IMO-Herbstzirkus 2022: Funktionalgleichungen
Christian Bernert

1 Eine Aufgabe aus dem Kosovo und sehr viele Monster

1.1 Vorrede: Worum es geht

In diesem Brief geht es um *Funktionalgleichungen*. Für einige von euch ist das vielleicht ein ganz neuer Aufgabentyp. Ihr kennt sicherlich Gleichungen und Gleichungssysteme. Gesucht sind jeweils Variablen/Unbekannte, die eine oder mehrere gegebene Gleichungen erfüllen.
Bei einer Funktionalgleichung gibt es auch Variablen/Unbekannte (meistens nur eine), die eine Gleichung erfüllen sollen. Allerdings stehen diese Unbekannten nicht für *Zahlen*, sondern für *Funktionen*.
Ein typisches Beispiel einer Funktionalgleichung ist die folgende Aufgabe aus der IMO 2015:

> **Aufgabe:** Finde alle Funktionen $f : \mathbb{R} \to \mathbb{R}$, die
>
> $$f(x + f(x + y)) + f(xy) = x + f(x + y) + yf(x)$$
>
> für alle $x, y \in \mathbb{R}$ erfüllen.

Diese Aufgabe ist für den Einstieg vielleicht etwas zu schwer, um sie auf Anhieb zu lösen, aber wir können hier direkt einige wichtige Merkmale von Funktionalgleichungen sehen, auf die man bereits beim Lesen der Aufgabenstellung genau achten sollte:

- Was für eine Funktion ist gesucht? Ist eine Funktion $f : A \to B$ gesucht, so heißt A der *Definitionsbereich* und B der *Wertebereich* von f. In unserem Fall sind beide Mengen die Menge der reellen Zahlen $\mathbb{R}$, typische andere Beispiele sind die Menge der ganzen Zahlen $\mathbb{Z}$, die Menge der positiven ganzen Zahlen $\mathbb{N}$ oder $\mathbb{Z}_{>0}$, die Menge der rationalen Zahlen $\mathbb{Q}$, die Menge der positiven reellen Zahlen $\mathbb{R}_{>0}$ usw. Im Allgemeinen müssen auch Definitions- und Wertebereich nicht übereinstimmen.

  Wichtig: Dass B der Wertebereich genannt wird, bedeutet *nicht*, dass alle Werte in B auch tatsächlich angenommen werden. Zum Beispiel ist die konstante Funktion mit $f(x) = 42$ für alle $x \in \mathbb{R}$ auch eine Funktion $f : \mathbb{R} \to \mathbb{R}$ (die allerdings in diesem Fall nicht die obige Funktionalgleichung erfüllt).

- Für welche Zahlen gilt die Funktionalgleichung? Hier gilt diese für alle $x, y \in \mathbb{R}$. Typischerweise ist der gesamte Definitionsbereich erlaubt, aber es gibt auch andere Beispiele. Man sollte hier besonders gut Acht geben, denn ist der erlaubte Bereich etwa nur $\mathbb{R}_{>0}$, so darf z.B. 0 nicht eingesetzt werden. Das macht eine solche Aufgabe oft deutlich schwieriger.

- Was ist das *Ergebnis* einer Funktionalgleichung? Von einer Gleichung oder einem Gleichungssystem kennen wir es, dass es eine, mehrere, unendlich viele, aber auch keine Lösung geben kann. Genauso ist es auch bei Funktionalgleichungen. Man sieht vielleicht mit bloßem Auge, dass die Funktion $f(x) = x$ die Bedingung erfüllt. Mit etwas mehr Aufwand findet man vielleicht auch die Lösung $f(x) = 2 - x$. Eine mögliche Antwort auf diese Aufgabe (ob diese richtig ist, sei einmal dahingestellt) wäre also, dass genau diese beiden Funktionen die Bedingung erfüllen. Eine vollständige Lösung besteht dann einerseits aus einer *Probe*, dass diese Funktionen auch tatsächlich die Bedingung erfüllen, und andererseits aus einem *Beweis*, dass es keine weiteren Lösungen gibt.

  Die Probe ist dabei meist recht einfach und erfordert im schlimmsten Fall eine kurze Rechnung, der schwierige Teil ist natürlich der Nachweis, dass damit bereits alle Lösungen gefunden wurden.

1.2 Eine Beispiel-Aufgabe

Wir zeigen nun anhand einer relativ einfachen Aufgabe aus der Olympiade des Kosovo, wie die Lösung einer Funktionalgleichung aussehen kann.

Aufgabe: Finde alle Funktionen $f : \mathbb{R} \to \mathbb{R}$ mit

$$f(f(x-y) - yf(x)) = xf(y)$$

für alle $x, y \in \mathbb{R}$.

Lösung: Wir beginnen damit, *spezielle Werte einzusetzen*. Dabei bietet es sich an, dies so geschickt zu tun, dass sich die in der Funktionalgleichung auftretenden Terme möglichst *vereinfachen*. Setzen wir etwa $y = 0$ ein, so sehen wir

$$f(f(x)) = xf(0) \tag{1}$$

für alle $x \in \mathbb{R}$. Da hier die unbekannte reelle Zahl $f(0)$ vorkommt und wir uns Schreibarbeit sparen wollen, können wir hierfür eine neue Variable einführen, etwa $a = f(0)$, es gilt also $f(f(x)) = ax$ für alle $x \in \mathbb{R}$. Setzen wir in (1) auch noch $x = 0$, erhalten wir insbesondere $f(a) = 0$. Setzen wir in (1) stattdessen $x = a$, erhalten wir $a^2 = f(f(a)) = f(0) = a$, also $a = a^2$ und damit $a = 0$ oder $a = 1$. Wir haben also am Ende eine Gleichung für die Unbekannte a gefunden und müssen nun zwei Fälle unterscheiden.
1. Fall: $a = 0$, also $f(0) = 0$. Dann sagt (1) einfach $f(f(x)) = 0$ für alle $x \in \mathbb{R}$. Setzen wir jetzt $x = f(z)$ ein, erhalten wir

$$f(f(f(z) - y) - yf(f(z))) = f(y)f(z)$$

für alle $y, z \in \mathbb{R}$. Dies sieht furchtbar kompliziert aus, warum haben wir dies gemacht? Das Ziel war, die linke Seite zu vereinfachen. Dadurch, dass wir für x einen Funktionswert von f eingesetzt haben, nämlich $f(z)$, folgt jetzt $f(x) = f(f(z)) = 0$. Auf der linken Seite bleibt dann nur noch $f(f(f(z) - y))$, was zwar immer noch kompliziert aussieht, aber aus dem gleichen Grund 0 ist! Also folgt $f(y)f(z) = 0$ für alle $y, z \in \mathbb{R}$ und insbesondere mit $y = z$ sofort $\boxed{f(y) = 0}$ für alle $y \in \mathbb{R}$. Probe: Für diese Funktion wird die Ausgangsgleichung zu $0 = 0$, ist also stets erfüllt, wir haben damit eine Lösung unserer Funktionalgleichung gefunden und gezeigt, dass es in diesem Fall keine weitere gibt.

2. Fall: $a = 1$, also $f(0) = 1$. Dann sagt (1) einfach $f(f(x)) = x$, insbesondere erhalten wir durch Einsetzen von $x = 0$ auch $f(1) = 0$.
Jetzt setzen wir $x = 1$ in die Ausgangsgleichung ein und erhalten

$$f(f(1-y) - yf(1)) = f(y)$$

für alle $y \in \mathbb{R}$. Dies sieht wieder zunächst kompliziert aus, wegen $f(1) = 0$ ist die linke Seite aber einfach $f(f(1-y))$, was wie eben hergeleitet einfach $1 - y$ ist.
Also steht dort tatsächlich sofort $\boxed{f(y) = 1 - y}$ für alle $y \in \mathbb{R}$. Wir haben also bereits gezeigt, dass höchstens diese Funktion die Funktionalgleichung in diesem Fall erfüllen kann. Es fehlt allein noch die Probe: Setzen wir $f(x) = 1 - x$ überall in die Ausgangsgleichung ein, steht dort

$$1 - (1 - (x - y) - y(1 - x)) = x(1 - y)$$

und nach Ausmultiplizieren bleibt auf beiden Seiten $x - xy$, also tatsächlich das gleiche, auch diese zweite Funktion erfüllt damit die Funktionalgleichung.
Insgesamt hat unsere Aufgabe also genau zwei Lösungen, nämlich $f(x) = 0$ für alle $x \in \mathbb{R}$ und $f(x) = 1 - x$ für alle $x \in \mathbb{R}$.

1.3 Cauchy und seine Monster

Wir beginnen mit einer kurzen Zwischenbemerkung zum Begriff der *Funktion*: Eine Funktion $f: A \to B$ kann a priori wirklich eine beliebige Vorschrift sein, die jedem Element aus A genau ein Element aus B zuordnet. Diese muss weder linear, noch quadratisch, noch ein Polynom, noch überhaupt eine Funktion sein, die eine besonders elegante Form hat.

Beispiele für Funktionen $f: \mathbb{R} \to \mathbb{R}$ sind also auch $f(x) = \begin{cases} 0 & x \neq 0 \\ 42 & x = 0 \end{cases}$ oder $f(x) = \begin{cases} 0 & x \in \mathbb{Q} \\ 42 & x \notin \mathbb{Q} \end{cases}$.
Insbesondere folgt etwa aus $f(x) = 0$ für alle $x \in \mathbb{Q}$ noch lange nicht, dass das gleiche auch für alle $x \in \mathbb{R}$ gilt.

Naturgemäß werden derartige Funktionen selten Lösungen zu „schönen" Funktionalgleichungen sein – zumindest in unseren bisherigen Beispielen waren die Lösungen allesamt sehr einfache Funktionen–, aber darauf sollte man sich erstens nicht verlassen (es gibt auch durchaus Olympiade-Aufgaben, bei denen sehr merkwürdige Funktionen als Lösung herauskommen) und zweitens muss man die Existenz solcher Lösungen in einem vollständigen Beweis natürlich trotzdem ausschließen, selbst wenn man sich sehr sicher ist, dass es keine solche verrückte Lösung gibt. Ein typisches Phänomen ist also, dass man vielleicht glaubt, dass die einzige Lösung $f: \mathbb{R} \to \mathbb{R}$ die Funktion $f(x) = x^2 + 1$ ist und bereits $f(x) = x^2 + 1$ für alle $x \in \mathbb{Z}$ oder vielleicht sogar für alle $x \in \mathbb{Q}$ gezeigt hat. Es wäre aber ein Trugschluss zu glauben, dass man damit bereits fertig oder auch nur so gut wie fertig ist. Oft genug ist der schwierigste Teil der Aufgabe dann noch, dies auch wirklich auf alle reellen Zahlen fortzusetzen.

Mit den Methoden, wie man so etwas angeht, werden wir uns beim nächsten Mal beschäftigen. Heute diskutieren wir zum Abschluss noch ein sehr bekanntes und fundamentales Beispiel einer Funktionalgleichung, bei der es gar nicht so einfach ist, alle Lösungen zu finden.

Aufgabe: Finde alle Funktionen $f: \mathbb{R} \to \mathbb{R}$, die die *Cauchy-Funktionalgleichung*[a]

$$f(x + y) = f(x) + f(y)$$

für alle $x, y \in \mathbb{R}$ erfüllen.

