

Problem Books in Mathematics

Edited by P. Winkler

Problem Books in Mathematics

Series Editor: Peter Winkler

Pell's Equation

by *Edward J. Barbeau*

Polynomials

by *Edward J. Barbeau*

Problems in Geometry

by *Marcel Berger, Pierre Pansu, Jean-Pic Berry, and Xavier Saint-Raymond*

Problem Book for First Year Calculus

by *George W. Bluman*

Exercises in Probability

by *T. Cacoullos*

Probability Through Problems

by *Marek Capiński and Tomasz Zastawniak*

An Introduction to Hilbert Space and Quantum Logic

by *David W. Cohen*

Unsolved Problems in Geometry

by *Hallard T. Croft, Kenneth J. Falconer, and Richard K. Guy*

Berkeley Problems in Mathematics, (Third Edition)

by *Paulo Ney de Souza and Jorge-Nuno Silva*

**The IMO Compendium: A Collection of Problems Suggested for the
International Mathematical Olympiads: 1959-2004**

by *Dušan Djukić, Vladimir Z. Janković, Ivan Matić, and Nikola Petrović*

Problem-Solving Strategies

by *Arthur Engel*

Problems in Analysis

by *Bernard R. Gelbaum*

Problems in Real and Complex Analysis

by *Bernard R. Gelbaum*

(continued after index)

Christopher G. Small

Functional Equations and How to Solve Them

 Springer

Christopher G. Small
Department of Statistics & Actuarial Science
University of Waterloo
200 University Avenue West
Waterloo N2L 3G1
Canada
cgsmall@uwaterloo.ca

Series Editor:
Peter Winkler
Department of Mathematics
Dartmouth College
Hanover, NH 03755
USA
Peter.winkler@dartmouth.edu

Mathematics Subject Classification (2000): 39-xx

Library of Congress Control Number: 2006929872

ISBN-10: 0-387-34534-5 e-ISBN-10: 0-387-48901-0
ISBN-13: 978-0-387-34534-5 e-ISBN-13: 978-0-387-48901-8

Printed on acid-free paper.

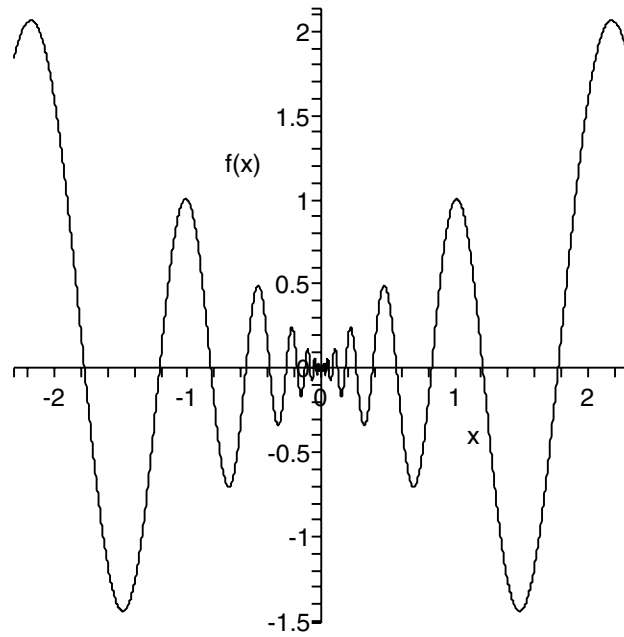
© 2007 Springer Science+Business Media, LLC

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science+Business Media, LLC, 233 Spring Street, New York, NY 10013, USA), except for with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.

The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

9 8 7 6 5 4 3 2 1

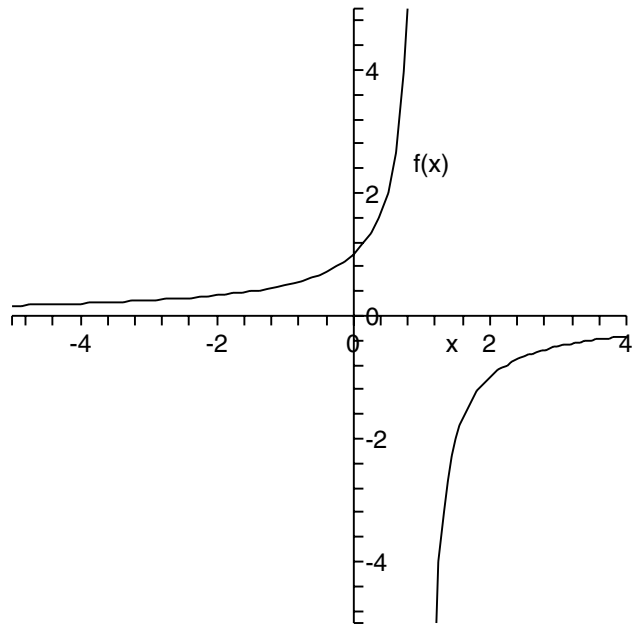
springer.com



$$f(x) + f(2x) + f(3x) = 0$$

for all real x .

This functional equation is satisfied by the function $f(x) \equiv 0$, and also by the strange example graphed above. To find out more about this function, see Chapter 3.



$$f(f(f(x))) = x$$

Can you discover a function $f(x)$ which satisfies this functional equation?

Contents

Preface	ix
1 An historical introduction	1
1.1 Preliminary remarks	1
1.2 Nicole Oresme	1
1.3 Gregory of Saint-Vincent	4
1.4 Augustin-Louis Cauchy	6
1.5 What about calculus?	8
1.6 Jean d'Alembert	9
1.7 Charles Babbage	10
1.8 Mathematics competitions and recreational mathematics	16
1.9 A contribution from Ramanujan	21
1.10 Simultaneous functional equations	24
1.11 A clarification of terminology	25
1.12 Existence and uniqueness of solutions	26
1.13 Problems	26
2 Functional equations with two variables	31
2.1 Cauchy's equation	31
2.2 Applications of Cauchy's equation	35
2.3 Jensen's equation	37
2.4 Linear functional equation	38
2.5 Cauchy's exponential equation	38
2.6 Pexider's equation	39
2.7 Vincze's equation	40
2.8 Cauchy's inequality	42
2.9 Equations involving functions of two variables	43
2.10 Euler's equation	44
2.11 D'Alembert's equation	45
2.12 Problems	49

3	Functional equations with one variable	55
3.1	Introduction	55
3.2	Linearization	55
3.3	Some basic families of equations	57
3.4	A menagerie of conjugacy equations	62
3.5	Finding solutions for conjugacy equations	64
3.5.1	The Koenigs algorithm for Schröder's equation	64
3.5.2	The Lévy algorithm for Abel's equation	66
3.5.3	An algorithm for Böttcher's equation	66
3.5.4	Solving commutativity equations	67
3.6	Generalizations of Abel's and Schröder's equations	67
3.7	General properties of iterative roots	69
3.8	Functional equations and nested radicals	72
3.9	Problems	75
4	Miscellaneous methods for functional equations	79
4.1	Polynomial equations	79
4.2	Power series methods	81
4.3	Equations involving arithmetic functions	82
4.4	An equation using special groups	87
4.5	Problems	89
5	Some closing heuristics	91
6	Appendix: Hamel bases	93
7	Hints and partial solutions to problems	97
7.1	A warning to the reader	97
7.2	Hints for Chapter 1	97
7.3	Hints for Chapter 2	102
7.4	Hints for Chapter 3	107
7.5	Hints for Chapter 4	113
8	Bibliography	123
	Index	125

Preface

Over the years, a number of books have been written on the theory of functional equations. However, few books have been published on solving functional equations which arise in mathematics competitions and mathematical problem solving. The intention of this book is to go some distance towards filling this gap.

This work began life some years ago as a set of training notes for mathematics competitions such as the William Lowell Putnam Competition for undergraduate university students, and the International Mathematical Olympiad for high school students. As part of the training for these competitions, I tried to put together some systematic material on functional equations, which have formed a part of the International Mathematical Olympiad and a small component of the Putnam Competition. As I became more involved in coaching students for the Putnam and the International Mathematical Olympiad, I started to understand why there is not much training material available in systematic form. Many would argue that there is no theory attached to functional equations that are encountered in mathematics competitions. Each such equation requires different techniques to solve it. Functional equations are often the most difficult problems to be found on mathematics competitions because they require a minimal amount of background theory and a maximal amount of ingenuity. The great advantage of a problem involving functional equations is that you can construct problems that students at all levels can understand and play with. The great disadvantage is that, for many problems, few students can make much progress in finding solutions even if the required techniques are essentially elementary in nature. It is perhaps this view of functional equations which explains why most problem-solving texts have little systematic material on the subject. Problem books in mathematics usually include some functional equations in their chapters on algebra. But by including functional equations among the problems on polynomials or inequalities the essential character of the methodology is often lost.

As my training notes grew, so grew my conviction that we often do not do full justice to the role of theory in the solution of functional equations. The

result of my growing awareness of the interplay between theory and problem application is the book you have before you. It is based upon my belief that a firm understanding of the theory is useful in practical problem solving with such equations. At times in this book, the marriage of theory and practice is not seamless as there are theoretical ideas whose practical utility is limited. However, they are an essential part of the subject that could not be omitted. Moreover, today's theoretical idea may be the inspiration for tomorrow's competition problem as the best problems often arise from pure research. We shall have to wait and see.

The student who encounters a functional equation on a mathematics contest will need to investigate solutions to the equation by finding all solutions (if any) or by showing that all solutions have a particular property. Our emphasis is on the development of those tools which are most useful in giving a family of solutions to each functional equation in explicit form.

At the end of each chapter, readers will find a list of problems associated with the material in that chapter. The problems vary greatly in difficulty, with the easiest problems being accessible to any high school student who has read the chapter carefully. It is my hope that the most difficult problems are a reasonable challenge to advanced students studying for the International Mathematical Olympiad at the high school level or the William Lowell Putnam Competition for university undergraduates. I have placed stars next to those problems which I consider to be the harder ones. However, I recognise that determining the level of difficulty of a problem is somewhat subjective. What one person finds difficult, another may find easy.

In writing these training notes, I have had to make a choice as to the generality of the topics covered. The modern theory of functional equations can occur in a very abstract setting that is quite inappropriate for the readership I have in mind. However, the abstraction of some parts of the modern theory reflects the fact that functional equations can occur in diverse settings: functions on the natural numbers, the integers, the reals, or the complex numbers can all be studied within the subject area of functional equations. Most of the time, the functions I have in mind are real-valued functions of a single real variable. However, I have tried not to be too restrictive in this. The reader will also find functions with complex arguments and functions defined on natural numbers in these pages. In some cases, equations for functions between circles will also crop up. Nor are functional inequalities ignored.

One word of warning is in order. You cannot study functional equations without making some use of the properties of limits and continuous functions. The fact is that many problems involving functional equations depend upon an assumption of such as continuity or some other regularity assumption that would usually not be encountered until university. This presents a difficulty for high school mathematics contests where the properties of limits and continuity cannot be assumed. One way to get around this problem is to substitute another regularity condition that is more acceptable for high school mathematics. Thus a problem where a continuity condition is natural may well get

by with the assumption of monotonicity. Although continuity and monotonicity are logically independent properties (in the sense that neither implies the other) the imposition of a monotonicity condition in a functional equations problem may serve the same purpose as continuity. Another way around the problem is to ask students to provide a weaker conclusion that is not “finished” by invoking continuity. Asking students to determine the nature of a function on the rational numbers is an example of this. Neither solution to this problem is completely satisfactory. Fortunately, there are enough problems which can be posed and solved using high school mathematics to serve the purpose. More advanced contests such as the William Lowell Putnam Competition have no such restrictions in imposing continuity or convexity, and expect the student to treat these assumptions with mathematical maturity.

Some readers may be surprised to find that the chapter on functional equations in a single variable follows that on functional equations in two or more variables. However this is the correct order. An equation in two or more variables is formally equivalent to a family of simultaneous equations in one variable. So equations in two variables give you more to play with. I have had to be very selective in choosing topics in the third chapter, because much of the academic literature is devoted to establishing uniqueness theorems for solutions within particular families of functions: functions that are convex or real analytic, functions which obey certain order conditions, and so on. It would be easy to simply ignore the entire subject if it were not for the fact that functional equations in a single variable are commonplace in mathematics competitions. So I have done my best to present those tools and unifying concepts which occur periodically in such problems in both high school and university competitions. Chapter 3 has been written with a confidence that advanced high school students will adapt well to the challenge of a bit of introductory university level mathematics. The chapters can be read more-or-less independently of each other. There are some results in later chapters which depend upon earlier chapters. However, the reader who wishes to sample the book in random order can probably piece together the necessary information. The fact that it is possible to write a book whose chapters are not heavily dependent is a consequence of the character of functional equations. Unlike some branches of mathematics, the subject is wide, providing easier access from a number of perspectives. This makes it an excellent area for competition problems. Even tough functional equations are relatively easy to state and provide lots of “play value” for students who may not be able to solve them completely.

Because this is a book about problem solving, the reader may be surprised to find that it begins with a chapter of the history of the subject. It is my belief that the present way of teaching mathematics to students puts much emphasis on the tools of mathematics, and not enough on the intellectual climate which gave rise to these ideas. Functional equations were posed and solved for reasons that had much to do with the intellectual challenges of

their times. This book attempts to provide a small glimpse of some of those reasons.

I have learned much about functional equations from other people. This book also owes much to others. So this preface would not be complete without some mention of the debts that I owe. I have learned much from the work of Janos Aczél, Distinguished Professor Emeritus at the University of Waterloo. The impact of his work and that of his colleagues is to be found throughout the following pages in places too numerous to mention. The initial stages of this monograph were written at the instigation of Pat Stewart and Richard Nowakowski. Sadly, Pat Stewart is no longer with us, and is missed by the mathematical community. Thank you, Patrick and Richard. Finally, I would like to thank Professor Ed Barbeau, who generously sent some of his correspondence problems to me. His encouragement and assistance are much appreciated.

An historical introduction

1.1 Preliminary remarks

In high school algebra, we learn about algebraic equations involving one or more unknown real numbers. Functional equations are much like algebraic equations, except that the unknown quantities are functions rather than real numbers. This book is about functional equations: their role in contemporary mathematics as well as the body of techniques that is available for their solution. Functional equations appear quite regularly on mathematics competitions. So this book is intended as a toolkit of methods for students who wish to tackle competition problems involving functional equations at the high school or university level.

In this chapter, we take a rather broad look at functional equations. Rather than focusing on the solutions to such equations—a topic for later chapters—we show how functional equations arise in mathematical investigations. Our entry into the subject is primarily, but not solely, historical.

1.2 Nicole Oresme

Mathematicians have been working with functional equations for a much longer period of time than the formal discipline has existed. Examples of early functional equations can be traced back as far as the work of the fourteenth century mathematician Nicole Oresme who provided an indirect definition of linear functions by means of a functional equation. Of Norman heritage, Oresme was born in 1323 and died in 1382. To put these dates in perspective, we should note that the dreaded Black Death, which swept through Europe killing possibly as much as a third of the population, occurred around the middle of the fourteenth century. Although the origins of the Black Death are unclear, we know that by December of 1347, it had reached the western Mediterranean through the ports of Sicily, then Sardinia, then the port city

of Marseilles. It reached Paris in the spring of 1348, having spread throughout much of the country.

The year 1348 is also of some significance in mathematics, because that is the year that Nicole Oresme is recorded in a list of scholarship holders at the University of Paris. Thus it appears that Oresme was studying at the University of Paris from some time in the early 1340s up to the time the Black Death arrived in the city itself. Today, perhaps, we might well wonder how, in the face of this calamitous disease, scholars such as Oresme were able to flourish. However, the Black Death was more disruptive for the generation that followed Oresme. He and his colleagues had completed much of their education by the time that the Black Death's devastating effects were felt. By 1355, Oresme had obtained the Master of Theology degree, and was soon thereafter appointed Grand Master at the College of Navarre, one of the colleges of the University of Paris, founded in 1304. Nicole Oresme was arguably the greatest European mathematician of the fourteenth century. He died in 1382 in Lisieux.

Although he lived in a medieval world in which the writings of Aristotle were the dominant influence on natural philosophy—what we would now call the natural sciences—Oresme's scholarly work foreshadowed the work of later writers in the Renaissance and Enlightenment periods, who broke away from Aristotle to reformulate the laws of mechanics and, by so doing, create the field of classical physics. In 1352, Oresme wrote a major treatise on uniformity and difformity of intensities, entitled *Tractatus de configurationibus qualitatum et motuum*. In this important work, Oresme established the definition of a *functional relationship* between two variables, and the idea (well ahead of René Descartes) that one can express this relationship geometrically by what we would now call a graph.¹ In Part 1 he wrote

Therefore, every intensity which can be acquired successively ought to be imagined by a straight line perpendicularly erected on some point of the space or subject of the intensible thing, e.g., a quality. For whatever ratio is found to exist between intensity and intensity, in relating intensities of the same kind, a similar ratio is found to exist between line and line, and vice versa.²

¹ Part of Oresme's efforts were dedicated to proving the so-called *mean speed theorem*, also known as the *Merton theorem* because of its associations with the work at Merton College, Oxford. William of Heytesbury, a Fellow of Merton College summarised this result in his treatise *Rules for Solving Sophisms* in 1335 by saying “the moving body, acquiring or losing . . . [velocity] uniformly during some period of time, will traverse distance exactly equal to what it would traverse in an equal period of time if it were moved uniformly at its mean degree [of velocity].” An important consequence of this theorem is that a body undergoing a constant acceleration (such as a freely falling body) will traverse a distance which is a quadratic function of time.

² This is the 1968 translation of Marshall Clagett. See Oresme [1968].

Of central interest in his treatise is the idea of uniform motion and “uniformly difform motion,” the latter denoting the motion of a particle undergoing uniform acceleration.³ Also considered was “difformly difform motion,” where the acceleration itself varied. In the section on quadrangular quality, Oresme took care to define his notion of uniform difformity (i.e., linearity) as follows.

A uniform quality is one which is equally intense in all parts of the subject, while a quality uniformly difform is one in which if any three points [of the subject line] are taken, the ratio of the distance between the first and the second to the distance between the second and the third is as the ratio of the excess in intensity of the first point over that of the second point to the excess of that of the second point over that of the third point, calling the first of those three points the one of greatest intensity.

As Aczél [1984] and Aczél and Dhombres [1985] have noted, the passage defines a linear function (i.e., a quality which is uniformly difform) through a functional equation. In modern terminology, we would have three distinct⁴ real numbers x , y , and z , say, which are described in the passage above as three points of the subject line. Associated with x , y and z , we have a variable (i.e., the “intensity” of the quality at each point of the subject line) which we can write as $f(x)$, $f(y)$, and $f(z)$, respectively. The function f is defined to be linear, or “uniformly difform” if

$$\frac{y - x}{z - y} = \frac{f(y) - f(x)}{f(z) - f(y)} \quad \text{for all distinct values of } x, y, z. \quad (1.1)$$

What makes Oresme’s definition a functional equation is that f is treated abstractly: one may plug any function into this equation to see whether the equation is satisfied for all possible x , y and z . We can compare this with the standard definition to be found in most introductory modern textbooks which say that a linear function is one of the form

$$f(x) = ax + b \quad \text{for some } a, b. \quad (1.2)$$

³ Although in modern mathematical terms we would probably prefer to translate the word “uniform” as “constant” here. Thus we could say in the case of uniform motion that a particle which undergoes motion with a constant velocity changes its position uniformly, i.e., linearly, as a function of time. The preferred choice of terminology can be left to the taste of the reader.

⁴ By “distinct” here, I mean that no pair of the three numbers can be equal. For the purposes of the definition it is sufficient to require that $y \neq z$. However, the geometric language in which Oresme frames his definition clearly points to the interpretation given. Note also that we need to take a small liberty with the text and interpret the word “distance” as “signed distance” in the modern sense. Trying to resolve this ambiguity by ordering the points and requiring the function to be increasing is too artificial.

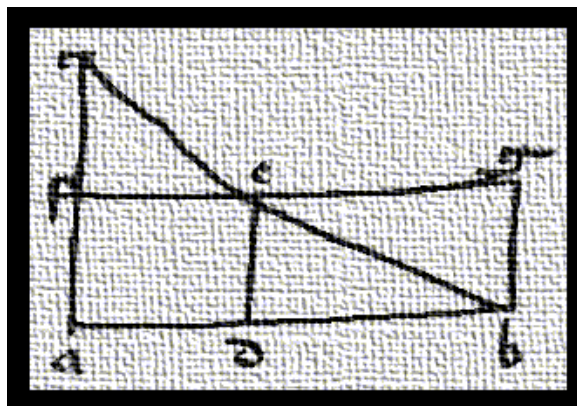


Fig. 1.1. One of Oresme’s “uniformly difform” (i.e., linear) functions, defined by a functional equation. Redrawn and adapted from a manuscript diagram in his *Tractatus de configurationibus qualitatum et motuum*. An illustration for the Mean Speed Theorem.

Oresme’s equation in (1.1) is a functional equation. The definition in (1.2) with $a \neq 0$ is its solution. Note that Oresme’s definition does not allow the constant linear function where $a = 0$. This is faithful to his intentions here, which distinguish uniformly difform functions from uniform functions according to the choices $a \neq 0$ and $a = 0$, respectively.

1.3 Gregory of Saint-Vincent

Over the next few hundred years, functional equations were used but no general theory of such equations arose. Notable among such mathematicians was Gregory of Saint-Vincent (1584–1667), whose work on the hyperbola made implicit use of the functional equation $f(xy) = f(x) + f(y)$, and pioneered the theory of the logarithm.

Saint-Vincent’s result appeared in his great 1647 treatise entitled *Opus Geometricum quadraturae circuli et sectionum conii*. If the title of this work appears long, the treatise itself, at about 1250 pages, was much longer! It deals with methods for calculating areas and with the properties of conic sections. In particular, Saint-Vincent shows how it is possible to calculate the area under an hyperbola such as $y = x^{-1}$ as in Figure 1.3. In modern times, the area under a curve such as an hyperbola is a topic usually left for the theory of integration. However, Saint-Vincent made great progress on the problem using purely geometric arguments.

In particular, Saint-Vincent’s argument was based upon the following geometrical principle.

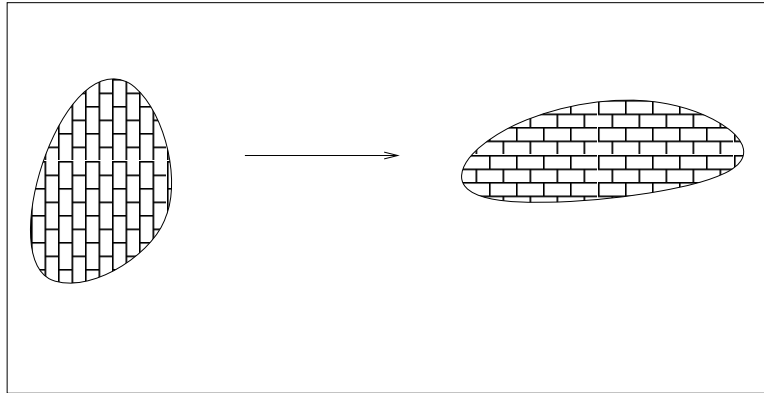


Fig. 1.2. Simultaneous stretching and shrinking of a planar region in perpendicular directions. If a region is simultaneously stretched and shrunk in perpendicular directions by the same factor, the area will be unchanged.

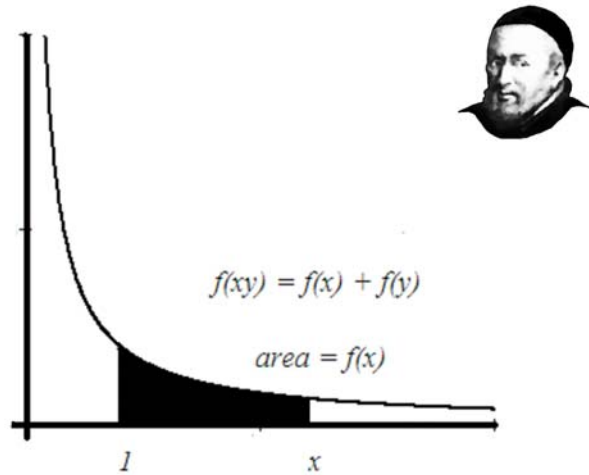


Fig. 1.3. Gregory of Saint-Vincent pioneered the theory of logarithms by recognising that the area under an hyperbola satisfies a functional equation. Using the geometric argument for constant area shown above, Gregory of Saint-Vincent was able to derive the functional equation now associated with the logarithm.

If a planar region is stretched horizontally by a given factor, and simultaneously shrunk vertically by the same factor, then the resulting region will have an area which is equal to that of the original region.

For example, in Figure 1.2, we see that a planar region has been scaled vertically and horizontally by stretching with a factor of 2 horizontally, and shrinking by a factor of 2 vertically. The second region has the same area as the first.

Now let us see how this geometrical principle applies to the area under the hyperbola $y = x^{-1}$. Let $f(x)$ denote the shaded area on the interval from 1 to x shown in Figure 1.3, and consider the corresponding shaded area under the same hyperbola erected on the interval from y to xy for any $y > 1$ say. Comparing the two shaded regions, Saint-Vincent noted that they differ by a scale factor of y along the x -axis, and by a scale factor of y^{-1} along the y -axis. Thus the areas of the two regions must be the same.

