetzungen an F in Satz 9.5, und wir haben $H(G(X)) = X$. Nach dem Satz gilt also

$$[X^n]G(X) = \frac{1}{n}\operatorname{Res} H(X)^{-n} = \frac{1}{n}[X^{-1}]X^{-n}F(X)^n = \frac{1}{n}[X^{n-1}]F(X)^n\,.$$

□

Bemerkung 9.7. Korollar 9.6 hat folgende Verallgemeinerung.
Sei zusätzlich $H(X) \in R[\![X]\!]$. Dann gilt für $n \geq 1$

$$[X^n]H(G(X)) = \frac{1}{n}[X^{n-1}]H'(X)F(X)^n\,.$$

Nach all der grauen Theorie nun ein paar Beispiele. Es ist klar, daß die Inversionsformel normalerweise nur dann nutzbringend angewendet werden kann, wenn man die Koeffizienten der Potenzen der Reihe $F(X)$ kennt. Dafür sollte $F(X)$ von recht einfacher Bauart sein.

Beispiele 9.8.

(1) Die Catalan-Zahlen. In Kapitel 3 haben wir ausführlich die Catalan-Zahlen behandelt. Insbesondere haben wir in Kor. 3.4 folgende Rekursion aufgestellt.

$$c_0 = 1\,, \qquad c_{n+1} = \sum_{k=0}^{n} c_k c_{n-k}\,.$$

Wir wollen nun sehen, wie man aus dieser Rekursion mit einer GEF die Zahlen c_n bekommt.

Sei also $C(X) = \sum_{n=0}^{\infty} c_n X^n = 1 + X + 2X^2 + 5X^3 + \dots$ die GEF der Catalan-Zahlen. Wenn wir die obige Rekursionsgleichung mit X^{n+1} multiplizieren und über alle $n \in \mathbb{N}$ summieren, erhalten wir die Gleichung

$$C(X) - 1 = XC(X)^2\,.$$

An diesem Punkt haben wir nun mehrere Möglichkeiten fortzufahren.

1. Wir lösen die quadratische Gleichung für $C(X)$ und bekommen

$$C(X) = \frac{1 - \sqrt{1-4X}}{2X}\,.$$

(Das Vorzeichen + kommt nicht in Frage, da $C(X)$ eine Potenzreihe ist.) Daraus ergibt sich dann

$$c_n = [X^n]C(X) = -\tfrac{1}{2}[X^{n+1}](1-4X)^{1/2} = -\frac{1}{2}\binom{1/2}{n+1}(-4)^{n+1} = \frac{(-4)^n}{n+1}\binom{-1/2}{n}$$

und schließlich

$$c_n = \frac{1}{n+1}\binom{2n}{n},$$

denn

$$\begin{aligned}\binom{-1/2}{n} &= \frac{(-1/2)(-3/2)\dots(-(2n-1)/2)}{n!} \\ &= (-2)^{-n}\frac{1\cdot 3\cdot\dots\cdot(2n-1)\cdot 2\cdot 4\cdot\dots\cdot 2n}{n!\cdot 2^n n!} = (-4)^{-n}\binom{2n}{n}.\end{aligned}$$

2. Wir wenden die Lagrangesche Inversionsformel an auf $\tilde{C}(X) = XC(X)$ (für die Reihe G in Kor. 9.6 muß ja $v(G) = 1$ gelten). Unsere Gleichung für $C(X)$ läßt sich (nach Multiplikation mit X) umschreiben als

$$\tilde{C}(X) = \frac{X}{1-\tilde{C}(X)};$$

wir können also Kor. 9.6 anwenden mit $F(X) = 1/(1-X)$ und bekommen

$$c_n = [X^{n+1}]\tilde{C}(X) = \frac{1}{n+1}[X^n]\frac{1}{(1-X)^{n+1}} = \frac{1}{n+1}\binom{2n}{n}$$

(verwende $[X^n]1/(1-X)^{k+1} = \binom{n+k}{k}$.)

3. Wir wenden die Inversionsformel an auf $\tilde{C}(X) = C(X)-1$. Dann haben wir

$$\tilde{C}(X) = X(1+\tilde{C}(X))^2,$$

hier ist also $F(X) = (1+X)^2$, und wir bekommen für $n \geq 1$

$$c_n = [X^n]\tilde{C}(X) = \frac{1}{n}[X^{n-1}](1+X)^{2n} = \frac{1}{n}\binom{2n}{n-1} = \frac{1}{n+1}\binom{2n}{n}.$$

(2) Das Beispiel der Catalan-Zahlen läßt sich verallgemeinern. Seien die Zahlen a_n definiert durch

$$a_0 = 1, \qquad a_{n+1} = \sum_{k_1+\dots+k_m=n} a_{k_1}\dots a_{k_m},$$

wobei $m \geq 1$ eine natürliche Zahl ist. Mit $A(X) = \sum_n a_n X^n$ ergibt sich die Gleichung

$$A(X) - 1 = XA(X)^m \quad \text{oder} \quad \tilde{A}(X) = X(1+\tilde{A}(X))^m,$$

wenn wir $\tilde{A}(X) = A(X) - 1$ setzen. Die Inversionsformel liefert dann für $n \geq 1$

$$a_n = [X^n]\tilde{A}(X) = \frac{1}{n}[X^{n-1}](1+X)^{mn} = \frac{1}{n}\binom{mn}{n-1} = \frac{1}{(m-1)n+1}\binom{mn}{n},$$

wobei der letzte Ausdruck auch für $n = 0$ gültig ist.

Eine mögliche kombinatorische Interpretation der Zahlen a_n ist als die Anzahl der Zerlegungen eines konvexen $((m-1)n+2)$-Ecks in konvexe $(m+1)$-Ecke durch sich nicht schneidende Diagonalen.

10. Exponentielle Erzeugende Funktionen und die Exponentialformel

Nachdem wir uns bisher auf gewöhnliche EF konzentriert haben, soll es in diesem Kapitel nun um exponentielle EF gehen. Diese Art von EF ist besonders gut für Probleme geeignet, bei denen es um kombinatorische Strukturen auf (endlichen) Mengen mit unterscheidbaren Elementen geht (wie zum Beispiel markierte Bäume im Gegensatz zu freien Bäumen, bei denen die Ecken nicht von vorne herein unterschieden werden).

Wir werden also im Folgenden Typen von kombinatorischen Strukturen betrachten, die auf endlichen Mengen „leben“. Jeder solche Typ von Struktur wird einen Großbuchstaben als Namen bekommen. Dabei ist es durchaus möglich, daß eine gegebene Menge keine solche Struktur tragen kann. Beispiele sind

S	—	endliche Menge (d.h. keine zusätzkiche Struktur)
$\tilde{S}$	—	nichtleere endliche Menge
$S^{(k)}$	—	k-elementige Menge
Π	—	Permutation der Menge
D	—	fixpunktfreie Permutation
C	—	Orientierter (nichtleerer) Zykel
$C^{(k)}$	—	Orientierter Zykel der Länge k
W	—	Wurzelbaum (dessen Ecken die Elemente der Menge sind)

Wenn A so ein Typ von kombinatorischer Struktur ist, dann schreiben wir

$$a_n = \#\{A\text{-Strukturen auf einer } n\text{-elementigen Menge}\}$$
$$A(X) = E_a(X) = \sum_{n=0}^{\infty} a_n \frac{X^n}{n!}$$

(dabei ist a der zu A gehörende Kleinbuchstabe). Zum Beispiel haben wir dann folgende EEF.

$$S(X) = e^X\,, \qquad \tilde{S}(X) = e^X - 1\,, \qquad S^{(k)}(X) = \frac{X^k}{k!}\,,$$
$$\Pi(X) = \frac{1}{1-X}\,, \qquad C(X) = \log\frac{1}{1-X}\,, \qquad C^{(k)}(X) = \frac{X^k}{k}\,.$$

Wir haben dann folgende Zusammenhänge.

Lemma 10.1. *A und B seien Typen kombinatorischer Strukturen wie oben.*

(1) *Eine $(A+B)$-Struktur auf M sei entweder eine A-Struktur oder eine B-Struktur. Dann gilt*

$$(A+B)(X) = A(X) + B(X)\,.$$

(2) *Eine $(A\cdot B)$-Struktur auf M sei gegeben durch eine Zerlegung $M = M_1 \uplus M_2$ und eine A-Struktur auf M_1 und eine B-Struktur auf M_2. Dann gilt*

$$(A \cdot B)(X) = A(X) \cdot B(X)\,.$$

(3) *Ein Spezialfall von Teil (2) ist*

$$B(X) = X\,A(X)\,,$$

wenn $B = S^{(1)} \cdot A$ ist. Eine B-Struktur auf M entsteht also, indem man ein Element $m \in M$ auswählt und die restliche Menge $M \setminus \{m\}$ mit einer A-Struktur versieht.

(4) *Es gelte $a_0 = 0$. Eine B-Struktur auf M sei gegeben durch eine ungeordnete Zerlegung von M in k nichtleere Teilmengen sowie A-Strukturen auf jeder der Teilmengen. Dann gilt*

$$B(X) = \frac{1}{k!}A(X)^k\,.$$

(5) *Eine Ergänzung zu (1): Seien $A_1, A_2, \ldots$ Typen von Strukturen, so daß die Reihe $\sum_{k=1}^{\infty} A_k(X)$ im Ring der formalen Potenzreihen konvergiert. Sei $A = \sum_{k=1}^{\infty} A_k$ (d.h. eine A-Struktur auf M besteht aus der Wahl eines k und einer A_k-Struktur auf M). Dann gilt*

$$A(X) = \sum_{k=1}^{\infty} A_k(X)\,.$$

Beweis: (1) Klar.
(2) Sei $C = A \cdot B$. Dann gilt (mit einer n-elementigen Menge M)

$$c_n = \sum_{T\subset M} a_{\#T} b_{\#(M\setminus T)} = \sum_{k=0}^{n} \binom{n}{k} a_k b_{n-k}\,.$$

Nach Lemma 8.4 (2) gilt dann $C(X) = A(X)B(X)$.
(3) Klar (beachte $S^{(1)}(X) = X$).
(4) Da $a_0 = 0$, kommt es nicht darauf an, ob wir beliebige oder nichtleere Teilmengen betrachten. Wenn wir eine geordnete Zerlegung von M haben, $M = M_1 \uplus \cdots \uplus M_k$ und jede der Teilmengen mit einer A-Struktur versehen, dann ist die zugehörige EEF nach Teil (2) (mit einer trivialen Induktion) gerade $A(X)^k$. Da die Teilmengen alle nichtleer (und daher paarweise verschieden) sein müssen, wenn sie eine A-Struktur haben, gehört die entsprechende ungeordnete Zerlegung von M zu genau $k!$ verschiedenen geordneten Zerlegungen. Daraus folgt die Behauptung.
(5) Die Konvergenzbedingung bedeutet, daß es nur endlich viele k gibt, so daß eine Menge der Größe $\le n$ eine A_k-Struktur haben kann. Um die Behauptung für den Koeffizienten von X^n zu beweisen, können wir uns also auf endlich viele der A_k beschränken. Die Behauptung folgt dann aus Teil (1). □

Die folgende Proposition behandelt eine weitere Möglichkeit, aus zwei Typen von kombinatorischen Strukturen einen neuen zu basteln.