Lösung (Beginn):
Einsetzen von $x = y = 0$ zeigt $f(0) = 0$.
Einsetzen von $x = y$ zeigt $f(2x) = 2f(x)$ für alle $x \in \mathbb{R}$.
Einsetzen von $y = 2x$ zeigt $f(3x) = 3f(x)$ für alle $x \in \mathbb{R}$.
Dies können wir wiederholen und per Induktion leicht $f(nx) = nf(x)$ für alle $x \in \mathbb{R}$ und alle $n \in \mathbb{N}$ zeigen. Setzen wir stattdessen $y = -x$ ein, sehen wir $f(-x) = -f(x)$ und somit gilt $f(nx) = nf(x)$ auch für alle $x \in \mathbb{R}$ und $n \in \mathbb{Z}$.

Wir können noch etwas weiter gehen: Ist $q = \frac{r}{s} \in \mathbb{Q}$, so folgt wie eben hergeleitet jetzt

$$f(qx) = f\left(\frac{rx}{s}\right) = \frac{f(rx)}{s} = \frac{rf(x)}{s} = qf(x)$$

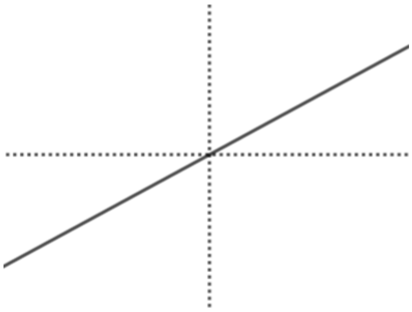
für alle $x \in \mathbb{R}$, also gilt $f(qx) = qf(x)$ sogar für alle $x \in \mathbb{R}$ und $q \in \mathbb{Q}$.
Insbesondere folgt sofort $f(q) = qf(1)$ für alle $q \in \mathbb{Q}$. Wären also nur Lösungen $f: \mathbb{Q} \to \mathbb{Q}$ der Cauchy-Funktionalgleichung gesucht, wären wir bereits fertig: Eine solche Funktion muss von der Form $f(x) = cx$ für eine beliebige Konstante c sein, und man sieht leicht, dass auch jede solche lineare Funktion die Cauchy-Funktionalgleichung erfüllt.

[a] *Augustin-Louis Cauchy* (1789–1857) war ein sehr bedeutender französischer Mathematiker, der sich unter anderem als erster systematisch mit den Lösungen dieser Gleichung beschäftigt hat.

Die Struktur dieses Arguments ist sehr typisch und bei vielen Aufgaben nützlich: Mittels vollständiger Induktion wird eine Aussage für alle natürlichen oder sogar alle ganzen Zahlen gezeigt. Mit etwas Glück kann man das dann auf alle rationalen Zahlen fortsetzen. Auch wenn man wie oben beschrieben damit noch lange nicht auf alle reellen Zahlen schließen kann, hat man damit doch schon einiges über die Funktion herausgefunden, was dann auch im weiteren Verlauf der Lösung nützlich sein kann.

Als nächstes könnte man nun versuchen, zu zeigen, dass $f(x) = cx$ auch für alle reellen Zahlen gelten muss. Dies ist aber im Allgemeinen einfach falsch. Es gilt nämlich der folgende

Satz (über Monster): Neben den linearen Funktionen $f(x) = cx$ gibt es unendlich viele „Monster“, also Lösungen $f : \mathbb{R} \to \mathbb{R}$ der Cauchy-Funktionalgleichung, die nicht von dieser Form sind. Ist f ein Monster, dann ist der Graph von f als Teilmenge der $x-y$-Ebene *dicht*, d.h. zu jedem beliebig kleinen Kreis in der Ebene gibt es einen Punkt $(x, f(x))$ des Funktionsgraphen, der im Inneren dieses Kreises liegt.



(a) Graph einer linearen Funktion

(b) Graph eines Monsters

Während der erste Teil des Satzes also enttäuschend ist (es gibt sehr viele verrückte Lösungen der Cauchy-Gleichung, die wir nicht verstehen), ist der zweite Teil in der Praxis sehr wichtig und hilfreich. Er sagt uns nämlich, dass wir unter sehr schwachen Zusatzbedingungen aus der Cauchy-Gleichung doch die Linearität von f folgern können:

Nützliche Folgerung: Erfüllt $f : \mathbb{R} \to \mathbb{R}$ die Cauchy-Gleichung und *mindestens eine* der folgenden Bedingungen, so ist $f(x) = cx$ für alle $x \in \mathbb{R}$ (mit $c = f(1)$).

a) f ist auf einem beliebigen Intervall nach oben oder nach unten beschränkt.

b) f ist auf einem beliebigen Intervall monoton.

c) f ist auf einem beliebigen Intervall stetig[a].

d) f ist multiplikativ: $f(xy) = f(x)f(y)$

Beweis: Die erste Aussage impliziert sofort, dass der Graph von f nicht dicht in der Ebene liegen kann. Die zweite und dritte Aussage sind einfach stärker als die erste. Auch aus der vierten folgt die erste, denn sie impliziert insbesondere $f(x^2) = f(x)^2 \geq 0$, also ist f z.B. auf dem Intervall $[0, 1]$ nach unten durch 0 beschränkt. $\square$

[a] Was das genau bedeutet, können und wollen wir hier nicht diskutieren.

In der Praxis ist dies häufig dann nützlich, wenn man aus einer komplizierteren Gleichung durch geschicktes Einsetzen/Umformen gefolgert hat, dass auch die Cauchy-Gleichung erfüllt ist. Kann man dann aus der komplizierten Gleichung durch andere Überlegungen noch weitere Informationen, etwa $f(xy) = f(x)f(y)$ gewinnen, so muss man sich über die Monster keine Sorgen machen und kann auf die Linearität der Funktion schließen.

Zum Abschluss wollen wir noch den Beweis des Monster-Satzes skizzieren. Auch wenn ein strenger Beweis der Existenz der Monster etwas subtil ist, ist die Idee eigentlich recht simpel.

Beweisidee des Monster-Satzes, 1. Teil: Die Existenz der Monster ergibt sich aus folgender Überlegung: Wir haben $f(x)$ bereits (nach Wahl von $f(1)$) für alle rationalen Werte von x bestimmt. Über andere Werte können wir zunächst nichts aussagen. Wir wählen also eine beliebige irrationale Zahl, z.B. π, und definieren $f(\pi)$ beliebig, z.B. $f(\pi) = \sqrt{2}$. Jetzt können wir über einige weitere Werte von f etwas aussagen, nämlich für alle rationalen Vielfachen von π und sogar für alle Kombinationen aus diesen und rationalen Zahlen wegen $f(q_1+q_2\pi) = f(q_1)+f(q_2\pi) = q_1 f(1)+q_2 f(\pi)$ für alle rationalen Zahlen q_1, q_2. Nun gibt es aber natürlich irrationale Zahlen, die nicht von der Form $q_1 + q_2\pi$ sind, z.B. $\sqrt{2}$. Wir definieren nun $f(\sqrt{2})$ beliebig und stellen fest, dass jetzt alle Funktionswerte der Form

$$f(q_1 + q_2\pi + q_3\sqrt{2}) = q_1 f(1) + q_2 f(\pi) + q_3 f(\sqrt{2})$$

für $q_1, q_2, q_3 \in \mathbb{Q}$ eindeutig definiert sind. Dieses Spiel können wir nun immer so weiterführen. Der einzige Haken ist noch, dass wir nie fertig werden, nicht einmal wirklich nach unendlich vielen Schritten, denn selbst zu jeder unendlichen Folge von irrationalen Zahlen $x_1, x_2, x_3, \ldots$ gibt es immer noch reelle Zahlen, die sich nicht als endliche Summe der Form $q_1x_1 + q_2x_2 + \ldots$ schreiben lassen (dies liegt daran, dass die reellen Zahlen überabzählbar sind). Im Prinzip ist das kein Problem, denn wir müssen die Funktionswerte ja nicht wirklich nacheinander bestimmen, sondern können uns erst die (nun überabzählbare Menge) geeigneter reeller Zahlen als Basis wählen und dann alle Funktionswerte gleichzeitig festlegen. Um das formal zu begründen, benötigt es aber das sogenannte *Auswahlaxiom*, daher wollen wir hier nicht weiter darauf eingehen.[a] □

[a]In der Sprache höherer Mathematik geht das eben vorgestellte Argument so (nur lesen, wenn du so etwas schonmal gesehen hast): Wir wählen eine Basis von $\mathbb{R}$ als Vektorraum über $\mathbb{Q}$ (bei der Wahl dieser Basis benötigen wir das Auswahlaxiom). Die Cauchy-Funktionalgleichung sagt nichts anderes, als dass $f : \mathbb{R} \to \mathbb{R}$ eine $\mathbb{Q}$-lineare Abbildung von $\mathbb{Q}$-Vektorräumen ist. Dazu können wir den Wert von f aber auf jedem Element der Basis beliebig wählen (in unserem Beispiel waren die ersten Elemente der Basis $1, \pi, \sqrt{2}, \ldots$).

Beweisidee des Monster-Satzes, 2. Teil: Den zweiten Teil des Satzes können wir ganz leicht geometrisch begründen, ohne etwas über diese verrückte Konstruktion zu benutzen. Entscheidend ist die Beobachtung, dass aus der Cauchy-Gleichung und dem bereits hergeleiteten $f(2x) = 2f(x)$ sofort $f\left(\frac{x+y}{2}\right) = \frac{f(x)+f(y)}{2}$ folgt. Geometrisch bedeutet das, dass der Mittelpunkt zwischen zwei Punkten auf dem Graphen wieder auf dem Graphen liegt. Und noch besser: Liegt der Mittelpunkt und einer der Eckpunkte einer Strecke auf dem Graphen, dann auch der andere Eckpunkt. Zu Punkten P und Q auf dem Graphen liegt also auch der Mittelpunkt der Strecke PQ sowie der Spiegelpunkt von P an Q auf dem Graphen.

Nun folgt der Rest aus einer einfachen geometrischen Überlegung: Entweder liegen alle Punkte des Graphen auf einer Geraden, dann ist f linear. Andernfalls gibt es drei Punkte, die nicht auf einer Geraden liegen. Durch wiederholte Anwendung der Mittelpunktkonstruktion können wir nun leicht eine dichte Teilmenge des Graphen im Inneren des von den drei Punkten aufgespannten Dreiecks erzeugen. Durch die Spiegelungskonstruktion können wir dies dann Schritt für Schritt auf die gesamte Ebene fortsetzen. □

1.4 Aufgaben

Die Aufgaben sollten etwa nach Schwierigkeit geordnet sein und alle mit den Inhalten dieses Briefes lösbar sein.

Aufgabe 1.1: a) Finde alle Funktionen $f : \mathbb{Q} \to \mathbb{Q}$, die die *Jensen-Funktionalgleichung*

$$f\left(\frac{x+y}{2}\right) = \frac{f(x)+f(y)}{2}$$

für alle $x, y \in \mathbb{Q}$ erfüllen.
b) Finde alle Funktionen $f : \mathbb{R} \to \mathbb{R}$, die die *Jensen-Funktionalgleichung*

$$f\left(\frac{x+y}{2}\right) = \frac{f(x)+f(y)}{2}$$

für alle $x, y \in \mathbb{R}$ erfüllen.