The area of the shaded region with base from y to xy is $f(xy) - f(y)$. This follows immediately from the fact that the region under the hyperbola from y to xy is exactly that which is obtained by removing the region from 1 to y from the region from 1 to xy . Thus, using Saint-Vincent's scaling argument, we have

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

or equivalently that

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

We now recognise this equation as the distinctive functional equation for the family of logarithms. However, the theoretical work which links this functional equation to the family of logarithms had to wait for the work of Augustin-Louis Cauchy.

Along with his work on conics and the calculation of area, Saint-Vincent is also remembered for the contributions from the second part of his treatise *Opus geometricum*, where he studied infinite series. With his work on areas, the method of exhaustion, and series, Gregory of Saint-Vincent was one of the early pioneers of the modern methods of calculus and analysis.

1.4 Augustin-Louis Cauchy

Although Nicole Oresme's definition of linearity can be interpreted as an early example of a functional equation, it does not represent a starting point for the theory of functional equations. The subject of functional equations is more properly dated from the work of A. L. Cauchy. Born in 1789 in Paris, France, Cauchy's early years coincided with the French Revolution. To put his birthdate in context, we should recall that the French Revolution is generally dated to the ten-year period 1789–1799, starting roughly with the storming of the Bastille in 1789. In 1799, when the young Cauchy was ten years old, general

Napoleon Bonaparte led a coup against the *Directoire* to begin a period of direct military rule of France. As Cauchy's family had royalist sympathies, they left Paris and did not return until 1800. A strong monarchist himself, Augustin-Louis Cauchy later ran counter to the republican and Napoleonic trends in France during his early years. A brilliant mathematician, Cauchy worked in many areas of mathematics. However, he is primarily known for his work on calculus, and is recognised as one of the founders of the modern theory of mathematical analysis.

The functional equation that is particularly associated with Cauchy is

$$f(x + y) = f(x) + f(y) \quad (1.3)$$

for all real x and y , and is now called *Cauchy's equation*. It is required to find all real-valued functions f satisfying equation (1.3). Now the reader can immediately notice that Cauchy's equation is satisfied by any function of the form

$$f(x) = ax,$$

the constant a being an arbitrary real number. However, our ability to find a simple solution to this equation is only a small part of the story. We must also ask whether the family of functions of the form $f(x) = ax$ is the *complete set of all solutions* to equation (1.3). It seems reasonable that such linear functions are the only solutions to (1.3). However, this turns out to be true only if some mild restriction is imposed upon the function f . For example, functions of the form $f(x) = ax$ are the only solutions to (1.3) among the class of functions which are bounded on some interval of the form $(-c, c)$, where $c > 0$. Alternatively, it can be shown that $f(x) = ax$ form the only class of solutions among the continuous real-valued functions on the real line. We investigate this equation and its solutions in detail in Chapter 2.

What was the motivation for Cauchy's investigation of the full set of solutions to (1.3)? To understand this, we must examine Cauchy's rigorous derivation of the general statement of the *binomial theorem*. For centuries, mathematicians have known the formula

$$(1 + x)^n = 1 + \binom{n}{1}x + \binom{n}{2}x^2 + \cdots + \binom{n}{n-1}x^{n-1} + x^n \quad (1.4)$$

for nonnegative integer values of n . The binomial coefficients of this sum, namely

$$\binom{n}{i} = \frac{n(n-1)(n-2)\cdots(n-i+1)}{i!} \quad (1.5)$$

are the well-known entries in Pascal's triangle.

In fact, the binomial coefficients defined in (1.5) are meaningfully defined by the formula when n is replaced with any real number z , as long as i remains a nonnegative integer. It was Isaac Newton who, in 1676, demonstrated how to extend the binomial formula in (1.4) in order to expand $(1 + x)^z$ in powers of x when z is an arbitrary real number. Newton's formula

$$(1+x)^z = 1 + zx + \binom{z}{2}x^2 + \binom{z}{3}x^3 + \dots \quad (1.6)$$

is an infinite sum, which reduces to the finite sum in (1.4) because terms after the $(z+1)$ st vanish when z is a nonnegative integer. Unfortunately, Newton's proof of (1.6) was not rigorous. So it was left to subsequent mathematicians to fill in the entire argument. Cauchy began by considering the right hand side of equation (1.6). The first problem that arises is whether this expression has any meaning at all. For example, if $z = -1$ and $x = 1$ we get

$$1 - 1 + 1 - 1 + 1 - \dots,$$

which does not converge to a finite real number as more and more terms are summed, despite the fact that the left hand side of (1.6) equals $1/2$. Cauchy rigorously demonstrated that when $|x| < 1$ the right hand side does converge to a (finite) real number for all real values of z . So he defined a function

$$f(z) = 1 + zx + \binom{z}{2}x^2 + \binom{z}{3}x^3 + \dots \quad (1.7)$$

and showed by combinatorial arguments and rules for multiplying two such infinite sums that

$$f(z+w) = f(z)f(w) \quad (1.8)$$

for all real z and w . This functional equation is called *Cauchy's exponential equation*, and we meet it again in the next chapter. It is tempting simply to take logarithms in order to turn it into (1.3). However, we have to be careful to check that $f(z) > 0$ before we can say that $\log f(z)$ is meaningful. As we show, a solution to Cauchy's exponential equation is either everywhere zero or is everywhere strictly positive. As the latter is the case here, we can take logarithms to obtain the equation

$$g(z+w) = g(z) + g(w)$$

where $g = \log f$. In order to show that $g(z) = az$ for some a , Cauchy had to demonstrate that f , and therefore g , are continuous functions. This turned out to be harder than he expected. Having verified that $g(z) = az$ for some a , he was able to conclude that $f(z) = b^z$ for some value of b . In particular, $f(1) = b$. It remained for Cauchy to observe that $b = 1+x$, a fact that is immediately deduced by setting $z = 1$ in (1.7).

1.5 What about calculus?

Readers who know some calculus may wonder why Cauchy's equation $f(x+y) = f(x) + f(y)$ cannot be solved by differentiating.

Substituting $y = c$, a constant, and differentiating with respect to x , we get

$$f'(x + c) = f'(x)$$

for all real c . This implies that f' is a constant function, and therefore that f is a linear function of the form $f(x) = ax + b$. Substituting back into Cauchy's equation, we determine that $b = 0$.

This argument is fine, as far as it goes. However, it makes a very big assumption, namely that f can be differentiated. Although many of us are used to making this assumption as a matter of course, we would prefer to prove results without any assumptions other than those which are strictly necessary. Functions which do not have derivatives at certain points are quite common. For example, the function $f(x) = |x|$ has a derivative everywhere except at $x = 0$. In higher mathematics, it is not uncommon to consider functions which have no derivative at any value of x . A good example of such a function is

$$f(x) = \begin{cases} 1 & \text{if } x \text{ is rational} \\ 0 & \text{if } x \text{ is irrational.} \end{cases}$$

This function is not continuous for any value of x . So it has no derivative for any value of x . It is even possible to construct continuous functions which do not have a derivative at any value x .

The point is that we do not wish to rule out functions from consideration by automatically assuming that we can differentiate $f(x)$. We will have occasion to assume that derivatives exist in subsequent functional equations. However, such assumptions should always be made sparingly, with an eye to removing them if they are unnecessary for the proof of the result.

1.6 Jean d'Alembert

Historically, Jean d'Alembert precedes Augustin-Louis Cauchy. However, in the context of functional equations, it seems more natural to consider his contributions after Cauchy.

Jean d'Alembert was a man of many names. The illegitimate son of an army officer, Louis-Camus Destouches, and a writer, Claudine Guérin de Tencin, he was born in Paris in 1717, while his father was abroad. Shortly after his birth, his mother abandoned him at the church of Saint-Jean-le-Rond. Following tradition, he was named Jean le Rond after the church, and placed in an orphanage. Upon the return of his father, he was removed from the orphanage, and placed with Mme. Rousseau, the wife of a glazier. Although Destouches continued to support his son financially, he chose not to publicly acknowledge his son. In 1738, Jean le Rond entered law school, where he was registered under the name Daremberg. He later changed this name to d'Alembert.

In their efforts to understand the principles of combinations of forces, mathematicians of the eighteenth century, such as d'Alembert, were led to the equation

$$g(x+y) + g(x-y) = 2g(x)g(y), \quad (1.9)$$

where $0 \leq y \leq x \leq \pi/2$. Equation (1.9) is now known as *d'Alembert's equation*. It is required to find all real-valued functions g satisfying equation (1.9).

Here, we encounter a greater level of difficulty in the analysis compared to Cauchy's equation. The equation is reminiscent of a trigonometric identity. Trying out our elementary trigonometric functions, we observe that $g(x) = \cos x$ works, but $g(x) = \sin x$ does not. Are there any other solutions? Playing with our solution a bit, we are tempted to look for solutions of the form

$$g(x) = b \cos ax$$

for suitably chosen constants a, b . However, letting $x = y = 0$ in equation (1.9) reduces it to the equation $g(0) = g(0)^2$, telling us that $g(0) = 0$ or $g(0) = 1$. These correspond to the cases $b = 0$ and $b = 1$ respectively. The constant a turns out to be arbitrary: if $g(x)$ is any solution to equation (1.9) then $g(ax)$ is also a solution.

This tells us that we can expand our single solution to include

$$g(x) = \begin{cases} 0 \\ \cos ax. \end{cases}$$

Are there any other solutions to d'Alembert's equation? It turns out that there are. Once again, it was Cauchy [1821] who managed to show that a full set of continuous solutions to (1.9) is given by

$$g(x) = \begin{cases} 0 \\ \cos ax \\ (b^x + b^{-x})/2, \quad b > 0. \end{cases} \quad (1.10)$$

The seeming incongruity between the second and third lines in (1.10) can be explained by the fact that $(b^x + b^{-x})/2$ can be written in terms of the hyperbolic cosine of x . The hyperbolic cosine shares many standard algebraic identities in common with the usual (circular) cosine, and this explains its ability to satisfy d'Alembert's equation as well. Indeed, if we choose b to be the famous constant $e \approx 2.71828$, then the resulting formula is the definition of the hyperbolic cosine.

1.7 Charles Babbage

One property that both Cauchy's and d'Alembert's equations have in common is that, although they involve functions of a single variable, the equations are formulated using two variables, namely x and y . A rather different class of functional equations was investigated by the British mathematician Charles Babbage.



Fig. 1.4. Charles Babbage (1791–1871).

Charles Babbage is best remembered as a founder of modern computing. Although his difference engine and his analytical engine—the latter programmable by punched cards—were not built during his lifetime, some of his other inventions were. Among his other inventions was the *cowcatcher*, that remarkable device that was attached to the front of trains to remove obstacles—such as cows—that might cause the train to be derailed. Today this device is most commonly seen in movies of the old west of the old Hollywood tradition where the steam locomotive, with cowcatcher, has achieved iconic status. His other inventions include skeleton keys. His mathematical studies of the postal system were also to lead to the idea of the flat rate postage stamp.

However, it is Charles Babbage the mathematician that we shall be concerned with here. In 1812, together with Robert Woodhouse, John Herschel, the son of the astronomer William Herschel, and George Peacock, he founded the Cambridge Analytical Society. It was the intention of the Analytical Society to promote the use of Leibniz' methodology for calculus over the Newtonian version which was entrenched in British mathematics at the time. In 1819, the Society was renamed the Cambridge Philosophical Society. It is by this name that it is known today.

On June 15, in 1815, Charles Babbage read a paper before the Royal Society of London that made his reputation as a mathematician of high originality and ability.⁵

Babbage opened this paper by noting that many mathematical calculations consist of two parts: the direct calculation and its inverse. For example, raising a number to a power is a direct calculation, whereas the extraction of a root is its inverse operation. In most cases, the inverse operation is of much greater difficulty than the direct operation. Babbage remarked that

It is this inverse method with respect to functions, which I at present propose to consider.

More explicitly, we may say that if a function is given, then we may determine some particular equation that the function satisfies; this would be a direct calculation applied to the function. The inverse problem would be to start with a given equation and to determine the family of functions satisfying that equation.

From this broad outline, Babbage's argument proceeds by way of a sequence of examples of functional equations and their solutions, of increasing generality and complexity. The first part of the paper considers equations of the form

$$F[x, f(x), f(\alpha_1(x)), f(\alpha_2(x)), \dots, f(\alpha_n(x))] = 0, \quad (1.11)$$

where functions F and $\alpha_1, \dots, \alpha_n$ are given. It is required to find all functions f satisfying the equation for each particular choice of F and α . For example, the first problem treated in the paper is to determine all functions f satisfying the functional equation

$$f(x) = f[\alpha(x)] \quad (1.12)$$

for given function $\alpha(x)$. Upon inspection it is easily seen that for many functions α there are infinitely many functions f satisfying this equation, and that these functions are quite diverse in nature. If a particular solution f_0 has been found, then all functions of the form

$$f(x) = \sigma[f_0(x)], \quad (1.13)$$

where σ is arbitrary, will all satisfy the functional equation, although the family of such functions may not be the complete collection of all solutions.

⁵ An essay towards a calculus of functions, *Philosophical Transactions of the Royal Society of London* **105**, 389–423. This paper was the first of several on the subject. It was followed by An essay towards the calculus of functions, part II, *Phil. Trans.* **106**, 179–256, which appeared in 1816. In 1817, he published two papers entitled Observations on the analogy which subsists between the calculus of functions and the other branches of analysis, *Phil. Trans.* **107**, 197–216, and Solutions of some problems by means of the calculus of functions, *Journal of Science*, **2**, 371–79. In 1822, there followed the paper Observations on the notation employed in the calculus of functions, *Trans. Camb. Phil. Soc.* **1**, 63–76.

Example 1.1. Of particular interest in equation (1.12) are special cases such as

$$f(x) = f\left(\frac{x}{x-1}\right), \quad (1.14)$$

where $\alpha(x) = x/(x-1)$. In this case, the function $\alpha(x)$ is an *involution*, in the sense that if it is applied twice to any value $x \neq 1$, we get the identity:

$$x \xrightarrow{\alpha} \frac{x}{x-1} \xrightarrow{\alpha} \frac{\left(\frac{x}{x-1}\right)}{\left(\frac{x}{x-1}\right) - 1} = x.$$

Better known examples of involutions include $x \mapsto -x$ and $x \mapsto x^{-1}$. A general family of solutions to (1.12) when $\alpha(x)$ is an involution is

$$f(x) = \tau[x, \alpha(x)], \quad (1.15)$$

where $\tau(u, v)$ is an arbitrary symmetric function of u and v . (That is, $\tau(u, v) = \tau(v, u)$.) An interesting function which is a solution to (1.14) is $f(x) = \cos \log(x-1)$. The reader can consider how to express this particular solution in the form given in (1.15).

An extension of this problem is to find solutions to two or more simultaneous functional equations, such as

$$f(x) = f[\alpha(x)], \quad f(x) = f[\beta(x)] \quad (1.16)$$

for given functions $\alpha(x)$ and $\beta(x)$.

Example 1.2. The pair of simultaneous equations

$$f(x) = f(-x), \quad f(x) = f\left(\frac{x}{\sqrt{x^2-1}}\right)$$

has, as particular solutions,

$$f(x) = \sigma\left(\frac{x^4}{x^2-1}\right)$$

where, once again, σ is arbitrary.

In the second part of the paper, Babbage studied families of equations which, up to that point, had not received any systematic treatment in the literature. These n^{th} order functional equations, as he called them, are of the form

$$F[x, f(x), f^2(x), \dots, f^n(x)] = 0, \quad (1.17)$$

where, once again, F is given and f is to be determined. Here

$$f^2(x) = f[f(x)], \quad f^3(x) = f[f^2(x)],$$

and so on. In this case, the function F is known, and one seeks to find all functions f such that equation (1.17) holds for all real x . For example, a solution $f(x)$ to the equation

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

would be an *iterative n^{th} root* of the identity function $x \mapsto x$. In particular, we can recognise that an iterative square root of the identity function is an involution as discussed above. Babbage noted that if $\tau(u, v)$ is any arbitrary antisymmetric function of u and v ,⁶ then any involution $f(x)$ must satisfy

$$\tau[x, f(x)] = 0. \quad (1.18)$$

Such an equation can be used to find involutions.

Example 1.3. Setting $\tau(u, v) = u^2 - v^2$ gives us two involutions, namely

$$f(x) = x \quad (\text{the trivial involution}) \quad \text{and} \quad f(x) = -x.$$

A less obvious example is $\tau(u, v) = (u^2 + v) - (v^2 + u)$, from which we obtain the equation

$$(x^2 - x) + f(x) - [f(x)]^2 = 0$$

This is a quadratic in $f(x)$ that is solved by

$$f(x) = \frac{1 + \sqrt{1 - 4x(1-x)}}{2} \quad \text{and} \quad f(x) = \frac{1 - \sqrt{1 - 4x(1-x)}}{2}$$

both of which are easily checked to be involutions using the identity

$$f(x)[1 - f(x)] = x(1 - x).$$

In a similar manner, many other solutions to $f^2(x) = x$ can be obtained.

Using such solutions, Babbage noticed a method for generating many more. If $f(x)$ is an involution, then so is

$$g(x) = \phi^{-1}\{f[\phi(x)]\} \quad (1.19)$$

for any one-to-one function ϕ . This is because

$$\begin{aligned} g^2(x) &= \phi^{-1}\{f^2[\phi(x)]\} \\ &= \phi^{-1}[\phi(x)] \\ &= x. \end{aligned}$$

Problem 12 of Babbage's paper asks for solutions to the equation

⁶ That is, $\tau(u, v) = -\tau(v, u)$.

$$f^2(x) = \alpha(x) \tag{1.20}$$

for any given function $\alpha(x)$. Suppose we search for solutions of the form

$$f(x) = \phi^{-1}\{\beta[\phi(x)]\},$$

where β is some given function—but not an involution unless $\alpha(x) = x$ —and $\phi(x)$ is a one-to-one function to be determined. Then

$$\begin{aligned} \alpha(x) &= f^2(x) \\ &= \phi^{-1}\{\beta^2[\phi(x)]\}, \end{aligned}$$

which implies that $\phi(x)$ must be a solution to the functional equation

$$\phi[\alpha(x)] = \beta^2[\phi(x)]. \tag{1.21}$$

This is a type of functional equation called a *conjugacy equation*. We discuss equations of this sort in Chapter 3. Although the reduction of (1.20) to a conjugacy equation does not amount to a solution, it is easier in some cases to solve an equation such as (1.21) than equation (1.20). The hope is that we can reduce a given equation involving an iterative root to a conjugacy equation of a particularly simple sort that we can solve. This is one of the topics that we discuss in Chapter 3, where we develop some methods to solve special conjugacy equations due to Abel and Schröder.

Example 1.4. It is easy to check that a solution to

$$f^2(x) = x^{-1}, \quad x \neq 0 \tag{1.22}$$

is given by

$$f(x) = -x^{\text{sgn}(x)}, \tag{1.23}$$

where $\text{sgn}(x)$ denotes the signum function of x whose value is +1 when x is positive and -1 when x is negative.

Example 1.5. Let us find a solution to

$$f^2(x) = \frac{x}{x+1}.$$

Following Babbage's method, we choose

$$\beta(x) = x + \frac{1}{2}.$$

Thus $\beta^2(x) = x + 1$. We search for a function $\phi(x)$ such that

$$f(x) = \phi^{-1}\{\beta[\phi(x)]\}$$

will work. So $\phi(x)$ must be a solution to the conjugacy equation

$$\phi\left(\frac{x}{x+1}\right) = \phi(x) + 1.$$

This is also a bit hard to solve. However, we can write it as

$$\phi\left(\frac{1}{1+x^{-1}}\right) \frac{1}{1+\phi(x)} = 1$$

which is immediately seen by inspection to be solved by $\phi(x) = x^{-1}$. Therefore our solution has the form

$$\begin{aligned} f(x) &= \phi^{-1}\{\beta[\phi(x)]\} \\ &= \left(x^{-1} + \frac{1}{2}\right)^{-1} \\ &= \frac{2x}{2+x}. \end{aligned}$$

It is easily checked that this is a solution.

1.8 Mathematics competitions and recreational mathematics

The subject of functional equations has also found a niche in recreational mathematics and in mathematics competitions.

The idea of throwing out mathematical challenges and puzzles for other enthusiasts to solve is very old. For example, during the Italian Renaissance, it was commonplace for mathematicians to issue public challenges to others for the solutions to certain problems. Nowadays, a similar function is provided by numerous journals dedicated—in whole or in part—to problem solving. Many of these journals encourage the publication of problems which require only high school mathematics, or undergraduate mathematics at most. For example, in the United States and Canada, journals such as *Mathematics Magazine*, published by the Mathematical Association of America, devote space to mathematics problems. The journal, *Crux Mathematicorum*, published by the Canadian Mathematical Society, is dedicated entirely to publishing original problems. Its full title, *Crux Mathematicorum with Mathematical Mayhem* reflects the fact that it is a union of two previous journals, *Crux Mathematicorum* and *Mathematical Mayhem*. As an independent journal, *Mathematical Mayhem* was written by, and for, students.

Among problems appearing in such journals, functional equations often appear because their solutions often require only elementary manipulations without advanced theory. Another forum in which high school or undergraduate students can participate is in mathematics competitions. Many countries offer a mathematical olympiad system for high school students with a system

of competitions that culminates in a national olympiad. Beyond that, the International Mathematical Olympiad (IMO) involves more than 80 countries, each country being represented by an elite group of six students. Functional equations often appear among the questions posed at the IMO, and are regularly shortlisted for consideration each year.

At the university level, the choices for students are fewer. However, in North America, the William Lowell Putnam Competition is held regularly on the first Saturday in December at more than 300 universities and colleges in the United States and Canada. Problems involving functional equations made an early appearance in the history of this competition.

In early competition problems, the study of functional equations has been classified more as university mathematics, in part because the subject has obvious affinities with topics that are studied in more advanced university courses, such as differential equations, integral equations, and the theory of representations of groups. So it is not surprising that many of the early examples of functional equations can be found on the William Lowell Putnam Competition. Many of these problems are only accessible to students with some background in calculus and differential equations, which for the purposes of this book, we have split off as a separate discipline. For example, Problem A3 from the third Putnam competition in 1940 asked for all functions $f(x)$ such that

$$\int [f(x)]^n dx = \left[\int f(x) dx \right]^n$$

when constants of integration are suitably chosen. Here, we choose not to classify this problem as functional equations because the equation involves calculus.

The very first problem on the Putnam competition where a functional equation appears is in the fourth Putnam competition held in 1941.

Example 1.6 (Fourth Putnam Competition, Problem B14(i)). Show that any solution $f(t)$ of the functional equation

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

(x and y real) is such that

$$f''(t) = \pm m^2 f(t),$$

where $m \geq 0$ is a constant. We assume the existence and continuity of the second derivative.

The reader will notice the spirit of d'Alembert's equation runs through this example, although the equation is different. This time the problem involves a functional equation. However, both the conclusion and the solution involve calculus.

The first example on the Putnam competition which involves the kinds of elementary methods (with no calculus) such as Cauchy used to solve his equation and that of d'Alembert is Problem 2 from the morning session of the seventh Putnam competition held in 1947.

Example 1.7 (Seventh Putnam Competition, Problem A2). A real-valued continuous function satisfies, for all real x and y , the functional equation

$$f\left(\sqrt{x^2 + y^2}\right) = f(x)f(y).$$

Prove that

$$f(x) = [f(1)]^{x^2}.$$

We do not consider the solution to this problem until later. However, it should come as no surprise that Cauchy's equation is fundamental in its solution.⁷ Thus we defer more discussion until the next chapter.

This early example illustrates one of the difficulties with such functional equations as competition problems. The examiners for that year either expected students to recognise—perhaps without proof—the solution to Cauchy's equation, or expected the students to duplicate all or part of Cauchy's methodology. As we show in the next chapter, this methodology is by no means difficult to follow. However, it is a mathematically sophisticated approach to the problem that is very different in style to the kinds of solutions to differential equations that the students would have been used to. Perhaps it is an act of faith to put a problem like this on a competition for students at any level. Fortunately, there are many functional equations which can be solved without using the mathematically sophisticated style of the solution to this problem. Equations which can be solved without appealing to concepts such as continuity are often fair game for high school mathematics competitions.

Thereafter, functional equations arise sporadically over the years on the Putnam. There are polynomial functional equations, in which the class of solutions is restricted to polynomials of arbitrary degree. See, for example, Problem B7 from the sixteenth competition in 1956. There are also problems such as B6 from the fifteenth competition in 1955 which involve a restriction on the domain of the function, in this case to positive integers. The task of enumerating these problems is not a particularly fruitful exercise, because the classification of the problems is often artificial: functional equations defined on positive integers are often equivalent to problems involving sequences; functional equations involving polynomials may have more in common with

⁷ The statement of this problem illustrates one of the hazards of putting functional equations on mathematics competitions. It is easy to leave out the necessary regularity required to ensure the conclusion. In this case, the function $f(x) = 0$ for all x satisfies the equation. However, for $x = 0$, the value of $f(1)^{x^2} = 0^0$ is indeterminate. Thus the statement of the problem should rule out this solution.

other polynomial problems than with the theory of functional equations, and so on. However, there can be no doubt that in 1959 there were two unambiguous functional equations on the Putnam. Problem B3 in 1963 involves differentiability. Because the equations themselves do not involve derivatives, we may safely classify it as a functional equations problem and not a problem in differential equations.