Proposition 10.2. *Seien A und B Typen kombinatorischer Strukturen mit $a_0 = 0$. Eine $B(A)$-Struktur auf einer Menge M sei gegeben durch eine Partition von M (d.h. eine ungeordnete Zerlegung von M in nichtleere Teilmengen, die*

sogenannten „Blöcke“ der Partition), eine A-Struktur auf jeder der Teilmengen und eine B-Struktur auf der Menge der Blöcke. Dann gilt

$$(B(A))(X) = B(A(X))\,.$$

Beweis: Sei zunächst $B = S^{(k)}$, d.h. wir betrachten nur Partitionen in k Blöcke. Dann gilt nach Lemma 10.1 (4), daß $(S^{(k)}(A))(X) = \frac{1}{k!}A(X)^k = S^{(k)}(A(X))$. Wir können B schreiben als $B = \sum_{m=1}^{N} B_m$ (mit $N \in \mathbb{N} \cup \{\infty\}$), wobei jedes B_m von der Form $S^{(k)}$ ist und jedes $S^{(k)}$ genau b_k-mal vorkommt. Dann gilt offensichtlich auch $B(A) = \sum_{m=1}^{N} B_m(A)$. Nach Lemma 10.1 (5) folgt nun

$$(B(A))(X) = \sum_{m=1}^{N} (B_m(A))(X) = \sum_{m=1}^{N} B_m(A(X)) = B(A(X))\,.$$

(Beachte, daß die Behauptung für $B_m = S^{(k)}$ bereits gezeigt war.) □

Bemerkung 10.3.

1. Die Voraussetzung $a_0 = 0$ ist notwendig, um die Definiertheit von $B(A(X))$ zu sichern.
2. Man kann sich eine $B(A)$-Struktur folgendermaßen vorstellen: Man nimmt eine Menge mit B-Struktur und ersetzt jedes Element durch eine Menge mit A-Struktur.

Ein wichtiger Spezialfall liegt vor, wenn $B = S$ ist, wenn also auf der Menge der Blöcke keine zusätzliche Struktur vorliegt.

Korollar 10.4 (Exponentialformel)**.** *Es gilt*

$$(S(A))(X) = e^{A(X)}\,.$$

Sei $B = S(A)$. Dann lassen sich die Zahlen a_n und b_n durch folgende Rekursionsformeln auseinander berechnen.

$$b_n = \sum_{k=1}^{n} \binom{n-1}{k-1} a_k b_{n-k}\,, \qquad b_0 = 1$$

$$a_n = b_n - \sum_{k=1}^{n-1} \binom{n-1}{k-1} a_k b_{n-k}\,.$$

Beweis: Es sind nur die Rekursionsformeln zu beweisen. Die zweite Formel ist nur eine Umformung der ersten. Zum Beweis der ersten differenzieren wir die Gleichung $B(X) = \exp(A(X))$. Wir erhalten

$$B'(X) = A'(X)e^{A(X)} = A'(X)B(X) \quad \Longrightarrow \quad XB'(X) = (XA'(X))B(X)\,.$$

Nach Lemma 8.4 (2) bedeutet das

$$nb_n = \sum_{k=0}^{n} \binom{n}{k} k a_k b_{n-k} = \sum_{k=1}^{n} n \binom{n-1}{k-1} a_k b_{n-k}\,.$$

Daß $b_0 = 1$ ist, folgt aus $B(X) = \exp(A(X)) = 1 + A(X) + \dots$ und $A(0) = 0$. □

In vielen Fällen kann man den Zusammenhang zwischen $B = S(A)$ und A so interpretieren, daß eine A-Struktur eine Zusammenhangskomponente einer B-Struktur ist. Zum Beispiel hat man $B =$ Graphen (mit gegebener Eckenmenge)

und A = zusammenhängende Graphen. Mit der Exponentialformel kann man dann die Zahlen a_n aus den $b_n = 2^{\binom{n}{2}}$ berechnen.

Beispiele 10.5.

(1) Eine Permutation kann man als disjunkte Vereinigung von orientierten Zykeln auffassen: $\Pi = S(C)$. Tatsächlich gilt auch

$$\Pi(X) = \frac{1}{1-X} = \exp\left(\log \frac{1}{1-X}\right) = e^{C(X)}\,.$$

(Denn $c_n = (n-1)!$ für $n \geq 1$.)

(2) Eine Permutation ist auch gegeben durch eine Teilmenge (ihre Fixpunkte) und eine fixpunktfreie Permutation auf dem Rest: $\Pi = S \cdot D$. Also gilt

$$D(X) = \frac{\Pi(X)}{S(X)} = \frac{e^{-X}}{1-X}\,,$$

und daraus

$$d_n = n! \sum_{k=0}^{n} \frac{(-1)^k}{k!}\,.$$

Vergleiche Abschnitt 4.3.

(3) Wie viele Involutionen (d.h. Permutationen σ mit $\sigma^2 = \mathrm{id}$) gibt es auf einer n-elementigen Menge? Die Bedingung an die Permutation bedeutet, daß in der Zykelzerlegung nur Zykel der Länge 1 oder 2 auftreten dürfen. Also gilt $I = S(C^{(1)} + C^{(2)})$ und daher

$$I(X) = e^{X+\frac{1}{2}X^2}\,.$$

In dieser EEF stecken alle Informationen über die Anzahlen i_n, auch wenn es dafür keine einfache geschlossene Formel gibt. Zum Beispiel liefert uns Kor. 10.4 die Rekursion

$$i_0 = 1\,, \quad i_1 = 1\,, \quad i_{n+2} = i_{n+1} + (n+1)i_n\,.$$

(4) Allgemeiner können wir fragen nach der Anzahl der Permutationen σ mit $\sigma^k = \mathrm{id}$. Hier müssen die Längen der Zykel Teiler von k sein. Also ist die EEF

$$I_k(X) = \exp(\sum_{d|k} C^{(d)}(X)) = \exp\left(\sum_{d|k} \frac{X^d}{d}\right)\,.$$

(5) Wie viele Partitionen hat eine Menge? In diesem Fall ist die „innere“ Struktur einfach eine nichtleere Menge, d.h. es gilt $P = S(\tilde{S})$ und daher

$$\sum_{n=0}^{\infty} B(n)\frac{X^n}{n!} = P(X) = e^{e^X - 1}\,.$$

Die Zahlen $B(n)$ sind die *Bellschen Exponentialzahlen*, die uns bereits in Abschnitt 4.4 begegnet sind.

(6) Genauer können wir nach der Anzahl der Partitionen in k Blöcke fragen. Hier haben wir $P_k = S^{(k)}(\tilde{S})$, also

$$\sum_{n=0}^{\infty} S(n,k)\frac{X^n}{n!} = P_k(X) = \frac{1}{k!}(e^X - 1)^k\,.$$

Die Zahlen $S(n,k)$ sind die *Stirling-Zahlen 2. Art*, siehe ebenfalls Abschnitt 4.4.

(7) Wenn es Stirling-Zahlen 2. Art gibt, muß es natürlich auch *Stirling-Zahlen 1. Art* geben. Sie werden mit $s(n,k)$ bezeichnet (jedenfalls von manchen Autoren) und zählen die Permutationen von n Elementen, die genau k Zykel haben. Also haben wir $\Pi_k = S^{(k)}(C)$ und daher

$$\sum_{n=0}^{\infty} s(n,k)\frac{X^n}{n!} = \Pi_k(X) = \frac{1}{k!}\left(\log\frac{1}{1-X}\right)^k\,.$$

(8) Beim dritten Beweis des Satzes von Cayley in Kapitel 6 hatten wir gesehen, daß man eine Abbildung $M \to M$ interpretieren kann als eine disjunkte Vereinigung von Zykeln (d.h. eine Permutation), deren Punkte durch Wurzelbäume ersetzt sind. Hier gilt also $A = \Pi(W)$ und damit

$$\sum_{n=0}^{\infty} n^n\frac{X^n}{n!} = A(X) = \frac{1}{1-W(X)}\,.$$

Zusammen mit $w_n = n^{n-1}$ (siehe Kapitel 6 oder auch Beispiel (9) unten) ergibt das die witzige Beziehung

$$n^n = \sum_{k=0}^{n-1}\binom{n}{k}k^k(n-k)^{n-k-1}$$

(für $n \geq 1$).