Aufgabe 1.2: Finde alle Funktionen $f : \mathbb{R} \to \mathbb{R}$, sodass für beliebige reelle Zahlen x, y gilt:

$$f(x+y) + f(x)f(y) = f(xy) + f(x) + f(y)$$

Aufgabe 1.3: Finde alle Funktionen $f : \mathbb{R} \to \mathbb{R}$, die für alle reellen Zahlen x die folgenden drei Bedingungen erfüllen:
(i) $f(-x) = -f(x)$
(ii) $f(x+1) = f(x) + 1$
(iii) $f\left(\frac{1}{x}\right) = \frac{f(x)}{x^2}$, falls $x \neq 0$

2 Wichtige Strategien und etwas von der kurzen Liste

Nachdem wir im ersten Kapitel bisher nur zu einer relativ einfachen Aufgabe eine vollständige Lösung gesehen haben, wollen wir jetzt anhand einiger weiterer Aufgaben verschiedene grundlegende Strategien diskutieren, die bei der Lösung einer Funktionalgleichung hilfreich sein können.

2.1 Symmetrien und eine neue Notation

Wir beginnen mit einer Aufgabe aus Usbekistan.

Aufgabe: Finde alle Funktionen $f : \mathbb{R} \to \mathbb{R}$ mit

$$f(x^3) + f(y^3) = (x + y)(f(x^2) + f(y^2) - f(xy))$$

für alle $x, y \in \mathbb{R}$.

Lösung: Zunächst sehen wir, dass hier $f(x) = x$ eine Lösung ist, denn wir erkennen die bekannte Faktorisierung

$$x^3 + y^3 = (x + y)(x^2 + y^2 - xy).$$

Etwas allgemeiner ist auch $f(x) = cx$ für beliebige c eine Lösung, da hier beide Seiten nur mit c multipliziert werden. Es ist immer eine gute Strategie, möglichst viele Lösungen einer Funktionalgleichung zu erraten. Auch wenn man dafür alleine meist nicht viele Punkte bekommen wird (solange der Beweis fehlt, dass es keine weiteren gibt), hilft es bei der Lösungsfindung. Zum Beispiel wissen wir nun, dass wir gar nicht versuchen brauchen, $f(1)$ zu berechnen, denn für jedes c gibt es eine Lösung mit $f(1) = c$. Der einzige Funktionswert, von dem wir hoffen können, ihn auszurechnen, ist $f(0)$. Tatsächlich zeigt Einsetzen von $x = y = 0$ sofort $f(0) = 0$.

Dies können wir nutzen und nur $y = 0$ einsetzen, woraufhin wir $f(x^3) = xf(x^2)$ erhalten. Eine andere Möglichkeit, bei der die kompliziert aussehende rechte Seite der Funktionalgleichung sehr einfach wird, ist $y = -x$ einzusetzen. So erhalten wir $f(x^3) + f(-x^3) = 0$ und da x^3 eine beliebige reelle Zahl sein kann, folgt $f(-x) = -f(x)$ für beliebige x. Diese *Symmetrie* können wir nun ausnutzen, indem wir in der Ausgangsgleichung y durch $-y$ ersetzen. Wir erhalten also

$$f(x^3) - f(y^3) = (x - y)(f(x^2) + f(y^2) + f(xy))$$

für beliebige $x, y \in \mathbb{R}$. Addieren wir dies zur Ausgangsgleichung, fällt einiges weg und übrig bleibt

$$2f(x^3) = 2xf(x^2) + 2xf(y^2) - 2yf(xy).$$

Wir wissen aber bereits $f(x^3) = xf(x^2)$ (das könnten wir hieraus auch noch einmal durch Einsetzen von $y = 0$ erhalten), also folgt

$$xf(y^2) = yf(xy)$$

für beliebige x, y. Setzen wir hier nun einfach $y = 1$, so folgt sofort $f(x) = xf(1)$, also ist f tatsächlich eine der oben bereits als Lösungen identifizierten linearen Funktionen. □

Wir haben nun schon einige Male Werte in Funktionalgleichungen eingesetzt, etwa $x = y = 0$ oder $x = 1, y = 2$ oder auch $x = y$ oder $x = -y$. In dieser Lösung haben wir aber auch einmal y durch $-y$ ersetzt. Natürlich sollten wir hier nicht schreiben, dass wir $y = -y$ eingesetzt haben, das wäre für $y \neq 0$ ja einfach falsch. Um bei solchen Argumenten nicht den Überblick zu verlieren, führen wir jetzt eine sehr nützliche neue Notation ein.

Notation: Wir schreiben $P(x, y)$ für die Funktionalgleichung mit den Variablen x, y in der Aufgabenstellung, also hier z.B.

$$P(x,y): \quad f(x^3)+f(y^3)=(x+y)(f(x^2)+f(y^2)-f(xy))$$

Einsetzen von $x = y = 0$ wäre dann die Gleichung $P(0,0): \quad 2f(0) = 0$.
Das eben noch problematische Einsetzen von $-y$ für y ist jetzt ganz leicht:

$$P(x,-y): \quad f(x^3)-f(y^3)=(x-y)(f(x^2)+f(y^2)+f(xy))$$

Wollten wir etwa x und y vertauschen, können wir jetzt einfach $P(y, x)$ schreiben und müssen nicht das unsinnige „Wir setzen $x = y$ und $y = x$ ein.“ schreiben. Wichtig: Will man diese Notation verwenden, sollte man diese am Beginn der Lösung einmal einführen.

Gerade und ungerade Funktionen: Der entscheidende Schritt in der soeben diskutierten Lösung war die *Symmetrie* $f(-x) = -f(x)$. Geometrisch bedeutet dies, dass der Graph der Funktion f punktsymmetrisch zum Ursprung liegt (insbesondere folgt $f(0) = 0$). Da ein Polynom $P(x)$ genau dann diese Eigenschaft hat, wenn in P nur ungerade Potenzen von x vorkommen (warum?), nennen wir eine Funktion mit dieser Symmetrie auch *ungerade*. Natürlich folgt daraus aber nicht, dass f ein Polynom sein muss. Zum Beispiel sind alle Monster aus dem letzten Kapitel ungerade Funktionen, aber auch die Funktion $f(x) = \sin(x)$ ist ungerade.

Wenig überraschend gibt es dann auch noch eine zweite Eigenschaft aus der gleichen Kategorie: Eine Funktion f heißt *gerade*, wenn $f(-x) = f(x)$ für alle x gilt. Geometrisch bedeutet dies, dass der Graph achsensymmetrisch zur y-Achse ist und ein Polynom $P(x)$ hat genau dann diese Eigenschaft, wenn in P nur gerade Potenzen von x vorkommen.

2.2 Injektivität und Surjektivität

Wo wir gerade dabei sind, neue Dinge zu definieren, machen wir gleich damit weiter. Jetzt interessieren wir uns dafür, welche Werte eine Funktion annimmt und wie oft sie dies tut.

Injektive, surjektive und bijektive Funktionen:
Nimmt f jeden Wert im Wertebereich mindestens einmal an, heißt f *surjektiv*.
Nimmt f jeden Wert im Wertebereich höchstens einmal an, heißt f *injektiv*.
Nimmt f jeden Wert im Wertebereich genau einmal an, ist also sowohl injektiv als auch surjektiv, so heißt f *bijektiv*.
Eine Funktion ist offenbar genau dann bijektiv, wenn sie eine Umkehrfunktion besitzt.

Beispiele:

- Die Funktion $f_1 : \mathbb{R} \to \mathbb{R}, x \mapsto x^2$ ist nicht surjektiv, denn der Wert -1 wird nicht angenommen. Sie ist auch nicht injektiv, denn der Wert 1 wird zweimal angenommen: $f_1(1) = f_1(-1) = 1$.

- Betrachten wir stattdessen $f_2 : \mathbb{R} \to \mathbb{R}_{\geq 0}, x \mapsto x^2$, also die gleiche Funktion nur mit anderem Wertebereich, so ist f surjektiv, denn jede nichtnegative Zahl hat eine Quadratwurzel. Allerdings ist sie immer noch nicht injektiv.

- Schränken wir diese Funktion weiter ein zu $f_3 : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}, x \mapsto x^2$; so ist sie jetzt injektiv und surjektiv mit der Umkehrfunktion $f_3^{-1} : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}, x \mapsto \sqrt{x}$.

Moral: Ob eine Funktion injektiv oder surjektiv ab, hängt also sehr stark davon ab, was ihr Definitions- und Wertebereich ist. Jede Funktion wird surjektiv, wenn wir ihren Wertebereich auf die Menge der tatsächlich angenommenen Werte einschränken.

Wir wollen uns nun am Beispiel einer Aufgabe aus der kurzen Liste der IMO 2002 anschauen, wie man in der Praxis *Injektivität* oder *Surjektivität* einer Funktion zeigen kann und wie man diese nutzen kann, um Funktionalgleichungen zu lösen:

Aufgabe: Finde alle Funktionen $f : \mathbb{R} \to \mathbb{R}$ mit der Eigenschaft

$$f(f(x) + y) = 2x + f(f(y) - x)$$

für alle $x, y \in \mathbb{R}$.

Lösung: Der Ausdruck $2x$ auf der rechten Seite, der außerhalb von f steht, lässt uns auf Surjektivität hoffen (zumindest zeigt er sofort, dass f nicht konstant und auch nicht beschränkt sein kann). Tatsächlich lässt sich dies auch leicht zeigen: Setzen wir in unser eben gelernten Notation

$$P(x, y) : \quad f(f(x) + y) = 2x + f(f(y) - x),$$

so zeigt

$$P(x, -f(x)) : \quad f(0) = 2x + f(f(-f(x)) - x)$$

für alle $x \in \mathbb{R}$. Dies sieht zwar furchtbar kompliziert aus, bedeutet aber

$$f(f(-f(x)) - x) = f(0) - 2x.$$

Da die rechte Seite sicherlich alle reellen Werte annimmt, gilt dies auch für die linke Seite und damit insbesondere für f (da die linke Seite von der Form $f(\dots)$ ist).
Die Surjektivität von f können wir nun auf verschiedene Weisen nutzen:

Erste Möglichkeit, die Surjektivität zu nutzen: Aufgrund der Surjektivität gibt es insbesondere eine Nullstelle von f, also ein c mit $f(c) = 0$. Das ist sicherlich nützlich zu wissen, denn c einzusetzen, vereinfacht nun einige Ausdrücke deutlich. Zum Beispiel zeigt jetzt

$$P(c, y) : \quad f(y) = 2c + f(f(y) - c)$$

für alle $y \in \mathbb{R}$. Aufgrund der Surjektivität von f folgt damit $z = 2c + f(z - c)$ für alle $z \in \mathbb{R}$, denn für jedes solche z können wir y mit $f(y) = z$ finden. Schreiben wir noch $z = x + c$, so folgt $\boxed{f(x) = x - c}$ für beliebige x und eine Probe bestätigt, dass jede solche Funktion (also für beliebige Konstanten c) eine Lösung ist. □

Wir haben also schon zwei Möglichkeiten gesehen, Surjektivität zu nutzen:

- um die Existenz einer Nullstelle $f(c) = 0$ zu zeigen.
- um in einem Ausdruck, in dem eine der Variablen nur noch als Argument von f vorkommt, dieses f zu „kürzen", also in unserem Fall statt $f(y)$ einfach z zu schreiben

Auch, wenn wir die Aufgabe schon gelöst haben, wollen wir uns noch eine zweite Methode anschauen, sie ausgehend von der Surjektivität zu lösen.