The following example is question number 5 from the 1972 International Mathematical Olympiad, and illustrates how the historical work on functional equations has influenced competitive mathematical problem solving.

Example 1.8. Let f and g be real-valued functions defined for all real values of x and y , and satisfying the equation

$$f(x+y) + f(x-y) = 2f(x)g(y) \quad (1.24)$$

for all x, y . Prove that if f is not identically zero, and if $|f(x)| \leq 1$ for all x , then $|g(y)| \leq 1$ for all y .

Solution. As $|f(x)| \leq 1$ for all x , it follows that $|f|$ has a least upper bound M . That is, there exists a real value M such that $|f(x)| \leq M$ for all x , and for all $m < M$ there is some x_0 such that $|f(x_0)| > m$. Now suppose that there exists a point y_0 such that $|g(y_0)| > 1$. We prove a contradiction. Using (1.24) and the triangle inequality, we get

$$\begin{aligned} 2|f(x)||g(y_0)| &= |f(x+y_0) + f(x-y_0)| \\ &\leq |f(x+y_0)| + |f(x-y_0)| \\ &\leq 2M \end{aligned}$$

So

$$|f(x)| \leq \frac{M}{|g(y_0)|} < M.$$

This is a contradiction, because M has been defined as the least upper bound of $|f(x)|$. \square

The functional equations (1.24) and (1.9) are special cases of a more general functional equation, namely

$$f(x+y) + f(x-y) = g(x)h(y). \quad (1.25)$$

Equations of this kind were of interest to applied mathematicians of the eighteenth century. The right-hand side denotes the situation where the two variables x and y “separate” into the factors $g(x)$ and $h(y)$ as shown.

Earlier, we looked at the work of Charles Babbage, and his results on iterative roots. Problems involving iterative roots form an interesting class of functional equations that commonly arise in mathematics competitions. A typical form for such a problem is the following.

Example 1.9 (Putnam 1979, Problem A2). Establish necessary and sufficient conditions on the constant k for the existence of a continuous, real-valued function $f(x)$ satisfying

$$f(f(x)) = kx^9 \quad (1.26)$$

for all real x .

We do not solve this problem here, but postpone consideration of such examples to Chapter 3, where we consider iterative roots in greater detail. More common examples of iterative roots occur where function composition using $f(x)$ appears on both sides of the equation. The example given above is notable because it is of the simple form that Babbage considered.

Many functional equations involving iterative square roots are closely related to particular recursions. Consider, for example, the functional equation

$$f[f(x)] = x + f(x). \quad (1.27)$$

Starting at any value x , we can construct a sequence $x, f(x), f[f(x)], \dots$ by successive application of the function f . Letting f_i be the i th value in this sequence, we see that the f_i s satisfy the recursion

$$f_{i+2} = f_i + f_{i+1}. \quad (1.28)$$

In particular, the famous Fibonacci numbers

$$f_1 = 1, \quad f_2 = 1, \quad f_3 = 2, \quad f_4 = 3, \quad f_5 = 5, \quad f_6 = 8, \quad f_7 = 13, \dots$$

satisfy this recursion. However other sequences satisfying (1.28) can be chosen by arbitrarily specifying f_1 and f_2 .

Now such recursions are well studied. For example, it is well known that the general form of the solution to (1.28) is given by

$$f_i = a \left(\frac{1 + \sqrt{5}}{2} \right)^i + b \left(\frac{1 - \sqrt{5}}{2} \right)^i \quad (1.29)$$

where the constants a and b can be chosen arbitrarily so to determine the starting values f_1 and f_2 . For example, we can obtain Binet's formula for the Fibonacci sequence by setting $a = 1/\sqrt{5}$ and $b = -1/\sqrt{5}$. Other values of a and b give generalized versions of the Fibonacci sequence with different starting values. The quantity

$$\Phi = \frac{1 + \sqrt{5}}{2} \quad (1.30)$$

is the *golden ratio*, which can be obtained as the limiting form of the ratio of successive terms of the Fibonacci sequence. As $\Phi^{-1} = (\sqrt{5} - 1)/2$, the formula for f_i can also be written as $f_i = a\Phi^i + b(-\Phi)^{-i}$.

From (1.29), it is not hard to find a solution to (1.28): it is immediate that $f(x) = \Phi x$ is a solution to the functional equation. However, the solution is not unique. Another solution is the function $f(x) = -\Phi^{-1}x$. Are there any other functions?

An example of this kind of functional equation and the issue of uniqueness arose as a question on the 1988 Putnam competition.

Example 1.10 (Putnam 1988, Problem A5). Prove that there exists a unique function f defined on the set of positive real numbers such that

$$f[f(x)] = 6x - f(x) \quad (1.31)$$

and $f(x) > 0$ for all $x > 0$.

Solution. For each x we construct the sequence

$$f_1 = x, \quad f_2 = f(x), \quad f_3 = f[f(x)], \quad \dots \quad (1.32)$$

The sequence satisfies the recursion $f_{i+2} = 6f_i - f_{i+1}$. Clearly, $f_i = 0$ for all i satisfies the recursion. However, this possibility is ruled out by the fact that $f(x) > 0$ for all $x > 0$. Let us look for a solution of the form $f_i = y^i$, where $y \neq 0$. For this to hold, y must be a solution to the characteristic equation of the recursion, namely

$$y^2 + y - 6 = 0. \quad (1.33)$$

This has the roots $y = -3$ and $y = 2$. Checking back, we find that $f_i = 2^i$ and $f_i = (-3)^i$ both satisfy the recursion. Now if the sequence $\{f_i\}$ satisfies the recursion, then so does any multiple $\{af_i\}$. Moreover, if $\{g_i\}$ also satisfies the recursion, then so does $\{f_i + g_i\}$. From these two facts we deduce that

$$f_i = a2^i + b(-3)^i$$

satisfies the recursion for arbitrary real values a and b . Indeed, this is the general family of solutions to the recursion. To demonstrate this, note that the recursion is completely determined by the arbitrary starting values f_1 and f_2 . For any choice of f_1 and f_2 we can solve for a and b to obtain that sequence.

Now, recall that the original function $f(x)$ is required to be strictly positive. However, if b were nonzero, then eventually f_i would become negative as i becomes large, no matter what value of a is chosen. (Why is this?) So this requires that $b = 0$. Therefore $f_i = a2^i$. This solution to the recursion corresponds uniquely to the function $f(x) = 2x$, which is seen to solve the equation. \square

1.9 A contribution from Ramanujan

The theory of *nested radicals* is closely related to the theory of recursions. So it comes as no surprise that some nested radicals can be studied using the

methodology of functional equations. In this section, we consider a famous example due to Ramanujan. Problems involving the calculation of nested radicals occasionally appear in problem-solving journals and on mathematics competitions.

As the name implies, a nested radical is an expression in which a radical is contained within one or more other radicals. One of the most famous of all such expressions is François Viète's lovely formula

$$\pi = \frac{2}{1} \times \frac{2}{\sqrt{2}} \times \frac{2}{\sqrt{2+\sqrt{2}}} \times \frac{2}{\sqrt{2+\sqrt{2+\sqrt{2}}}} \times \cdots \quad (1.34)$$

We do not prove that the right-hand side converges to π as displayed. However, let us consider the infinite nested radical associated with this formula. In order for Viète's formula to converge, the successive factors must get closer and closer to one. Therefore

$$2 = \sqrt{2 + \sqrt{2 + \sqrt{2 + \sqrt{2 + \cdots}}}}$$

where the right-hand side is interpreted as the limiting value of the denominators of the factors in (1.34).

Nested radicals such as these can be easily evaluated. However, some others cannot be evaluated as easily. For example, in 1911 the Indian mathematician Ramanujan posed the problem of evaluating

$$\sqrt{1 + 2\sqrt{1 + 3\sqrt{1 + 4\sqrt{\cdots}}}} \quad (1.35)$$

Interestingly, this nested radical was to appear in 1966 as Problem A6 on the 27th Putnam competition, where students were asked to prove that the nested radical evaluates to 3. The problem is not a functional equation itself, and the official solution did not involve functional equations. However, iterations are closely connected to the theory of functional equations. There is a functional equation hidden here, which we uncover below.

The story of Ramanujan and his legendary mathematical abilities has been well told in popular literature.⁸ Born in Erode, India, in 1887, Ramanujan received a basic education in mathematics. Beyond that, he taught himself mathematics from a book by George Shoobridge Carr. The deficiencies of this book were arguably the source of some of Ramanujan's problems in presenting his own work to other mathematicians. Be that as it may, Ramanujan's mathematical talent was not to be suppressed, and he pressed on with his mathematical investigations by studying divergent series and continued fractions

⁸ See, for example, the book *The Man Who Knew Infinity*, by Robert Kanigel, published by Scribners, New York, and written for readers with a minimum of mathematics necessary to explain some of Ramanujan's work.

in particular. By 1911, Ramanujan had made contacts in the mathematical community, and began to publish and solve problems in the *Journal of the Indian Mathematical Society*, which is where he posed the problem of evaluating the nested radical in (1.35), given above. It was typical of such problems that they would be published in the *Journal*, and that solutions would be published subsequently upon receipt. However, Ramanujan's problem, though simple to state, remained unsolved until Ramanujan solved the problem himself. His clever insight into the problem was through a generalisation of the nested radical to the formula

$$x + n + a = \sqrt{ax + (n + a)^2 + x \sqrt{a(x + n) + (n + a)^2 + (x + n) \sqrt{\dots}}} \quad (1.36)$$

Thus the nested radical posed by Ramanujan is simply the special case where $x = 2$, $a = 0$, and $n = 1$. Presumably, Ramanujan expected the original nested radical to be solved by elementary methods. For example

$$\begin{aligned} 3 &= \sqrt{9} \\ &= \sqrt{1 + 2\sqrt{16}} \\ &= \sqrt{1 + 2\sqrt{1 + 3\sqrt{25}}} \end{aligned}$$

and so on. The continuation of this recursion depends upon the ability to decompose each perfect square on the extreme right as

$$n^2 = 1 + (n - 1) \sqrt{(n + 1)^2}.$$

Thus Ramanujan's challenge cannot be said to have required any new mathematics. Nevertheless, the problem went unsolved for a short while.

Let us look at this nested radical, to see how it relates to the topic of this book. The nested radical in (1.35) contains a hidden functional equation. Suppose we generalize (1.35) by defining

$$f(x) = \sqrt{1 + x \sqrt{1 + (x + 1) \sqrt{\dots}}} \quad (1.37)$$

It is not clear that the function $f(x)$ is well defined. However, if it is meaningful, then squaring both sides gives us the functional equation

$$[f(x)]^2 = 1 + x f(x + 1). \quad (1.38)$$

The reader will recognise that this is an example of a functional equation in a single variable, such as Babbage considered. In particular, it is a *nonlinear* functional equation. The use of the word *nonlinear* is not a reference to the nonlinearity of f . Rather, this refers to the fact that the equation involves $[f(x)]^2$, and so cannot be written as a linear expression in $f(x)$ and $f(x + 1)$.

The reader should not be under any misapprehensions about the function $f(x)$. While the functional equation looks complex, it has a very simple solution. Looking at Ramanujan's formula in (1.38) we see that $f(x) = x + 1$ solves the equation. However, we cannot expect this solution to be the only solution to our functional equation. So in order to deduce that $f(x) = x + 1$, we will have to extract additional information about $f(x)$ from our nested radical representation in (1.37). Looking at (1.36), we see that there is a family of functional equations here, satisfying

$$[f(x)]^2 = ax + (n + a)^2 + xf(x + n)$$

with solution

$$f(x) = x + n + a.$$

We examine the solution to Ramanujan's problem and the solution to functional equation (1.38) in Chapter 3.

1.10 Simultaneous functional equations

We can obtain functional equations in a single variable easily from functional equations in two or more variables. For example, Cauchy's equation $f(x + y) = f(x) + f(y)$ produces the functional equation $f(2x) = 2f(x)$ in a single variable when the substitution $y = x$ is made. It is clear that the class of solutions of the latter equation is at least as large as the class of solutions of the former equation. Indeed, it is easy to check that the equation $f(2x) = 2f(x)$ has nonlinear continuous solutions that do not satisfy Cauchy's equation.

By contrast, in the following example, we see that linearity can be characterized.

Example 1.11. Let $f : \mathbf{R} \rightarrow \mathbf{R}$ satisfy the equations

$$f(x^2) = [f(x)]^2, \tag{1.39}$$

$$f(x + 1) = f(x) + 1. \tag{1.40}$$

Prove that $f(x) = x$ for all real x .

Solution. From (1.39) we see that $f(x) \geq 0$ for all $x \geq 0$. In addition, by mathematical induction on (1.40) we get

$$f(x + n) = f(x) + n \tag{1.41}$$

for all positive integers n . Substituting $x - 1$ for x in (1.40) and applying mathematical induction again, we see that (1.41) holds for all integers n and all real x .

Now applying (1.39) we get $[f(x - 1)]^2 = [f(1 - x)]^2$. Using (1.41) on this identity yields

$$[1 + f(-x)]^2 = [f(x) - 1]^2.$$

We can expand this, noting in passing that $[f(-x)]^2 = [f(x)]^2$ to obtain the fact that

$$f(-x) = -f(x) \quad (1.42)$$

for all real values of x . Equation (1.42) immediately implies that $f(0) = 0$. So, this fact and (1.41) together imply that $f(n) = n$ for all integers n .

We are part way there. All we need to do is fill in the gaps and prove this equation for noninteger values. As shown in future examples, extending such identities to all real values often involves using arguments involving limits or continuity. This problem is no exception.

We next prove a statement that is weaker than the one we want. We show that if $x \in [n, n + 1]$, then $f(x) \in [n, n + 1]$. Suppose $x - n \geq 0$. Then using (1.41) we see that $f(x) - n = f(x - n) \geq 0$. Thus $f(x) \geq n$. Similarly if $(n + 1) - x \geq 0$, we obtain $(n + 1) - f(x) = (n + 1) + f(-x) = f(n + 1 - x) \geq 0$. Therefore we have proved that $f(x)$ is in the interval $[n, n + 1]$ whenever x is. But this means that

$$|f(x) - x| \leq 1 \quad (1.43)$$

for all real x .

Now set $k = 2^n$, where n is a positive integer. Let $x \geq 0$. Then $f(x) \geq 0$, and

$$\begin{aligned} 1 &\geq |x^k - f(x^k)| \\ &= |x^k - f(x)^k| \\ &= |x - f(x)| \cdot |x^{k-1} + x^{k-2}f(x) + \cdots + f(x)^{k-1}| \\ &\geq |x - f(x)| x^{k-1}. \end{aligned}$$

Therefore

$$|x - f(x)| \leq x^{-k+1}. \quad (1.44)$$

However, this statement is true for all such k . As k becomes large, x^{-k+1} goes to zero when $x > 1$. So $f(x) = x$ when $x > 1$. Using (1.41) we can extend this statement to all real x . \square

1.11 A clarification of terminology

Those who work on functional equations speak of *equations in one, two or more variables*. Such an expression is not to be confused with the expression *functions of one or more variables*. For example, Cauchy's equation in (1.3) involves a function of a single real variable. However, it is a functional equation in two variables, namely the x and y that appear in (1.3). Although most of the equations that we study in this book involve functions of one real variable, we shall occasionally consider functions of two or more. See Section 2.9.

1.12 Existence and uniqueness of solutions

It is unrealistic to hope that every functional equation has one and only one solution. Often there are many solutions to a given functional equation. For example, consider the equation $f(x+1) = f(x)$. We can construct infinitely many solutions as follows. Let g be any real-valued function defined on the interval $[0, 1)$. Define $f(x) = g(x - [x])$, where $[x]$ denotes the greatest integer less than or equal to x . Then f satisfies the functional equation because $(x+1) - [x+1] = x - [x]$. Similarly, the equation $f(f(x)) = x$ has infinitely many solutions, among which is the obvious solution $f(x) = x$. On the other hand, in Chapter 3 we show that the equation $f(f(x)) = -x$ has no solution if f is required to be a continuous function.

In trying to solve functional equations, extraneous solutions may appear in the derivation. So it is important that we go back to our original equation and check out our supposed solution by plugging it back in. If a derivation shows that only one function could be the solution, and that does not work, then we have shown that there can be no solution to the original functional equation. On the other hand, we should not stop once we have found a single solution. A full understanding of the problem does not come until we know that the solution set is complete.

1.13 Problems

1. Find all possible choices of real-valued functions $f(x)$ and $g(y)$ such that

$$f(x) + g(y) = \log(1 + x + xy + y)$$

for all positive x and y .

2. Let $f(x)$ be a real-valued function defined for all positive x , satisfying $f(x+y) = f(xy)$ for all positive x, y . Prove that f is a constant function.
3. Let f be a real-valued function such that, for all real x and y ,

$$f(x+y)f(x-y) = [f(x)f(y)]^2.$$

Prove that there exists a function g such that $f(x) = g(x^2)$.

4. Determine all functions $f : \mathbf{R} \rightarrow \mathbf{R}$ which satisfy

$$f(a+x) - f(a-x) = 4ax$$

for all real a and x .

5. Determine as many solutions as you can to each of the following functional equations:

- a) $f(x)f(x+1) = f(2x+1)$
 b) $f(x)f(x+1) = \frac{1}{2}[f(1) + f(2x+1)]$
 c) $f(x)f(x+1) = f(x) - f(x+1)$
 d) $f(x)f(x+1) = f[f(x)+x]$.

6. Without using Cauchy's solution to d'Alembert's equation given in (1.10), show directly that the only solution $g: \mathbf{R} \rightarrow \mathbf{R}$ to d'Alembert's equation

$$g(x+y) + g(x-y) = 2g(x)g(y),$$

for all real x and y , that satisfies

$$\lim_{x \rightarrow \infty} g(x) = 0$$

is the function $g(x) = 0$ for all x .

7. Consider two continuous functions $f, g: \mathbf{R} \rightarrow \mathbf{R}$, satisfying the functional equations

$$f(x+y) = f(x)f(y) - g(x)g(y) \quad (1.45)$$

$$g(x+y) = g(x)f(y) + g(y)f(x), \quad (1.46)$$

where $f(0) \neq 0$. Prove the following properties of f and g .

- a) $g(0) = 0$ and $f(0) = 1$.
 b) Let

$$h(x) = [f(x)]^2 + [g(x)]^2.$$

Show that h satisfies Cauchy's exponential equation

$$h(x+y) = h(x)h(y).$$

In addition, show that if $f(x)$ and $g(x)$ are solutions, so are $a^x f(x)$ and $a^x g(x)$, where $a > 0$.

- c) Suppose there exists a constant K such that

$$[f(x)]^2 + [g(x)]^2 \leq K$$

for all x . Prove that

$$[f(x)]^2 + [g(x)]^2 = 1.$$

- d) (The final part of this question uses calculus. Those unfamiliar with techniques of differentiation may wish to skip this part.) Suppose that f and g are differentiable, and that $f'(0) = 0$. Prove that

$$[f(x)]^2 + [g(x)]^2 = 1.$$

8. Prove that $f(n) = 1 - n$ is the only integer-valued function defined on the integers that satisfies the following conditions.
- $f[f(n)] = n$, for all integers n ;
 - $f[f(n+2) + 2] = n$, for all integers n ;
 - $f(0) = 1$.

9. ★ Let $f(n)$ be a function defined on the set of all positive integers and having all its values in the same set. Prove that, if

$$f(n+1) > f[f(n)]$$

for each positive integer n , then $f(n) = n$ for all n .

10. ★ The following two questions use techniques that are specific to polynomial functional equations.

- a) Find all polynomials $q(x)$, with real coefficients, satisfying the equation

$$q(x^2) = q(x)q(x+1)$$

for all real x .

- b) Find all nonconstant polynomials $p(x)$, with real coefficients, satisfying the equation

$$p(x^2) = p(x)p(x-1)$$

for all real x .

11. ★ Let f be a real-valued function defined for all real numbers x such that, for some positive constant a , the equation

$$f(x+a) = \frac{1}{2} + \sqrt{f(x) - [f(x)]^2}$$

holds for all real x .

- Prove that f is periodic (i.e., there exists a positive number b such that $f(x+b) = f(x)$ for all x).
- For $a = 1$, give an example of a nonconstant function with the required properties.

12. ★ Let f be a continuous function defined on the interval $[0, 1]$, such that for each $x \in (0, 1)$ there is an h satisfying $0 \leq x - h < x + h \leq 1$ and

$$f(x) = \frac{f(x-h) + f(x+h)}{2}.$$

Prove that $f(x) = cx + d$ for some choice of real numbers c and d .

13. Suppose that the functions $f : \mathbf{R} \rightarrow \mathbf{R}$ and $g : \mathbf{R} \rightarrow \mathbf{R}$ satisfy

$$f[x + g(y)] = 2x + y + 5.$$

Find an explicit expression for $g[x + f(y)]$.

14. \star Suppose $f : \mathbf{R} \rightarrow \mathbf{R}$ is a one-to-one continuous function satisfying the following.
- a) $f[2x - f(x)] = x$, for all real x .
 - b) There exists a real number x_0 such that $f(x_0) = x_0$.
- Prove that $f(x) = x$ for all real x .

Functional equations with two variables

2.1 Cauchy's equation

Let us begin by restating and solving Cauchy's functional equation. Let $f : \mathbf{R} \rightarrow \mathbf{R}$ be a continuous function satisfying

$$f(x + y) = f(x) + f(y) \quad (2.1)$$

for all real x and y . We show that there exists a real number a such that $f(x) = ax$ for all $x \in \mathbf{R}$.

It is straightforward to show by mathematical induction that (2.1) implies

$$f(x_1 + x_2 + \cdots + x_n) = f(x_1) + f(x_2) + \cdots + f(x_n) \quad (2.2)$$

for all $x_1, \dots, x_n \in \mathbf{R}$. A special case of this is found by setting $x_1 = \cdots = x_n = x$, say. Then (2.2) becomes

$$f(nx) = n f(x) \quad (2.3)$$

for all positive integers n and for all real x . Let $x = (m/n)t$, where m and n are positive integers. Then $nx = mt$. So

$$\begin{aligned} f(nx) &= f(mt) \\ n f(x) &= m f(t) \\ n f\left(\frac{m}{n}t\right) &= m f(t). \end{aligned}$$

But this can be written as

$$f\left(\frac{m}{n}t\right) = \frac{m}{n} f(t) \quad (2.4)$$

for all $t \in \mathbf{R}$. Thus we have proved that



Fig. 2.1. Augustin-Louis Cauchy (1789–1857).

$$f(qt) = qf(t) \quad (2.5)$$

for all real t and all positive rationals q .

What about $q = 0$? We can extend (2.5) to include $q = 0$ in the following way. Returning to (2.1), we see that

$$\begin{aligned} f(y) &= f(y+0) \\ &= f(y) + f(0). \end{aligned}$$

So $f(0) = 0$. From this, we immediately have $f(0t) = 0f(t)$. Therefore (2.5) is true for all nonnegative rational q .

What about $q < 0$? Once again, returning to (2.1), we obtain

$$\begin{aligned} 0 &= f(0) \\ &= f[q + (-q)] \\ &= f(q) + f(-q). \end{aligned}$$

Therefore $f(-q) = -f(q)$. From this fact, we get for $q < 0$,

$$\begin{aligned} f(qt) &= f[-(-q)t] \\ &= -f[(-q)t] \\ &= -(-q)f(t) \\ &= qf(t). \end{aligned}$$

Therefore,

$$f(qt) = qf(t) \quad (2.6)$$

for all real values of t and all rational values of q .

Now suppose we substitute $t = 1$ into (2.6). Letting $f(1) = a$, we deduce that $f(q) = aq$ for all rational numbers q . We can summarize what has been discovered so far in the following proposition.

Proposition 2.1. *Let $f : \mathbf{R} \rightarrow \mathbf{R}$ satisfy Cauchy's equation*

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

for all real values x and y . Then there exists a real number a such that

$$f(q) = aq$$

for all rational numbers q .

We are almost there. All we need to do is draw the same conclusion when we replace the rational number q by any real number x . To make this leap, we finally make use of the assumption that f is a continuous function. Note that Proposition 2.1 made no such assumption on f . The following result is just the tool for this final step of the argument.

Proposition 2.2. *Suppose that $f : \mathbf{R} \rightarrow \mathbf{R}$ and $g : \mathbf{R} \rightarrow \mathbf{R}$ are continuous functions such that $f(q) = g(q)$ for all rational numbers q . Then $f(x) = g(x)$ for all real numbers x .*

Proof. This result follows from the fact that any real number can be approximated arbitrarily closely by rational numbers. For example, we can write x with an infinite decimal expansion, and let q_i be the rational number obtained by terminating the decimal expansion of x after i digits to the right of the decimal point. Then

$$\lim_{i \rightarrow \infty} q_i = x.$$

Because f and g are continuous, we can “pass” limits through the functions. So

$$\begin{aligned} f(x) &= f\left(\lim_{i \rightarrow \infty} q_i\right) \\ &= \lim_{i \rightarrow \infty} f(q_i). \end{aligned}$$

Similarly,

$$g(x) = \lim_{i \rightarrow \infty} g(q_i).$$

However, by assumption the functions f and g agree on all rational numbers. So $f(q_i) = g(q_i)$ for all $i = 1, 2, \dots$. It follows that $f(x) = g(x)$ for all real x . \square

Putting Propositions 2.1 and 2.2 together gives us the result we need.