(9) Hier folgt ein vierter Beweis des Satzes von Cayley. Es gilt $W = S^{(1)}\cdot S(W)$, denn ein Wurzelbaum besteht aus einem ausgezeichneten Element (der Wurzel) und dazu einer Menge von an der Wurzel „angehängten“ Teilbäumen. Das ergibt die Funktionalgleichung

$$W(X) = Xe^{W(X)}\,.$$

Dies ist nun genau so ein Fall, wo sich die Lagrangesche Inversionsformel wunderbar anwenden läßt. Wir haben nämlich

$$w_n = \left[\frac{X^n}{n!}\right]W(X) = \left[\frac{X^{n-1}}{(n-1)!}\right]e^{nX} = n^{n-1}\,.$$

(10) Schließlich wollen wir noch die Anzahl der 2-regulären Graphen mit n Ecken bestimmen („2-regulär“ heißt, daß jede Ecke den Grad 2 hat). So ein (wie immer einfacher, schlingenloser) 2-regulärer Graph ist die disjunkte Vereinigung von nicht-orientierten Zykeln (Kreisen) der Länge ≥ 3. Sei $\tilde{C}$ die Struktur, die zu solchen Zykeln gehört. Dann gilt

$$\tilde{c}_0 = \tilde{c}_1 = \tilde{c}_2 = 0\,, \qquad \tilde{c}_n = \tfrac{1}{2}(n-1)! \quad \text{für } n \geq 3\,,$$

also $\tilde{C}(X) = \frac{1}{2}(\log(1-X)^{-1} - X - \frac{1}{2}X^2)$ und daher

$$G_2(X) = \exp\left(\tfrac{1}{2}\log\frac{1}{1-X} - \tfrac{1}{2}X - \tfrac{1}{4}X^2\right) = \frac{e^{-\frac{1}{2}X-\frac{1}{4}X^2}}{\sqrt{1-X}}\,.$$

Daraus ließe sich auch wieder eine Rekursionsformel für die Anzahlen $g_{2,n}$ herleiten oder auch die asymptotische Formel

$$g_{2,n} = \frac{e^{-3/4}n!}{\sqrt{\pi n}}(1 + O(1/n))\,.$$

11. P-rekursive Folgen und D-finite Potenzreihen

In Kapitel 8 haben wir gesehen, daß Folgen, die einer linearen Rekursionsgleichung mit konstanten Koeffizienten genügen, rationale Gewöhnliche Erzeugende Funktionen haben. In diesem Kapitel soll es nun darum gehen, Folgen zu studieren, die einer linearen Rekursionsgleichung mit *Polynomkoeffizienten* genügen. Beispiele dafür sind $a_n = n!$ mit der Gleichung $a_{n+1} = (n+1)a_n$ oder $a_n = \binom{2n}{n}$ mit der Gleichung $(n+1)a_{n+1} = 2(2n+1)a_n$.

Wir wollen jedoch nicht mit den Folgen beginnen, sondern auf der anderen Seite, nämlich den zugehörigen Potenzreihen. Der Grund dafür ist, daß bei der Formulierung der relevanten Eigenschaften für Folgen eine Komplikation auftritt, die damit zu tun hat, daß Polynome Nullstellen haben können. Bei den Potenzreihen tritt das Problem nicht auf, da die Polynome nicht ausgewertet werden.

Von jetzt ab sei K ein Körper der Charakteristik 0, z.B. $\mathbb{Q}$, $\mathbb{R}$ oder $\mathbb{C}$.

Der Körper $K(X)$ der rationalen Funktionen über K (das ist der Quotientenkörper des Polynomrings $K[X]$; seine Elemente sind Quotienten $P(X)/Q(X)$ von Polynomen) ist in natürlicher Weise eingebettet in den Körper $K(\!(X)\!)$ der Laurent-Reihen über K ($K(\!(X)\!)$ ist ein Körper, da K ein Körper ist, vergleiche Bem. 9.2). Insbesondere ist damit $K(\!(X)\!)$ ein $K(X)$-Vektorraum.

Lemma 11.1. *Für eine (Potenz- oder) Laurent-Reihe $F \in K(\!(X)\!)$ sind folgende Eigenschaften äquivalent:*

(1) *Es gibt Polynome $P_0, P_1, \dots, P_d \in K[X]$ mit $P_d \neq 0$, so daß*

$$P_d(X)F^{(d)}(X) + \cdots + P_1(X)F'(X) + P_0(X)F(X) = 0\,.$$

(Hier bezeichnet $F^{(n)}$ die n-te Ableitung von F.)

(2) *Es gibt Polynome $P_0, P_1, \dots, P_d, Q \in K[X]$ mit $P_d \neq 0$, so daß*

$$P_d(X)F^{(d)}(X) + \cdots + P_1(X)F'(X) + P_0(X)F(X) = Q(X)\,.$$

(3) *Sei $V(F) = \operatorname{span}_{K(X)}(F, F', F'', \dots)$ der von F und allen seinen Ableitungen erzeugte $K(X)$-Untervektorraum von $K(\!(X)\!)$. Dann gilt*

$$\dim_{K(X)} V(F) < \infty\,.$$

Definition 11.2. Eine Reihe $F \in K(\!(X)\!)$, die obige äquivalente Eigenschaften hat, heißt *D-finit* (engl. *D-finite* als Abkürzung von *differentiably finite*; das bezieht sich auf Eigenscahft (3)).

Beweis (Lemma 11.1): (1)$\Longrightarrow$(2) ist trivial.
(2)$\Longrightarrow$(1): Sei $m = \deg(Q) + 1$. Wenn wir die Gleichung in (2) m-mal ableiten, bekommen wir eine Gleichung

$$P_d(X)F^{(d+m)}(X) + \cdots + P_0^{(m)}(X)F(X) = Q^{(m)}(X) = 0$$

wie in (1).
(1)$\Longrightarrow$(3): Sei $V = \operatorname{span}_{K(X)}(F, F', \dots, F^{(d-1)})$. Die Gleichung in (1) impliziert, daß $F^{(d)} \in V$ ist (da $P_d \neq 0$, können wir durch P_d dividieren; dann ist $F^{(d)}$ als $K(X)$-Linearkombination der niedrigeren Ableitungen ausgedrückt). Es folgt (mit der Regel fürs Ableiten eines Produkts und der Tatsache, daß $K(X)$ unter Differentiation invariant ist) $F^{(d+1)} \in V + K(X) \cdot F^{(d)} = V$, also ist V invariant unter Differentiation. Da $F \in V \subset V(F)$, folgt $V = V(F)$ und damit $\dim V(F) = \dim V \leq d$.
(3)$\Longrightarrow$(1): Da $\dim V(F) < \infty$, gibt es ein kleinstes $d \in \mathbb{N}$, so daß $F, F', \dots, F^{(d)}$ linear abhängig sind über $K(X)$. Es folgt

$$F^{(d)}(X) = R_{d-1}(X)F^{(d-1)}(X) + \cdots + R_0(X)F(X)$$

mit $R_j \in K(X)$. Durch Multiplikation mit einem gemeinsamen Nenner der R_j erhalten wir eine Gleichung wie in (1). □

Beispiele für D-finite Potenzreihen sind

$$e^X\,, \quad \sin(X)\,, \quad \cos(X)\,, \quad \log\frac{1}{1-X}\,, \quad \sqrt{1-4X}\,, \quad \frac{1}{\sqrt{1-X}}e^{-X/2-X^2/4}\,.$$

Die Exponentielle Erzeugende Funktion der Bell-Zahlen, $F(X) = \exp(e^X - 1)$, ist jedoch *nicht* D-finit. Um das zu sehen, beachte man, daß

$$F^{(d)}(X) = (e^{dX} + a_{d,d-1}e^{(d-1)X} + \cdots + a_{d,1}e^X + a_{d,0})F(X)$$

ist. Damit ist $V(F) = \operatorname{span}_{K(X)}(1, e^X, e^{2X}, \dots) \cdot F$, und da e^X *transzendent* ist (zu algebraischen (d.h. nicht-transzendenten) Potenzreihen siehe weiter unten in diesem Kapitel), ist das nicht endlich-dimensional.

Wir würden jetzt gerne in ähnlicher Weise die Menge $K^{\mathbb{N}}$ aller Folgen in K als einen $K(X)$-Vektorraum auffassen, indem wir $R(X) \cdot a = (R(n)a_n)_{n\in\mathbb{N}}$ setzen. Das Problem dabei ist, daß R an Stellen $n \in \mathbb{N}$ Pole haben kann, so daß $R(n)$ dann nicht definiert ist. Auf der anderen Seite hat so eine rationale Funktion R aber nur endlich viele Pole, also ist $R(n)a_n$ jedenfalls für $n \gg 0$ (also hinreichend großes n) definiert. Das motiviert folgende Begriffsbildung.

Definition 11.3. Zwei Folgen $a, b \in K^{\mathbb{N}}$ heißen *äquivalent*, kurz $a \sim b$, wenn $a_n = b_n$ für alle hinreichend großen $n \in \mathbb{N}$ gilt. Die Äquivalenzklassen $[a]$ heißen *Folgenkeime*; die Menge aller Folgenkeime sei mit $G(K)$ bezeichnet (von engl. *germ*).

Dies ist analog zu dem aus der Funktionentheorie geläufigen Begriff der Funktionskeime an einer Stelle. Hier werden zwei Folgen identifiziert, wenn sie „in einer Umgebung von ∞“ übereinstimmen.

Proposition 11.4.

(1) *Mit den Definitionen* $[a]+[b] = [a+b]$ *und* $[a][b] = [ab]$ *(wobei* $(ab)_n = a_nb_n$*) wird* $G(K)$ *zu einem kommutativen Ring mit 1.*

(2) *Die Abbildung* $K(X) \ni R(X) \longmapsto [r] \in G(K)$, *wobei* $r_n = 0$, *falls* n *ein Pol von* R *ist, und* $r_n = R(n)$ *sonst, ist ein (dann injektiver, weil* $K(X)$ *ein Köper ist) Ringhomomorphismus. Insbesondere ist* $G(K)$ *eine* $K(X)$*-Algebra und damit ein* $K(X)$*-Vektorraum.*

BEWEIS: (1) Es gilt offensichtlich $a \sim b \iff (a-b) \sim 0$. Außerdem ist $I = [0] = \{a \in K^{\mathbb{N}} \mid a \sim 0\}$ ein Ideal im Ring $K^{\mathbb{N}}$. Nach den üblichen Sätzen aus der Algebra ist damit $G(K) \cong K^{\mathbb{N}}/I$ ein Ring mit Addition und Multiplikation wie angegeben.
(2) Da $(R_1 + R_2)(n) = R_1(n) + R_2(n)$ und $(R_1R_2)(n) = R_1(n)R_2(n)$ für alle $n \gg 0$ gilt, ist die Abbildung ein Ringhomomorphismus. Der Rest ist klar. □

Wir definieren noch den Shiftoperator.

Definition 11.5. Für eine Folge $a \in K^{\mathbb{N}}$ sei $Sa = (a_{n+1})_{n\in\mathbb{N}}$. Diese Operation ist offenbar mit der Äquivalenz von Folgen verträglich, also setzen wir $S[a] = [Sa]$. Auf den Folgenkeimen ist dieser *Shiftoperator* bijektiv; damit ist $S^{-1}[a] = [Ta]$ wohldefiniert (wobei $(Ta)_0 = 0$ (oder sonst irgend etwas) und $(Ta)_{n+1} = a_n$).

Nach all diesen Vorbereitungen können wir jetzt endlich die zu Lemma 11.1 analogen Aussagen für Folgen formulieren.