Zweite Möglichkeit, die Surjektivität zu nutzen: Wir erinnern uns an die gegebene Gleichung

$$P(x,y): \quad f(f(x)+y) = 2x + f(f(y)-x)$$

und wissen bereits, dass f surjektiv ist. Einsetzen von $x=0$ zeigt

$$f(f(0)+y) = f(f(y)).$$

Wüssten wir, dass f auch noch injektiv ist, dann würde hieraus schon $f(y) = f(0)+y$ folgern (denn Injektivität bedeutet ja gerade, dass aus $f(a) = f(b)$ schon $a = b$ folgt, wir dürfen also außen ein f „kürzen“). Wir müssen also nur noch zeigen, dass f auch tatsächlich injektiv ist. Nehmen wir also $f(a) = f(b)$ an. In $P(x,a)$ und $P(x,b)$ steht nun rechts das gleiche, also folgt auch

$$f(f(x)+a) = f(f(x)+b)$$

für alle $x \in \mathbb{R}$. Nun können wir die Surjektivität nutzen und $f(x+a) = f(x+b)$ für alle $x \in \mathbb{R}$ folgern oder – nach etwas Kosmetik – $f(x) = f(x+b-a)$ für alle $x \in \mathbb{R}$.
Wir erinnern uns, dass wir Injektivität, also $a = b$, zeigen wollten. Wäre nun aber $a \neq b$, so wäre $p = b-a \neq 0$ und die Gleichung $f(x) = f(x+p)$ sagt aus, dass f *periodisch* mit der Periodenlänge p ist.

Dies lässt sich nun aber leicht zum Widerspruch führen: Vergleichen wir $P(x,y)$ und $P(x+p,y)$, so ändert sich auf der linken Seite wegen $f(x) = f(x+p)$ nichts. Es folgt also

$$2x + f(f(y)-x) = 2(x+p) + f(f(y)-x-p).$$

Hier unterscheiden sich die beiden Argumente von f aber wieder um p, also folgt auch $f(f(y)-x) = f(f(y)-x-p)$. Damit bleibt aber nur noch $2x = 2(x+p)$ und damit $p = 0$ übrig. Damit muss dann aber doch $a = b$ gelten und die Injektivität ist bewiesen. $\square$

Auch hier sind wir sehr typisch vorgegangen:

- Aus der Annahme, dass f nicht injektiv ist, haben wir gefolgert, dass f sogar periodisch ist. Das ist natürlich viel stärker, denn eine periodische Funktion kann nicht injektiv sein, aber nicht jede nicht-injektive Funktion ist periodisch. Diese „Upgrade“-Strategie lässt sich aber in vielen Situationen anwenden.
- Aus der Periodizität haben wir einen Widerspruch bekommen, indem wir eine Variable um die Periodenlänge verschoben haben. Dies ist vor allem dann vielversprechend, wenn die Variable nicht nur im Inneren von f vorkommt, sondern auch – wie hier im Ausdruck $2x$ auf der rechten Seite – außerhalb.
- Unter Annahme der Injektivität waren wir sehr schnell fertig, der schwierige Teil war der Beweis der Injektivität. Das kann auch genau umgekehrt sein: Manchmal ist es sehr leicht, Injektivität oder Surjektivität zu zeigen, aber nicht so klar, wie man diese geschickt verwenden kann.

2.3 Eine schwierige Aufgabe

Zum Abschluss wollen wir eine recht schwierige Aufgabe lösen und werden im Lauf der Lösung fast alle bisher gelernten Konzepte benutzen. (Wie auch bei allen anderen hier besprochenen Aufgaben lohnt es sich natürlich, erst selbst über die Aufgabe nachzudenken. Selbst wenn man sie nicht vollständig löst, ist es ja interessant zu sehen, wie weit man selbst gekommen wäre.)

Aufgabe: Finde alle Funktionen $f : \mathbb{R} \to \mathbb{R}$ mit

$$f(x^2 + y + f(y)) = 2y + f(x)^2$$

für alle $x, y \in \mathbb{R}$.

Lösung: Wir beginnen mit der Beobachtung, dass sicherlich $f(x) = x$ eine Lösung ist. Schreiben wir wie inzwischen üblich $P(x, y) : \quad f(x^2 + y + f(y)) = 2y + f(x)^2$, so folgt aus $P(0, y)$ sofort $f(y + f(y)) = 2y + f(0)^2$. Da die rechte Seite alle Werte annimmt, ist f surjektiv. Weiter liefert Vergleich von $P(x, y)$ und $P(-x, y)$ sofort $f(x)^2 = f(-x)^2$.

Achtung: Daraus folgt noch nicht, dass f gerade oder ungerade ist, denn es könnte ja sein, dass für manche x die Gleichung $f(-x) = f(x)$ gilt, für andere aber $f(-x) = -f(x)$. Ein Beispiel wäre die Betragsfunktion $f(x) = |x|$.

Aus der Surjektivität folgt aber sicherlich die Existenz einer Nullstelle, also eines c mit $f(c) = 0$. Wie eben hergeleitet folgt dann auch sofort $f(-c) = 0$. Vergleich von $P(0, c)$ und $P(0, -c)$ zeigt jetzt aber $2c = -2c$, also $c = 0$, d.h. $f(0) = 0$. Wir haben sogar mehr gezeigt, nämlich dass $c = 0$ die einzige Nullstelle von f ist (wenn man so will eine Art „Mini-Injektivität“).
Dies können wir nun nutzen: $P\left(x, -\frac{f(x)^2}{2}\right)$ zeigt $f\left(x^2 - \frac{f(x)^2}{2} + f\left(-\frac{f(x)^2}{2}\right)\right) = 0$ und damit schon $x^2 - \frac{f(x)^2}{2} + f\left(-\frac{f(x)^2}{2}\right) = 0$. Dies sieht zwar kompliziert aus, ist aber qualitativ interessant: Es sagt nämlich, dass aus $f(a)^2 = f(b)^2$ schon $a^2 = b^2$ folgt.

Damit können wir nun zeigen, dass f ungerade ist: Nach Surjektivität muss es ja zu jedem x ein y geben mit $f(y) = -f(x)$. Aus der eben hergeleiteten Beziehung folgt dann aber schon $y^2 = x^2$, also $y = x$ oder $y = -x$. Ist $x \neq 0$, so ist $y = x$ ausgeschlossen und es folgt $y = -x$, also tatsächlich $f(-x) = -f(x)$. Für $x = 0$ ist dies ohnehin richtig.

Die Funktion f ist also ungerade. Nebenbei haben wir gezeigt, dass f injektiv und damit bijektiv ist (warum?).

Aus $P(x, 0)$ erhalten wir $f(x)^2 = f(x^2)$. Damit wird unsere Ausgangsgleichung jetzt zu $f(x^2 + y + f(y)) = 2y + f(x^2)$, für alle $z \geq 0$ und alle y gilt also

$$f(z + y + f(y)) = 2y + f(z). \tag{2}$$

Indem wir benutzen, dass f ungerade ist, können wir dies auch für alle z bekommen: Denn ist $z \leq 0$, so folgt $f(-z - y + f(-y)) = -2y + f(-z)$ und damit unter Benutzung von $f(-x) = -f(x)$ sofort $f(z + y + f(y)) = 2y + f(z)$ wie gewünscht.
Jetzt können wir in (2) $z = -y$ einsetzen und erhalten $f(f(y)) = 2y - f(y)$, äquivalent $f(f(y)) + f(y) = 2y$. Setzen wir in (2) nun $y = f(t)$ ein, erhalten wir

$$2f(t) + f(z) = f(z + f(t) + f(f(t))) = f(z + 2t)$$

für alle $z, t \in \mathbb{R}$. Insbesondere folgt mit $z = 0$ sofort $f(2t) = 2f(t)$ und damit jetzt $f(z + 2t) = f(z) + f(2t)$, also auch $f(x + y) = f(x) + f(y)$ für alle $x, y \in \mathbb{R}$.

Unsere Funktion f erfüllt also die Cauchy-Funktionalgleichung! Und da wir bereits $f(x^2) = f(x)^2 \geq 0$ und damit $f(t) \geq 0$ für alle $t \geq 0$ wissen, folgt wie im letzten Kapitel gezeigt schon, dass f kein Monster und damit von der Form $f(x) = cx$ ist.
Einsetzen in $f(x^2) = f(x)^2$ zeigt dann $c^2 = c$, also $c = 0$ oder $c = 1$. Für $c = 0$ wäre f aber nicht surjektiv, also muss $c = 1$ und damit $\boxed{f(x) = x}$ für alle x gelten, was ja auch bereits zu Beginn als Lösung erkannt wurde. □

2.4 Aufgaben

Aufgabe 2.0: a) Man kann jedes Polynom als Summe einer geraden und einer ungeraden Funktion schreiben (wie?). Zeige, dass das gleiche für eine beliebige Funktion $f : \mathbb{R} \to \mathbb{R}$ gilt.

b) Zeige, dass jede Funktion $f : \mathbb{Z} \to \mathbb{Z}$ als Summe von zwei injektiven Funktionen geschrieben werden kann.

Das ist natürlich nicht wirklich eine Aufgabe zu Funktionalgleichungen, sondern eher eine Übung zu den in diesem Kapitel neu eingeführten Begriffen, deshalb steht sie an Stelle 0 *(was aber nicht unbedingt bedeutet, dass sie einfacher ist als die anderen Aufgaben).*

Aufgabe 2.1: Finde alle Funktionen $f : \mathbb{R} \to \mathbb{R}$ mit

$$(x-y)\left(f(x^2)+f(y^2)+\frac{1}{2}f(2xy)\right)=f(x^3)-f(y^3).$$

Aufgabe 2.2: Finde alle Funktionen $f : \mathbb{R} \to \mathbb{R}$ mit

$$f(x+y)=f(x-y)+f(f(1-xy))$$

für alle reellen Zahlen x, y.

Aufgabe 2.3: Finde alle Funktionen $f : \mathbb{R} \to \mathbb{R}$ mit

$$f(x+yf(x))+f(xy)=f(x)+f(2022y)$$

für alle $x, y \in \mathbb{R}$.

3 Und nun zu etwas ganz anderem...

Bisher haben wir uns vor allem mit Aufgaben beschäftigt, bei denen Funktionen $f : \mathbb{R} \to \mathbb{R}$ gesucht waren. In diesem Abschnitt kümmern wir uns um etwas andere Typen von Funktionalgleichungen, die auch häufig in Olympiaden vorkommen.
Ganz zum Schluss sehen wir dann auch noch, dass dabei der Fantasie der Aufgabenstellerin (fast) keine Grenzen gesetzt sind...

3.1 Funktionen auf den ganzen oder den natürlichen Zahlen

Wir beginnen mit einer Aufgabe aus der kurzen Liste der IMO 2015:

Aufgabe: Finde alle Funktionen $f : \mathbb{Z} \to \mathbb{Z}$ mit

$$f(x - f(y)) = f(f(x)) - f(y) - 1$$

für alle $x, y \in \mathbb{Z}$.