Theorem 2.3. *Let $f : \mathbf{R} \rightarrow \mathbf{R}$ be a continuous function satisfying Cauchy's equation*

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

for all real values x and y . Then there exists a real number a such that

$$f(x) = ax$$

for all real numbers x .

Proof. From Proposition 2.1, we see that there exists a real number a such that $f(q) = aq$ for all rational numbers q . But $f(x)$ and $g(x) = ax$ are both continuous functions. By Proposition 2.2, $f(x) = g(x)$ for all real values x . \square

Theorem 2.3 has a number of variants. For example, Problem 1 at the end of the chapter asks the reader to solve Cauchy's equation when the domain of f is restricted to be the nonnegative real numbers. A less immediate generalization is the following.

Theorem 2.4. *Let $f : \mathbf{R} \rightarrow \mathbf{R}$ satisfy Cauchy's equation. Suppose in addition that there exists some interval $[c, d]$ of real numbers, where $c < d$, such that f is bounded below on $[c, d]$. In other words, there exists a real number A such that $f(x) \geq A$ for all $c \leq x \leq d$. Then there exists a real number a such that $f(x) = ax$ for all real numbers x .*

Problem 2 at the end of the chapter takes the reader through the steps of the proof. The assumption on f in Theorem 2.4 may not look as natural as the continuity assumption in Theorem 2.3. However, every continuous function is bounded on closed finite intervals by the extreme value theorem. So the assumption in Theorem 2.4 is weaker. Another class of functions that is bounded on finite intervals is the class of monotone functions. We consider the consequences of this in Section 2.2.

Example 2.5. If we ask for all solutions to Cauchy's equation when there is no boundedness or continuity condition imposed, then many unusual functions other than $f(x) = ax$ will be found to satisfy the equation. The proof of this fact is beyond the scope of this discussion here. However, the basic idea is not too difficult. Consider all real numbers x of the form $p + q\sqrt{2}$ where p and q are rational numbers. It can be seen that any such number is uniquely represented in the form $p + q\sqrt{2}$ in the sense that if

$$p_1 + q_1 \sqrt{2} = p_2 + q_2 \sqrt{2},$$

where p_1, p_2 and q_1, q_2 are rational, then $p_1 = p_2$ and $q_1 = q_2$. (Why?) We write the set of all such numbers as $\mathbf{Q}[\sqrt{2}]$. Note that $\mathbf{Q}[\sqrt{2}]$ is closed under addition: if x and y are in $\mathbf{Q}[\sqrt{2}]$ then so is $x + y$.

Define a function $f : \mathbf{Q}[\sqrt{2}] \rightarrow \mathbf{R}$ by

$$f(p + q\sqrt{2}) = ap + bq, \quad (2.7)$$

where $a\sqrt{2} \neq b$. The fact that elements of $\mathbf{Q}[\sqrt{2}]$ have a unique representation shows that this equation makes f a well-defined function. Now f satisfies a restricted form of Cauchy's equation: $f(x + y) = f(x) + f(y)$ for all x and y in $\mathbf{Q}[\sqrt{2}]$. Nevertheless, f is not the restriction of a linear function to the set $\mathbf{Q}[\sqrt{2}]$.

In order to find a function $f : \mathbf{R} \rightarrow \mathbf{R}$ that satisfies Cauchy's equation but is not linear, we need to look for an extension of the function f above from $\mathbf{Q}[\sqrt{2}]$ to the entire real line. That is where all the hard work is. Hamel [1905] proved that there exist subsets H of real numbers such that any real number x can be uniquely written as a finite linear combination of elements of H with rational coefficients!¹ The reader who wishes to explore this subject further is referred to the appendix of this book on Hamel bases.

2.2 Applications of Cauchy's equation

As we noted in the previous section, any continuous function $f : \mathbf{R} \rightarrow \mathbf{R}$ is bounded (both above and below) on a closed finite interval, and the boundedness condition of Theorem 2.4 is more general than the continuity condition of Theorem 2.3. However, continuity and boundedness are sometimes hard conditions to prove.² In this section, we consider two results which are similar to Theorems 2.3 and 2.4. These results are not as powerful because the conditions imposed are rather strong. However, in some cases, it may be easier to check the conditions than those of our theorems in the previous section.

In the first of these applications, we replace the continuity condition of Theorem 2.3 with an assumption of monotonicity. We show that the conclusion still follows.

Proposition 2.6. *Suppose $f : \mathbf{R} \rightarrow \mathbf{R}$ satisfies Cauchy's equation*

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

for all x and y , and is also monotone increasing in the sense that

$$f(x) \leq f(y) \quad \text{for all } x \leq y.$$

Then $f(x) = ax$ for some $a \geq 0$.

First proof. Of course, monotone increasing can be replaced by monotone decreasing with much the same conclusion except that $a \leq 0$. There are two

¹ In our setting restricted to $\mathbf{Q}[\sqrt{2}]$, above, the elements of $\mathbf{Q}[\sqrt{2}]$ can be written as a finite linear combination of elements of $\{1, \sqrt{2}\}$ with rational coefficients.

² Even Cauchy had some trouble proving continuity when he was proving his version of the binomial theorem.

ways to prove this proposition. The first, and much longer way is to start from Proposition 2.1 which shows that $f(q) = aq$, $a \geq 0$ for all rational numbers q . After that, the monotonicity assumption can be invoked much as we invoked continuity to extend the statement from the rational numbers to the real numbers. For any real number x , we can write

$$aq = f(q) \leq f(x) \leq f(r) = ar$$

for all rational numbers q, r such that $q < x < r$. Now consider a sequence of rational numbers $q_1 < q_2 < \cdots < x$, and another sequence of rational numbers $r_1 > r_2 > \cdots > x$ such that

$$\lim_{n \rightarrow \infty} q_n = \lim_{n \rightarrow \infty} r_n = x.$$

Then

$$\begin{aligned} ax &= \lim_{n \rightarrow \infty} aq_n \\ &= \lim_{n \rightarrow \infty} f(q_n) \\ &\leq \lim_{n \rightarrow \infty} f(r_n) \quad (\text{monotonicity!}) \\ &= \lim_{n \rightarrow \infty} ar_n \\ &= ax. \end{aligned}$$

So by the Squeeze Theorem, we conclude that $f(x) = ax$. \square

Second proof. However, a much easier proof follows directly from Theorem 2.4. Because the function f is monotone increasing, it follows that for any interval $[c, d]$ of real numbers, we can set $A = f(c)$. Then

$$f(x) \geq A \quad \text{for all } c \leq x \leq d.$$

Thus the function satisfies the conditions of Theorem 2.4. We can conclude that $f(x) = ax$ for some real number a . We can see that $a \geq 0$. This follows immediately from the fact that f is a monotone increasing function. \square

Proposition 2.7. *Suppose $f : \mathbf{R} \rightarrow \mathbf{R}$ satisfies the pair of equations*

$$f(x + y) = f(x) + f(y) \tag{2.8}$$

$$f(xy) = f(x)f(y) \tag{2.9}$$

for all x and y . Then either

$$f(x) = 0 \text{ for all } x \quad \text{or} \quad f(x) = x \text{ for all } x.$$

Note that the function has no continuity, boundedness, or other assumption imposed other than the two equations above. Once again, there are two ways to prove this. In this case, both are fairly easy. First, let us note that for any

number $z \geq 0$ there exists a $w \geq 0$ such that $z = w^2$. Let us set $x = w$ and $y = w$ in equation (2.9), to observe that $f(z) = [f(w)]^2 \geq 0$. We can conclude that

$$f(z) \geq 0 \text{ for all } z \geq 0. \quad (2.10)$$

First proof. For our first method of proof, let us suppose that $x \leq y$. If we replace y by $y - x$ in equation (2.8), we obtain

$$f(y) = f(x) + f(y - x) \geq f(x),$$

using property (2.10) to establish that $f(y - x) \geq 0$. It follows that $f : \mathbf{R} \rightarrow \mathbf{R}$ must be a monotone increasing function as in Proposition 2.6, above. Applying that result, we conclude that $f(x) = ax$ for some real number a . If we insert this formula into (2.9) using $x = y = 1$, we obtain $a = a^2$. This equation in a has two solutions: $a = 0$ or $a = 1$. Plugging these solutions back in to (2.8) and (2.9), we see that they both work. \square

Second proof. The second method of proof also uses (2.10), but in a more direct manner. It follows from (2.10) that for all $0 \leq c < d$, we have $f(x) \geq 0$ for any $x \in [c, d]$. Thus the function is bounded below on $[c, d]$ by $A = 0$ as in Theorem 2.4. So we can conclude that $f(x) = ax$ for some choice of real number a . The argument that $a = 0$ or $a = 1$ follows as in the first proof. \square

2.3 Jensen's equation

Jensen's equation has the form

$$f\left(\frac{x+y}{2}\right) = \frac{f(x) + f(y)}{2} \quad (2.11)$$

and can be thought of as a version of Cauchy's equation using averages. Once again, the function f is usually assumed to be continuous. For simplicity we assume that the domain of f is the entire real line.

The solution to the equation is easily obtained from the results of the previous section. Let $g(x) = f(x) - f(0)$. Then

$$g\left(\frac{x+y}{2}\right) = \frac{g(x) + g(y)}{2}. \quad (2.12)$$

Moreover, g is a continuous function with $g(0) = 0$. Setting $y = 0$ in (2.12), we obtain

$$g\left(\frac{x}{2}\right) = \frac{g(x)}{2}. \quad (2.13)$$

As (2.13) holds true for all x , we can substitute $x + y$ for x to obtain

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

Substituting this back into the left-hand side of (2.12) reduces (2.12) to the equation

$$g(x+y) = g(x) + g(y)$$

which is Cauchy's equation. Therefore, there exists a constant a such that $g(x) = ax$ for all real x . Letting $f(0) = b$, we see that the general form for f must be

$$f(x) = ax + b. \quad (2.14)$$

2.4 Linear functional equation

A plausible extension of both Cauchy's and Jensen's equations is

$$f(ax + by + c) = pf(x) + qf(y) + r. \quad (2.15)$$

For which choices of constants a , b , c , p , q , and r does there exist a continuous function $f : \mathbf{R} \rightarrow \mathbf{R}$ satisfying (2.15), and what is the general form of the solution?

The reader who wishes a challenge can solve this problem directly, finding thereby that the class of functions is not particularly rich. The reader who wants a few hints to get started can look at Problem 11 at the end of the chapter.

2.5 Cauchy's exponential equation

Cauchy's exponential equation has the form

$$f(x+y) = f(x)f(y), \quad (2.16)$$

where $f : \mathbf{R} \rightarrow \mathbf{R}$ is assumed to be continuous and not identically zero. We encountered this equation in Chapter 1, where it arose as the functional equation used in proving the general form of the binomial theorem.

Suppose there exists a value x_0 such that $f(x_0) = 0$. Cauchy's exponential equation implies that

$$\begin{aligned} f(x) &= f[(x-x_0) + x_0] \\ &= f(x-x_0)f(x_0) \\ &= 0. \end{aligned}$$

So any function which satisfies (2.16) and vanishes at a point must vanish everywhere. As f is assumed to be some function other than the function $f \equiv 0$, we see that we have a contradiction. So $f(x) \neq 0$ for all real values x .

In addition, (2.16) implies that

$$f(x) = \left[f\left(\frac{x}{2}\right) \right]^2. \quad (2.17)$$

As the right-hand side is strictly positive, we conclude that $f(x) > 0$ for all real x .

Having established this, the rest is easy. As f is strictly positive, we can take logarithms. Define $g(x) = \log f(x)$. Then g is continuous and satisfies Cauchy's equation. So $g(x) = ax$ for some real value a . We conclude that there exists a real number $b > 0$ such that $f(x) = b^x$ for all real x .

Before closing this section, it is worth mentioning one final equation of this type studied by Cauchy. This is the equation

$$f(xy) = f(x)f(y), \quad (2.18)$$

where x and y are positive real numbers. Letting $z = \log x$ and $w = \log y$ we see that this equation reduces to

$$g(z+w) = g(z)g(w), \quad (2.19)$$

where $f(x) = g(\log x)$. But this is Cauchy's exponential equation. If $f(x)$ is continuous for all positive x , then $g(z)$ must be continuous for all real z . We can conclude that either f is a power function of the form $f(x) = x^k$ for some real number k , or is equal to zero everywhere. Power functions of this kind have a role to play when we consider Euler's equation in Section 2.10.

2.6 Pexider's equation

Pexider's equation has the form

$$f(x+y) = g(x) + h(y). \quad (2.20)$$

We are required to find all continuous functions $f, g, h : \mathbf{R} \rightarrow \mathbf{R}$ satisfying (2.20) for all real values of x and y .

Pexider's equation is a natural generalization of Cauchy's equation. However, as we show, it does not lead to a much richer class of solutions. Put $y = 0$ in (2.20), and set $h(0) = b$. Then

$$g(x) = f(x) - b. \quad (2.21)$$

Similarly, putting $x = 0$ and $g(0) = a$, we have

$$h(y) = f(y) - a. \quad (2.22)$$

So (2.20) becomes

$$f(x+y) = f(x) + f(y) - a - b. \quad (2.23)$$

Now let $f_0(z) = f(z) - a - b$ for all real z . Then f_0 satisfies

$$f_0(x+y) = f_0(x) + f_0(y), \quad (2.24)$$

which is Cauchy's equation. The continuity of f implies the continuity of f_0 . So solving Cauchy's equation for f_0 we obtain $f_0(z) = cz$ for some real value c . So our solution to Pexider's equation in (2.20) is

$$f(z) = cz + a + b \quad (2.25)$$

$$g(x) = cx + a \quad (2.26)$$

$$h(y) = cy + b, \quad (2.27)$$

where a, b, c are arbitrary.

2.7 Vincze's equation

The next equation that we consider is a generalization of Pexider's equation. Suppose we wish to find all solutions in functions f, g, h and k to the equation

$$f(x+y) = g(x)k(y) + h(y) \quad (2.28)$$

for all real x and y . Once again, we impose an appropriate condition on the functions such as continuity. See Vincze [1962].

Let $k(0) = a$ and $h(0) = b$. Now if $a = 0$, then f is a constant function. We dispense with this case henceforth, and assume that $a \neq 0$. Putting $y = 0$ in (2.28) we get

$$g(x) = [f(x) - b]/a. \quad (2.29)$$

So we can rewrite (2.28) as

$$f(x+y) = \phi(y)f(x) + \psi(y), \quad (2.30)$$

where

$$\phi(y) = k(y)/a \quad (2.31)$$

and

$$\psi(y) = h(y) - \frac{bk(y)}{a}. \quad (2.32)$$

With these substitutions, we have $\phi(0) = 1$ and $\psi(0) = 0$.

Now, let us define the function

$$\chi(y) = f(y) - f(0). \quad (2.33)$$

Then (2.30) can be rewritten as

$$\chi(x+y) = \phi(y)\chi(x) + \chi(y). \quad (2.34)$$

However, addition is commutative. Therefore,

$$\phi(y)\chi(x) + \chi(y) = \chi(x+y) = \chi(y+x) = \phi(x)\chi(y) + \chi(x). \quad (2.35)$$

So we obtain

$$[\phi(y) - 1]\chi(x) = [\phi(x) - 1]\chi(y). \quad (2.36)$$

This equation can be solved by breaking it into two cases.

First Case: $\phi(y) = 1$ for all y . In this case, from (2.34) we see that χ satisfies Cauchy's equation $\chi(x+y) = \chi(x) + \chi(y)$. So there exists a constant d such that $\chi(x) = dx$ for all x . From (2.33) we determine that

$$f(x) = dx + c, \quad (2.37)$$

where $c = f(0)$. Plugging (2.37) into (2.30) and using the fact that $\phi \equiv 1$ we can determine that

$$\psi(y) = dy. \quad (2.38)$$

From these solutions, we can obtain expressions for our original unknown functions, namely

$$f(x) = dx + c \quad (2.39)$$

$$k(y) = a \text{ for all } y \quad (2.40)$$

$$g(x) = (dx + c - b)/a \quad (2.41)$$

$$h(y) = dy + b. \quad (2.42)$$

Second Case: $\phi(y_0) \neq 1$ for some y_0 . Set $y = y_0$ in equation (2.36). In addition, set

$$\frac{\chi(y_0)}{\phi(y_0) - 1} = s \quad (2.43)$$

say. Then (2.36) becomes

$$\chi(x) = s[\phi(x) - 1]. \quad (2.44)$$

Now we have excluded the case where f is a constant function from consideration here. This implies that χ is not the zero function, and therefore that $s \neq 0$. Let us put equation (2.44) into equation (2.34). We obtain

$$s[\phi(x+y) - 1] = s\phi(y)[\phi(x) - 1] + s[\phi(y) - 1]. \quad (2.45)$$

As $s \neq 0$, we can cancel s and reduce the equation to

$$\phi(x + y) = \phi(x) \phi(y) \quad (2.46)$$

which is Cauchy's exponential equation.

So there must exist a positive constant t such that

$$\phi(y) = t^y. \quad (2.47)$$

From this solution for ϕ we can obtain a solution to f using equations (2.44) and (2.33). We must have

$$f(x) = s t^x + c. \quad (2.48)$$

Then using (2.30) we see that ψ can be written as

$$\psi(y) = c[1 - t^y]. \quad (2.49)$$

Having found solutions for f , ϕ and ψ we can finally write out solutions in terms of our original functions. We get

$$f(x) = s t^x + c \quad (2.50)$$

$$k(y) = a t^y \quad (2.51)$$

$$g(x) = (s t^x + c - b)/a \quad (2.52)$$

$$h(y) = c + (b - c) t^y. \quad (2.53)$$

Having solved Vincze's equation, the reader may naturally wish to try to solve an equation of the form

$$f(x + y) = g(x) k(y) + j(x) h(y). \quad (2.54)$$

However, we have reached the limit for which this kind of generality is profitable. Equation (2.54) contains many interesting special cases, but unfortunately there are too many solutions to the equation.

2.8 Cauchy's inequality

Just for a change of pace let us try a functional inequality known as *Cauchy's inequality*.³ Let us consider all solutions to

$$f(x + y) \leq f(x) + f(y) \quad (2.55)$$

for all real values of x and y . Now it turns out—not unexpectedly—that there are too many solutions to this inequality for us to specify a particular family of functions as the general solution. However, let us include an additional inequality on the function, namely

³ I know that this is a book about equations. But this one is too good to miss!

$$f(x) \leq x \quad (2.56)$$

for all real x . What solutions exist satisfying Cauchy's inequality and this additional constraint?

First, we set $x = y = 0$ in equation (2.55) to obtain

$$f(0) \leq 2f(0) \quad \text{or} \quad f(0) \geq 0. \quad (2.57)$$

Because we also have $f(0) \leq 0$ by (2.56), it follows that $f(0) = 0$. Next, set $y = -x$ in (2.55). We get

$$f(0) \leq f(x) + f(-x)$$

or

$$f(x) \geq -f(-x) \quad (2.58)$$

using $f(0) = 0$. Combining (2.56) with (2.58) we have

$$f(x) \geq -f(-x) \geq -(-x) = x.$$

Combined with the inequality in (2.56), we can conclude that $f(x) = x$.

2.9 Equations involving functions of two variables

Another way to extend Cauchy's equation is to consider functions of two or more variables. In this section, we consider the two-variable case. The extension to more variables is routine.

Suppose $f(x, y)$ is a real-valued function of two real variables x and y , which satisfies

$$f(x_1 + x_2, y) = f(x_1, y) + f(x_2, y) \quad (2.59)$$

$$f(x, y_1 + y_2) = f(x, y_1) + f(x, y_2) \quad (2.60)$$

for all real x, y, x_1, x_2, y_1 , and y_2 . In order to solve this equation, we also impose a continuity condition on f . It is sufficient to require that f be a continuous function of two variables. Although this assumption is a natural one, it turns out to be stronger than we need. In fact, it is sufficient to demand that f be continuous in each of its variables, a weaker condition, that says that if we fix x , then $f(x, y)$ is continuous in y and vice versa.

Treating y as fixed in (2.59), we can apply Theorem 2.3 to equation (2.59). From that theorem, we see that f can be written in the form

$$f(x, y) = c(y)x$$

for all real x and y . Plugging this expression into (2.60), we get

$$c(y_1 + y_2)x = c(y_1)x + c(y_2)x \quad (2.61)$$

for all x, y_1 , and y_2 . Setting $x = 1$, we obtain Cauchy's equation again. As $f(x, y)$ is continuous in the variable y , it follows that c is a continuous function. So we can conclude that there exists some real number c_0 such that $c(y) = c_0 y$ for all y . From this we can conclude that

$$f(x, y) = c_0 x y \quad (2.62)$$

for all real x and y .

2.10 Euler's equation

In this section, we continue our investigation of functions of two variables by considering a family of functional equations associated with the property of homogeneity. Let k be any real number. For given k , the equation

$$f(tx, ty) = t^k f(x, y) \quad (2.63)$$

for all positive x, y , and t , is called *Euler's equation*. A function $f(x, y)$ satisfying Euler's equation is said to be a *homogeneous function of degree k* . For example, the function

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

is homogeneous of degree one, the function

$$f(x, y) = \frac{x}{y}$$

has degree zero, and the function

$$f(x, y) = x^2 + 2y^2 + 7xy$$

has degree two.

How can we find the class of all functions which satisfy Euler's equation for a given choice of k ? This looks difficult, but is really quite easy. Let us substitute $t = x^{-1}$ into (2.63). It then becomes

$$f(x, y) = x^k f(1, x^{-1}y). \quad (2.64)$$

Suppose that we let $g(z) = f(1, z)$. We can write

$$f(x, y) = x^k g(x^{-1}y) \quad (2.65)$$

to obtain the solution to Euler's equation. It can be seen that any function g defined on the positive real axis determines a homogeneous function of degree k , and vice versa.

The simplicity of (2.65) suggests that it might be reasonable to generalize Euler's equation to



Fig. 2.2. Jean le Rond d'Alembert (1717–1783).

$$f(tx, ty) = h(t) f(x, y), \quad (2.66)$$

where the function $h(t)$ is given. However, it turns out that the generality of this modified Euler equation is a bit of an illusion here. Problem 7 at the end of the chapter asks the reader to show that if h is continuous (or bounded below on a closed interval) then h can be written in the form $h(t) = t^k$, provided that f is not the function which is zero everywhere.

2.11 D'Alembert's equation

It is now time to return to the analysis of D'Alembert's equation

$$g(x+y) + g(x-y) = 2g(x)g(y) \quad (2.67)$$

that was briefly discussed in Section 1.6 of Chapter 1.

We consider the derivation of the full solution to this equation under the assumption that $g(x)$ is continuous. Note first that if $g(x) = 0$ for all x , then g satisfies the equation. From now on, we assume that there is at least one value at which g does not vanish.

Next set $y = 0$. The equation reduces to

$$2g(x) = 2g(x)g(0). \quad (2.68)$$

Assume that x is chosen so that $g(x) \neq 0$. Then we can cancel $g(x)$ to conclude that $g(0) = 1$. On the other hand, setting $x = 0$, equation (2.67) reduces to

$$\begin{aligned} g(y) + g(-y) &= 2g(0)g(y) \\ &= 2g(y) \end{aligned} \tag{2.69}$$

which implies that $g(y) = g(-y)$ for all y .

Next, let us set $x = y$. Using (2.68) we get

$$g(2x) + 1 = 2[g(x)]^2.$$

Equivalently, we have

$$\left[g\left(\frac{x}{2}\right) \right]^2 = \frac{g(x) + 1}{2}. \tag{2.70}$$

An equation such as this is already quite promising in suggesting the kind of double angle formula satisfied by $\cos x$. However, it is also satisfied by the function $\cosh x$. So we must be careful to determine the full set of solutions.

Because $g(x)$ is assumed to be continuous, and because $g(0) = 1$, there exists some value $t > 0$ such that $g(x) > 0$ for all x in the closed interval $[-t, t]$. We now break the argument up into two cases.

First, we consider the case where $0 < g(t) \leq 1$. So there exists some number $0 \leq s \leq \pi/2$ such that

$$g(t) = \cos s. \tag{2.71}$$

From this statement, it is a straightforward argument by mathematical induction, using (2.70) to show that

$$g\left(\frac{t}{2^m}\right) = \cos\left(\frac{s}{2^m}\right) \tag{2.72}$$

for $m = 0, 1, 2, 3, \dots$. To see this, note that the base case for the induction with $m = 0$ is simply (2.71). Suppose the result holds up to m . Then

$$\begin{aligned} \left[g\left(\frac{t}{2^{m+1}}\right) \right]^2 &= \left[g\left(\frac{t/2^m}{2}\right) \right]^2 \\ &= \frac{g(t/2^m) + 1}{2} \text{ by equation (2.70)} \\ &= \frac{\cos(s/2^m) + 1}{2} \text{ by induction hypothesis} \\ &= \cos^2\left(\frac{s}{2^{m+1}}\right). \end{aligned} \tag{2.73}$$

Both $g(t/2^{m+1})$ and $\cos(s/2^{m+1})$ are positive. So we may take square roots of both sides to conclude that

$$g\left(\frac{t}{2^{m+1}}\right) = \cos\left(\frac{s}{2^{m+1}}\right)$$

as required. This completes the induction step, and the proof of equation (2.72) is complete.