Lemma 11.6. *Für eine Folge* $a \in K^{\mathbb{N}}$ *sind folgende Eigenschaften äquivalent.*

(1) *Es gibt Polynome* $P_0, P_1, \dots, P_d \in K[X]$ *mit* $P_d \neq 0$, *so daß für alle* $n \in \mathbb{N}$ *gilt*

$$P_d(n)a_{n+d} + \cdots + P_1(n)a_{n+1} + P_0(n)a_n = 0\,.$$

(2) *Es gibt Polynome* $P_0, P_1, \dots, P_d \in K[X]$ *mit* $P_d \neq 0$, *so daß für alle* $n \gg 0$ *gilt*

$$P_d(n)a_{n+d} + \cdots + P_1(n)a_{n+1} + P_0(n)a_n = 0\,.$$

(3) *Es gibt Polynome* $P_0, P_1, \dots, P_d \in K[X]$ *mit* $P_d \neq 0$, *so daß für den Folgenkeim* $[a]$ *gilt*

$$P_d(X) \cdot S^d[a] + \cdots + P_1(X) \cdot S[a] + P_0(X) \cdot [a] = 0\,.$$

(4) *Sei* $V([a]) = \operatorname{span}_{K(X)}([a], S[a], S^2[a], \dots)$ *der von* $[a]$ *und allen seinen Shifts erzeugte* $K(X)$*-Untervekotrraum von* $G(K)$. *Dann gilt*

$$\dim_{K(X)} V([a]) < \infty\,.$$

Definition 11.7. Eine Folge a (oder ihr Folgenkeim $[a]$) heißt *P-rekursiv* (polynomial linear rekursiv), wenn sie obige Eigenschaften hat.

BEWEIS (LEMMA 11.6): (1)$\Longrightarrow$(2) ist trivial.
(2)$\Longrightarrow$(1): Die Gleichung in (2) gelte für $n \geq m$. Man multipliziere sie mit $n(n-1)\dots(n-m+1)$, dann gilt sie für alle $n \in \mathbb{N}$.
(2) $\Longleftrightarrow$ (3) folgt aus den Definitionen.
(3) $\Longleftrightarrow$ (4) geht analog zu Lemma 11.1. Man beachte, daß das Bild von $K(X)$ in $G(K)$ unter dem Shiftoperator invariant ist und daß man die Beziehung $S([a][b]) = (S[a])(S[b])$ hat. □

Das erste Hauptresultat dieses Kapitels zeigt, daß die Eigenschaften „P-rekursiv“ und „D-finit“ zwei Seiten derselben Medaille sind.

Satz 11.8. *Für eine Folge $a \in K^{\mathbb{N}}$ sind folgende Aussagen äquivalent.*

(1) *a ist P-rekursiv.*
(2) *Die GEF F_a ist D-finit.*
(3) *Die EEF E_a ist D-finit.*

Beweis: (1)$\Longrightarrow$(2): Nach Voraussetzung gibt es Polynome $P_0, \dots, P_d \in K[X]$ und $m \in \mathbb{N}$, so daß für $n \ge m$

$$P_d(n)a_{n+d} + \cdots + P_1(n)a_{n+1} + P_0(n)a_n = 0\,.$$

Indem wir eventuell m vergrößern, können wir $P_0 \ne 0$ annehmen. Wir ersetzen nun n durch $n-d$ und erhalten

$$\tilde{P}_d(n)a_n + \cdots + \tilde{P}_1(n)a_{n-d+1} + \tilde{P}_0(n)a_{n-d} = 0$$

für alle $n \ge m+d$, wobei wir $\tilde{P}_j(n) = P_j(n-d)$ gesetzt haben.

Wir erinnern uns nun an Lemma 8.4 (1). Mit dem dort angegebenen Lexikon zur Übersetzung zwischen Folgen und GEF erhalten wir

$$\Big(\tilde{P}_d\big(X\frac{d}{dX}\big) + \tilde{P}_{d-1}\big(X\frac{d}{dX}\big)X + \cdots + \tilde{P}_0\big(X\frac{d}{dX}\big)X^d\Big)F_a(X) = Q(X)\,,$$

wobei Q ein Polynom (vom Grad $< m+d$) ist und in der großen Klammer ein linearer *Differentialoperator* mit Polynomkoeffizienten steht. (Dabei gilt die Rechenregel $(d/dX)X = X(d/dX)+1$, wie man durch Anwenden auf eine Reihe F leicht sieht.) Wenn k der Grad und α der Leitkoeffizient von P_0 (und damit auch von $\tilde{P}_0$) sind, dann enthält dieser Differentialoperator den Term $\alpha X^{k+d}(d/dX)^k$, also ist der Operator nichttrival, d.h. F_a ist D-finit.

(2)$\Longrightarrow$(1): In ähnlicher Weise übersetzt sich eine Differentialgleichung

$$P_d(X)F_a^{(d)}(X) + \cdots + P_1(X)F_a'(X) + P_0(X)F_a(X) = 0$$

in die Gleichung

$$\begin{aligned}\Big(&P_d(T)(X+1)(X+2)\dots(X+d)S^d\\ &+ P_{d-1}(T)(X+1)(X+2)\dots(X+d-1)S^{d-1}\\ &+ \cdots + P_1(T)(X+1)S + P_0(T)\Big)\cdot a = 0\,,\end{aligned}$$

wobei $(Ta)_0 = 0$ und $(Ta)_{n+1} = a_n$ und die große Klammer wiederum als Operator auf $K^{\mathbb{N}}$ aufzufassen ist. Wenn wir darauf eine geeignete Potenz S^t anwenden, läßt sich die resultierende Gleichung in der Form

$$\Big(Q_r(X)S^r + \cdots + Q_1(X)S + Q_0(X)\Big)\cdot[a] = 0$$

schreiben (beachte die Rechenregel $SX = (X+1)S$); dies hat die Form der Bedingung in Lemma 11.6 (3). Man überzeugt sich davon, daß der Leitkoeffizient α von P_d einen nicht-verschwindenden Term $\alpha X^d S^{d-k+t}$ in der großen Klammer erzeugt (wo k der Grad von P_d ist).

(2) $\Longleftrightarrow$ (3): Dies folgt aus der Äquivalenz von (1) und (2) und der Aussage

$$(a_n)\text{ P-rekursiv} \quad\Longleftrightarrow\quad \Big(\frac{a_n}{n!}\Big)\text{ P-rekursiv}\,,$$

die leicht zu sehen ist. □

Als Beispiel betrachten wir die Folge (a_n) der Anzahlen 2-regulärer Graphen, deren EEF

$$F(X) = \frac{1}{\sqrt{1-X}} \exp(-\tfrac{1}{2}X - \tfrac{1}{4}X^2)$$

ist, wie wir gesehen haben. Wir bestimmen die logarithmische Ableitung

$$\frac{F'(X)}{F(X)} = -\frac{1}{2}\frac{-1}{1-X} - \frac{1}{2} - \frac{1}{2}X = \frac{X^2}{2(1-X)}\,.$$

Also haben wir die Differentialgleichung

$$2(1-X)F'(X) - X^2F(X) = 0\,.$$

Für die Folge $b_n = a_n/n!$ ergibt sich dann gemäß obigem Beweis:

$$2(n+1)b_{n+1} - 2nb_n - b_{n-2} = 0 \quad \text{für alle } n \geq 0,$$

wobei $b_{-1} = b_{-2} = 0$ gesetzt werden muß. Multiplikation mit $n!$ liefert dann die Rekursion

$$a_{n+1} = na_n + \binom{n}{2} a_{n-2} \quad \text{für alle } n \geq 0$$

mit $a_{-2} = a_{-1} = 0$ und $a_0 = 1$.

Die nun folgenden Resultate zeigen, daß D-finite Reihen bzw. P-rekursive Folgen einigermaßen schöne Abgeschlossenheitseigenschaften haben.

Satz 11.9.

(1) *Die D-finiten Laurent-Reihen bilden eine $K(X)$-Unteralgebra von $K(\!(X)\!)$. Außerdem gilt: $F \in K(\!(X)\!)$ D-finit $\implies$ F' D-finit.*
(2) *Die P-rekursiven Folgenkeime in $G(K)$ bilden eine $K(X)$-Unteralgebra. Außerdem gilt: $[a] \in G(K)$ P-rekursiv $\implies$ $S[a]$ P-rekursiv.*

Beweis: Zunächst eine Vorbemerkung. Seien V und W zwei $K(X)$-Untervektorräume in einer $K(X)$-Algebra A. Wir setzen

$$V \cdot W = \operatorname{span}_{K(X)}(\{v \cdot w \mid v \in V, w \in W\})\,.$$

Dann gilt $\dim_{K(X)} V \cdot W \leq (\dim_{K(X)} V) \cdot (\dim_{K(X)} W)$. (Denn die Elemente $v \cdot w$, wo v eine Basis von V und w eine Basis von W durchläuft, erzeugen bereits $V \cdot W$.)

(1) Wir verwenden das Kriterium in Lemma 11.1 (3). Seien $F, G \in K(\!(X)\!)$ D-finit. Wir müssen zeigen, daß dann auch $F+G$ und $F \cdot G$ D-finit sind und daß rationale Reihen $R \in K(X)$ D-finit sind. Nun gilt offenbar

$$V(F+G) \subset V(F) + V(G)$$
$$V(F \cdot G) \subset V(F) \cdot V(G)$$
$$V(R) \subset K(X)$$

(denn $(F \cdot G)^{(n)} = \sum_{k=0}^{n} \binom{n}{k} F^{(k)} G^{(n-k)}$), also

$$\dim V(F+G) \leq \dim V(F) + \dim V(G) < \infty$$
$$\dim V(F \cdot G) \leq (\dim V(F)) \cdot (\dim V(G)) < \infty$$
$$\dim V(R) \leq 1\,.$$

Außerdem ist $V(F') \subset V(F)$, also $\dim V(F') \le \dim V(F) < \infty$.

(2) Analog zum Teil (1) verwenden wir hier das Kriterium in Lemma 11.6 (4). Seien also $[a], [b] \in G(K)$ P-rekursiv. Wir müssen zeigen, daß dann auch $[a]+[b]$ und $[a]\cdot[b]$ P-rekursiv sind und daß das Bild von $K(X)$ in $G(K)$ aus P-rekursiven Folgenkeimen besteht. Sei $[r] \in G(K)$ das Bild von $R(X) \in K(X)$. Dann gilt wie oben

$$\begin{aligned} V([a]+[b]) &\subset V([a]) + V([b]) \\ V([a]\cdot[b]) &\subset V([a]) \cdot V([b]) \\ V([r]) &\subset K(X)\cdot[1]\,, \end{aligned}$$

also folgt die Behauptung (hier gilt $S^n([a]\cdot[b]) = (S^n[a])(S^n[b])$). Außerdem ist wieder $V(S[a]) \subset V([a])$, also ist $S[a]$ auch P-rekursiv. □

Für die weiteren Ergebnisse müsen wir eine weitere Klasse von Laurent-Reihen einführen.