Lösung: Wie üblich bezeichnen wir mit $P(x, y)$ die gegebene Funktionalgleichung.
Beide Seiten sehen recht kompliziert aus. Es liegt also nahe, etwas derart einzusetzen, dass sich zumindest eine der beiden Seiten vereinfacht. Auf der linken wäre das für $x = f(y)$ der Fall, rechts für $y = f(x)$. Das zweite ist vielversprechender, denn so wird zwar die linke Seite kompliziert, ist aber von der Form $f(\dots)$. Wir erhalten also

$$P(x, f(x)) : \quad f(x - f(f(x)) = -1.$$

Hier ist die linke Seite zwar kompliziert, wir folgern aber immerhin, dass f den Wert -1 annimmt. Wir können dies auch einfacher umformulieren: Wir wählen ein festes b mit $f(b) = -1$. Dieses können wir nun natürlich wieder einsetzen! Speziell folgt

$$P(x, b) : \quad f(x + 1) = f(f(x)).$$

Dies sieht nun schon recht vielversprechend aus, denn nun wird die Ausgangsgleichung zu $f(x - f(y)) = f(x + 1) - f(y) - 1$. Setzen wir hier $x = f(y) - 1$ ein, werden beide Seiten einfacher und wir erhalten

$$f(-1) = f(f(y)) - f(y) - 1 = f(y + 1) - f(y) - 1.$$

Auch wenn wir $f(-1)$ noch nicht kennen, folgt daraus sofort $f(y + 1) = f(y) + C$ für irgendeine Konstante C. Da wir nur auf den ganzen Zahlen arbeiten, folgt daraus sofort mit Induktion (genauer gesagt mit zwei Induktionen, einmal in die positive und einmal in die negative Richtung), dass f linear ist, also von der Form $f(x) = Cx + D$ für gewisse Konstanten C und D.
Um zu untersuchen, für welche Werte von C und D dies tatsächlich eine Lösung ist, setzen wir in die Ausgangsgleichung ein. Diese wird dann zu

$$C(x - Cy - D) + D = C(Cx + D) + D - Cy - D - 1.$$

Damit dies für alle x und y stimmt, müssen hier die Koeffizienten von x, die von y und die konstanten auf beiden Seiten übereinstimmen, also muss $C = C^2$ und $-CD + D = CD - 1$ gelten. Ist $C = 0$, so folgt $D = -1$ und damit $\boxed{f(x) = -1}$ für alle $x \in \mathbb{Z}$.

Ansonsten ist $C = 1$ und es folgt $D = 1$ und damit $\boxed{f(x) = x + 1}$ für alle $x \in \mathbb{Z}$.
Für diese beiden Funktionen ist die Gleichung dann auch tatsächlich erfüllt, es gibt also genau diese beiden Lösungen.

Es war wirklich entscheidend, dass f hier nur auf den ganzen Zahlen definiert war. Nur so konnten wir aus $f(x+1) = f(x) + C$ sofort mit Induktion auf die Linearität $f(x) = Cx + D$ schließen (zum Vergleich denke man zum Beispiel an die Diskussion der Cauchy-Funktionalgleichung zurück). Ersetzen wir in der obigen Aufgabe überall $\mathbb{Z}$ durch $\mathbb{Q}$, so wird die Aufgabe sofort zu einer deutlich schwierigeren (wenn nicht unlösbaren) Aufgabe.

Natürlich ist eine Folge $f : \mathbb{Z} \to B$ nichts anderes als eine (in beide Richtungen unendliche) Folge von Elementen aus B. Wir könnten also statt der Funktions-Notation $f(n)$ auch die Folgen-Notation f_n verwenden. Trotz dieser Äquivalenz gibt es aber Aufgaben dieses Typs, die eindeutig eher „Folgen-Aufgaben“ sind und solche, die eher „Funktionalgleichungs-Aufgaben“ sind, etwa wenn Ausdrücke wie $f(f(n))$ vorkommen.

Eine zusätzliche Schwierigkeit bei Funktionalgleichungen auf den ganzen Zahlen ist es oft herauszufinden, ob es sich eher um eine „Algebra“- oder eine Zahlentheorie-Aufgabe handelt. Im vorigen Beispiel war das allerdings relativ klar, genauso wie im nächsten, einer Aufgabe aus der Asiatisch-Pazifischen Mathematik-Olympiade. Wir zeigen zwei sehr unterschiedliche Lösungen, die beide wichtige Strategien bei der Lösung solcher „Zahlentheorie-Funktionalgleichungen“ verdeutlichen.

Aufgabe: Bestimme alle Funktionen $f : \mathbb{N} \to \mathbb{N}$ mit

$$f(a) + b \mid a^2 + f(a)f(b)$$

für alle $a, b \in \mathbb{N}$.

Erste Lösung: Sei $P(a, b)$ die gegebene Bedingung. Es wäre vor allem nützlich, wenn die Zahl auf der linken Seite eine einfachere Struktur hätte. Dazu können wir $b = f(a)$ einsetzen und erhalten

$$P(a, f(a)) : \quad 2f(a) \mid a^2 + f(a)f(f(a))$$

und insbesondere $f(a) \mid a^2$. Für $a = 1$ folgt hier sofort $f(1) = 1$.
Ähnlich nützlich ist diese Bedingung auch dann, wenn a^2 wenige Teiler hat, also z.B. wenn $a = p$ eine Primzahl ist. So folgt sofort $f(p) \in \{1, p, p^2\}$ für beliebige Primzahlen p. Einsetzen von $a = b = p$ zeigt jetzt aber $P(p, p) : \quad f(p) + p \mid p^2 + f(p)^2$. Für $f(p) = 1$ wäre das $p + 1 \mid p^2 + 1$ und damit wegen $p + 1 \mid p^2 - 1$ auch $p + 1 \mid 2$, was wegen $p + 1 \geq 3$ ein Widerspruch ist.
Für $f(p) = p^2$ wäre das $p^2 + p \mid p^2 + p^4$, also $p + 1 \mid p(p^2 + 1)$. Da p und $p + 1$ teilerfremd sind, folgt dann aber auch wieder $p + 1 \mid p^2 + 1$, also erneut ein Widerspruch!

Somit bleibt nur noch ein Fall und es folgt $f(p) = p$ für alle Primzahlen p. Nun könnte man versuchen, dieses Argument auf Zahlen der Form p^2 oder pq auszudehnen. Wir können aber von hier aus viel einfacher zum Ziel kommen. Dazu betrachten wir

$$P(p, b) : \quad b + p \mid p^2 + pf(b) = p(p + f(b))$$

für beliebiges b und jede Primzahl p. Wählen wir p insbesondere zu b teilerfremd (also z.B. $p > b$), so folgt sogar $b + p \mid p + f(b)$ und damit $b + p \mid f(b) - b$. Dies war geschickt, denn nun steht rechts eine von p unabhängige ganze Zahl. Diese soll nun aber beliebig große Teiler haben (nämlich $b + p$ für beliebige $p > b$), was nur für die ganze Zahl 0 erfüllt ist. Es folgt also $\boxed{f(b) = b}$ für beliebige b und eine kurze Probe zeigt, dass dies tatsächlich eine Lösung ist, denn es gilt sicherlich $a + b \mid a^2 + ab$ für beliebige $a, b \in \mathbb{N}$.

Zweite Lösung: Sei
$$P(a,b): \quad f(a)+b \mid a^2+f(a)f(b)$$
die gegebene Bedingung. Wir beginnen damit, einfache Werte einzusetzen. Da 0 nicht erlaubt ist, beginnen wir mit $P(1,1): \quad f(1)+1 \mid f(1)^2+1$.
Da außerdem $f(1)+1 \mid f(1)^2-1$ gilt, folgt sofort $f(1)+1 \mid 2$ und damit insbesondere $f(1)+1 \leq 2$, also $f(1) \leq 1$ und damit $f(1)=1$. Wir haben also bereits einen Funktionswert bestimmt und können diesen nun natürlich wieder verwenden:
$$P(1,a): \quad a+1 \mid f(a)+1.$$
(Obacht: Hier haben wir für a die Zahl 1 und für b die (neue) Zahl a eingesetzt. Das kann man machen, um die Notation zu vereinheitlichen, also in allen Gleichungen, in denen nur eine Variable vorkommt, die gleiche Variable a zu verwenden. Man sollte sich aber natürlich dabei nicht selbst verwirren.)
Insbesondere folgt sofort $f(a) \geq a$ für alle a. Kehren wir das noch einmal um, erhalten wir
$$P(a,1): \quad f(a)+1 \mid a^2+f(a)$$
und damit nach Abziehen von $f(a)+1$ rechts auch $f(a)+1 \mid a^2-1$. Dies liefert nun natürlich für $a \geq 2$ auch die obere Schranke $f(a) \leq a^2-2$, diese ist aber für die meisten a noch recht weit entfernt von der unteren Schranke $f(a) \geq a$. Zumindest für $a=2$ stimmen sie aber überein und wir erhalten $f(2)=2$.
Nun ist es naheliegend, das Spiel durch Einsetzen von 2 als einer der Variablen zu wiederholen. Allein dadurch werden wir aber nicht zum Ziel kommen. Kombinieren lassen sich die Aussagen nur dann sinnvoll, wenn rechts oder links jeweils der gleiche Ausdruck herauskommt. Dies lässt sich nun z.B. wie folgt erreichen: Wir haben bereits $f(a)+1 \mid a^2-1$ und wegen $f(a)+1 \mid f(a)^2-1$ dann auch $f(a)+1 \mid f(a)^2-a^2$. Betrachten wir stattdessen
$$P(a,2): \quad f(a)+2 \mid a^2+2f(a),$$
so folgt analog wegen $f(a)+2 \mid f(a)^2+2f(a)$ auch $f(a)+2 \mid f(a)^2-a^2$.
Wir haben es also geschafft, zwei Teiler des gleichen Ausdrucks $f(a)^2-a^2$ zu finden. Da $f(a)+1$ und $f(a)+2$ als aufeinanderfolgende Zahlen auch auf jeden Fall teilerfremd sind, folgt
$$(f(a)+1)(f(a)+2) \mid f(a)^2-a^2.$$
Nun ist aber (das haben wir gewonnen!) die linke Seite sicherlich größer als $f(a)^2-a^2$ und wir wissen auch schon, dass $f(a)^2-a^2 \geq 0$ gilt. Damit ist diese Teilbarkeit nur möglich, wenn $f(a)^2=a^2$ gilt, denn eine positive Zahl kann nur Teiler haben, die höchstens so groß sind wie sie selbst. Da a beliebig war und alles positiv ist, haben wir damit $f(a)=a$ für alle a gezeigt und wie oben zeigt eine Probe, dass dies in der Tat eine Lösung ist.

Wie auch bei anderen Zahlentheorie-Aufgaben mit Teilbarkeiten war es in der zweiten Lösung also eine hilfreiche Strategie, die durch etwas teilbaren Ausdrücke so lange umzuformen, bis sie möglichst klein sind im Vergleich zum Ausdruck, durch den sie teilbar sein sollen. Auf diese Weise erhält man aus der Teilbarkeit dann Ungleichungen oder sogar direkt eine Gleichheit. Man hätte diese Lösung auch finden können, indem man nach $f(1)$ und $f(2)$ zunächst versucht, $f(3)$, dann $f(4)$, dann $f(5)$ etc. zu berechnen und schließlich das Argument zu verallgemeinern.
In der ersten Lösung sind wir dagegen anders vorgegangen und haben direkt unendlich viele Funktionswerte (bei Primzahlen) berechnet.