Next, let us return to equation (2.67), and make the substitution $x = ny$. This gives us

$$g[(n+1)y] = 2g(y)g(ny) - g[(n-1)y]. \quad (2.74)$$

Equations (2.72) and (2.74) can be used to construct a second argument from induction on n that

$$g\left(\frac{nt}{2^m}\right) = \cos\left(\frac{ns}{2^m}\right) \quad (2.75)$$

for all integers $m, n \geq 0$. To see this, we note that the base case $n = 0$ follows from the fact that $g(0) = 1$, and the case $n = 1$ has already been established by the earlier induction argument. Suppose the result holds up to n . Then

$$\begin{aligned} g\left[\frac{(n+1)t}{2^m}\right] &= 2g\left(\frac{t}{2^m}\right)g\left(\frac{nt}{2^m}\right) - g\left[\frac{(n-1)t}{2^m}\right] \quad \text{by (2.74)} \\ &= 2\cos\left(\frac{s}{2^m}\right)\cos\left(\frac{ns}{2^m}\right) - \cos\left[\frac{(n-1)s}{2^m}\right] \quad \text{by induction hyp.} \\ &= 2\cos\left(\frac{s}{2^m}\right)\cos\left(\frac{ns}{2^m}\right) \\ &\quad - \left\{\cos\left(\frac{ns}{2^m}\right)\cos\left(\frac{s}{2^m}\right) + \sin\left(\frac{ns}{2^m}\right)\sin\left(\frac{s}{2^m}\right)\right\} \\ &= \cos\left(\frac{ns}{2^m}\right)\cos\left(\frac{s}{2^m}\right) - \sin\left(\frac{ns}{2^m}\right)\sin\left(\frac{s}{2^m}\right) \\ &= \cos\left[\frac{(n+1)s}{2^m}\right]. \end{aligned} \quad (2.76)$$

This completes the induction step, and equation (2.75) is proved.

Since both s and t are greater than zero, we can write $s = at$ for some $a > 0$. So equation (2.75) becomes

$$g\left(\frac{nt}{2^m}\right) = \cos\left(a\frac{nt}{2^m}\right) \quad (2.77)$$

for all $n, m \geq 0$. Using the fact that both g and \cos are even functions, we can extend (2.77) so that

$$g(x) = \cos(ax)$$

for all x which can be written in the form $x = \pm n2^{-m}t$ for some $m, n \geq 0$. Numbers $\pm n2^{-m}$ are called the *dyadic rationals*. It is well known, and is not proved here, that the set of dyadic rationals is dense. Thus any real number can be written as a limit of a sequence of dyadic rationals. Using an argument which parallels Proposition 2.2, we see that $g(x) = \cos(ax)$ for all real x .

The *second case* that needs to be considered is where $g(t) > 1$. We define the hyperbolic cosine function to be

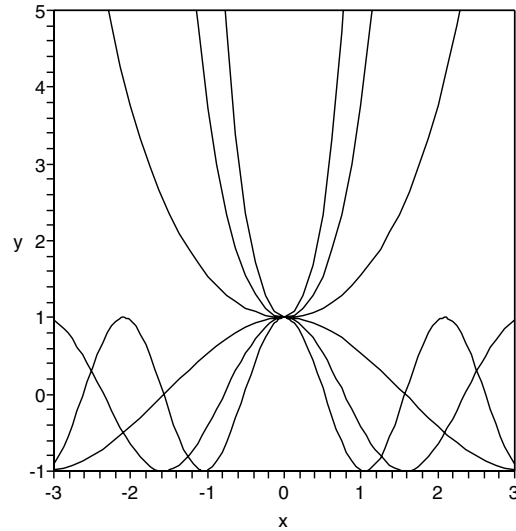


Fig. 2.3. A few solutions to d'Alembert's equation.

$$\cosh x = \frac{e^x + e^{-x}}{2}. \quad (2.78)$$

It is straightforward to see that there exists some $s > 0$ such that $g(t) = \cosh s$. From here on, the argument in this second case parallels the argument given in the first case above. So we omit the details. Note that the hyperbolic cosine function $\cosh x$ satisfies the same sorts of trigonometric identities that the circular cosine function $\cos x$ satisfies, if allowance is made for the occasional change of sign in the formula. The rule for sign changes is known as Osborne's rule: if the trigonometric identity has the product of two hyperbolic sines or two hyperbolic tangents, the sign on that term is reversed. For example, the summation formula for $\cos x$ is

$$\cos(u + v) = \cos u \cos v - \sin u \sin v.$$

The summation formula for $\cosh x$ is

$$\cosh(u + v) = \cosh u \cosh v + \sinh u \sinh v,$$

where

$$\sinh x = \frac{e^x - e^{-x}}{2}.$$

Upon completion of this part of the proof, we deduce that $g(x) = \cosh(ax)$ for some $a > 0$.

Figure 2.3 displays some of the functions that we have found to be solutions to d'Alembert's equation in this section. Plotted are three circular functions $\cos nx$, and three hyperbolic functions $\cosh nx$, for $n = 1, 2, 3$.

2.12 Problems

1. Let $f : \mathbf{R}^+ \rightarrow \mathbf{R}$ be a continuous function, satisfying $f(x+y) = f(x) + f(y)$ for all $x, y \geq 0$. Prove that $f(x) = ax$ for some real number a .
2. This problem asks the reader to fill in the details of the proof of Theorem 2.4, above. Let $f : \mathbf{R} \rightarrow \mathbf{R}$ be a continuous function satisfying Cauchy's equation. Suppose in addition that there exists some interval $[c, d]$ of real numbers, where $c < d$, such that f is bounded below on $[c, d]$.
 - a) Show that $f(nx) = nf(x)$ for all real x .
 - b) Define $p = d - c$. Show that f is bounded below on the interval $[0, p]$. (However, it need not be bounded below by the same constant as on the interval $[c, d]$.)
 - c) Define the function

$$g(x) = f(x) - \frac{f(p)}{p}x.$$

Prove that g is also bounded below on the interval $[0, p]$ and satisfies Cauchy's equation.

- d) Show that g is periodic with period p in the sense that $g(x+p) = g(x)$ for all real x . Conclude from this, and the fact that g is bounded below on the interval $[0, p]$ that g is bounded below on the entire real line $(-\infty, +\infty)$.
- e) Suppose that there exists some x_0 for which $g(x_0) \neq 0$. Prove a contradiction, by showing that the sequence of values $g(nx_0)$, $n = \pm 1, \pm 2, \pm 3, \dots$ is not bounded below.
- f) Conclude that $g(x) = 0$ for all real x , and therefore that $f(x) = ax$ for all real x , where $a = f(p)/p$.

3. Find all solutions in continuous functions $f, g, h : \mathbf{R} \rightarrow \mathbf{R}$ to the functional equation

$$f(x+y) = g(x)h(y).$$

4. Consider continuous functions $f : \mathbf{R} \rightarrow \mathbf{R}$ such that

$$f(x_1) + f(x_2) + f(x_3) = f(y_1) + f(y_2) + f(y_3)$$

whenever $x_1 + x_2 + x_3 = y_1 + y_2 + y_3 = 0$. Prove that there exist real numbers a and b such that $f(x) = ax + b$ for all real x .

5. ★ *Cauchy's equation on a circle.* Let \mathbf{T} be the set of real numbers modulo one. That is, $x = y \pmod{1}$ if and only if $x - y$ is an integer. (We can conveniently think of \mathbf{T} as the set of points of the interval $[0, 1)$ with the right-hand end “wrapped” around to the left-hand end of the interval.) Consider a continuous function $f : \mathbf{T} \rightarrow \mathbf{T}$ such that

$$f(x + y) = f(x) + f(y) \pmod{1}.$$

Note that a function f is continuous on \mathbf{T} if f is continuous on the interval $[0, 1)$ and

$$\lim_{x \rightarrow 1^-} f(x) = f(0).$$

Show that there exists an integer n such that $f(x) = nx \pmod{1}$ for all $x \in \mathbf{T}$.

6. State and prove a version of Problem 4 above, for a continuous function f from \mathbf{T} to \mathbf{T} .
7. ★ Suppose that there exists a function $h(t)$ defined for all positive t and a positive function $f(x, y)$ defined for all positive x and y such that

$$f(tx, ty) = h(t) f(x, y) \tag{2.79}$$

which is the modified Euler equation of Section 2.10 above. Suppose, in addition, that h is a continuous function. Show that $h(t) = t^k$ for some choice of k .

8. Suppose $f : \mathbf{R} \rightarrow \mathbf{R}$ has the property that there exists a constant K such that

$$|f(x) - f(y)| \leq K(x - y)^2.$$

Prove that f is a constant function.

9. The property associated with Cauchy's inequality, namely

$$f(x + y) \leq f(x) + f(y)$$

is known as *subadditivity* of the function $f(x)$. Suppose we reverse the other inequality and assume that $f(x) \geq x$ for all x . Does the conclusion $f(x) = x$ follow in this case?

10. ★ Determine which functions $f(x)$ mapping the rationals to the rationals satisfy the functional equation

$$f[x + f(y)] = f(x) f(y).$$

11. In this problem, we attack the linear functional equation of Section 2.4. Suppose $f : \mathbf{R} \rightarrow \mathbf{R}$ is a continuous function satisfying

$$f(ax + by + c) = pf(x) + qf(y) + r \quad (2.80)$$

for some choice of constants $a, b, c, p, q,$ and r . The reader is warned that the steps below contain many hints. So the greater challenge is to solve the linear functional equation without reading the steps below.

- a) Prove the following equations

$$f(0) = p(-c/a) + qf(0) + r \quad (2.81)$$

$$f(u) = pf\left(\frac{u-c}{a}\right) + qf(0) + r \quad (2.82)$$

$$f(v) = pf(-c/a) + qf(v/b) + r \quad (2.83)$$

$$f(u+v) = pf\left(\frac{u-c}{a}\right) + qf(v/b) + r \quad (2.84)$$

- b) Using (2.81)–(2.84), prove that $f(u+v) + f(0) - f(u) - f(v) = 0$.
 c) Conclude that f must have the form $f(x) = sx + t$ for some constants s and t .
 d) Show that such a solution will work if and only if $p = a$ and $q = b$.

12. Find all continuous $f : \mathbf{R} \rightarrow \mathbf{R}$ such that

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

for all real x and y .

13. Find all continuous functions f such that for all real x and y

$$f(x+y)f(x-y) = [f(x)f(y)]^2.$$

14. ★ For any real numbers x and y , we define the *quasiarithmetic mean* of x and y with generating function f to be a function of the form

$$m(x, y) = f^{-1}\left[\frac{f(x) + f(y)}{2}\right],$$

where f is a continuous strictly increasing or strictly decreasing function. For example, the ordinary arithmetic mean is obtained by setting $f(x) = x$ and the geometric mean by setting $f(x) = \log x$. Other means that can be constructed include the harmonic mean ($f(x) = x^{-1}$) and the root mean square ($f(x) = x^2$). These last three definitions are usually defined only for x and y positive.

- a) A question of immediate interest is when distinct functions f and g give distinct quasiarithmetic means. Suppose that

$$f^{-1} \left[\frac{f(x) + f(y)}{2} \right] = g^{-1} \left[\frac{g(x) + g(y)}{2} \right].$$

Prove that there exist constants $a \neq 0$ and b such that

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

In addition, prove the converse. Thus prove that two quasiarithmetic means are equivalent if and only if their generating functions are linearly related.

- b) A quasiarithmetic mean is said to be *translative* if it has the property that

$$m(x + t, y + t) = m(x, y) + t$$

for all real values x, y , and t . Obviously, the arithmetic mean ($f(x) = x$) is translative. Is it the only one? It turns out that there is another family of translative quasiarithmetic means.

Solve

$$f^{-1} \left[\frac{f(x + t) + f(y + t)}{2} \right] = f^{-1} \left[\frac{f(x) + f(y)}{2} \right] + t$$

to find all possible generating functions f .

- c) A quasiarithmetic mean is said to be *homogeneous* if it has the property that

$$m(tx, ty) = t m(x, y)$$

for all positive values x, y and t . Obviously, the geometric mean ($f(x) = \log x$) is homogeneous. Find all others.

15. ★ (The following functional equation requires calculus.) Find every twice-differentiable real-valued function f with domain the set of all real numbers and satisfying the functional equation

$$[f(x)]^2 - [f(y)]^2 = f(x + y) f(x - y)$$

for all real numbers x and y .

16. ★ For each positive integer n , let f_n be a real-valued symmetric function of n real variables. Suppose that for all n and for all real numbers x_1, \dots, x_n, y , it is true that

- $f_n(x_1 + y, \dots, x_n + y) = f_n(x_1, \dots, x_n) + y$.
- $f_n(-x_1, \dots, -x_n) = -f_n(x_1, \dots, x_n)$.
- $f_{n+1}(f_n(x_1, \dots, x_n), \dots, f_n(x_1, \dots, x_n), x_{n+1}) = f_{n+1}(x_1, \dots, x_{n+1})$.

Prove that

$$f_n(x_1, \dots, x_n) = \frac{x_1 + \dots + x_n}{n}.$$

(Note that a function f is said to be symmetric if its value is independent of the order of its variables: $f(x, y) = f(y, x)$ for two variables, and so on.)

Functional equations with one variable

3.1 Introduction

In this chapter, we consider functional equations involving a single variable. We have already considered these equations in Chapter 1, in the work of Charles Babbage. It is now time to examine some of these equations in greater detail. We begin by considering a useful technique known as linearization, which Babbage used to solve for iterative roots. After that, we look at some examples of conjugacy equations, which also arise from attempts to linearize functions.

3.2 Linearization

Sometimes, rather complicated-looking problems can be made quite simple. For example, suppose that we are asked to find a solution to the equation

$$f(x^2) - f(x) = 1 \tag{3.1}$$

for $x > 1$. We can tell by looking at this equation that if f is increasing, it must do so rather slowly, as squaring x only makes f increase by one! It is rather tempting to try to change the problem to a function which increases faster. Let $F(x) = f(a^x)$, where $a > 0$. Equivalently $F(\log_a x) = f(x)$. Then for $x > 0$ the function F satisfies

$$F(2x) - F(x) = 1 \tag{3.2}$$

for all $x > 0$. But this equation reminds us of the properties of a logarithm. The function $F(x) = \log_2 x$ satisfies equation (3.2). So

$$f(x) = \log_2 \log_a x \tag{3.3}$$

satisfies equation (3.1) for all $a > 0$.

This technique is known as *linearization*. It can be used to turn some more complicated functional equations into simpler ones. In the example above, the nonlinearity was in the domain of the function. However, in the following equation the nonlinearity is in the range of f .

$$f(x+1) = [f(x)]^2. \quad (3.4)$$

We assume that $f(x) \neq 0$ for all x . We can make this equation linear by letting $F(x) = \log_a f(x)$, where, once again, $a > 0$. (Note that if the equation holds true for all x , then f must be strictly positive.) Then the equation becomes $F(x+1) = 2F(x)$, for which a solution $F(x) = 2^x$ can easily be found. Thus we can conclude that a solution to the original equation is

$$f(x) = a^{2^x}. \quad (3.5)$$

More generally, we seek to linearize a functional equation by replacing a function f by a function F , where

$$\rho(F[\phi(x)]) = f(x). \quad (3.6)$$

The functions ρ and ϕ are specially selected for the purposes of linearizing the equation. Unfortunately, many functional equations cannot be simplified by linearization techniques. However, it is worth watching carefully for complexity in an equation that can be eliminated by a simple transformation.

The following problem is part of a shortlisted question from the 1997 International Mathematical Olympiad.

Example 3.1. Do there exist functions $f : \mathbf{R} \rightarrow \mathbf{R}$ and $g : \mathbf{R} \rightarrow \mathbf{R}$ such that

$$f[g(x)] = x^2 \quad \text{and} \quad g[f(x)] = x^4 \quad (3.7)$$

for all $x \in \mathbf{R}$?

Solution. For the moment, let us set aside the question of whether f and g can be found for all real x . We concentrate on constructing the functions on the restricted domain $[1, \infty)$.

From the two equations we get

$$f(x^4) = f(g[f(x)]) = [f(x)]^2. \quad (3.8)$$

We can linearize the equation $f(x^4) = [f(x)]^2$ by defining a function F such that

$$f(x) = a^{F(\log_b x)}, \quad (3.9)$$

where $a, b > 0$. Making this substitution in (3.8), and letting $y = \log_b x$, we get

$$F(4y) = 2F(y) \quad (3.10)$$

for all $x \geq 0$. This is easily solved with a function proportional to $F(x) = \sqrt{x}$. Thus

$$f(x) = a\sqrt{\log_b x} \quad (3.11)$$

for all $x \geq 1$ is a candidate. Now working the other way, we have

$$g(x^2) = g(f[g(x)]) = [g(x)]^4. \quad (3.12)$$

We linearize using

$$g(x) = c^{G(\log_a x)} \quad (3.13)$$

as before to obtain the functional equation

$$G(2y) = 4G(y). \quad (3.14)$$

A solution to this equation is $G(x) = x^2$. So

$$g(x) = c^{(\log_a x)^2} \quad (3.15)$$

for $x \geq 1$ is a candidate for a solution for g . In order to check properly we have to go back and plug these functions into the original functional equations to see if they work for any choice of a , b , c , and d . Plugging (3.11) and (3.15) into (3.7) we see that

$$f(x) = 2\sqrt{\log_2 x} \quad \text{and} \quad g(x) = 16^{(\log_2 x)^2} \quad (3.16)$$

is one possible pair of solutions to (3.7).

It remains to extend f and g to the entire real line. This can be done, so as to satisfy (3.7). This is Problem 8 at the end of the chapter. \square

3.3 Some basic families of equations

One of the simplest families of functional equations that we can write has the form

$$f(x) = f[\alpha(x)] \quad (3.17)$$

for all real x , where $\alpha : \mathbf{R} \rightarrow \mathbf{R}$ is a specified function. What are the restrictions on f if it is to satisfy such an equation?

If we do not assume that f is continuous, then the complete family of solutions is easy to write down. First of all, let us write

$$\alpha^1(x) = \alpha(x) \quad \text{and} \quad \alpha^{n+1}(x) = \alpha[\alpha^n(x)] \quad (3.18)$$

for $n = 1, 2, \dots$. For convenience, we also define α^0 to be the function $\alpha^0(x) = x$. We call the sequence

$$\alpha(x), \alpha^2(x), \alpha^3(x), \dots \quad (3.19)$$

the *splinter* of x . Applying (3.17) a total of n times in succession, we see that

$$f(x) = f[\alpha^n(x)]. \quad (3.20)$$

Thus the function f is constant on the splinter of x .

We say that real numbers x and y are *equivalent under iteration by α* provided that there exist two integers $n, m \geq 0$ such that

$$\alpha^n(x) = \alpha^m(y).$$

We write $x \sim_\alpha y$ when this is true. It is easy to check that \sim_α is an *equivalence relation* on \mathbf{R} . That is, the relation \sim_α satisfies the following three properties.

1. *Reflexivity*: $x \sim_\alpha x$ for all real x ;
2. *Symmetry*: $x \sim_\alpha y$ implies that $y \sim_\alpha x$;
3. *Transitivity*: $x \sim_\alpha y$ and $y \sim_\alpha z$ implies that $x \sim_\alpha z$.

For each real number x , let us define the set $A_\alpha(x)$ to be the set of all real values y for which $y \sim_\alpha x$. That is,

$$A_\alpha(x) = \{y \in \mathbf{R} : y \sim_\alpha x\}.$$

The sets $A_\alpha(x)$ are the *equivalence classes* for the equivalence relation \sim_α , and are often called the *orbits* of α . It follows from (3.20) that the function f must be constant on each of the orbits of α . So a necessary and sufficient condition that f satisfy (3.17) is that we can factor f so that

$$f(x) = g[A_\alpha(x)]$$

where g is some real-valued function defined on the collection of all orbits.

Example 3.2. For example, the equation $f(x) = f(x + 1)$ is of this kind with $\alpha(x) = x + 1$. In this case, the set A_x consists of all y whose fractional part equals the fractional part of x . Equivalently, $y \in A_x$ if and only if $x \equiv y \pmod{1}$. So the class of solutions f to the functional equations are the functions

$$f(x) = g(x - [x])$$

where $g : \mathbf{R} \rightarrow \mathbf{R}$ is arbitrary. Here, $[x]$ represents the greatest integer less than or equal to x .

An interesting problem is to determine the family of continuous functions f which are solutions to (3.17). The assumption of continuity is often strong enough to force the solutions to (3.17) to be constant. For example, suppose that there exists some real number x_0 such that

$$\lim_{n \rightarrow \infty} \alpha^n(x) = x_0 \quad (3.21)$$

for all real x . Then

$$f(x_0) = f \left[\lim_{n \rightarrow \infty} \alpha^n(x) \right].$$

However, if f is a continuous function, then we can “pull” the limit to the outside. So

$$\begin{aligned} f(x_0) &= f \left[\lim_{n \rightarrow \infty} \alpha^n(x) \right] \\ &= \lim_{n \rightarrow \infty} f[\alpha^n(x)] \\ &= f(x) \end{aligned}$$

for all x . In other words, f must be constant. Note that this argument assumes that x_0 lies in the domain of f . A completely different situation can arise if x_0 is not in the domain.

If the splinter $\alpha^n(x)$ has a limit we can determine its value by checking for fixed points under iteration by α when α is a continuous function. For example, suppose x_0 is the limit of $\alpha^n(x)$ as in (3.21). Then

$$\begin{aligned} x_0 &= \lim_{n \rightarrow \infty} \alpha^n(x) \\ &= \lim_{n \rightarrow \infty} \alpha^{n+1}(x) \\ &= \lim_{n \rightarrow \infty} \alpha[\alpha^n(x)] \\ &= \alpha \left[\lim_{n \rightarrow \infty} \alpha^n(x) \right] \\ &= \alpha(x_0). \end{aligned}$$

Note that we can “push” the limit through α because the function α is continuous. Also, note that being a fixed point under iteration by α is a necessary but not sufficient condition for the convergence of $\alpha^n(x)$.

In general, if f is a continuous function satisfying (3.17), then f will not only be constant on the orbits $A_\alpha(x)$ for each x , but will also be *constant on the closure* of each of these orbits. Rather informally, we can say that the closure of any orbit $A_\alpha(x)$ is the set of points in $A_\alpha(x)$ together with the set of all limit points from sequences chosen with values from $A_\alpha(x)$. So f will be a constant function if the closure of $A_\alpha(x)$ and $A_\alpha(y)$ have a point in common for any x and y .

To illustrate these ideas, let us consider problem A6 from the 1996 Putnam Competition.

Example 3.3. Let a be any positive real number. Find, with proof, all continuous functions $f : \mathbf{R} \rightarrow \mathbf{R}$ such that $f(x) = f(x^2 + a)$, for all real x .

Solution. First note that

$$f(-x) = f[(-x)^2 + a] = f(x^2 + a) = f(x). \quad (3.22)$$

Thus it suffices to determine the value of the function for $x \geq 0$, as the relation above immediately determines the value of $f(x)$ for negative values of x . Henceforth, assume that $x \geq 0$.

For each x we need to examine the limit points of the orbit $A_\alpha(x)$ where

$$\alpha(x) = x^2 + a. \quad (3.23)$$

Solving $\alpha(x_0) = x_0$, we observe that we have the following cases:

1. $a = 1/4$. In this case, $\alpha(x_0) = x_0$ has the unique solution $x_0 = 1/2$. If $0 \leq x \leq 1/2$, then

$$x < \alpha(x) \leq 1/2.$$

So in this case, the splinter $\alpha^n(x)$ is an increasing sequence bounded above by x_0 . Therefore, the sequence $\alpha^n(x)$ has a limit. As x_0 is the only fixed point of α , the splinter of any $x < 1/2$ must have $x_0 = 1/2$ as its limit. So the closure of every orbit $A_\alpha(x)$ will contain x_0 . It follows that f must be constant on the interval $[0, 1/2]$.

On the other hand, if $x > 1/2$, then we also see that $\alpha^n(x)$ is an increasing sequence, which must be unbounded. So the orbit of any point $x > 1/2$ is the set

$$A_\alpha(x) = \{\alpha^n(x) : n = 0, \pm 1, \pm 2, \dots\},$$

where

$$\alpha^{-1}(x) = \sqrt{x - a} > x_0.$$

Therefore, when $x > 1/2$, the sequence $\alpha^{-n}(x)$ is a decreasing sequence which is bounded below by x_0 . For reasons similar to the previous case, this sequence must converge to x_0 . Thus we can conclude that the closure of this orbit contains the point $x_0 = 1/2$ as well. So f must be constant on the interval $[1/2, \infty)$. It follows that f is constant on the interval $[0, \infty)$. As $f(-x) = f(x)$, the function f is constant on the entire real line.

2. $a < 1/4$. When a is in this range, the equation $x^2 - x + a = 0$ has two distinct solutions, say

$$x_1 = (1 - \sqrt{1 - 4a})/2 \quad \text{and} \quad x_2 = (1 + \sqrt{1 - 4a})/2.$$

If $x \leq x_1$ the splinter $\alpha^n(x)$ is an increasing sequence (although not necessarily strictly) with limit point x_1 . So x_1 is in the closure of $A_\alpha(x)$. Additionally, if $x_1 < x < x_2$, then the splinter $\alpha^n(x)$ is monotone decreasing with limit point x_1 . Therefore f must be constant on the interval $[0, x_2]$.