Definition 11.10. Eine Reihe $F \in K(\!(X)\!)$ heißt *algebraisch*, wenn sie folgende äquivalente Eigenschaften hat.

(1) Es gibt Polynome $P_0, P_1, \dots, P_d \in K[X]$ mit $P_d \neq 0$, so daß
$$P_d(X)F(X)^d + \cdots + P_1(X)F(X) + P_0(X) = 0\,.$$
(2) Sei $A(F) = \operatorname{span}_{K(X)}(1, F, F^2, \dots)$. Dann ist $\dim_{K(X)} A(F) < \infty$.

Die Äquivalenz der beiden Eigenschaften sieht man auf die gleiche Art und Weise wie in Lemma 11.1 oder 11.6.

Die folgende Proposition faßt einige wichtige Eigenschaften algebraischer Reihen zusammen.

Proposition 11.11.

(1) *Wenn $F \in K(\!(X)\!)$ algebraisch ist, dann ist $A(F)$ ein Körper.*
(2) *Wenn $F \in K(\!(X)\!)$ algebraisch ist, dann gilt $V(F) \subset A(F)$.*
(3) *Die algebraischen Reihen in $K(\!(X)\!)$ bilden einen Unterkörper.*
(4) *Seien $F_1, \dots, F_n \in K(\!(X)\!)$ algebraisch. Dann ist*
$$A(F_1, \dots, F_n) = A(F_1) \cdot \ \dots \ \cdot A(F_n)$$
ein Körper, und zwar der kleinste Körper, der $K(X)$ und $F_1, \dots, F_n$ enthält. Er hat endliche $K(X)$-Dimension.

Beweis: (1) Nach Definition ist $A(F)$ jedenfalls eine $K(X)$-Algebra (nämlich die von F erzeugte $K(X)$-Unteralgebra von $K(\!(X)\!)$). Es ist also nur noch zu zeigen, daß jedes $0 \neq G \in A(F)$ in $A(F)$ einen Kehrwert hat. Dazu zeigen wir zunächst, daß $F^{-1} \in A(F)$ ist (falls $F \neq 0$). Wenn $F \neq 0$ ist, dann können wir in Eigenschaft (1) von Def. 11.10 annehmen, daß $P_0 \neq 0$ ist (andernfalls dividieren wir so oft durch F, bis das der Fall ist). Nach Multiplikation mit F^{-1} erhalten wir

$$F(X)^{-1} = -\frac{P_1(X)}{P_0(X)} - \frac{P_2(X)}{P_0(X)}F(X) - \cdots - \frac{P_d(X)}{P_0(X)}F(X)^{d-1} \in A(F)\,.$$

Sei nun $0 \neq G \in A(F)$ beliebig. Dann gilt offenbar $A(G) \subset A(F)$, also (wie eben gezeigt) $G^{-1} \in A(G) \subset A(F)$.

(2) Wir zeigen $F' \in A(F)$. Dann folgt induktiv $F^{(n)} \in A(F)$ für alle $n \geq 0$, also $V(F) \subset A(F)$. Wir können in Def. 11.10 (1) den Grad d minimal wählen. Wir differenzieren die Gleichung dann und bekommen

$$\begin{aligned} &P_d'(X)F(X)^d + \cdots + P_1'(X)F(X) + P_0'(X) \\ &\quad = -(dP_d(X)F(X)^{d-1} + \cdots + 2P_2(X)F(X) + P_1(X))F'(X)\,. \end{aligned}$$

Da d minimal gewählt war, ist die Klammer auf der rechten Seite nicht null, also haben wir

$$F'(X) = -\frac{P_d'(X)F(X)^d + \cdots + P_1'(X)F(X) + P_0'(X)}{dP_d(X)F(X)^{d-1} + \cdots + 2P_2(X)F(X) + P_1(X)} \in A(F)$$

(denn nach (1) ist $A(F)$ ein Körper).

(3) Es gilt $A(F+G) \subset A(F) \cdot A(G)$ und $A(F \cdot G) \subset A(F) \cdot A(G)$, also sind mit F und G auch $F+G$ und $F \cdot G$ algebraisch. Wir haben bereits in Teil (1) gesehen, daß für algebraisches $F \neq 0$ gilt, daß $F^{-1} \in A(F)$; damit ist auch F^{-1} algebraisch.

(4) Nach Definition ist $A(F_1, \ldots, F_n)$ die von $F_1, \ldots, F_n$ erzeugte $K(X)$-Unteralgebra von $K(\!(X)\!)$ (und damit die kleinste $K(X)$-Unteralgebra, die $F_1, \ldots, F_n$ enthält). Es ist also nur noch zu zeigen, daß für $0 \neq G \in A(F_1, \ldots, F_n)$ auch $G^{-1} \in A(F_1, \ldots, F_n)$ ist. Da $A(F_1, \ldots, F_n)$ aber eine Unteralgebra ist, haben wir $G^{-1} \in A(G) \subset A(F_1, \ldots, F_n)$. □

Mit diesen Vorberitungen können wir zeigen, daß viele Reihen D-finit sind.

Satz 11.12. *Sei $F \in K(\!(X)\!)$. Dann gilt:*

(1) *Ist F algebraisch, dann ist F D-finit.*
(2) *Ist F'/F algebraisch, dann ist F D-finit.*
(3) *Wenn es algebraische Reihen $A_0, \ldots A_d \in K(\!(X)\!)$ gibt mit $A_d \neq 0$, so daß*

$$A_d(X)F(X)^{(d)} + \cdots + A_1(X)F'(X) + A_0(X)F(X) = 0\,,$$

dann ist F D-finit.

Beweis: (1) folgt aus (2), denn für algebraisches F ist auch $F'/F \in A(F)$, also algebraisch. (Alternativ kann man auch direkt Prop. 11.11 (2) benutzen.)
(2) ist der Spzeialfall $d = 1$ von (3)
(3) Sei $K_1 = A(A_0, A_1, \ldots, A_d)$. Dann ist K_1 ein Körper, der unter Differentiation invariant ist (denn $G \in K_1$ impliziert $G' \in A(G) \subset K_1$). Mit demselben Argument wie in Lemma 11.1 folgt aus der gegebenen Differentialgleichung dann, daß

$$\dim_{K_1} \operatorname{span}_{K_1}(F, F', F'', \ldots) < \infty$$

ist. Da $\dim_{K(X)} K_1$ nach Prop. 11.11 (4) endlich ist, folgt

$$\begin{aligned} \dim_{K(X)} \operatorname{span}_{K(X)}(F, F', F'', \ldots) &\leq \dim_{K(X)} \operatorname{span}_{K_1}(F, F', F'', \ldots) \\ &= (\dim_{K(X)} K_1) \cdot \dim_{K_1} \operatorname{span}_{K_1}(F, F', F'', \ldots) \\ &< \infty\,. \end{aligned}$$

□

Aus den bisher bewiesenen Sätzen folgt nun zum Beispiel sofort, daß die Reihe $F(X) = (1-X)^{-1/2}\exp(-X/2 - X^2/4)$ D-finit ist, ohne daß wir die Differentialgleichung für F explizit aufstellen müssten (denn F ist Produkt einer algebraischen Reihe mit einer Reihe, deren logarithmische Ableitung sogar ein Polynom ist).

Das Beispiel e^{e^X-1} zeigt, daß die Komposition zweier D-finiter Reihen nicht notwendig wieder D-finit ist. Es gilt aber immerhin folgender Satz.

Satz 11.13. *Sei $F \in K(\!(X)\!)$ D-finit und $G \in XK[\![X]\!]$ algebraisch. Dann ist $F(G)$ D-finit.*

Beweis: Sei $W = \operatorname{span}_{A(G)}(F(G), F'(G), F''(G), \dots)$. Dann gilt $V(F(G)) \subset W$, denn $F(G) \in W$, und W ist abgeschlossen unter Differentiation: Es gilt $(F^{(n)}(G))' = G' \cdot F^{(n+1)}(G)$. Da $\dim_{K(X)} A(G) < \infty$, genügt es also zu zeigen, daß $\dim_{A(G)} W < \infty$ ist. Dazu betrachten wir eine Gleichung für F wie in Lemma 11.1 (1):

$$P_d(X)F^{(d)}(X) + \cdots + P_1(X)F'(X) + P_0(X)F(X) = 0$$

Wir ersetzen X durch $G(X)$ und erhalten

$$P_d(G(X))F^{(d)}(G(X)) + \cdots + P_1(G(X))F'(G(X)) + P_0(G(X))F(G(X)) = 0\,.$$

Dies zeigt, daß $F^{(d)}(G) \in W_1 = \operatorname{span}_{A(G)}(F(G), F'(G), \dots, F^{(d-1)}(G))$ ist. Wie üblich folgt dann $F^{(d+1)}(G) = \frac{1}{G'}(F^{(d)}(G))' \in W_1$ usw., also schließlich $W = W_1$ und $\dim_{A(G)} W \le d$. □

Das nächste Ergebnis (das wir hier aber nicht beweisen wollen), befaßt sich damit, wann der Kehrwert einer D-finiten Reihe wieder D-finit ist.

Satz 11.14. *Sei $0 \ne F \in K(\!(X)\!)$. Dann gilt*

$$F \text{ D-finit und } F^{-1} \text{ D-finit} \iff F'/F \text{ algebraisch.}$$

(Die Richtung von rechts nach links folgt aus Satz 11.12 (2), denn für die logarithmische Ableitung gilt $(F^{-1})'/F^{-1} = -F'/F$.)

Daraus folgt zum Beispiel, daß Reihen wie

$$\frac{1}{\cos X}\,, \qquad \tan X\,, \qquad \frac{X}{e^X - 1}$$

(letztere ist die EEF der Bernoulli-Zahlen, die in vielen Zusammenhängen, besonders in der Zahlentheorie, immer wieder auftauchen) nicht D-finit sind. Für ihre Koeffizienten gibt es demnach auch keine linearen Rekursionsgleichungen fester Länge, deren Koeffizienten Polynome sind.