3.2 Funktionen auf den positiven reellen Zahlen

In der folgenden Aufgabe, die aus Kasachstan stammt, ist eine Funktion $f : \mathbb{R}^+ \to \mathbb{R}^+$ gesucht. Dies macht die Sache zunächst schwieriger, denn wir dürfen z.B. keine der Variablen auf Null setzen. Die Einschränkung des Wertebereichs kann aber auch hilfreich sein, denn aus der Positivität der Funktionswerte folgen auch wieder bestimmte Ungleichungen. Insgesamt sind bei solchen Aufgaben oft (aber nicht immer!) komplett andere Argumente hilfreich als bei den Funktionalgleichungen auf ganz $\mathbb{R}$, die wir im letzten Kapitel gesehen haben.

Aufgabe: Finde alle Funktionen $f : \mathbb{R}^+ \to \mathbb{R}^+$ mit

$$f(x)^2 = f(xy) + f(x + f(y)) - 1$$

für alle $x, y \in \mathbb{R}^+$.

Beweis: Sei wie üblich $P(x, y)$ die Gleichung aus der Aufgabe. Aus

$$P(1, x) : \quad f(1)^2 = f(x) + f(1 + f(x)) - 1$$

erhalten wir zwar eine komplizierte Gleichung, aber wegen $f(1 + f(x)) > 0$ immerhin sofort $f(x) < f(1)^2 + 1$. Insbesondere sind die Funktionswerte von f also nach oben beschränkt!
Wir beobachten auch: Sind alle Funktionswerte von f kleiner als C, so ist die rechte Seite in der Ausgangsgleichung kleiner als $2C - 1$ und damit $f(x)^2$ kleiner als $2C - 1$ (insbesondere ist $2C - 1 > 0$), also $f(x)$ kleiner als $\sqrt{2C - 1}$. Ist also C eine obere Schranke, so auch $\sqrt{2C - 1}$. Für große C ist dies eine deutlich bessere Schranke. Genauer gesagt gilt $\sqrt{2C - 1} < C$ genau dann, wenn $2C - 1 < C^2$ und damit wenn $(C - 1)^2 > 0$ gilt, also wann immer $C \neq 1$ gilt.
Wir können nun wie folgt argumentieren: Ist $M > 0$ die kleinste obere Schranke an die Funktionswerte von f (also die kleinste Zahl M mit $f(x) \leq M$ für alle x), so ist auch $\sqrt{2M - 1}$ eine obere Schranke an die Funktionswerte von f und aufgrund der Minimalität von M folgt $\sqrt{2M - 1} \leq M$ und damit $(M - 1)^2 \leq 0$, also $M = 1$.

Aus der Existenz irgendeiner oberen Schranke haben wir also gefolgert, dass sogar $f(x) \leq 1$ für alle x gilt. Da die konstante Funktion $f(x) = 1$ sicherlich eine Lösung ist, lässt sich dies auch nicht weiter verbessern.

Um nun eine Abschätzung in die andere Richtung zu bekommen, erinnern wir uns an das Argument zu Beginn, welches wir nun umkehren können: Wegen $f(x + f(y)) \leq 1$ folgt aus der Ausgangsgleichung auch $f(x)^2 \leq f(xy)$ für beliebige $x, y > 0$. Für festes x läuft hier $z := xy$ durch alle reellen Zahlen, also können wir stattdessen auch $f(x)^2 \leq f(z)$ für beliebige $x, z > 0$ folgern.

Jetzt müssen wir nur noch einmal scharf hinsehen: Hier steht, dass $\sqrt{f(z)}$ für beliebige z eine obere Schranke an f ist. Da $M = 1$ die kleinste obere Schranke war, folgt daraus aber $\sqrt{f(z)} \geq M = 1$ und damit $f(z) \geq 1$ für alle z, insgesamt also $\boxed{f(x) = 1}$ für alle $x > 0$. Dies ist damit die einzige Lösung.

Wir haben in dieser Lösung die folgende fundamentale Eigenschaft der reellen Zahlen benutzt: *Jede nach oben beschränkte Menge reeller Zahlen hat eine kleinste obere Schranke.* Mit $\mathbb{Q}$ statt $\mathbb{R}$ wäre diese Aussage falsch! Die Menge $\{x \in \mathbb{Q} : x^2 < 2\}$ ist nämlich nach oben beschränkt, es gibt aber keine kleinste rationale Zahl M mit $x < M$ für alle x in dieser Menge (wir würden gerne $M = \sqrt{2}$ hinschreiben, aber das ist ja nicht rational). Auf dieser *Supremums-Eigenschaft* von $\mathbb{R}$ basiert das gesamte Gebiet der Analysis!

3.3 Etwas aus dem Kuriositätenkabinett

Zum Abschluss zeigen wir noch an einem Beispiel aus der Amerikanischen IMO-Vorauswahlklausur, was man mit ein wenig Kreativität alles aus dem Konzept „Funktionalgleichung“ machen kann.

Aufgabe: Finde alle Funktionen $\theta : \mathbb{Z}[x] \to \mathbb{Z}$ mit der folgenden Eigenschaft: Für alle Polynome $p, q \in \mathbb{Z}[x]$ gilt
1) $\theta(p + 1) = \theta(p) + 1$ und
2) Ist $\theta(p) \neq 0$, dann ist $\theta(pq)$ durch $\theta(p)$ teilbar.

Lösung: Zunächst müssen wir verstehen, was die Aufgabe von uns möchte. Wir sollen eine Funktion θ finden, die jedem Element aus $\mathbb{Z}[x]$, also jedem Polynom in einer Variablen x mit ganzzahligen Koeffizienten eine ganze Zahl zuordnet. Zum Beispiel könnte jedem Polynom p der Wert $p(0)$ zugeordnet werden. Tatsächlich erfüllt diese Funktion auch die beiden Eigenschaften: Es gilt nämlich stets, dass $p(x) + 1$ an der Stelle $x = 0$ den Wert $p(0) + 1$ annimmt (klar!) und ist $p(0) \neq 0$, dann ist sicherlich $p(0)$ ein Teiler von $(pq)(0) = p(0)q(0)$. Fallen uns noch mehr solche Funktionen ein? Klar, wir können auch stattdessen ein Polynom p auf den Wert $p(1)$ abbilden. Allgemeiner können wir für jedes $n \in \mathbb{Z}$ die Funktion $\theta_n : \mathbb{Z}[x] \to \mathbb{Z}$ betrachten, die p auf $p(n)$ abbildet, also $\theta_n(p) = p(n)$. Zum Beispiel bildet θ_{42} jedes Polynom auf den Funktionswert bei $x = 42$ ab.
Man sieht nun leicht genau wie bei θ_0, dass jede solche Funktion θ_n die Bedingung der Aufgabenstellung erfüllt. Fallen uns spontan keine weiteren Lösungen ein, so können wir versuchen zu zeigen, dass alle Lösungen von dieser Form sind.
Wie können wir das tun, wo es doch unendlich viele Lösungen gibt? Praktisch ist es in einer solchen Situation, eine gewisse oBdA-Annahme machen zu können, die uns auf einen der Fälle reduziert. Oben haben wir gesehen, dass das Argument für θ_0 und θ_n das gleiche war. Allgemeiner zeigt diese Überlegung folgendes: Ist θ eine Lösung, dann für beliebiges c auch die „um c verschobene Funktion“ $\theta'(p) = \theta(p(x - c))$ (rechts wird dabei θ auf das Polynom $p(x - c)$ angewendet, also p bei $x - c$ ausgewertet).
Ist $\theta(x) = c$, so folgt aus (1) schon $\theta(x - c) = 0$ und damit $\theta'(x) = 0$. Was wir daraus lernen, ist, dass wir aus einer Lösung mit $\theta(x) = c$ auch eine mit $\theta(x) = 0$ konstruieren können (und umgekehrt). Wir können also o.B.d.A. $\theta(x) = 0$ annehmen. Können wir zeigen, dass dann $\theta = \theta_0$ gilt, so folgt durch Rückverschiebung auch, dass jede Lösung von der Form θ_n für ein n sein muss (denn diese gehen bei Verschiebung ineinander über).
Wir wollen also $\theta(p) = p(0)$ für alle Polynome p zeigen. Nach (1) genügt es dafür, $\theta(p) = 0$ für alle Polynome p mit $p(0) = 0$ zu zeigen. Sei also P ein solches Polynom.
Irgendwann sollten wir die zweite Bedingung benutzen, allerdings können wir diese ja gerade nicht auf P anwenden. Stattdessen wollen wir benutzen, dass $x - a \mid P(x) - P(a)$ für beliebige $a \in \mathbb{Z}$ gilt. Wenden wir also 2) mit $p = x - a$ und $q(x) = \frac{P(x)-P(a)}{x-a}$ an, so folgt: Ist $\theta(x - a) \neq 0$, so ist $\theta(P - P(a))$ durch $\theta(x - a)$ teilbar.
Indem wir (1) und $\theta(x) = 0$ anwenden, können wir dies übersetzen als: Ist $a \neq 0$, so ist $\theta(P) - P(a)$ durch a teilbar. Wegen $P(0) = 0$ ist aber auch $P(a)$ durch a teilbar, also ist auch $\theta(P)$ durch a teilbar. Da $a \neq 0$ beliebig war, folgt jetzt allerdings $\theta(P) = 0$ wie behauptet. Somit ist gezeigt, dass im Fall $\theta(x) = 0$ die einzige Lösung $\theta = \theta_0$ ist und im allgemeinen Fall alle Lösungen von der Form θ_n für ein $n \in \mathbb{Z}$ sind.

Das sah vielleicht recht kompliziert aus, was aber vor allem daran liegt, dass es sehr ungewohnt ist, Funktionen zu betrachten, die andere Funktionen (hier Polynome) als Argumente haben. An der Uni ist das ein sehr wichtiges Prinzip, wenn man zum Beispiel die Eigenschaften der „Funktion“ (dann meist eher *Operator* genannt) untersucht, die eine (differenzierbare) Funktion $f(x)$ auf die Ableitungsfunktion $f'(x)$ abbildet.

3.4 Aufgaben

Diesmal sind die ersten beiden Aufgaben vielleicht ungefähr gleich schwer (vor allem aber schwer zu vergleichen, weil sie so unterschiedlich sind).

Aufgabe 3.1: Finde alle Funktionen $f : \mathbb{R}^+ \to \mathbb{R}^+$ mit

$$f(x)f(y) = 2f(x + yf(x))$$

für alle $x, y \in \mathbb{R}^+$.

Aufgabe 3.2: Finde alle Funktionen $f : \mathbb{N} \to \mathbb{N}$, so dass es für alle $a, b \in \mathbb{N}$ ein nicht entartetes Dreieck mit Seitenlängen

$$a, f(b) \quad \text{und} \quad f(b + f(a) - 1)$$

gibt.

Aufgabe 3.3: Gegeben sei eine natürliche Zahl n und ganze Zahlen $a_1, \ldots, a_n$. Eine Funktion $f : \mathbb{Z} \to \mathbb{R}$ habe die Eigenschaft, dass für alle $k, \ell \in \mathbb{Z}$ mit $\ell \neq 0$ stets

$$f(k + a_1\ell) + f(k + a_2\ell) + \cdots + f(k + a_n\ell) = 0$$

gilt. Zeige, dass f die konstante Nullfunktion ist.

4 Der letzte Schliff

In diesem letzten Kapitel besprechen wir noch zwei wichtige Themen bzw. Methoden.