Similarly, if $x \geq x_2$, the sequence $\alpha^{-n}(x)$ is decreasing with limit point x_2 . It follows that f is constant on the interval $[x_2, \infty)$.

Putting these facts together and using $f(-x) = f(x)$ we conclude that f is constant on the entire real line.

3. $a > 1/4$. In this case, α has no fixed points. For every real x , the splinter $\alpha^n(x)$ is a strictly increasing sequence that is not bounded above. Moreover, the function

$$\alpha^{-1}(x) = \sqrt{x - a}$$

is only defined if $x \geq a$. So the orbit $A_\alpha(x)$ of any point x contains no limit points.

To find all functions f we divide the nonnegative real numbers into the intervals

$$[0, \alpha(0)], [\alpha(0), \alpha^2(0)], [\alpha^2(0), \alpha^3(0)], \dots$$

and so on. Let g be any continuous function on $[0, \alpha(0)]$ such that $g(0) = g(\alpha(0))$. Now the reader can check that, for any positive x , there exists a unique $y \in [0, \alpha(0))$ and a nonnegative integer n such that $x = \alpha^n(y)$. We define f in such a way that $f(x) = g(y)$. We extend f to negative x by defining $f(x) = f(-x)$ when $x < 0$.

It can be checked that any f so defined satisfies the functional equation and is continuous. Furthermore, any continuous solution to the functional equation can be represented in this form. \square

This problem is a reminder to us that unless we can enforce some strong condition such as constancy, we are likely to find that a functional equation in a single variable has infinitely many solutions. The standard way to enforce a unique solution is to impose additional constraints on the function. In some cases, those additional constraints can be in the form of additional functional equations.

An interesting exception to this rule occurs when a functional equation involves an *iterative root* of the identity, also known as a *cyclic function*. Suppose there exists some positive integer n such that $\alpha^n(x) = x$. Clearly, an involution is a special case where $n = 2$. The reader will recall functions of this sort from Chapter 1, where we reviewed the work of Charles Babbage. In Chapter 1, it was the unknown function f which was presumed to be an iterative root. In the present case, the function $\alpha(x)$ given in the equation is the iterative root. Because $\alpha(x)$ is cyclic, the splinter is finite. The following example illustrates a method for solving functional equations such as this.

Example 3.4. Consider, for example, the equation

$$2f(x) + f(x^{-1}) = x,$$

where $x \neq 0$. We replace x by x^{-1} in this equation to obtain a second equation. Letting $\alpha(x) = x^{-1}$ we note that $\alpha^2(x) = x$. So we also have

$$2f(x^{-1}) + f(x) = x^{-1}.$$

Multiplying the first equation by 2, and subtracting the second equation from the first, we get

$$f(x) = \frac{2x^2 - 1}{3x}.$$

It can be checked that this satisfies the original equation.

Example 3.5. More generally, we may suppose we have an equation involving the terms $f(\alpha^j(x))$, where $\alpha^n(x) = x$. Suppose that the equation is linear in the n unknowns

$$f(x), f(\alpha(x)), \dots, f(\alpha^{n-1}(x)).$$

By successively substituting $\alpha^j(x)$, $j < n$ for x in the original equation, we obtain a set of n equations in the n unknowns $f(\alpha^j(x))$, $0 \leq j < n$. We can solve for the n unknowns using standard methods of linear algebra.

Problem 11 at the end of the chapter is of this kind, and appeared as Problem B2 on the 1971 Putnam competition.

3.4 A menagerie of conjugacy equations

The form of equation (3.17) precludes the possibility that there are any one-to-one functions which are solutions to it. However, (3.17) has a number of generalizations which can have one-to-one solutions, and which are sufficiently important to be given special names.

The equation

$$f[\alpha(x)] = s f(x) \tag{3.24}$$

is called Schröder's equation. For a given function α , we are required to find one or more functions f and a real number $s \neq 1$ such that (3.24) holds.

We have to be a bit careful about the domain of definition of any solution to (3.24). For example, suppose there exists a fixed point of the function α . If x_0 is such a value, then making the substitution $\alpha(x_0) = x_0$ in Schröder's equation leads us to the identity $f(x_0) = s f(x_0)$. As $s \neq 1$, we must either conclude that $f(x_0) = 0$ or that x_0 is outside the natural domain of definition of the function f . The latter is quite common in solving Schröder's equation. To avoid this problem, we can assume that the equation $\alpha(x) = x$ has no solutions on some interval on which a function $f(x)$ is to be constructed. A common assumption is that $0 < \alpha(x) < x$ for all x in some interval $(0, a)$ where $0 < a \leq \infty$, and $0 < s < 1$. These assumptions are not as restrictive as they look, because the equation can often be transformed (e.g., by substituting $x = \alpha^{-1}(y)$, etc.) to an equation of this kind. If we can find a positive solution to (3.24) with any choice of $0 < s < 1$, then we can easily find solutions for other values of s . For example, if $f(x)$ solves (3.24), then $[f(x)]^p$ solves (3.24) with s replaced by s^p .

A kind of inverse equation to Schröder's equation is obtained as follows. Let f be a solution to (3.24) which admits an inverse: a function $g = f^{-1}$ such that $f(x) = y$ if and only if $g(y) = x$. Then g satisfies the equation

$$g(sx) = \alpha[g(x)]. \tag{3.25}$$

Equation (3.25) is referred to as Poincaré's equation.

The equation

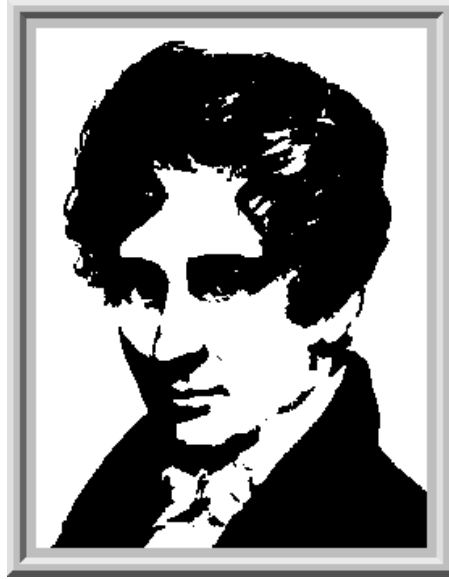


Fig. 3.1. Niels Henrik Abel (1802–1829)

$$f[\alpha(x)] = f(x) + a \quad (3.26)$$

is called Abel's equation. The real number a is assumed to be nonzero, and is either given or to be determined along with f . As with Schröder's equation, we must once again be cautious about the domain of definition of f . If we choose a value x_0 such that $\alpha(x_0) = x_0$, then we see that x_0 cannot lie in the domain of f . Otherwise we would obtain the equation $f(x_0) = f(x_0) + a$, which is clearly impossible if a is assumed to be a nonzero value. So the restriction of the domain of definition of $f(x)$ to values x such that $\alpha(x) \neq x$ is required.

The equation

$$f[\alpha(x)] = [f(x)]^p, \quad (3.27)$$

where $p \neq 1$, is Böttcher's equation. For this equation, we are interested in nonnegative functions $f(x)$.

One additional equation worthy of note is the *commutativity equation* defined by

$$f[\alpha(x)] = \alpha[f(x)]. \quad (3.28)$$

For example, Problem 1.11 in Chapter 1 involved two simultaneous commutativity equations.

All the equations that we have considered above are special cases of a family of equations called *conjugacy equations*, which take the form

$$f[\alpha(x)] = \beta[f(x)] \quad (3.29)$$

for appropriately chosen functions α and β . For example, when $\alpha = \beta$ we have the commutativity equation, when $\beta(x) = sx$ we have Schröder's equation, and so on. We do not attempt to solve conjugacy equations in their full generality below. However, we do look for solutions to their special cases.

When (3.29) is satisfied by some one-to-one function f , we say that α and β are *conjugates* of each other. When Schröder's or Abel's equations can be solved in this way, then the function α is shown to be conjugate to a linear function. Typically, linear functions are easier to study than nonlinear ones. So it is very important in mathematical sciences to be able to establish that *a nonlinear function is conjugate to a linear one*. When this is the case, we are able to linearize complex mathematical relationships and study them using linear methods. For example, Abel's equation

$$f[\alpha(x)] = x + a$$

uses a function f to establish a conjugacy between a given transformation $x \mapsto \alpha(x)$ and a simple shift transformation $x \mapsto x + a$. As the latter transformation is particularly simple, we are often able to solve problems with the shift transformation and then use the conjugacy to “transfer” the solution back to α . If α is a nonlinear transformation, then Abel's equation, and its solution f have the effect of linearizing $x \mapsto \alpha(x)$. Of course, the solution f to Schröder's equation also linearizes $x \mapsto \alpha(x)$ by establishing a conjugacy between this transformation and the scale change $x \mapsto sx$. Which of the two linearizations is more appropriate, if either, depends upon the particular character of α .

This idea of using conjugacy between two functions was essential to Babbage's solution to the problem of finding iterative roots: the iterative root of a shift transformation or a scale change is easy to define. We use conjugacy to transfer such solutions to more complicated functions. We return to consider iterative roots in Section 3.7.

3.5 Finding solutions for conjugacy equations

3.5.1 The Koenigs algorithm for Schröder's equation

First note that if $f(x)$ is any solution to Schröder's equation

$$f[\alpha(x)] = sf(x)$$

then so is any constant multiple of $f(x)$.

If the splinter of α behaves approximately as a geometric sequence, a solution can be found. The splinter $\alpha^n(x)$ is said to be *approximately geometric* if there exists a number $s \in (0, 1)$ such that

$$\lim_{n \rightarrow \infty} \frac{\alpha^n(x)}{s^n} \tag{3.30}$$

exists, is finite, and nonzero. In this case we say that the splinter has rate s .

On a domain of values x where the splinter of x is approximately geometric with rate s independent of x , a solution to Schröder's equation (3.24) is given by

$$f(x) = \lim_{n \rightarrow \infty} \frac{\alpha^n(x)}{s^n} \quad (3.31)$$

for the particular choice of s which is a witness to the property (3.30). This is easily checked by direct substitution into Schröder's equation. We have

$$\begin{aligned} f[\alpha(x)] &= \lim_{n \rightarrow \infty} \frac{\alpha^n[\alpha(x)]}{s^n} \\ &= s \lim_{n \rightarrow \infty} \frac{\alpha^{n+1}(x)}{s^{n+1}} \\ &= s f(x). \end{aligned}$$

The particular method given in (3.31) for generating a solution is known as the *Koenigs algorithm*. See Problem 6 for an example of a functional equation that can be solved using the Koenigs algorithm.

It is possible to generalize this solution. Provided that the limit

$$s = \lim_{n \rightarrow \infty} \frac{\alpha^{n+1}(x)}{\alpha^n(x)} \quad (3.32)$$

exists for all x on some domain, and is independent of x , then we can check that for each x_0 in that domain

$$f(x) = \lim_{n \rightarrow \infty} \frac{\alpha^n(x)}{\alpha^n(x_0)} \quad (3.33)$$

is a solution to Schröder's equation for each x_0 , provided the limit exists. We have

$$\begin{aligned} f[\alpha(x)] &= \lim_{n \rightarrow \infty} \frac{\alpha^{n+1}(x)}{\alpha^n(x_0)} \\ &= \lim_{n \rightarrow \infty} \left[\frac{\alpha^{n+1}(x)}{\alpha^n(x)} \cdot \frac{\alpha^n(x)}{\alpha^n(x_0)} \right] \\ &= \lim_{n \rightarrow \infty} \frac{\alpha^{n+1}(x)}{\alpha^n(x)} \cdot \lim_{n \rightarrow \infty} \frac{\alpha^n(x)}{\alpha^n(x_0)} \\ &= s f(x). \end{aligned}$$

This function is called the *principal solution* to Schröder's equation. Note that the choice of x_0 need not bother us unduly here. If we change the value of x_0 , the effect of this is to multiply the principal solution by a constant. So the principal solution can be said to be uniquely defined except for an arbitrary multiple.

3.5.2 The Lévy algorithm for Abel's equation

We consider the special case of Abel's equation where $a = 1$, so that

$$f[\alpha(x)] = f(x) + 1.$$

Note that if $f(x)$ is any solution to Abel's equation (3.26), then so is $f(x) + c$ for any constant c .

If the function $\alpha(x)$ is approximately geometric, it may be most appropriate to transform Abel's equation to Schröder's equation and to find the principal solution as outlined in the previous section. Alternatively, it may be that the function $\alpha(x)$ behaves approximately as does the function $x \mapsto x + a$. In this case, the most convenient formulation of the functional equation is as a form of Abel's equation.

However, suppose there exists an x_0 such that

$$\lim_{n \rightarrow \infty} \frac{\alpha^{n+1}(x_0) - \alpha^n(x_0)}{\alpha^{n+1}(x) - \alpha^n(x)} = 1 \quad (3.34)$$

for all x . Then if the limit

$$f(x) = \lim_{n \rightarrow \infty} \frac{\alpha^n(x) - \alpha^n(x_0)}{\alpha^{n+1}(x) - \alpha^n(x)} \quad (3.35)$$

exists, it is a solution to Abel's equation $f[\alpha(x)] = f(x) + 1$. To see this, write

$$\begin{aligned} f[\alpha(x)] &= \lim_{n \rightarrow \infty} \frac{\alpha^{n+1}(x) - \alpha^n(x_0)}{\alpha^{n+2}(x) - \alpha^{n+1}(x)} \\ &= \lim_{n \rightarrow \infty} \frac{\alpha^{n+1}(x) - \alpha^{n+1}(x_0)}{\alpha^{n+2}(x) - \alpha^{n+1}(x)} + \lim_{n \rightarrow \infty} \frac{\alpha^{n+1}(x_0) - \alpha^n(x_0)}{\alpha^{n+2}(x) - \alpha^{n+1}(x)} \\ &= f(x) + 1. \end{aligned}$$

The limit in the second term is seen to be one by replacing x with $\alpha(x)$ in (3.34). Equation (3.35) is known as the Lévy algorithm for Abel's equation.

3.5.3 An algorithm for Böttcher's equation

If $f(x)$ is any solution to Böttcher's equation (3.27), then so is the function $[f(x)]^q$ for any power q . Böttcher's equation may be the natural way to formulate a linearization equation when the function $\alpha(x)$ behaves approximately as a power function.

A solution to Böttcher's equation can be obtained if the limit

$$f(x) = \lim_{n \rightarrow \infty} [\alpha^n(x)]^{p^{-n}} \quad (3.36)$$

exists. This can be checked by plugging into (3.27).

3.5.4 Solving commutativity equations

It is easily seen that the functions $f_n(x) = \alpha^n(x)$ satisfy the commutativity equation

$$f[\alpha(x)] = \alpha[f(x)]$$

for $n = 0, 1, 2, \dots$. However, there may well be many other solutions. One way to generate a solution is through a solution to the corresponding Schröder, Abel, or Böttcher equation. For example, suppose that g satisfies Schröder's equation: $g[\alpha(x)] = s g(x)$. Furthermore, suppose that g is one-to-one with inverse g^{-1} . Then for any constant c the function

$$f(x) = g^{-1}[c g(x)] \quad (3.37)$$

must satisfy the commutativity equation. This follows easily using the fact that g^{-1} satisfies Poincaré's equation. We have

$$\begin{aligned} f[\alpha(x)] &= g^{-1}[c g(\alpha(x))] \\ &= g^{-1}[c s g(x)] \\ &= \alpha[g^{-1}(c g(x))] \\ &= \alpha[f(x)]. \end{aligned}$$

Similarly, if g satisfies Abel's equation $g[\alpha(x)] = g(x) + a$ then, for every constant c , the function

$$f(x) = g^{-1}[g(x) + c] \quad (3.38)$$

will satisfy the commutativity equation.

Finally, if $g[\alpha(x)] = [g(x)]^p$ then the function

$$f(x) = g^{-1}\{[g(x)]^c\} \quad (3.39)$$

will satisfy the commutativity equation for all c .

It can be seen that these arguments extend. In general, if f is a solution to the conjugacy equation

$$f[\alpha(x)] = \beta[f(x)]$$

then solutions to the commutativity equation $f[\alpha(x)] = \alpha[f(x)]$ can be found by extending the function β .

3.6 Generalizations of Abel's and Schröder's equations

In this section we consider two generalizations of Schröder's and Abel's equations which fall outside the class of conjugacy equations.

The equation

$$f[\alpha(x)] = f(x) + h(x), \quad (3.40)$$

where $h(x)$ is a given function, is a generalization of Abel's equation. As with Abel's equation, an arbitrary constant can be added to any solution to create another solution. It can be checked that for any x_0 , the function

$$f(x) = \sum_{n=0}^{\infty} \{h[\alpha^n(x_0)] - h[\alpha^n(x)]\} \quad (3.41)$$

is a solution to this equation provided that the infinite sum converges.

Example 3.6. For example, as a solution to the equation

$$f(2x) = f(x) + x^{-1}, \quad x > 0$$

this algorithm gives

$$f(x) = \sum_{n=0}^{\infty} (2^{-n} x_0^{-1} - 2^{-n} x^{-1}) \quad (3.42)$$

$$= c - 2x^{-1} \quad (3.43)$$

where $c = 2x_0^{-1}$ is arbitrary.

The equation

$$f[\alpha(x)] = g(x) f(x), \quad (3.44)$$

where $g(x)$ is a given function, is a generalization of Schröder's equation. It can be checked that for any x_0 ,

$$f(x) = \prod_{n=0}^{\infty} \left\{ \frac{g[\alpha^n(x_0)]}{g[\alpha^n(x)]} \right\} \quad (3.45)$$

is a solution to this equation provided that the infinite product converges.

Example 3.7. As an illustration of this formula, the algorithm solves the equation

$$f(x/2) = \cos x \cdot f(x) \quad (3.46)$$

by functions of the form

$$f(x) = \prod_{n=0}^{\infty} \left[\frac{\cos(2^{-n} x_0)}{\cos(2^{-n} x)} \right] \quad (3.47)$$

$$= d x \csc 2x \quad (3.48)$$

where the constant $d = x_0^{-1} \cos x_0 \sin x_0$ is arbitrary. Note that to obtain step (3.48) above, we need to use the formula

$$\prod_{n=1}^{\infty} \cos(2^{-n} x) = \frac{\sin x}{x}. \quad (3.49)$$

3.7 General properties of iterative roots

In this section, we return to Babbage's problem of finding the iterative roots of a function that was discussed in Chapter 1. Here we examine some general questions about iterative roots. We also return to Babbage's original problem, and see how the conjugacy equations of Abel and Schröder can be used directly to find iterative roots.

First, we recap the basic definitions. Let $h(x)$ be a given function. Any solution $f(x)$ to the equation

$$f[f(x)] = h(x) \quad (3.50)$$

is called an *iterative square root* of the function $h(x)$.

Define

$$f^1(x) = f(x) \quad \text{and} \quad f^{n+1}(x) = f[f^n(x)] \quad (3.51)$$

for $n \geq 1$. We shall call a solution to the equation

$$f^n(x) = h(x) \quad (3.52)$$

an *iterative n^{th} root* of $h(x)$. In the particular case where $h(x) = x$ for all x , we call equation (3.52) the Babbage equation. When f is a solution to the Babbage equation with $n = 2$, then we call f an involution.

In general, a solution to (3.52) may not exist, and even when it does exist, it usually is not unique. For example, the function $f(x) = -x$ is an involution. But so is the function $f(x) = a - x$ for every real number a . Indeed, involutions are plentiful, easily constructed, and all satisfy the functional equation $f[f(x)] = x$. In general, any function of the form

$$f(x) = g^{-1}[-g(x)] \quad (3.53)$$

will be an involution.

Fortunately, however, we are often in a position to say something about the properties of $f(x)$ even when we cannot determine it completely. For example, suppose that h maps the set of real numbers *onto* itself. That is, suppose that the range of the function h is all of \mathbf{R} . It follows from this that the range of f must also be all of \mathbf{R} (i.e., that f itself must be onto).

Another property that f "inherits" from h is the property of being one-to-one. For if h is one-to-one and $f(x) = f(y)$, it follows that $f^n(x) = f^n(y)$. So $h(x) = h(y)$, which implies that $x = y$. A case of particular importance is when f is assumed to be a continuous function. A one-to-one continuous function is necessarily either strictly increasing (i.e., $f(x) < f(y)$ whenever $x < y$) or strictly decreasing (i.e., $f(x) > f(y)$ whenever $x < y$). The proof of this fact is left as Problem 5 for the reader.

Putting these results together, we obtain the following result.

Proposition 3.8. *Let h be a one-to-one function.*

1. Suppose n is even and that there exists a pair of real numbers $x < y$ such that $h(x) > h(y)$. Then there is no continuous iterative n^{th} root of h .
2. Suppose that n is odd, and that f is a continuous iterative n^{th} root of h . Then h is continuous and strictly increasing or strictly decreasing as f is strictly increasing or strictly decreasing, respectively.

Proof. Suppose $f^n(x) = h(x)$ for some continuous function f . Then, as indicated above, the function f is strictly increasing or strictly decreasing. This implies that f^2 is strictly increasing, and therefore that $h = f^n$ is strictly increasing for all even values of n . However, this contradicts the statement in part 1 that $x < y$ and $h(x) > h(y)$. So in part 1, there can be no continuous solution f .

On the other hand, suppose n is odd, and f is monotone. Then, as $n - 1$ is an even number, f^{n-1} is strictly increasing (including $f^0(x) = x$). Therefore

$$h(x) = f^n(x) = f[f^{n-1}(x)] \quad (3.54)$$

is strictly increasing or strictly decreasing as f is respectively. \square

Using Proposition 3.8, we can now solve the following problem which appeared as Problem A2 on the 1979 William Lowell Putnam Competition.

Example 3.9. Establish necessary and sufficient conditions on the constant k for the existence of a continuous real-valued function $f(x)$ satisfying

$$f[f(x)] = kx^9$$

for all real x .

Solution. This can be written in the form $f^n(x) = h(x)$ with $n = 2$ and $h(x) = kx^9$. If $k < 0$, then $h(x) = kx^9$ is monotone strictly decreasing. By Proposition 3.8, there can be no such continuous function $f(x)$ because $n = 2$ is even. So we have $k \geq 0$. To prove that this is possible, it suffices to exhibit an $f(x)$ for any such k . It is easily seen that

$$f(x) = k^{1/4} x^3$$

works. \square

Can we actually construct iterative roots? In many cases yes. For example, suppose we can construct a one-to-one solution $g(x)$ to the Abel equation $g[h(x)] = g(x) + 1$ for a given function $h(x)$. Define

$$f(x) = g^{-1} \left[g(x) + \frac{1}{n} \right]. \quad (3.55)$$

Then it can be checked that $f(x)$ is an iterative n^{th} root of $h(x)$. Similarly, if we can solve the equation $g[h(x)] = s g(x)$ for some $s \neq 1$ with a one-to-one function g , then

$$f(x) = g^{-1} \left[s^{1/n} g(x) \right] \quad (3.56)$$

is an iterative n^{th} root of $h(x)$. The reader should recognize that what these formulas implicitly do is linearize the function $h(x)$ and apply an obvious n^{th} root to the linearized function.

Another way to tackle an iterated root equation $f^n(x) = h(x)$ is to apply the function f to both sides of the equation. We get

$$f[h(x)] = f[f^n(x)] = f^n[f(x)] = h[f(x)]. \quad (3.57)$$

This turns an iterated root equation into a commutativity equation. The solution to the iterated root equation will solve the commutativity equation but the converse usually is not true. However, the commutativity equation may be helpful in narrowing down the possible solutions, and it is often easy to analyze. The following example illustrates the idea.

Example 3.10. Find all integer-valued functions $f(n)$ taking values in the integers satisfying the equation $f[f(n)] = n + 1$.

Solution. Applying f to both sides, we obtain $f(n + 1) = f(n) + 1$. From this it is immediate that f can be written in the form $f(n) = n + m$ for some integer m . However, if this is the case, we have

$$f[f(n)] = n + 2m = n + 1$$

which has no solution in integers m . Therefore there is no integer-valued function $f(n)$ satisfying the equation. \square

Such a method can also be applied to the composition of two or more different functions. For example, we could have simultaneous equations of the form

$$f[g(x)] = h_1(x) \quad \text{and} \quad g[f(x)] = h_2(x), \quad (3.58)$$

where h_1 and h_2 are given, and f and g are unknown. We apply g to the first equation to obtain

$$g[h_1(x)] = g(f[g(x)]) = h_2[g(x)] \quad (3.59)$$

and apply f to the second to obtain

$$f[h_2(x)] = f(g[f(x)]) = h_1[f(x)]. \quad (3.60)$$

The result is a pair of simultaneous conjugacy equations to solve.

Let us close this section by considering the following problem. It was part of the same shortlisted problem at the 1997 International Mathematical Olympiad as was given in Example 3.1 earlier.

Example 3.11. Do there exist functions $f : \mathbf{R} \rightarrow \mathbf{R}$ and $g : \mathbf{R} \rightarrow \mathbf{R}$ such that

$$f[g(x)] = x^2 \quad \text{and} \quad g[f(x)] = x^3 \quad (3.61)$$

for all $x \in \mathbf{R}$?