Ein weiteres wichtiges Resultat betrifft Diagonalfolgen. Wenn wir eine „Folge“ $a(n_1, n_2, \dots, n_m)$ in mehreren Variablen haben, dann können wir dazu die Diagonalfolge $a_n = a(n, n, \dots, n)$ bilden. Die zugehörige Operation an den Erzeugenden Funktionen heißt entsprechend.

Definition 11.15. Sei $F(X_1, \dots, X_m) \in K[\![X_1, \dots, X_m]\!]$ eine Potenzreihe in m Variablen. Dann heißt die Reihe

$$(\operatorname{diag} F)(X) = \sum_{n=0}^{\infty} ([X_1^n \dots X_m^n]F)X^n \in K[\![X]\!]$$

die *Diagonalreihe* von F.

Nun gilt folgendes (ebenfalls ohne Beweis).

Satz 11.16. *Sei $F(X_1,\dots,X_m)\in K[\![X_1,\dots,X_m]\!]$ eine rationale Potenzreihe in m Variablen (d.h. $F(X_1,\dots,X_m)=P(X_1,\dots,X_m)/Q(X_1,\dots,X_m)$ mit Polynomen $P,Q\in K[X_1,\dots,X_m]$ und $Q(0,\dots,0)=1$). Dann ist die Diagonalreihe $\operatorname{diag}F\in K[\![X]\!]$ D-finit.*

Im Falle $m=2$ ist $\operatorname{diag}F$ *sogar algebraisch; das gilt jedoch im allgemeinen nicht mehr, wenn $m\ge 3$ ist.*

Als typisches Beispiel können wir die Reihe

$$F(X_1,X_2)=\frac{1}{1-X_1-X_2}=\sum_{m,n=0}^{\infty}\binom{m+n}{n}X_1^mX_2^n$$

betrachten. Für ihre Diagonale gilt

$$(\operatorname{diag}F)(X)=\sum_{n=0}^{\infty}\binom{2n}{n}X^n=\frac{1}{\sqrt{1-4X}}\,;$$

das ist offensichtlich eine algebraische Reihe. Dagegen ist

$$F(X)=\Big(\operatorname{diag}\frac{1}{1-X_1-X_2-X_3}\Big)(X)=\sum_{n=0}^{\infty}\frac{(3n)!}{n!^3}X^n$$

zwar D-finit (die Koeffizienten erfüllen eine lineare Rekursionsgleichung der Länge 1), aber nicht algebraisch. Letzteres liegt (z.B.) an der Asymptotik (benutze die Stirlingsche Formel)

$$\frac{(3n)!}{n!^3}=\frac{3^{3n}n^{3n}e^{-3n}\sqrt{6\pi n}}{(n^ne^{-n}\sqrt{2\pi n})^3}\big(1+O(1/n)\big)=\Big(\frac{\sqrt{3}}{2\pi n}+O(n^{-2})\Big)27^n\,.$$

Daraus folgt, daß $F(27z)-\frac{\sqrt{3}}{2\pi}\log\frac{1}{1-z}$ für $z\to 1$ beschränkt ist, d.h. F hat eine logarithmische Singularität bei $1/27$, was bei einer algebraischen Funktion nicht passieren kann.

12. Algorithmischer Beweis kombinatorischer Identitäten

Dieses letzte Kapitel der Vorlesung ist im wesentlichen eine Wiederholung meines Habilitationsvortrags vom 30. Juni 1999. Der Einfachheit halber füge ich deswegen hier das Skript dieses Vortrags ein.

12.1. **Das Problem.** In diesem Vortrag geht es um *Identitäten*. Was ist eine Identität? Das ist eine Aussage, die zwei Dinge gleich setzt, in den meisten Fällen ein kompliziertes Ding mit einem einfachen Ding. Ein Beispiel ist

$$6\cdot 9=42\,.$$

Links vom Gleichheitszeichen steht etwas Kompliziertes, nämlich ein Produkt, während rechts etwas Einfaches steht, nämlich eine Zahl. Sie haben natürlich sofort gesehen, daß diese Identität *falsch* ist (es wird die einzige falsche Identität in diesem Vortrag bleiben). Warum konnten Sie das sofort sehen? Weil Sie (wie

wir alle) in der Grundschule einen Algorithmus gelernt haben, mit dem sich komplizierte Ausdrücke wie auf der linken Seite in einfache eindeutige *Normalformen* umwandeln lassen. Die Normalform der linken Seite ist 54 und damit von der rechten Seite (die bereits in Normalform ist) verschieden.

Andere Arten von Identitäten, die sich routinemäßig entscheiden lassen, werden zum Beispiel repräsentiert von

$$(x+y+z)(x^2+y^2+z^2-xy-yz-zx) = x^3+y^3+z^3-3xyz$$
$$\frac{2\tan\varphi}{1+\tan^2\varphi} = \sin 2\varphi$$
$$\int_0^t \sqrt{\frac{x}{1-x}} = \arcsin\sqrt{t} - \sqrt{t(1-t)}\,.$$

Es ist eine andere Sache, zu gegebener linker Seite eine einfache rechte Seite zu finden. Im ersten der obigen drei Beispiele gibt es noch Normalformen (die rechte Seite ist ein Beispiel dafür). Auch für trigonometrische Ausdrücke wie im zweiten Beispiel lassen sich noch Normalformen finden (das ist nicht mehr ganz so offensichtlich). Das dritte Beispiel führt auf die Frage, wann sich das unbestimmte Integral einer „elementaren“ Funktion wieder als elementare Funktion schreiben läßt. Dieses Problem hat eine algorithmische Lösung (Risch 1970), die inzwischen in Systeme wie Maple oder Mathematica eingebaut ist. Wir alle haben uns daran gewöhnt, unsere Integrale von solch einem System ausrechnen zu lassen.

Die Identitäten, um die es hier geht, sind in gewisser Weise diskrete Analoga zu der Integral-Identität oben (eigentlich eher zu bestimmten Integralen mit Parametern). Wir wollen folgende drei Probleme betrachten.

(1) Gilt $\sum_k F(n,k) = f(n)$ für alle $n \in \mathbb{N}$? Zum Beispiel:
$$\sum_k \binom{n}{k}^2 = \binom{2n}{n}$$

(2) Gilt $\sum_k F(n,k) = \sum_k G(n,k)$ für alle $n \in \mathbb{N}$? Zum Beispiel:
$$\sum_k \binom{n}{k}^3 = \sum_k \binom{n}{k}^2 \binom{2k}{n}$$

(3) Läßt sich $\sum_k F(n,k)$ in „geschlossener Form“ hinschreiben? Zum Beispiel:
$$\sum_k (-1)^k \frac{(4n+k)!(4n-k)!}{(n+k)!^4 (n-k)!^4} = ?$$

Die Summe erstreckt sich dabei über alle ganzen Zahlen, wenn keine Einschränkungen angegeben sind. Wir müssen natürlich präzisieren, welche Art von Funktionen F, G, f wir betrachten wollen.

Definition. L sei ein Körper der Charakteristik 0.

(a) Eine Funktion $f : \mathbb{N} \longrightarrow L$ heißt *hypergeometrisch*, wenn sie eine Gleichung
$$p_0(n) f(n) = p_1(n) f(n+1) \quad \text{für alle } n \in \mathbb{N}$$

erfüllt mit Polynomen $p_0, p_1 \in L[X]$, die nicht beide null sind.

Typische Beispiele sind

$$f(n) = n!\,, \quad \frac{1}{n+1}\binom{2n}{n}, \quad \frac{1}{x}\binom{x+n}{n}^{-1}$$

(letzteres mit $L = \mathbb{Q}(x)$), hingegen ist zum Beispiel

$$f(n) = n^n$$

nicht hypergeometrisch.

(b) Eine Funktion $f : \mathbb{N} \longrightarrow L$ heißt in (hypergeometrischer) *geschlossener Form* darstellbar, wenn sie eine endliche Summe von hypergeometrischen Funktionen ist.

Zum Beispiel haben die Fibonacci-Zahlen über $L = \mathbb{Q}(\sqrt{5})$ die geschlossene Form

$$f(n) = \frac{1}{\sqrt{5}}\left(\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n\right).$$

(c) Eine Funktion $F : \mathbb{N} \times \mathbb{Z} \longrightarrow L$ heißt *einfach hypergeometrisch*, falls sie sich schreiben läßt als

$$F(n,k) = P(n,k)x^k y - \prod_{i=1}^{m}(a_i n + b_i k + c_i)!^{e_i}\,,$$

wobei $P \in L[X,Y]$, $a_i, b_i, e_i \in \mathbb{Z}$ und $x, y, c_i \in L$. (Die Interpretation der Fakultät ist ein bißchen problematisch, wenn c_i keine ganze Zahl ist. Wir können uns statt dessen $(a_i n + b_i k + c_i)!/c_i!$ vorstellen, was eine rationale Funktion von c_i ist.)

Typische Beispiele sind

$$F(n,k) = (-1)^k\binom{2n}{k}^3, \quad (3k-2n)\binom{n}{k}^2\binom{2k}{k}.$$

Wenn wir für unsere Probleme spezifizieren, daß F und G einfach hypergeometrisch und f hypergeometrisch sein sollen, dann haben wir eine präzise gestellte Aufgabe vor uns (in Problem (3) ist natürlich der gerade definierte Begriff der geschlossenen Form gemeint).

12.2. **Die Theorie.** Wie können wir eine Identität wie

$$\sum_k F(n,k) = \sum_k \binom{n}{k}^2 = \binom{2n}{n} = f(n)$$

beweisen? Die rechte Seite f ist hypergeometrisch; im Beispiel gilt

$$2(2n+1)f(n) - (n+1)f(n+1) = 0 \quad \text{für alle } n \in \mathbb{N}.$$

Wenn wir also zeigen können, daß die linke Seite ebenfalls diese Rekursionsgleichung erfüllt, dann sind wir fertig, denn $\sum_k F(0,k) = 1 = f(0)$, und durch die Rekursion sind alle anderen Werte eindeutig bestimmt.

Nun können wir sicher nicht erwarten, daß jeder Ausdruck der Form $\sum_k F(n,k)$ einer hypergeometrischen Rekursion genügt (sonst wären ja alle diese Summen

in geschlossener Form ausdrückbar), aber wir können wenigstens hoffen, daß $f(n) = \sum_k F(n,k)$ einer allgemeineren Gleichung

$$p_0(n)f(n) + p_1(n)f(n+1) + p_2(n)f(n+2) + \cdots + p_d(n)f(n+d) = 0$$

für alle $n \in \mathbb{N}$ genügt, wobei die $p_i \in L[X]$ sind und $p_d \neq 0$ ist. Solche Funktionen nennt man *P-rekursiv* (von „polynomial linear rekursiv“). In diesem zweiten Teil werden wir zeigen, daß dem wirklich so ist.