4.1 Die punktweise Falle

Wir beginnen zum Aufwärmen mit einer Aufgabe aus einer Taiwanesischen Auswahlklausur.

Aufgabe: Finde alle Funktionen $f : \mathbb{R} \to \mathbb{R}$ mit

$$f(f(x)+y) = f(x^2-y) + 4(y-2)(f(x)+2)$$

für alle $x, y \in \mathbb{R}$.

Lösung: Sei wie üblich $P(x,y)$ die gegebene Bedingung. Dies sieht recht kompliziert aus. Wir sehen aber, dass y sowohl auf der linken Seite als auch im ersten Term auf der rechten Seite linear vorkommt und zwar mit unterschiedlichen Koeffizienten. Wir können für gegebenes x also ein y derart wählen, dass die beiden Ausdrücke gleich werden. Genauer gesagt benötigen wir $y = \frac{x^2-f(x)}{2}$, um $f(x)+y = x^2-y$ zu erreichen. Damit zeigt

$$P\left(x, \frac{x^2-f(x)}{2}\right): \quad 4\left(\frac{x^2-f(x)}{2} - 2\right)(f(x)+2) = 0$$

für beliebige x. Dies impliziert nun aber für alle x, dass der zweite Faktor 0 wird, dann ist $f(x) = -2$, oder dass der erste Faktor 0 wird, dann ist $f(x) = x^2-4$.
Tatsächlich zeigt eine kurze Probe auch, dass sowohl $\boxed{f(x) = x^2-4}$ als auch $\boxed{f(x) = -2}$ die Funktionalgleichung erfüllen. Fertig!
Fertig? Nein, denn wir sind in die *punktweise Falle* getappt! Wir haben gezeigt, dass für jedes x (mindestens) eine der Aussagen $f(x) = -2$ und $f(x) = x^2-4$ gelten muss. Was wir aber übersehen haben, ist die Möglichkeit, dass für manche x der eine Fall und für andere Werte der andere Fall eintritt.
Das sollte im Prinzip nicht so schwer sein, kann aber ein wenig fummelig werden, wenn man sich nicht geschickt anstellt, da es oft noch einige Sonderfälle zu berücksichtigen gilt (z.B. solche x mit $x^2-4=-2$, bei denen also beide Werte übereinstimmen). Ein mögliches sauberes Ende der Lösung sieht so aus:
Wir wissen schon, dass entweder $f(0) = -2$ oder $f(0) = -4$ gelten muss.
1. Fall: $f(0) = -4$. Dann wollen wir zeigen, dass $f(x) = x^2-4$ für alle x gilt. Angenommen, es gäbe x_0 mit $f(x_0) \neq x_0^2-4$. Insbesondere folgt dann $f(x_0) = -2$. Einsetzen zeigt

$$P(x_0, 2): \quad f(x_0^2-2) = -4.$$

Dann folgt $x_0^2-2 = 0$ (denn $f(z) = -4$ kann nur für $z = 0$ gelten), dann ist aber $f(x_0) = -2 = x_0^2-4$, Widerspruch! Somit ist in diesem Fall $f(x) = x^2-4$ für alle x.

2. Fall: $f(0) = -2$. Dann wollen wir $f(x) = -2$ für alle x zeigen. Einsetzen von $x = 0$ zeigt $P(0,x): \quad f(x-2) = f(-x)$ für alle $x \in \mathbb{R}$. Angenommen, es gäbe nun x_0 mit $f(x_0) \neq -2$. Dann folgt auch $f(-x_0-2) = f(x_0) \neq -2$. Also folgt

$$x_0^2 - 4 = f(x_0) = f(-x_0-2) = (-x_0-2)^2 - 4$$

und damit nach Auflösen $x_0 = -1$. Der einzige Wert mit $f(x) \neq -2$ ist also $x = -1$. Wir haben also $f(x) = \begin{cases} -2 & x \neq -1 \\ -3 & x = -1 \end{cases}$. Nun führt aber z.B. $P(-1, 42)$ direkt zu einem Widerspruch! Somit ist in diesem Fall $f(x) = -2$ für alle x.

Es gibt oft viele Wege, aus der punktweisen Falle zu entkommen. Wichtig ist vor allem, dass man dies nicht einfach vergisst und weiter, dass man dabei wirklich darauf achtet, sauber zu arbeiten. Manchmal entgehen einem sonst nämlich nicht nur Punkte, sondern auch unerwartete Lösungen, wie das folgende Beispiel aus der kurzen Liste der IMO 2001 eindrucksvoll zeigt...

Aufgabe: Finde alle Funktionen $f : \mathbb{R} \to \mathbb{R}$ mit

$$f(xy)(f(x) - f(y)) = (x - y)f(x)f(y)$$

für alle $x, y \in \mathbb{R}$.

Lösung: Das sieht zunächst recht harmlos aus. Wir sehen sofort, dass alle linearen Funktionen $f(x) = cx$ die Bedingung erfüllen und es liegt nahe zu vermuten, dass dies die einzigen sind. Wir nennen wieder $P(x, y)$ die gegebene Gleichung. Um ähnliche Terme auf beiden Seiten zu haben, können wir $y = 1$ setzen. Schreiben wir $b = f(1)$, so folgt

$$P(x, 1): \quad f(x)(f(x) - b) = b(x - 1)f(x),$$

also $f(x)^2 = bxf(x)$. Mit $x = 0$ folgt schon $f(0) = 0$ und allgemein folgt $f(x) = 0$ oder $f(x) = bx$. Fast fertig!?! Wir müssen nur noch ausschließen (das haben wir ja eben gelernt), dass manchmal $f(x) = 0$ und manchmal $f(x) = bx$ gilt.
Probieren wir das eine Weile, stellen wir aber fest, dass das schwierig wird und mit etwas mehr Zeit stellen wir vielleicht fest, dass z.B. $f(x) = \begin{cases} b & x = 1 \\ 0 & x \neq 1 \end{cases}$ oder $f(x) = \begin{cases} bx & x \geq 0 \\ 0 & x < 0 \end{cases}$ auch die Bedingung erfüllen.
Umgekehrt ist es schon so, dass wir nicht einfach beliebig auswählen können, wo $f(x) = 0$ und wo $f(x) = bx$ gilt, denn gilt z.B. $f(x) = 0$, so folgt aus der Gleichung $f(y) = 0$ oder $f(xy) = 0$. Um nun systematisch heranzugehen, setzen wir $A = \{x \in \mathbb{R} : f(x) \neq 0\}$. Für $x \in A$ gilt dann $f(x) = bx$. Sind x und y beide nicht in A, so ist $P(x, y)$ sicherlich erfüllt. Ist $x, y \in A$ und $x \neq y$, so ist $P(x, y)$ äquivalent zu $xy \in A$. Sei nun $x \in A, y \notin A$. Dann gilt $P(x, y)$ genau dann, wenn $f(xy) = 0$, also $xy \notin A$ gilt. Wir nehmen nun an, dass $b \neq 0$ gilt, denn sonst ist $A = \emptyset$ und $f(x) = 0$ für alle x. Dann ist insbesondere $1 \in A$. Unsere Bedingung ist nun einfach, dass für $x \in A, y \notin A$ immer $xy \notin A$ gilt und dass für $x, y \in A$ mit $x \neq y$ auch $xy \in A$ gilt.
Wir behaupten, dass dies nun genau dann gilt, wenn A eine multiplikative Untergruppe von $\mathbb{R}\backslash\{0\}$ ist, also abgeschlossen unter Bildung von Inversen und Multiplikation.
Zunächst die eine Richtung: Ist A eine Untergruppe und $x \in A, y \notin A$, so kann nicht $xy \in A$ gelten, denn sonst wäre $\frac{1}{x} \in A$ und dann auch $y = xy \cdot \frac{1}{x}$, Widerspruch! Die zweite Bedingung gilt sogar ohne die Einschränkung $x \neq y$.
Nun umgekehrt: Sei A eine Menge mit $1 \in A$ und der Bedingung, dass aus $x \in A, y \notin A$ immer $xy \notin A$ folgt. Dann wollen wir zeigen, dass A eine Untergruppe ist.
Zunächst folgt aus $x \in A$ sicherlich $\frac{1}{x} \in A$, denn sonst könnten wir $y = \frac{1}{x}$ wählen und hätten $1 = x \cdot \frac{1}{x} \notin A$. Sind nun $x, y \in A$, so auch xy, denn sonst wäre auch $\frac{1}{xy}$ nicht in A (wegen der Inversen), dann aber auch $\frac{1}{y} = x \cdot \frac{1}{xy}$ nach unserer Bedingung nicht, aber $\frac{1}{y}$ ist in A, wenn $y \in A$ gilt.

Die Lösungen sind also $\boxed{f(x) = 0}$ und $\boxed{\begin{cases} f(x) = bx & x \in A \\ 0 & x \notin A \end{cases}}$, wobei $b \in \mathbb{R}$ beliebig ist und $A \subset \mathbb{R}\backslash\{0\}$ eine beliebige multiplikative Untergruppe ist.
Mögliche Wahlen sind $A = \mathbb{R}\backslash\{0\}$, was zur Lösung $f(x) = bx$ vom Beginn führt, oder $A = \{1\}$, was zu der ersten merkwürdigen Lösung führt, die wir oben gefunden hatten, zur zweiten gehört $A = \mathbb{R}_{>0}$. Möglich sind aber z.B. auch $A = \mathbb{Q}\backslash\{0\}$, $A = \{2^n : n \in \mathbb{Z}\}$ usw. $\square$

4.2 Doppelt hält besser

Der Titel dieses Abschnitts erklärt sich nicht allein dadurch, dass wir in der Lösung der folgenden Aufgabe aus der kurzen Liste der IMO 2005 bei zwei Schritten jeweils zwei verschiedene Lösungswege zeigen. Vor allem bezieht er sich auf die Methode, die wir jetzt kennenlernen.

Aufgabe: Finde alle Funktionen $f : \mathbb{R} \to \mathbb{R}$ mit

$$f(x+y) + f(x)f(y) = f(xy) + 2xy + 1$$

für alle $x, y \in \mathbb{R}$.

Lösung: Sei $P(x, y)$ die gegebene Bedingung. Die erste grundsätzliche Schwierigkeit ist, dass f hier mit den vier verschiedenen Argumenten $x, y, xy, x+y$ vorkommt. Setzen wir also irgendwelche Werte von x und y ein, werden wir meistens eine Gleichung zwischen vier verschiedenen Funktionswerten von f bekommen, was wahrscheinlich nicht besonders hilfreich ist. Stattdessen versuchen wir, so einzusetzen, dass möglichst viele der Werte gleich sind. Einsetzen von $y = 0$ zeigt

$$P(x, 0): \quad (f(0) + 1)f(x) = f(0) + 1.$$

Wäre $f(0) \neq -1$, dann müsste $f(x) = 1$ für alle x gelten, dies ist aber sicherlich keine Lösung. Also folgt $f(0) = -1$.
Eine andere Möglichkeit, möglichst wenige verschiedene Ausdrücke zu erhalten, ist es $x = y = 2$ einzusetzen, denn dann gilt $x + y = xy = 4$ und übrig bleibt

$$P(2, 2): \quad f(2)^2 = 9,$$

also $f(2) = \pm 3$. Kennen wir $f(2)$, so können wir $x = y = 1$ einsetzen und erhalten

$$P(1, 1): \quad f(2) + f(1)^2 = f(1) + 3.$$

Nun gibt es vier Möglichkeiten: Ist $f(2) = 3$, dann folgt $f(1) = 0$ oder $f(1) = 1$. Ist $f(2) = -3$, dann folgt $f(1) = 3$ oder $f(1) = -2$.
Es gibt also vier Fälle, die wir nun der Reihe nach abarbeiten. (Dabei ist die Reihenfolge hier der Übersichtlichkeit zuliebe so dargestellt, dass wir die einfachen Fälle zuerst arbeiten. In der Praxis weiß man das natürlich im Vorhinein nicht.)
1. Fall: $f(1) = 1$. Dann zeigt $P(x, 1)$ sofort $f(x+1) = 2x + 1$, also $\boxed{f(x) = 2x - 1}$, was tatsächlich eine Lösung ist.