Solution. No. Suppose f and g were solutions. As the function $h_2(x) = x^3$ is a one-to-one function, it follows that f is a one-to-one function. Thus $f(1)$, $f(-1)$, and $f(0)$ are three distinct numbers. On the other hand, one of the associated conjugacy equations states that $[f(x)]^2 = f(x^3)$. Inserting $x = -1, 0, +1$ into this equation we get that each of $f(-1)$, $f(0)$, and $f(1)$ equals its square. Thus each of these three numbers must equal 0 or 1. By the pigeonhole principle, two of $f(-1)$, $f(0)$, and $f(1)$ must be the same. This contradicts the fact that f is a one-to-one function. \square

3.8 Functional equations and nested radicals

Let us return to the nested radical proposed by Ramanujan, and described in Chapter 1. We wish to compute

$$\sqrt{1 + 2\sqrt{1 + 3\sqrt{1 + 4\sqrt{\dots}}}} \quad (3.62)$$

As previously, we generalize this problem to solving the nested radical

$$f(x) = \sqrt{1 + x\sqrt{1 + (x+1)\sqrt{\dots}}} \quad (3.63)$$

Squaring both sides, we see that $f(x)$ must satisfy the functional equation

$$[f(x)]^2 = 1 + x f(x+1), \quad (3.64)$$

where $f(x) \geq 0$. Setting $x = 0$, we determine that $f(0) = 1$. Unfortunately, we cannot get $f(1)$ by recursion.

This equation does not fit into any family of equations that we have managed to solve so far. How can we find a solution, and how can we determine that any solution we find is the correct solution to the original nested radical? We can try and guess at a solution. Let us look for a polynomial $f(x)$ which solves this. (See the section on polynomial equations in the next chapter.) If $f(x)$ were a polynomial of degree n then the left-hand side of our functional equation would be of degree $2n$ and the right-hand side of degree $n+1$. So our polynomial would have to be of degree one. Let $f(x) = ax + b$. We get

$$(ax + b)^2 = 1 + x(ax + a + b) \quad (3.65)$$

or



$$\sqrt{1 + 2 \cdot \sqrt{1 + 3 \cdot \sqrt{1 + 4 \cdot \sqrt{1 + 5 \cdot \sqrt{1 + \dots}}}}} = 3$$

$$1729 = 1^3 + 12^3 = 9^3 + 10^3$$

Fig. 3.2. There are many formulas associated with the legendary Indian mathematician Srinivasa Ramanujan (1887–1920). Here are two of the simpler ones, the first of these being our main interest here. The second formula is associated with one of the most famous and touching stories of the collaboration between G. H. Hardy and Ramanujan. While in England, Ramanujan became sick with the disease that prematurely ended his life. He was admitted to a nursing home in Putney called Colinette House. Hardy, visiting him there, and apparently at a loss for substantial conversation, remarked that the taxicab number was 1729. “A rather dull number,” Hardy added. “No, Hardy,” replied Ramanujan. “It is a very interesting number. It is the smallest number expressible as the sum of two cubes in two different ways.” Nowadays, the number 1729 is known as the Hardy–Ramanujan number.

$$a^2 x^2 + (2ab)x + b^2 = ax^2 + (a+b)x + 1. \quad (3.66)$$

If this is to be true for all x , then the corresponding coefficients on each side must agree. We are rather fortunate that the assignment $a = 1$ and $b = 1$ works!

So $f(x) = x + 1$ is a solution. But is it the correct solution? That is, is the equality

$$x + 1 = \sqrt{1 + x\sqrt{1 + (x+1)\sqrt{\dots}}} \quad (3.67)$$

correct?

We can prove that it is. To do this note that with $f(x)$ defined as in (3.63), we have for all $x \geq 1$,

$$f(x) \geq \sqrt{x\sqrt{x\sqrt{x\dots}}} \quad (3.68)$$

$$= x^{(1/2)+(1/4)+(1/8)+\dots} \quad (3.69)$$

$$= x \quad (3.70)$$

$$\geq (x+1)/2. \quad (3.71)$$

Bounding on the other side, we have

$$f(x) \leq \sqrt{(x+1)\sqrt{(x+2)\sqrt{(x+3)\dots}}} \quad (3.72)$$

$$\leq \sqrt{(x+1)\sqrt{2(x+1)\sqrt{3(x+1)\dots}}} \quad (3.73)$$

$$= (x+1)\sqrt{1\sqrt{2\sqrt{3}\dots}} \quad (3.74)$$

$$\leq (x+1)\sqrt{1\sqrt{2\sqrt{4}\dots}} \quad (3.75)$$

$$= 2(x+1). \quad (3.76)$$

Thus

$$\frac{x+1}{2} \leq f(x) \leq 2(x+1). \quad (3.77)$$

The following result takes us home.

Proposition 3.12. *Let $f(x)$ satisfy the functional equation*

$$[f(x)]^2 = 1 + xf(x+1) \quad (3.78)$$

and the inequalities

$$\frac{x+1}{2} \leq f(x) \leq 2(x+1) \quad (3.79)$$

for all $x \geq 1$. Then $f(x) = x + 1$.

Proof. Replacing x by $x + 1$ in equation (3.79), we get

$$\frac{x+2}{2} \leq f(x+1) \leq 2(x+2). \quad (3.80)$$

From (3.78) we have

$$\frac{1}{2} + x f(x+1) < [f(x)]^2 < 2 + x f(x+1). \quad (3.81)$$

Applying (3.80) to (3.81), we find that

$$\frac{1+x(x+2)}{2} < [f(x)]^2 < 2[1+x(x+2)]. \quad (3.82)$$

Taking the square root of both sides, we have

$$2^{-1/2}(1+x) < f(x) < 2^{1/2}(1+x). \quad (3.83)$$

Compare (3.83) with (3.79). Our bound has improved on each side! Applying this argument k times we get

$$2^{-1/2^k}(1+x) < f(x) < 2^{1/2^k}(1+x). \quad (3.84)$$

Letting $k \rightarrow \infty$ gives us $1+x \leq f(x) \leq 1+x$ which proves the result. \square

3.9 Problems

1. Find, with proof, all continuous functions $f : \mathbf{R} \rightarrow \mathbf{R}$ such that $f(2x) = f(x)$ for all real x .
2. Give an example of a continuous function $f : \mathbf{R} \rightarrow \mathbf{R}$ such that $f(x+2) = f(x) + 2$ for all real x , which is not a monotone function.
3. \star Find, with proof, all continuous functions $f : \mathbf{R} \rightarrow \mathbf{R}$ such that $f(x) = f(\cos x)$ for all real x .
4. The gamma function $\Gamma(x)$, $x > 0$, is defined by

$$\Gamma(x) = \lim_{n \rightarrow \infty} \frac{n! n^x}{x(x+1)(x+2) \cdots (x+n)}.$$

(Compare this formula with those in Section 3.5. You may assume that the limit exists.) Find a functional equation involving $\Gamma(x)$ and $\Gamma(x+1)$.

5. Let $f : \mathbf{R} \rightarrow \mathbf{R}$ be a continuous one-to-one function. Prove that f is either strictly increasing (i.e., for all $x < y$ we have $f(x) < f(y)$) or strictly decreasing (i.e., for all $x < y$ we have $f(x) > f(y)$).

6. Use the Koenigs algorithm to find a function f other than the trivial zero function which satisfies the equation

$$f\left(\frac{sx}{1+x}\right) = sf(x),$$

where $0 < s < 1$. Specify carefully the domain over which f is defined.

7. Find as many functions f as you can which satisfy the commutativity equation

$$f\left(\frac{x}{x+1}\right) = \frac{f(x)}{f(x)+1}.$$

Relate your solution to the solution to the corresponding Abel equation

$$g\left(\frac{x}{x+1}\right) = g(x) + 1.$$

8. Complete the proof in the solution to the IMO shortlisted problem of Example 3.1.

9. Let $f : \mathbf{R} \rightarrow \mathbf{R}$ be a continuous function satisfying the pair of functional inequalities

$$f(x+a) \leq f(x) \leq f(x+b).$$

- a) Let $a = \sqrt{2}$ and $b = 1$. Prove that f is a constant function.
 b) State general conditions on a and b used in the proof above which ensure that f is constant.

10. Consider a polynomial $f(x)$ with real coefficients, having the property

$$f[g(x)] = g[f(x)]$$

for every polynomial $g(x)$ with real coefficients. Determine the nature of $f(x)$.

11. Let $f(x)$ be a real-valued function defined for all real x except for $x = 0$ and $x = 1$ and satisfying the functional equation

$$f(x) + f[(x-1)/x] = 1 + x.$$

Find all functions $f(x)$ satisfying these conditions.

12. ★ Does there exist a function $f : \mathbf{R} \rightarrow \mathbf{R}$ which simultaneously satisfies the following three conditions?

- a) There exists a positive number M such that $-M \leq f(x) \leq M$ for all x .

- b) $f(1) = 1$.
 c) If $x \neq 0$ then

$$f\left(x + \frac{1}{x^2}\right) = f(x) + \left[f\left(\frac{1}{x}\right)\right]^2.$$

13. Find all complex-valued functions f of a complex variable such that

$$f(z) + z f(1 - z) = 1 + z$$

for all z .

14. \star Suppose that $f : \mathbf{R} \rightarrow \mathbf{R}$ has the following properties,

- a) $f(x) \leq 1$, for all real x ;
 b) $f\left(x + \frac{13}{42}\right) + f(x) = f\left(x + \frac{1}{6}\right) + f\left(x + \frac{1}{7}\right)$.

Prove that f is periodic. In other words, prove that there exists a nonzero real number t such that for every real number x , we have $f(x + t) = f(x)$.

15. \star Show that there exists a function $f : \mathbf{R} \rightarrow \mathbf{R}$ which is continuous, and satisfies the equation

$$f(x) + f(2x) + f(3x) = 0$$

for all x , but which is not the zero function.

16. \star Prove that the equation

$$f^n(x) = x^{-1}$$

defined on the nonzero real numbers has an infinite number of solutions for each positive integer $n \geq 2$.

17. Show that there is a unique function $f : \mathbf{R} \rightarrow \mathbf{R}$, such that $f(x)$ is strictly positive, satisfying for all real x ,

- a) $f(2x) = [f(x)]^2$;
 b) $f(-x) = [f(x)]^{-1}$, and
 c) $f(x) \geq 1 + x$.

Determine this function.

18. \star Prove or disprove: there exists a function $f(n)$, defined for all positive integers n , taking values in the positive integers, such that

$$f^{f(n)}(n) = n + 1.$$

In this expression, $f^{f(n)}(n)$ denotes that f is applied $f(n)$ times to the positive integer n .

19. Suppose that $f : \mathbf{R} \rightarrow \mathbf{R}$ is continuous and satisfies the equation

$$f(x) \cdot f[f(x)] = 1$$

for all real x . Assuming that $f(1000) = 999$, find the value of $f(500)$.

20. ★ Find all continuous functions $f : \mathbf{R} \rightarrow \mathbf{R}$ which satisfy the equation

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

for all real x and y .

21. ★ For the previous problem, give an example of a function which is not continuous, and also satisfies the functional equation

$$f(x + y) = f[x + f(y)].$$

22. Find all continuous functions $f : \mathbf{R} \rightarrow \mathbf{R}$ satisfying

$$f[x + f(x)] = f(x)$$

for all real x .

23. ★ Let u , f , and g be functions, defined for all real numbers x , such that

$$\frac{u(x+1) + u(x-1)}{2} = f(x)$$

and

$$\frac{u(x+4) + u(x-4)}{2} = g(x).$$

Determine $u(x)$ in terms of f and g .

24. ★ Find all continuous solutions to the simultaneous equations

$$f[f(x)] = x$$

and

$$f(x+1) = \frac{f(x)}{f(x)+1},$$

where $x > 0$.

25. ★ Suppose that $f : \mathbf{R}^+ \rightarrow \mathbf{R}^+$ is a strictly decreasing continuous function that satisfies the equation

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

for all positive x and y . Prove that $f[f(x)] = x$.

Miscellaneous methods for functional equations

4.1 Polynomial equations

In this chapter, we consider sundry functional equations in which there is some specialized character to the problem which requires techniques that are not easily exported to other settings. The restrictions that we consider here are of various types and are typically much stronger than the assumptions of previous chapters. Finally, at the end of the chapter, we consider a functional equation whose solution requires some group theory.

Let us begin by looking at polynomial equations. We assume that the functional solution to the equation is known to be a polynomial. Clearly, the methods of previous chapters are applicable in this context. However, it is worth considering what additional methods are available if we know that the solution is a polynomial. In what follows, we usually assume that the polynomials have real coefficients, and are functions of real variables. However, sometimes we have need to evaluate these polynomials for values in the complex plane even when the functional equation is stated for real numbers. So we consider polynomials of a real or complex variable with real or complex coefficients. First, it may be possible to use the equation to determine the degree of the polynomial. Suppose we let $\deg f$ denote the degree of a polynomial $f(x)$. Let $(fg)(x) = f(x)g(x)$ and $(f \circ g)(x) = f[g(x)]$. Then

$$\deg [fg] = \deg f + \deg g \quad (4.1)$$

$$\deg f \circ g = \deg f \cdot \deg g. \quad (4.2)$$

Using these formulas, any polynomial equation using the operations of multiplication and composition leads to an equation for the degree.

Example 4.1. Consider an equation such as

$$f(x-1)f(x+1) = f(f(x)), \quad (4.3)$$

where f is a polynomial with real coefficients. Let $d = \deg f$. The left-hand side of the equation is of degree $2d$, whereas the right-hand side is of degree

d^2 . Thus $2d = d^2$, implying that $d = 0$ or $d = 2$. A polynomial of zero degree is constant, and the only constant $f(x) = c$ satisfying this equation is $c = 0$ or $c = 1$.

The case $d = 2$ can now be considered. It is left to the reader to verify that

$$f(x) = (x - 1)^2$$

in this case.

The next method for solving polynomial equations involves the use of the following principle.

Proposition 4.2. *Suppose $f(x)$ is a polynomial which is periodic in the sense that there exists some $a \neq 0$ for which $f(x + a) = f(x)$, for all real x . Then $f(x) = c$ for all x .*

The proof is routine. Now let us consider how this can be applied to functional equations. Suppose a solution, say $f_0(x)$ has been found to a given polynomial equation. We must ask whether this is the only solution, or whether there are others. We can write a general solution in the form

$$f(x) = f_0(x) + g(x),$$

where $g(x)$ is some polynomial whose form is to be determined. It may be possible to show that $g(x)$ satisfies the conditions of Proposition 4.2, and thereby to conclude that all solutions of the equation have the form $f_0(x) + c$.

Example 4.3. To illustrate this, consider all polynomials satisfying the equation

$$f(x + 1) - f(x - 1) = 6x^2 + 2.$$

By inspection, we can observe that $f_0(x) = x^3$ is a solution. But is it the only one? Letting a general solution have the form $f(x) = x^3 + g(x)$, we plug into our equation to obtain

$$g(x + 1) - g(x - 1) = 0 \quad \text{for all } x.$$

So $g(x)$ is a polynomial satisfying the conditions of Proposition 4.2 with $a = 2$. Therefore $f(x) = x^3 + c$. It is immediately seen that any value of c will work.

Functional equations can also be used to provide direct information about the roots of the polynomial. To illustrate this, here is an example from the 1995 Romanian Winter Competition.

Example 4.4. Let us search for all polynomial solutions to the equation

$$f(x^2) = f(x)f(x - 1) \tag{4.4}$$

which are not identically zero. From this equation, we cannot immediately infer any restriction on the degree of $f(x)$, as the associated equation for

$d = \deg f$ is $2d = 2d$, which is trivially satisfied. Instead, it is helpful to inquire as to the nature of the roots of f in the complex plane \mathbf{C} . Suppose $f(z) = 0$. Then it follows from the equation that $f(z^2) = f(z)f(z-1) = 0$. So z^2 is also a root of f . So if \mathcal{Z} is the set of all roots of f in the complex plane, then \mathcal{Z} must be closed under the mapping $z \mapsto z^2$. However, if $|z| > 1$ or if $0 < |z| < 1$, then the sequence

$$z, z^2, z^4, z^8, \dots$$

will produce distinct points in the plane. Because \mathcal{Z} is finite, this is impossible. Thus the roots of f can only occur when $z = 0$ or $|z| = 1$. Similarly, if $z - 1$ is a zero, then so is z^2 . Therefore, \mathcal{Z} is closed under the mapping $z \mapsto (z + 1)^2$. The rest of the problem is left to the reader to solve. See Problem 10 at the end of Chapter 1, and the hints to that problem at the end of the book.

4.2 Power series methods

Less restrictive than assuming that an unknown function is a polynomial is the assumption that a function has a *power series representation*.

Consider the equation

$$f\left(\frac{x+x^2}{2}\right) = \frac{1}{2}f(x). \quad (4.5)$$

The reader will recall that this is an example of Schröder's equation, as discussed in Chapter 3, with $\alpha(x) = (x+x^2)/2$. Let $\alpha^n(x) = (x+x^2)/2$. If $0 < x < 1$, then $\alpha^n(x)$ goes to zero at a rate that is approximately geometric with rate $1/2$.¹ Applying the Koenigs algorithm, we obtain

$$f(x) = \lim_{n \rightarrow \infty} 2^n \alpha^n(x) \quad (4.6)$$

as a solution to the equation. This limit is hard to express using well-known functions. However, $\alpha^n(x)$ is a polynomial of degree $2n$. This suggests that we should look for a solution to $f(x)$ as a *power series*

$$f(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \dots \quad (4.7)$$

for $0 \leq x < 1$, provided this infinite series converges. Plugging in $x = 0$, in (4.5), we see that $a_0 = 0$. As any constant multiple of a solution to (4.5) is also a solution to (4.5), we can regard a_1 as arbitrary and set it equal to one. Let us insert our power series back into (4.5). This gives us

$$\left(\frac{x+x^2}{2}\right) + a_2 \left(\frac{x+x^2}{2}\right)^2 + a_3 \left(\frac{x+x^2}{2}\right)^3 + \dots = \frac{1}{2}x + \frac{a_2}{2}x^2 + \frac{a_3}{2}x^3 + \dots \quad (4.8)$$

¹ See Section 3.5.1 of the previous chapter for the definition of this.

Now we expand out the left-hand side, and group terms with corresponding powers of x together:

$$\frac{1}{2}x + \left(\frac{1}{2} + \frac{a_2}{4}\right)x^2 + \left(\frac{a_2}{2} + \frac{a_3}{8}\right)x^3 + \dots = \frac{1}{2}x + \frac{a_2}{2}x^2 + \frac{a_3}{2}x^3 + \dots \quad (4.9)$$

Next, we need to invoke a very useful property of power series: two power series in x are equal for all x if and only if their corresponding coefficients are equal. Set $b_m = 2^{-m} a_m$. We obtain

$$b_0 = 0 \quad (4.10)$$

$$b_1 = 1/2 \quad (4.11)$$

together with

$$(2^{2m-1} - 1)b_{2m} = \binom{2m-1}{1}b_{2m-1} + \binom{2m-2}{2}b_{2m-2} + \dots + \binom{m}{m}b_m \quad (4.12)$$

and

$$(2^{2m} - 1)b_{2m+1} = \binom{2m}{1}b_{2m} + \binom{2m-1}{2}b_{2m-1} + \dots + \binom{m+1}{m}b_{m+1} \quad (4.13)$$

for $m \geq 1$. From these recursive formulas, the power series for $f(x)$ is determined:

$$f(x) = x + 2x^2 + \frac{8x^3}{3} + \frac{24x^4}{7} + \frac{416x^5}{105} + \dots \quad (4.14)$$

A plot of this function is shown in Figure 4.1.

4.3 Equations involving arithmetic functions

In the previous two sections we have considered functional equations with a restriction on the class of functions. In this section, we shall consider a completely different kind of restriction which radically changes the nature of the solution.

Consider the following from the Cauchy family of equations,

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

where x and y are strictly positive. In Chapter 2, we saw that the positive continuous solutions to this equation (if f does not vanish everywhere) are power functions of the form $f(x) = x^k$ for some real number k . Now suppose we restrict the domain of the function f to the positive integers. It is natural to make a shift in notation so that

$$f(nm) = f(n)f(m), \quad n, m \in \{1, 2, 3, \dots\}. \quad (4.15)$$

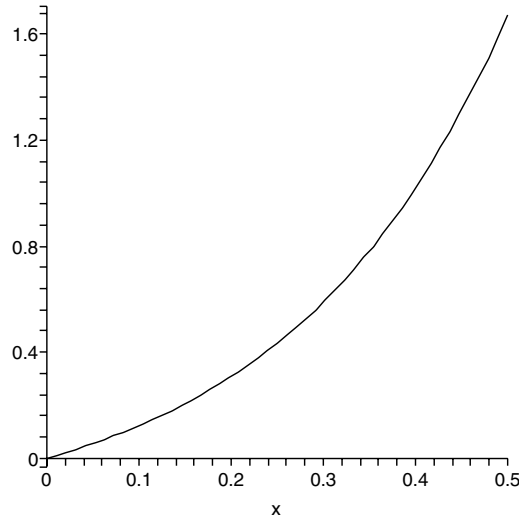


Fig. 4.1. A power series solution to a difficult Schröder equation

Such a function is said to be *completely multiplicative* if this equation is true for all n and m . Using induction, it is easy to see that any completely multiplicative function also satisfies

$$f(n_1 n_2 \cdots n_i) = f(n_1) f(n_2) \cdots f(n_i)$$

for all positive integers i .

What are the solutions to this equation when the domain of the function f is restricted in this way to the positive integers? It is clear that power functions $f(n) = n^k$ still satisfy equation (4.15). Are there other functions? We can no longer appeal to the same argument which we used in Chapter 2. Letting $n = m = 1$, we see that $f(1) = f(1)^2$ as before. So $f(1) = 0$ or $f(1) = 1$. As before, if $f(1) = 0$ then

$$\begin{aligned} f(n) &= f(n \cdot 1) \\ &= f(n) f(1) \\ &= 0 \end{aligned}$$

for all n . Thus f vanishes everywhere.

Let us restrict our attention to the case where $f(1) = 1$. Suppose the integer $n > 1$ has prime factorisation

$$n = p_1^{d_1} p_2^{d_2} \cdots p_j^{d_j},$$

where p_1, \dots, p_j are primes and d_1, \dots, d_j are positive integers. Applying (4.15) and induction, we see that

$$f(n) = f(p_1)^{d_1} f(p_2)^{d_2} \cdots f(p_j)^{d_j}. \quad (4.16)$$

Unlike Cauchy's multiplicative functional equation, it turns out that this is all we can prove, because any function that satisfies this property is completely multiplicative. Among all such functions the power functions are simply those for which $f(p) = p^k$ for every prime p .

The study of completely multiplicative functions is usually considered to be a branch of number theory rather than of functional equations. However, this is largely a matter of definition. Several common functions that are used in number theory are completely multiplicative. These include the following.

1. For any prime number p , the *Legendre symbol* is defined as

$$\left(\frac{n}{p}\right) = \begin{cases} 1 & \text{if } n \text{ is a square mod } p \\ 0 & \text{if } p \text{ divides } n \\ -1 & \text{if } n \text{ is not a square mod } p. \end{cases} \quad (4.17)$$

We can also write

$$\left(\frac{n}{p}\right) = n^{(p-1)/2} \pmod{p}.$$

Then

$$\left(\frac{nm}{p}\right) = \left(\frac{n}{p}\right) \left(\frac{m}{p}\right).$$

2. The *Liouville function* is defined to be

$$\lambda(n) = (-1)^{\Omega(n)} \quad (4.18)$$

where $\Omega(n)$, known as *Big Omega*, is the number of prime factors of n , with multiplicity counted. That is,

$$\lambda(p_1^{d_1} p_2^{d_2} \cdots p_j^{d_j}) = (-1)^{d_1 + d_2 + \cdots + d_j}.$$

Now let us turn to the role of multiplicative functions in mathematics competitions. Problem 6 on the 1998 International Mathematical Olympiad² held in Taiwan was

Example 4.5. Consider all functions f from the set \mathbf{N} of all positive integers into itself satisfying

$$f(n^2 f(m)) = m (f(n))^2, \quad (4.19)$$

for all m and n in \mathbf{N} . Determine the least possible value of $f(1998)$.

² On each IMO, problem 6 is traditionally thought of as the "gold medal" problem that is designed to separate gold medal winners from the bronze and silver medals.

Solution. At first sight this is a functional equation that does not seem to have much to do with completely multiplicative functions. Our first task will be to elicit the multiplicative property. Let $f(1) = a$. Setting $m = 1$, we obtain

$$f(an^2) = [f(n)]^2. \quad (4.20)$$

Then setting $n = 1$, we get

$$f(f(m)) = a^2 m. \quad (4.21)$$

So

$$\begin{aligned} [f(m)f(n)]^2 &= [f(m)]^2 f(an^2) && \text{by (4.20)} \\ &= f(m^2 f(f(an^2))) && \text{by (4.19)} \\ &= f(m^2 a^2 an^2) && \text{by (4.21)} \\ &= f(a(amn)^2) \\ &= (f(amn))^2 && \text{by (4.20) again.} \end{aligned}$$

Taking the square root of both sides, we conclude that

$$f(amn) = f(m)f(n). \quad (4.22)$$

Setting $n = 1$ in (4.22), we see that

$$\begin{aligned} f(am) &= f(m)f(1) \\ &= af(m). \end{aligned} \quad (4.23)$$

Putting (4.23) and (4.22) together we have

$$af(mn) = f(m)f(n) \quad (4.24)$$

for all positive integers m, n .