Dazu ist es nötig, den Begriff der P-Rekursivität in geeigneter Weise auf Funktionen in mehreren Variablen auszudehnen. Eine naive Verallgemeinerung führt zu Problemen. Schließlich hat sich herausgestellt, daß die weiter unten folgende Definition sinnvoll ist.

Bevor wir diese Definition formulieren können, müssen wir einige Objekte einführen.

Definition. Sei $r \geq 1$ eine ganze Zahl.

(a) Die nicht-kommutative L-Algebra $L\langle n_1, \ldots, n_r, N_1, \ldots, N_r\rangle$, die von den Variablen $n_1, \ldots, n_r$, $N_1, \ldots, N_r$ mit den Vertauschungsrelationen

$$n_i n_j = n_j n_i\,, \quad N_i N_j = N_j N_i\,, \quad N_i n_j = (n_j + \delta_{ij}) N_i$$

erzeugt wird, heißt $\mathcal{A}_r(L)$ oder einfach $\mathcal{A}_r$.

(b) Für $m \geq 0$ sei $\mathcal{A}_r^{\leq m} = \mathcal{A}_r^{\leq m}(L)$ der L-Unterraum von $\mathcal{A}_r$, der von allen Monomen $n_1^{e_1} \ldots n_r^{e_r} N_1^{e_{r+1}} \ldots N_r^{e_{2r}}$ mit $0 \leq e_j \leq m$ für alle $1 \leq j \leq 2r$ erzeugt wird. (Aus den Vertauschungsrelationen folgt, daß $\mathcal{A}_r^{\leq m}\mathcal{A}_r^{\leq n} \subset \mathcal{A}_r^{\leq m+n}$ ist; wir haben also eine Filtrierung auf $\mathcal{A}_r$ definiert.)

Sei nun $D = D_1 \times \cdots \times D_r$ mit $D_j = \mathbb{N}$ oder $\mathbb{Z}$ und $\mathcal{S}_D = \mathcal{S}_D(L)$ der Raum aller Funktionen $F : D \longrightarrow L$. Dieser Raum $\mathcal{S}_D$ wird ein $\mathcal{A}_r$-Modul, indem wir setzen

$$\begin{aligned}(n_j \cdot F)(\nu_1, \ldots, \nu_r) &= \nu_j F(\nu_1, \ldots, \nu_r)\\ (N_j \cdot F)(\nu_1, \ldots, \nu_r) &= F(\nu_1, \ldots, \nu_{j-1}, \nu_j + 1, \nu_{j+1}, \ldots, \nu_r)\,.\end{aligned}$$

Jetzt können wir endlich die angekündigte Definition formulieren.

Lemma und Definition. Für $F \in \mathcal{S}_D$ sind folgende Aussagen äquivalent.

(i) Es gibt $C \geq 0$, so daß für alle $m \geq 1$ gilt:

$$\dim_L \mathcal{A}_r^{\leq m}(L) \cdot F \leq C\, m^r\,.$$

(ii) Für jede $(r+1)$-elementige Teilmenge $X \subset \{n_1, \ldots, n_r, N_1, \ldots, N_r\}$ gibt es $0 \neq P_X \in L\langle X\rangle \subset \mathcal{A}_r(L)$ mit $P_X \cdot F = 0$.

Eine Funktion $F \in \mathcal{S}_D$, die diese Eigenschaften hat, heißt *holonom*. Die Menge der holonomen Funktionen in $\mathcal{S}_D$ sei mit $\mathcal{H}_D$ bezeichnet.

Wir wollen den Beweis wenigstens andeuten. Aus (i) folgt (ii), denn

$$\dim(\mathcal{A}_r^{\leq m}(L) \cap L\langle X\rangle) = (m+1)^{r+1} > Cm^r \geq \dim \mathcal{A}_r^{\leq m}(L) \cdot F$$

für m groß genug, also hat die Abbildung

$$\mathcal{A}_r^{\leq m}(L) \cap L\langle X\rangle \longrightarrow \mathcal{A}_r^{\leq m}(L) \cdot F\,, \quad P \longmapsto P \cdot F$$

nichttrivialen Kern. Umgekehrt sei μ so gewählt, daß alle P_X aus Eigenschaft (ii) in $\mathcal{A}_r^{\leq\mu}(L)$ liegen. Dann kann man die Relationen P_X dazu verwenden, hohe Potenzen zu eliminieren, so daß

$$\mathcal{A}_r(L) \cdot F \subset V_\mu \cdot F\,,$$

wobei V_μ der von allen Monomen $n_1^{e_1} \dots n_r^{e_r} N_1^{e_{r+1}} \dots N_r^{e_{2r}}$ erzeugt wird, so daß $e_j \geq \mu$ nur für höchstens r Elemente $j \in \{1, \dots, 2r\}$. Da $\dim(V_\mu \cap \mathcal{A}_r^{\leq m}(L)) \leq Cm^r$ für geeignetes C, folgt Eigenschaft (i).

Die Menge $\mathcal{H}_\mathbb{N}$ besteht also gerade aus den P-rekursiven Funktionen. Es gilt, daß $F \in \mathcal{S}_D$ genau dann holonom ist, wenn die triviale Fortsetzung $\tilde{F} \in \mathcal{S}_{\mathbb{Z}^r}$ holonom ist.

Aus den Eigenschaften (i) und (ii) oben lassen sich nun relativ leicht die folgenden Abgeschlossenheitsaussagen über holonome Funktionen gewinnen.

Satz.

(a) $\mathcal{H}_D$ ist ein $\mathcal{A}_r$-Untermodul von $\mathcal{S}_D$.
(b) Mit $F, G \in \mathcal{H}_D$ ist auch das punktweise Produkt $FG \in \mathcal{H}_D$.
(c) Ist $T : \mathbb{Z}^r \longrightarrow \mathbb{Z}^s$ eine $\mathbb{Z}$-lineare Abbildung, so folgt aus $F \in \mathcal{H}_{\mathbb{Z}^s}$, daß $F \circ T \in \mathcal{H}_{\mathbb{Z}^r}$ ist.
(d) Hat $F \in \mathcal{H}_{D\times\mathbb{Z}}$ in der letzten Variablen endlichen Träger (d.h. für alle $\nu \in D$ hat $\nu_{r+1} \mapsto F(\nu, \nu_{r+1})$ endlichen Träger), so ist $\Sigma F \in \mathcal{H}_D$, wobei $(\Sigma F)(\nu) = \sum_{\nu_{r+1}} F(\nu, \nu_{r+1})$.

Da $n!^e$ und $((n+c)!/c!)^e$ P-rekursiv sind, folgt daraus die erstrebte

Folgerung. Ist $F \in \mathcal{S}_{\mathbb{N}\times\mathbb{Z}}$ einfach hypergeometrisch mit endlichem Träger in der zweiten Variablen, so ist $f(n) = \sum_k F(n,k)$ P-rekursiv.

Diese Aussage geht zurück auf Arbeiten in den vierziger Jahren der US-Amerikanischen Nonne Sister Mary Celine Fasenmyer, die mit dieser Methode Rekursionen für Polynome wie etwa die Laguerre-Polynome

$$L_n(x) = \sum_k (-1)^k \binom{n}{k} \frac{x^k}{k!}$$

herleitete. Wieder ausgegraben und (nach einem ersten gescheiterten Anlauf) verallgemeinert wurde die Methode in den achtziger Jahren von Doron Zeilberger.

12.3. **Die Praxis.** Nun wissen wir also, daß unsere Summe $f(n) = \sum_k F(n,k)$ P-rekursiv ist. Um damit etwas anfangen zu können, müssen wir aber in der Lage sein, explizit eine Rekursionsgleichung für f zu finden.

Der erste Ansatz wäre, Sister Celines Methode zu folgen. Eigenschaft (ii) sagt uns, daß es einen Operator $P_{n,N,K} \neq 0$ (wir setzen der Einfachheit halber $n_1 = n, n_2 = k$ und $N_1 = N, N_2 = K$) gibt mit $P_{n,N,K} \cdot F = 0$. Da F einfach hypergeometrisch ist, sind alle $F(n+i, k+j)/F(n,k)$ rationale Funktionen von n und k. Man kann also ein $P_{n,N,K}$ von einem gewissen Grad in N und K mit unbestimmten Koeffizienten ansetzen und wird auf ein lineares Gleichungssystem über $L[n]$ geführt, das nach unserer Theorie für hinreichend hohen Grad

lösbar sein muß. Zum Beispiel findet man für $F(n,k) = \binom{n}{k}^2$ die Gleichung

$$\begin{aligned}&(n+1)\,(F(n,k) - 2F(n,k+1) + F(n,k+2))\\ &\quad - (2n+3)\,(F(n+1,k+1) + F(n+1,k+2)) + (n+2)F(n+2,k+2) = 0\end{aligned}$$

für alle $n \in \mathbb{N}$, $k \in \mathbb{Z}$, oder etwas kürzer

$$\big((n+1)(1-2K+K^2) - (2n+3)(K+K^2)N + (n+2)K^2N^2\big) \cdot F = 0\,.$$

Wenn wir diese Gleichung über alle $k \in \mathbb{Z}$ summieren, erhalten wir

$$-2(2n+3)f(n+1) + (n+2)f(n+2) = 0 \quad \text{für alle } n \in \mathbb{N},$$

woraus mit den Anfangswerten $f(0) = 1$, $f(1) = 2$ sehr schnell folgt, daß $f(n) = \binom{2n}{n}$ ist.

Diese Methode funktioniert, ist aber sehr langsam, da man im allgemeinen recht große Gleichungssysteme über $L[n]$ lösen muß. Hier hatte nun Zeilberger einen großen Einfall: Wir können den Operator $P_{n,N,K}$, den wir eben gefunden haben, schreiben als

$$P_{n,N,K} = P - (K-1)Q$$

mit

$$\begin{aligned}P &= -2(2n+3)N + (n+2)N^2 \quad \text{und}\\ Q &= -(n+1)(K-1) + (2n+3)(K+2)N - (n+2)(K+1)N^2\,.\end{aligned}$$

Also gilt, wenn wir $G = Q \cdot F$ setzen,

$$P \cdot F = (K-1) \cdot G$$

d.h. ausgeschrieben

$$-2(2n+3)F(n+1,k) + (n+2)F(n+2,k) = G(n,k+1) - G(n,k)\,.$$

Wenn wir diese Gleichung über k summieren, erhalten wir sofort

$$-2(2n+3)f(n+1) + (n+2)f(n+2) = 0\,,$$

da die rechte Seite eine Teleskopsumme liefert (G hat wie F endlichen Träger in k).