2. Fall: $f(1) = 3$, also $f(2) = -3$. Hier zeigt $P(x, 1)$ nun $f(x+1) = -2f(x) + 2x + 1$. Damit können wir nun Schritt für Schritt weitere Funktionswerte ausrechnen, etwa $f(3) = 11, f(4) = -15, f(5) = 39, f(6) = -67$ usw. Dies sieht nicht nach einer gescheiten Funktion aus und tatsächlich können wir einfach relativ beliebige andere Werte für x, y einsetzen, etwa $P(2, 3)$ und erhalten aus den bereits berechneten Funktionswerten einen Widerspruch. In diesem Fall gibt es also keine Lösung.

3. Fall: $f(1) = -2$, also $f(2) = -3$. Wieder zeigt $P(x, 1)$ jetzt $f(x+1) = 3f(x) + 2x + 1$ und wir können $f(n)$ für alle $n \in \mathbb{Z}$ ausrechnen, etwa $f(3) = -4, f(4) = -5, f(5) = -6$ usw. Dies sieht verdächtig nach $f(x) = -x - 1$ aus und tatsächlich zeigt eine Probe, dass dies eine Lösung der Funktionalgleichung ist. Mit unserer Rekursion für $f(x+1)$ können wir auch leicht $f(x) = -x - 1$ für alle ganzen Zahlen x zeigen. Wir wollen nun zeigen, dass dies sogar für alle reellen x gilt.

Erste Lösung in Fall 3: Wir erinnern uns an die Bedingung

$$P(x,y): \quad f(x+y)+f(x)f(y)=f(xy)+2xy+1$$

Wir haben insbesondere schon $f(-1)=0$ gezeigt. Jetzt zeigt

$$P(x,-1): \quad f(x-1)=f(-x)-2x+1.$$

Wir wissen aber auch (indem wir den Index in der Rekursion oben um 1 verschieben)

$$f(x)=3f(x-1)+2x-1=3f(-x)-4x+2.$$

Wir haben also eine Gleichung zwischen $f(x)$ und $f(-x)$, die allerdings nicht so schön symmetrisch ist wie die Beziehungen $f(-x)=\pm f(x)$ aus dem zweiten Kapitel.
Diese Asymmetrie kommt uns aber hier zugute! Denn wir können x durch $-x$ ersetzen und erhalten eine zweite Gleichung zwischen $f(x)$ und $f(-x)$, nämlich

$$f(-x)=3f(x)+4x+2.$$

Kombinieren liefert

$$f(x)=3(3f(x)+4x+2)-4x+2=9f(x)+8x+8$$

und damit $8f(x)=-8x-8$, also $\boxed{f(x)=-x-1}$. □

Die nächste Lösung sieht etwas komplizierter aus, die Methode ist aber oft sehr nützlich.

Zweite Lösung in Fall 3: Iterieren wir unsere Rekursion, erhalten wir

$$f(x+2)=3f(x+1)+2x+3=9f(x)+8x+6.$$

Jetzt können wir $y=2$ in die Ausgangsgleichung einsetzen und erhalten

$$P(x,2): \quad f(x+2)-3f(x)=f(2x)+4x+1$$

und nach Einsetzen der Formel für $f(x+2)$ und Auflösen $f(2x)=6f(x)+4x+5$. Das sieht hübsch aus, ist aber allein noch etwas wertlos. Natürlich können wir das gleiche aber auch mit anderen Zahlen statt 2 machen. Mit einer kurzen Rechnung erhalten wir

$$f(x+4)=9f(x+2)+8x+22=81f(x)+80x+76$$

und dann

$$P(x,4): \quad f(x+4)-5f(x)=f(4x)+8x+1$$

und nach Einsetzen und Auflösen $f(4x)=76f(x)+72x+75$. Andererseits können wir $f(4x)$ aber auch ausrechnen, indem wir die Formel für $f(2x)$ zweimal anwenden:

$$f(4x)=6f(2x)+8x+5=36f(x)+32x+35.$$

Entscheidend ist jetzt eigentlich nur, dass unsere beiden Ausdrücke für $f(4x)$ nicht identisch sind. So können wir sie gleichsetzen und erhalten

$$76f(x)+72x+75=36f(x)+32x+35$$

und nach Auflösen sofort $f(x)=-x-1$ wie gewünscht. □

Die gemeinsame Strategie in diesen beiden Lösungen war, **einen Ausdruck auf zwei verschiedene Weisen zu berechnen**. In der ersten Lösung war das der Ausdruck $f(-x)$, in der zweiten Lösung der Ausdruck $f(2x)$. Allgemeiner kann man versuchen, möglichst viele Gleichungen mit möglichst wenigen verschiedenen Ausdrücken zu erhalten.

4. Fall: $f(1) = 0$, also $f(2) = 3$. Wieder zeigt $P(x, 1)$ sofort

$$f(x + 1) = f(x) + 2x + 1.$$

Damit können wir wieder schrittweise f bei allen ganzen Zahlen ausrechnen, also etwa $f(3) = 8, f(4) = 15, f(5) = 24$ und man erkennt nun leicht das Muster und zeigt per Induktion $f(n) = n^2-1$ für alle ganzen Zahlen n. Eine Probe zeigt auch, dass $f(x) = x^2-1$ tatsächlich die Funktionalgleichung erfüllt. Wie zeigen wir dies für alle reellen x?
Wir können natürlich versuchen, die in Fall 3 benutzten Strategien wiederzuverwenden. Mit Induktion zeigen wir leicht $f(x + n) = f(x) + 2nx + n^2$ für alle $n \in \mathbb{Z}$ und $x \in \mathbb{R}$. Nun folgt

$$P(x, n): \quad f(x + n) + (n^2 - 1)f(x) = f(nx) + 2nx + 1$$

und Einsetzen führt zu $f(nx) = n^2 f(x) + n^2 - 1$, also z.B. wieder $f(2x) = 4f(x) + 3$ oder $f(4x) = 16f(x) + 15$. Die Strategie von eben klappt aber nicht, denn wir können zwar auch $f(4x) = 4f(2x) + 3 = 4(4f(x) + 3) + 3 = 16f(x) + 15$ rechnen, erhalten aber den gleichen Ausdruck noch einmal, also keine neue Information.
Man könnte jetzt natürlich versuchen, etwa $f(6x)$ auf zwei verschiedene Weisen auszurechnen, einmal direkt mit $n = 6$ und dann einmal aus den Ausdrücken für $f(2x)$ und $f(3x)$, aber auch hier werden wir zweimal den gleichen Ausdruck erhalten.
Auch die in der ersten Lösung oben benutzte Strategie funktioniert nicht mehr: Mit $n = -1$ haben wir $f(-x) = f(x)$, d.h. f ist gerade! Dies ist nun allerdings nicht mehr asymmetrisch, wir erhalten also durch Vertauschen von x und $-x$ keine zweite Gleichung.

Erste Lösung in Fall 4: Wollen wir noch einen anderen Ausdruck für $f(2x)$ erhalten, ist es eine Idee, $x = y$ zu setzen. Tatsächlich zeigt

$$P(x, x): \quad f(2x) + f(x)^2 = f(x^2) + 2x^2 + 1.$$

Der Nachteil ist, dass wir nun einen Term $f(x^2)$ erhalten. Wir haben also nicht wirklich einen zweiten Ausdruck für $f(2x)$, aber einen Ausdruck für $f(x^2)$ gefunden: Einsetzen von $f(2x) = 4f(x) + 3$ zeigt nämlich

$$f(x^2) = f(x)^2 + 4f(x) - 2x^2 + 2.$$

Dies ist nun allerdings wieder erst hilfreich, wenn wir eine weitere Möglichkeit finden, eine Gleichung mit $f(x^2)$ zu finden. Die entscheidende Idee ist nun die folgende:

$$P(x - 1, x + 1): \quad f(2x) + f(x - 1)f(x + 1) = f(x^2 - 1) + 2x^2 - 1 = f(x^2).$$

Hier können wir alle Ausdrücke auf der linken Seite „ausrechnen“ und erhalten

$$f(x^2) = f(x)^2 + 6f(x) + 4 - 4x^2.$$

Jetzt haben wir wirklich zwei verschiedene Ausdrücke für $f(x^2)$ und Gleichsetzen zeigt

$$f(x)^2 + 4f(x) - 2x^2 + 2 = f(x)^2 + 6f(x) + 4 - 4x^2$$

und nach Auflösen bleibt nur $f(x) = x^2 - 1$ wie gewünscht. $\square$

Zweite Lösung in Fall 4:
Die Symmetrie $f(-x) = f(x)$ führt zwar nicht unmittelbar zum Ziel, wir können sie aber wie im zweiten Kapitel gelernt einsetzen. Etwa zeigt jetzt

$$P(x, -y): \quad f(x-y) + f(x)f(y) = f(xy) - 2xy + 1$$

und Abziehen von $P(x, y)$ führt zu

$$f(x+y) - f(x-y) = 4xy.$$

Hier können wir aber einfach $x = y$ setzen und erhalten $f(2x) = 4x^2 - 1$, also $\boxed{f(x) = x^2 - 1}$ für beliebige x. □

Aus der recht langen Diskussion dieser Aufgabe können wir neben der neuen Methode noch etwas mitnehmen: *Eine Fallunterscheidung ist unser Freund, nicht unser Feind!* Am Anfang denkt man vielleicht, dass es ärgerlich ist, dass wir weder $f(1)$ noch $f(2)$ unmittelbar ausrechnen können. Allerdings haben wir gesehen, dass es zumindest in drei von vier Fällen auch eine Lösung gibt. Die Fallunterscheidung liefert uns also eine Möglichkeit, diese drei Lösungen *voneinander zu trennen*. In jedem Fall erhalten wir dann in gewisser Weise eine neue Aufgabe und die Lösung kann von Fall zu Fall sehr unterschiedlich sein.

4.3 Aufgaben

Aufgabe 4.1: Finde alle Funktionen $f : \mathbb{R} \to \mathbb{R}$ mit

$$x^2 f(x) + y f(y^2) = f(x+y) f(x^2 - xy + y^2)$$

für alle $x, y \in \mathbb{R}$.

Aufgabe 4.2: Finde alle Funktionen $f : \mathbb{R} \to \mathbb{R}$ mit

$$f(xf(x) + 2y) = f(x^2) + f(y) + x + y - 1$$

für alle $x, y \in \mathbb{R}$.

Aufgabe 4.3: Finde alle Funktionen $f : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ mit

$$f(x^3 + y) = f(x)^3 + \frac{f(xy)}{f(x)}$$

für alle $x, y > 0$.