Wenn wir beachten, daß $G(n,k) = R(n,k)F(n,k)$ ist mit einer rationalen Funktion R (das kommt daher, daß F einfach hypergeometrisch ist), dann können wir uns also alternativ die folgende Aufgabe stellen:

Gegeben F, finde $P \in L\langle n, N\rangle$ und $R \in L(n,k)$, so daß $P \cdot F = (K-1) \cdot G$ mit $G = RF$ ist!

Bevor wir diese Aufgabe allgemein lösen, betrachten wir einen Spezialfall. Wir nehmen an, $P = 1$. Dann kommt n gar nicht mehr explizit vor, und wir können das Problem formulieren als:

Gegeben $f : \mathbb{N} \longrightarrow L$ hypergeometrisch, finde $g : \mathbb{N} \longrightarrow L$ hypergeometrisch mit $g(k+1) - g(k) = f(k)$ für alle $k \in \mathbb{N}$, falls ein solches g existiert!

Gosper hat 1977 einen Algorithmus gefunden, der diese Aufgabe löst. Es ist leicht zu sehen, daß $g(k) = r(k)f(k)$ sein muß mit einer rationalen Funktion $r(k)$. Die wesentliche Idee Gospers war, das Problem auf eine lineare Gleichung für ein *Polynom* zu reduzieren, dessen Grad man a priori beschränken kann.

Zeilberger hat Gospers Algorithmus erweitert, so daß er die allgemeinere Aufgabe löst. (Man setzt P von einem gewissen Grad in N an; die unbestimmten Koeffizienten $p_j(n)$ treten dann zusammen mit den Koeffizienten des unbekannten Polynoms in Gospers Algorithmus linear in der zu lösenden Gleichung auf.)

In unserem Beispiel $F(n,k) = \binom{n}{k}^2$ liefert uns dieser Algorithmus direkt (und viel schneller) das Resultat

$$P = 2(2n+1) - (n+1)N \quad \text{und} \quad R(n,k) = \frac{k^2(3n-2k+3)}{(n-k+1)^2}\,.$$

Beachten Sie, daß man die Behauptung $P \cdot F = (K-1)\cdot(RF)$ leicht durch eine Routine-Rechnung nachprüfen kann! Wir haben

$$G(n,k) = R(n,k)F(n,k) = \frac{k^2(3n-2k+3)}{(n-k+1)^2}\frac{n!^2}{k!^2(n-k)!^2} = (3n-2k+3)\binom{n}{k-1}^2,$$

also

$$\begin{aligned}
&G(n,k+1) - G(n,k)\\
&\qquad = (3n-2k+1)\binom{n}{k}^2 - (3n-2k+3)\binom{n}{k-1}^2\\
&\qquad = \left((3n-2k+1)(n-k+1)^2 - (3n-2k+3)k^2\right)\frac{n!}{k!^2(n-k+1)!^2}\\
&\qquad = \left(2(2n+1)(n-k+1)^2 - (n+1)^3\right)\frac{n!}{k!^2(n-k+1)!^2}\\
&\qquad = 2(2n+1)\binom{n}{k}^2 - (n+1)\binom{n+1}{k}^2\\
&\qquad = 2(2n+1)F(n,k) - (n+1)F(n+1,k)\,.
\end{aligned}$$

Das gilt natürlich analog für jedes Resultat, das dieser Algorithmus liefert. Mit anderen Worten: Der Algorithmus *zertifiziert* sein Ergebnis! Man muß dem Computer also nicht blind vertrauen, sondern kann sich leicht selbst von der Richtigkeit des Ergebnisses überzeugen.

Nun haben wir eine vollständige Lösung unseres Problems (1) an der Hand.

1. Wende Zeilbergers Algorithmus auf $F(n,k)$ an. Man erhält P und R wie oben. Falls gewünscht, prüfe das Resultat.
2. Die Summe $\sum_k F(n,k)$ erfüllt die Rekursionsgleichung P. Prüfe nach, daß die rechte Seite $f(n)$ ebenfalls diese Gleichung erfüllt.
3. Prüfe genügend (aber endlich viele) Anfangswerte auf Gleichheit.

Ein Beispiel:

$$\sum_k \binom{n}{k}\frac{(-1)^k}{x+k} = \frac{1}{x}\binom{x+n}{n}^{-1} = \frac{n!}{x(x+1)\dots(x+n)}\,.$$

Hier ist also $F(n,k) = \binom{n}{k}\frac{(-1)^k}{x+k} \in \mathbb{Q}(x)$. Zeilbergers Algorithmus liefert

$$P = (n+1) - (n+x+1)N \quad \text{und} \quad R = \frac{k(x+k)}{(n-k+1)}\,.$$

Die rechte Seite $f(n)$ erfüllt die Rekursionsgleichung, und $\sum_k F(0,k) = 1/x = f(0)$, also ist die Identität bewiesen.

Das Problem (2) läßt sich ähnlich angehen, indem man den Algorithmus auf beide Seiten der Gleichung anwendet.

Man kann auch gewissermaßen umgekehrt an das Problem (1) herangehen und für P gleich die Rekursion einsetzen, der die rechte Seite genügt. Dann kann man Gospers Algorithmus auf die Funktion $\mathbb{Z} \ni k \mapsto (P \cdot F)(n,k)$ ansetzen (wobei n als Parameter behandelt wird). Wenn man Glück hat, ist der Algorithmus erfolgreich, und man ist schneller am Ziel. In der Praxis hat man offenbar meistens Glück, aber es gibt Fälle, wo diese Methode nicht funktioniert. Der Zeilberger-Algorithmus liefert dann eine Rekursion der Länge ≥ 2. (Im Spezialfall $f(n) = 1$ nennen die Autoren Wilf und Zeilberger den so gefundenen Beweis und sein Zertifikat R ganz bescheiden einen „WZ-Beweis“ und sein „WZ-Zertifikat“.)

Das führt uns zum Problem (3). Wenn wir eine Summe $f(n) = \sum_k F(n,k)$ vor uns haben und sich keine offensichtliche Vermutung über die Form von $f(n)$ anbietet (oft kann man $f(n)$ erraten, aber eben nicht immer), dann können wir in jedem Fall unseren Algorithmus anwenden. Bekommen wir eine Rekursion der Länge 1, dann können wir daraus die geschlossene Form leicht bestimmen.

Zum Beispiel sehen die Anfangswerte

$$1, 486, 3543750, 46003313664, 772679415543750, 15025186795291909236, \dots ,$$

die zu

$$F(n,k) = (-1)^k \frac{(4n+k)!(4n-k)!}{(n+k)!^4 (n-k)!^4} ,$$

gehören, eher abschreckend aus. Der Algorithmus gibt uns die Rekursion

$$P = 81(4n+1)(4n+3)(3n+1)^4(3n+2)^4 - 8(2n+1)^5(n+1)^5 N ,$$

woraus man leicht ableitet, daß

$$\sum_k (-1)^k \frac{(4n+k)!(4n-k)!}{(n+k)!^4 (n-k)!^4} = \frac{(4n)!(3n)!^4}{(2n)!^6 n!^4} = \binom{4n}{2n} \binom{3n}{n}^4 .$$

Wenn allerdings die Rekursion größere Länge hat, dann stehen wir vor dem Problem, daß wir feststellen müssen, ob diese Gleichung hypergeometrische Lösungen hat. Zum Glück gibt es auch hierfür einen Algorithmus, der von Petkovšek Anfang der neunziger Jahre gefunden wurde, so daß auch dieses Problem vollständig algorithmisch lösbar ist.

Als ein einfaches Beispiel sei

$$f(n) = \sum_k (-1)^k \binom{n}{k} \binom{3k}{n} = ?$$

betrachtet. Der Zeilberger-Algorithmus liefert uns

$$9(n+1)f(n) + 3(5n+7)f(n+1) + 2(2n+3)f(n+2) = 0 ,$$

eine Rekursion der Länge zwei. Petkovšeks Algorithmus sagt uns dann, daß $(-3)^n$ diese Rekursion löst (wir haben die Faktorisierung

$$9(n+1) + 3(5n+7)N + 2(2n+3)N^2 = (3(n+1) + 2(2n+3)N)(3+N)$$

in $\mathcal{A}_1$). Da $f(0) = 1$ und $f(1) = -3$, ist tatsächlich $f(n) = (-3)^n$. In diesem Fall hätte man das Ergebnis natürlich auch leicht erraten können.

Anders gelagert ist

$$f(n) = \sum_k \binom{n}{k}^3 = ?\,.$$

Hier bekommen wir

$$8(n+1)^2 f(n) + (7n^2 + 21n + 16) f(n+1) - (n+2)^2 f(n+2) = 0\,,$$

und Petkovšeks Algorithmus sagt uns, daß es keine hypergeometrischen Lösungen dieser Gleichung gibt. $f(n)$ ist also nicht in geschlossener Form darstellbar.

Dieselbe Rekursion erhält man übrigens für die Folge

$$g(n) = \sum_k \binom{n}{k}^2 \binom{2k}{n}\,.$$

Zusammen mit den beiden Anfangswerten

$$f(0) = g(0) = 1\,, \qquad f(1) = g(1) = 2$$

folgt dann $f(n) = g(n)$ für alle $n \in \mathbb{N}$.

Literatur

[1] H.-R. Halder, W. Heise: *Einführung in die Kombinatorik*, Carl Hanser Verlag, München, Wien (1976).

[2] J.H. van Lint, R.M. Wilson: *A course in combinatorics*, Cambridge University Press (1992).

[3] M. Petkovšek, H.S. Wilf, D. Zeilberger: $A = B$, A.K. Peters, Wellesley, MA (1996). Kann frei aus dem Internet bezogen werden.

[4] R.P. Stanley: *Enumerative Combintarics, Vol. I, Vol. II*, Cambdige studies in advanced mathematics **49**, **62**, Cambridge University Press (1997, 1999).

[5] H.S. Wilf: *generatingfunctionology*, 2nd ed., Academic Press (1994). Kann frei aus dem Internet bezogen werden.