The next step in the proof is to show that a divides $f(n)$ for every positive integer n . Suppose p is any prime number, and suppose that p^r and p^s are the highest powers of p which divide the numbers a and $f(n)$, respectively. Setting $m = n$ in (4.24) we get $af(n^2) = [f(n)]^2$. Continuing in this way by induction, we have

$$a^{k-1} f(n^k) = [f(n)]^k, \quad k = 1, 2, \dots \quad (4.25)$$

The highest power of p that divides $[f(n)]^k$ is p^{ks} . Similarly, the highest power of p dividing a^{k-1} is $p^{(k-1)r}$. So equation (4.25) implies that $(k-1)r \leq ks$ for every positive integer k . This can only be true if $r \leq s$. Now p is an arbitrary prime. So a must divide $f(n)$.

Set $g(n) = f(n)/a$. Because a divides $f(n)$, the function g takes values among the positive integers. In terms of g , equation (4.24) becomes

$$g(mn) = g(m)g(n) \quad (4.26)$$

which says that g is a completely multiplicative function. Note also that from (4.20) with $n = 1$, we get $f(a) = a^2$. In terms of g , this becomes

$$g(a) = a. \quad (4.27)$$

Therefore,

$$\begin{aligned} g(g(n)) &= g(a)g(g(n))/a \\ &= g(ag(n))/a \\ &= g(f(n))/a \\ &= \frac{f(f(n))}{a^2} \\ &= \frac{a^2 n}{a^2} \\ &= n. \end{aligned} \quad (4.28)$$

Thus g is a completely multiplicative involution on the positive integers.

Now let us return to our original functional equation given in the statement of the problem. In fact, the new function g also must satisfy this equation. To see this, we observe that

$$\begin{aligned} g(n^2 g(m)) &= g(n^2)g(g(m)) \\ &= [g(n)]^2 g(g(m)) \\ &= m [g(n)]^2. \end{aligned}$$

We need to prove one final property of g . Suppose p is any prime. Let $g(p) = tu$, where t and u are positive integers. Then

$$g(g(p)) = g(tu) = g(t)g(u).$$

Because $g(g(p)) = p$, it follows that at least one of $g(t)$ and $g(u)$ must be equal to one. Suppose, without loss of generality, that $g(t) = 1$. Then

$$t = g(g(t)) = g(1) = 1.$$

It follows from this argument that $g(p)$ is a prime for every prime p .

Recall that we are searching for the smallest possible value of $f(1998)$ where f is chosen from the class of functions satisfying this functional equation. However, for any such f , we have $g(1998) \leq f(1998)$, because a is a positive integer. Therefore it suffices to restrict attention to functions g which are completely multiplicative involutions as shown. Because g is completely multiplicative,

$$g(1998) = g(2 \cdot 3^3 \cdot 37) = g(2) \cdot [g(3)]^3 \cdot g(37)$$

where $g(2)$, $g(3)$, and $g(37)$ are prime. If we wish to minimise this, we would naturally assign the smallest primes to these values, with the smallest prime, namely 2 assigned to $g(3)$. For example, we can set $g(2) = 3$, $g(3) = 2$, and $g(37) = 5$. Therefore, the smallest possible value of $g(1998)$ is $2^3 \cdot 3 \cdot 5 = 120$.

This beautiful question was a worthy gold medal problem. The solution brings together the ideas of number theory, multiplicative functions, involutions, and algebra. There are many nontrivial steps. However, none of these steps requires the use of heavy mathematical tools.

4.4 An equation using special groups

We close this survey of functional equations by considering a problem where the solution needs some group theory. Consider the following problem, which was shortlisted for consideration at the International Mathematical Olympiad in 2001 in Washington, DC.

Example 4.6. Find all functions $f : \mathbf{R} \rightarrow \mathbf{R}$, satisfying

$$f(xy)(f(x) - f(y)) = (x - y)f(x)f(y) \quad (4.29)$$

for all x and y .

Solution. Set $y = 1$ in equation (4.29) to obtain

$$[f(x)]^2 = x f(x) f(1). \quad (4.30)$$

If $f(1) = 0$, then equation (4.30) implies that $f(x) = 0$ for all real x . So we restrict attention to the case where $f(1) = a \neq 0$.

Now, we could be forgiven for concluding that from here on, the solution is obvious. Dividing by $f(x)$ on both sides of equation (4.30), we would seem to get $f(x) = ax$. Because this solution works in the original equation (4.29), we might think we are finished.

However, the situation is not this simple, because we have assumed that $f(x) \neq 0$ when dividing both sides by $f(x)$. So we need to examine this more closely. Note first that setting $x = 0$ in (4.30) gives $f(0) = 0$. What can we say about the set of values x where $f(x) \neq 0$? Let

$$G = \{x \in \mathbf{R} : f(x) \neq 0\}.$$

We have already got

- $0 \notin G$, and $1 \in G$.

Equation (4.30) implies that

- $x^2 \in G$ for all $x \in G$.

Furthermore equation (4.29) implies that

- if x and y are elements of G , and $x \neq y$, then $xy \in G$.

Putting these last two statements together, we conclude that G is closed under multiplication.

Next, suppose $f(x) \neq 0$ and $f(y) = 0$. Once again appealing to equation (4.29), we find that $f(xy) = 0$. So we also conclude that

- if $x \in G$ and $y \notin G$, then $xy \notin G$.

A consequence of this is that

- if $x \in G$, then $x^{-1} \in G$.

For if x^{-1} were not in G , then $1 = xx^{-1}$ would also not be in G , which is a contradiction.

Let us summarise what we have discovered about G .

1. The set G is a collection of nonzero real numbers that is closed under multiplication xy and reciprocation $x \mapsto x^{-1}$ and that contains the multiplicative identity 1.

Now we are in a position to characterize all solutions to equation (4.29). Suppose $f(x)$ is any such solution which does not vanish everywhere. Then there exists a set G satisfying the properties listed above and a real number $a \neq 0$ such that

$$f(x) = \begin{cases} ax & \text{for } x \in G, \\ 0 & \text{for } x \notin G. \end{cases} \quad (4.31)$$

To see that this is true, note first that any such function satisfies the equation. In addition the set of values where f does not vanish has the properties given above. On this set G we may divide both sides of equation (4.30) by $f(x)$. From this we obtain the expression $f(x) = ax$, where $a = f(1)$.

A set G which is closed under multiplication, has the multiplicative identity 1 and the multiplicative inverse x^{-1} of every $x \in G$ is said to be a *multiplicative group*. In this particular case, the group operation is commutative, being the usual multiplication that we do on real numbers. Thus G is a *commutative group* or an *abelian group*. Obviously, the set of nonzero real numbers form such a group. However, there are many other choices for G including the rational numbers \mathbf{Q} . The beauty of this problem is that it illustrates the richness of the idea of a group. Although the problem itself has no reference to any special multiplicative group of real numbers, the idea of a group is fundamental to the solution.

After considerable discussion, the jury of the 2001 IMO in Washington decided not to use this lovely problem as one of the six problems for the contest. Some jury members, this author included, were concerned that students might solve the problem and fail to recognise that the solution is complete

in the form given. The concern was understandable because our solution is rather abstract. Perhaps I, and others on the jury, should have had more faith in the insights of some very talented students. We shall never know for sure.

4.5 Problems

1. Find all polynomials $f(x)$, $g(x)$, and $h(x)$ such that

$$|f(x)| - |g(x)| + h(x) = \begin{cases} -1 & \text{for } x < -1, \\ 3x + 2 & \text{for } -1 \leq x \leq 0, \\ -2x + 2 & \text{for } x > 0. \end{cases}$$

2. In each of the following cases, find all polynomials with real coefficients which satisfy the equation.

a) $f(x^2 + x) = f(x)f(x + 1)$.
 b) $f(g(x)) = f(x)g(x)$.

3. ★ Find all polynomials with real coefficients satisfying

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

for all triples x , y , and z such that $xy + yz + zx = 0$.

4. Let $f : \mathbf{N} \rightarrow \mathbf{N}$ be a function from the natural numbers to themselves, such that $f(n + 1) > f(n)$. Suppose that

$$f(f(n)) = 3n.$$

Evaluate $f(2001)$.

5. ★ In each of the following cases, find all functions $f : \mathbf{R} \rightarrow \mathbf{R}$ satisfying the equation.

a) $f(xf(x) + f(y)) = (f(x))^2 + y$,
 b) $f(f(x) + y) = f(x^2 - y) + 4f(x)y$.

6. ★ Find all functions $g : \mathbf{R} \rightarrow \mathbf{R}$ such that there exists a strictly increasing or a strictly decreasing function $f : \mathbf{R} \rightarrow \mathbf{R}$ such that

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

7. ★ Find all nondecreasing functions $f : \mathbf{R} \rightarrow \mathbf{R}$ such that $f(0) = 0$, $f(1) = 1$, and

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

for all x and y such that $x < 1$ and $y > 1$.

8. ★ Does there exist a function $s : \mathbf{Q} \rightarrow \{-1, 1\}$ such that, if x and y are *distinct* rational numbers satisfying $xy = 1$ or $x + y \in \{0, 1\}$, then $s(x)s(y) = -1$?

Some closing heuristics

We have discussed many functional equations in the previous chapters. However, the next functional equation you encounter may be completely different from any of these and may require radically different methods to solve. In that case, some real ingenuity may be necessary.

Nevertheless, there are a number of rules of thumb that can help you get started on a new functional equation. I offer the following suggestions in the hope that they will be of use to you.

Good luck!

1. Get to know the standard functional relationships satisfied by the usual suspects: linear functions, trigonometric functions, logarithms, powers, and the like. Don't forget reciprocals and functions of the form $1/(x+a)$, and so on.
2. Try plugging in a few simple functions such as linear functions or polynomials and solving for the coefficients.
3. Try solving for a few values such as $f(0)$, $f(1)$, $f(-1)$, and so on. If you can find enough of these you may be able to recognize the form of the function.
4. If the equation involves two variables, say x and y , try substitutions for y in terms of x (or vice versa) which
 - reduce the complexity of the expression,
 - make a term on the left-hand side equal to a term on the right-hand side so that the terms cancel.
5. Look for functional equations that are symmetric in x and y , say, on one side and asymmetric on the other side. You can make a new functional equation by switching the roles of x and y .
6. Try to determine from the functional equation whether the function is strictly positive, strictly negative, or has roots for particular values of x . This may give insight into the function. Check to see what happens to $f(x)$ if x is very large or very small. Check to see whether the functional

equation tells you if $f(x)$ is a monotone function. Check to see whether the function can be proved to be unbounded.

7. Watch out for the opportunity to linearize the equation.
8. If the equation involves $f[\alpha(x)]$, investigate the splinter $\alpha^n(x)$. If $\alpha^m(x) = x$ for some m , it may be possible to solve for $f(x)$ by applying the equation m times.
9. Watch out for periodic functions, that is, those for which $f(x+a) = f(x)$ for some positive a . A statement of periodicity is itself a functional equation, and can sometimes be derived from the given functional equation.

Appendix: Hamel bases

It would have been appropriate to mention Georg Hamel and his connection with the theory of functional equations in our historical survey of Chapter 1. However, the work of Hamel is technical, and best studied after a thorough analysis of Cauchy's contributions such as in Chapter 2. I have seen fit to place Hamel's work in this appendix because his contribution to the subject depends upon a highly speculative axiom in mathematics, namely the Axiom of Choice, which is not accepted by all mathematicians.

Georg Hamel was born in 1877 in Düren, Germany, and died in 1954 in Landshut, Germany. In 1897, Hamel went to the University of Berlin, where he was taught by Hermann Schwarz and Max Planck, to name two. Subsequently, he went to Göttingen University, where he studied with Felix Klein and David Hilbert. He was awarded a doctorate under the supervision of Hilbert in 1901. The subject of his dissertation was Hilbert's fourth problem. In 1905, he went to Brno. It was during the period of his work in Brno that his 1905 paper on Hamel bases was written. In this appendix, we briefly consider the concept of the Hamel basis for the real numbers and its connection to the theory of functional equations.

In Chapter 2, we saw that the only continuous solutions to the equation

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

are functions of the form $f(x) = ax$ for any $a \in \mathbf{R}$. It was Georg Hamel [1905] who managed to show that there exist many solutions to this functional equation which are not continuous, provided that one accepts the Axiom of Choice. Hamel's result may be stated as follows.

Proposition 6.1. *There exists an uncountable set \mathcal{U} of real numbers such that every real number x has a unique representation*

$$x = \sum_{i=1}^n q_i u_i, \tag{6.1}$$

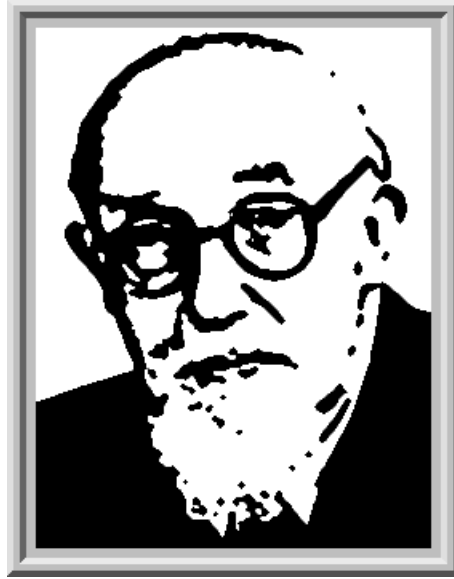


Fig. 6.1. Georg Hamel (1877–1954)

where n is a positive integer, q_1, \dots, q_n are rational numbers, and u_1, \dots, u_n are distinct elements of the set \mathcal{U} .

The representation of each real number x in terms of this basis is unique. This may be said to be the crucial point of the proposition. More precisely, suppose we can write x in two ways, namely

$$x = \sum_{i=1}^n q_i u_i, \quad x = \sum_{j=1}^m r_j v_j,$$

where $q_1, \dots, q_n, r_1, \dots, r_m$ are rational numbers and $u_1, \dots, u_n, v_1, \dots, v_m \in \mathcal{U}$. The ordering of terms is arbitrary. So let us assume that

$$u_1 < u_2 < \dots < u_n, \quad v_1 < v_2 < \dots < v_m.$$

Then the uniqueness of the representation in (6.1) implies that $n = m$, that $q_i = r_i$, and that $u_i = v_i$, for all $i = 1, 2, \dots, n$.

For the purposes of the work which follows, it is easiest to write the representation of each x as

$$x = \sum_{u \in \mathcal{U}} q_u u,$$

where the sum is over all elements of \mathcal{U} , but only finitely many coefficients q_u are nonzero.

The set \mathcal{U} is called a *Hamel basis* for \mathbf{R} . Hamel's proposition asserts that the set of real numbers, regarded as a vector space with the scalar field of rational numbers, has an uncountable basis. To see how this result can be useful in determining additional solutions to Cauchy's equation, let us consider some arbitrary function

$$\alpha : \mathcal{U} \rightarrow \mathbf{R}.$$

We define a function $f : \mathbf{R} \rightarrow \mathbf{R}$ as follows. For each $x \in \mathbf{R}$ we can uniquely write

$$x = \sum_{u \in \mathcal{U}} q_u u,$$

by Hamel's representation. Define

$$f(x) = \sum_{u \in \mathcal{U}} q_u \alpha(u).$$

Then the following holds.

Proposition. The functional f satisfies

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

for all real x and y .

Proof. Let

$$x = \sum_{u \in \mathcal{U}} q_u u$$

and

$$y = \sum_{u \in \mathcal{U}} r_u u$$

be the two unique representations of x and y as a linear combination of elements from the Hamel basis. Then

$$x + y = \sum_{u \in \mathcal{U}} (q_u + r_u) u$$

is the unique representation of $x + y$. Thus

$$\begin{aligned} f(x + y) &= \sum_{u \in \mathcal{U}} (q_u + r_u) \alpha(u) \\ &= \sum_{u \in \mathcal{U}} q_u \alpha(u) + \sum_{u \in \mathcal{U}} r_u \alpha(u) \\ &= f(x) + f(y). \end{aligned}$$

Note that the second equation holds because all sums are finite, and thus we may rearrange terms at will. \square

Because all functions $f(x)$ of this form will satisfy Cauchy's equation, it is natural to check whether our "tame" solutions (*e.g.*, continuous, monotone or locally bounded solutions) are of this form. They are, because for any $a \in \mathbf{R}$, we can choose the function $\alpha(u) = a u$. Then

$$\begin{aligned} f(x) &= \sum_{u \in \mathcal{U}} q_u \alpha(u) \\ &= \sum_{u \in \mathcal{U}} q_u (a u) \\ &= a \left(\sum_{u \in \mathcal{U}} q_u u \right) \\ &= f(x). \end{aligned}$$

Do we have other solutions? It is not hard to see that we must have other solutions. There are clearly other ways to define the function $\alpha(u)$. So we need only to show that each distinct definition of $\alpha(u)$ leads to a distinct choice of $f(x)$. Suppose we have two functions $\alpha_1(u)$ and $\alpha_2(u)$ which both determine the same function $f(x)$. Then for any $x = \sum q_u u$, we must have

$$\sum_{u \in \mathcal{U}} q_u \alpha_1(u) = \sum_{u \in \mathcal{U}} q_u \alpha_2(u).$$

In particular, we set $x = u_0$, say for any $u_0 \in \mathcal{U}$. Then from the uniqueness of the Hamel basis representation, we see that $q_{u_0} = 1$ and $q_u = 0$ for $u = u_0$ or $u \neq u_0$, respectively. So

$$\begin{aligned} \alpha_1(u_0) &= \sum_{u \in \mathcal{U}} q_u \alpha_1(u) \\ &= f(u_0) \\ &= \sum_{u \in \mathcal{U}} q_u \alpha_2(u) \\ &= \alpha_2(u_0). \end{aligned}$$

So just as there are many other functions $\alpha(u)$, there are many other solutions to Cauchy's functional equations other than the tame ones $f(x) = a x$.

What do these solutions look like? As we have seen from Chapter 2, any solution to Cauchy's equation which is bounded below (or bounded above) on any interval of positive length must be tame. So we must conclude that any "wild" solution to the equation is unbounded on every such interval. Therefore, it cannot be continuous at any point.

In attempting to visualise such a function, we find that the situation is much worse even than this. It requires the Axiom of Choice, a notoriously nonconstructive existence axiom, to demonstrate that such functions exist. This means that we have little hope of being able to construct an algorithm that would let a computer calculate such a function.

Hints and partial solutions to problems

7.1 A warning to the reader

It is always tempting after working on a problem for a while to turn to the answers at the back. However, the reader is warned not to turn too quickly to this section if a problem is a hard nut to crack. Often, a problem which seems to be almost impossible on a first run through will yield its secrets after you put it aside for a little while.

Although this chapter is said to contain hints and partial solutions, the reader is also warned that some of the partial solutions are just short of being complete proofs. So if you only want a hint, do not read too much of what follows.

7.2 Hints for Chapter 1

1. Note that there are multiple solutions: if $f(x)$ and $g(y)$ are solutions, then so are $f(x) + a$ and $g(y) - a$. Write

$$\log(1 + x + xy + y) = \log(1 + x) + \log(1 + y).$$

Then set $x = 0$ to determine $g(y)$ and $y = 0$ to determine $f(x)$.

2. Suppose that z and w are any two positive numbers with $z^2 > 4w$. Then the simultaneous equations

$$\begin{aligned}x + y &= z \\xy &= w\end{aligned}$$

have solutions in positive x and y . Thus $f(z) = f(w)$. To remove the restriction that $z^2 > 4w$, prove by induction that $f(x + n) = f(x)$ for all positive x and for all positive integers n .

3. Setting x and y to zero, we observe that $f(0)^2 = f(0)^4$. So $f(0) = 0$ or $f(0) = 1$. Show that if the former, then $f(x) = 0$ for all x . The conclusion will then follow. Show that if the latter case is true, then $f(x) = f(-x)$.
4. Set $a = x$ and show that $f(x) = b + x^2$, where b is arbitrary.
5. a) Try $f(x) = b^x$.
 b) Compare this equation with d'Alembert's equation.
 c) The function $f(x) = 0$ works. So does $f(x) = x^{-1}$. Can you extend this solution to a more general family?
 d) Once again, $f(x) = 0$ and works. Another solution is $f(x) = 1$. For less trivial examples, try a polynomial. What happens when $f(x) = x^2$? Note that if $f(x)$ is a polynomial then you can figure out its degree quickly by equating the degrees of each side of the equation.
6. Setting $x = y$, we see that

$$g(2x) + g(0) = 2[g(x)]^2.$$

Now apply limits to each side to obtain $g(0) = 0$. Plug in the values $y = 0$ in the original equation and use the fact that $g(0) = 0$.

7. a) Plug in $x = 0$ and $y = 0$ and solve.
 b) Square both sides of equations (1.45) and (1.46), and then add.
 c) It is necessary to show that the only bounded solution to Cauchy's exponential equation such that $h(0) > 0$ is the function $h(x) = 1$ for all x . Use $h(0) = h(x)h(-x)$ to show that $h(x) > 0$ for all x . Now suppose that there is some x_0 such that $h(x_0) \neq 1$. Use this fact to prove a contradiction by showing that $h(x)$ is not bounded.
 d) Differentiate Cauchy's exponential equation to y , and set $y = 0$ to derive the equation $h'(x) = h(x)h'(0)$. In addition, prove that $h'(0) = 0$. The result will follow.
8. First note that f must be a one-to-one function. For if $f(n) = f(m)$ it will follow by (a) that $n = m$. So we can write (a) as $f(n) = f^{-1}(n)$. From this, we get in (b)

$$\begin{aligned} f(n+2) + 2 &= f^{-1}(n) \\ &= f(n). \end{aligned}$$

Equivalently, we have $f(n+2) = f(n) - 2$. The result now follows easily from an argument by induction.

9. Suppose we can show that f is a strictly increasing function. That is, suppose that $f(n) < f(m)$ if and only if $n < m$. Then in particular, we would have $n + 1 > f(n)$ for all n using the property of the function given in the question. However, a strictly increasing function from the positive integers into the positive integers has the property that $f(n) \geq n$. It would follow that $n + 1 > f(n) \geq n$, and therefore that $f(n) = n$, for all n . To demonstrate that f is strictly increasing, first note that $f(1) < f(n)$ for all $n > 1$. To prove this, we need to recognize that $f(1)$ is the only value in the range of f for which we cannot construct a smaller value: $f(n) = f[(n-1) + 1] > f[f(n-1)]$ for all $n > 1$. Next, define the function $f_1(n) = f(n+1) - 1$ for $n = 1, 2, \dots$. Once again, prove that $f_1(1) < f_1(n)$ for all $n > 1$. Continue this argument.

10. a) The only constant solutions are $q(x) = 0$ and $q(x) = 1$ for all x . So we restrict consideration to the case where $q(x)$ is nonconstant. Note that for $x = 0$ and $x = -1$ the equation becomes

$$q(0) = q(0)q(1), \quad q(1) = q(-1)q(0).$$

Thus 0 is a root if and only if 1 is a root. Investigating functions of the form $q(x) = x^k(x-1)^k$ we observe that these are solutions. Are there any more?

Let us write

$$q(x) = ax^j(x-1)^k r(x)$$

where neither 0 nor 1 is a root of $r(x)$, and the leading coefficient of $r(x)$ is 1. Plugging into the equation, we see that

$$ax^{2j}(x-1)^k(x+1)^k r(x^2) = a^2 x^{j+k}(x+1)^j(x-1)^k r(x)r(x+1).$$

So $a = 1$. Because x cannot divide any of $r(x^2)$, $r(x)$, and $r(x+1)$, we conclude that $j = k$. Cancelling the appropriate factors on each side we see that $r(x^2) = r(x)r(x+1)$ where 0 and 1 are not roots of $r(x)$. We show that $r(x) \equiv 1$.

Suppose not. So $r(x)$ has a complex root, say $a \neq 0$. Because $r(a) = 0$, we have

$$r(a^2) = r(a)r(a+1) = 0.$$

So a^2 is also a complex root of $r(x)$. Continuing in this way, we conclude that a, a^2, a^4 , and so on are all roots of $r(x)$. But a polynomial can only have finitely many roots. If $|a| \neq 1$ then this sequence of roots would be distinct, which is impossible. So $|a| = 1$.

Similarly, if a is any such root, then

$$r((a-1)^2) = r(a-1)r(a) = 0.$$

So $(a-1)^2$ is also a root. It follows that $|a-1| = 1$. There are two points in the complex plane satisfying $|a| = 1$ and $|a-1| = 1$. These are

$$a = \frac{1}{2} \pm \frac{\sqrt{3}}{2} i.$$

However, we observed earlier that if a is a root then so is a^2 . But this property is not satisfied by the only two possibilities for roots listed above.

- b) In this case, the situation is very similar to the previous case up to a point. Note first that $p(1^2) = p(1)p(0)$. So $p(1) = 0$ or $p(0) = 1$. Suppose that $p(1) = 0$. Then $p(2^2) = p(2)p(1) = 0$. Continuing in this way, we deduce that $p(n) = 0$ for infinitely many positive integers n . As a nonconstant polynomial, $p(x)$ can have only finitely many roots. Thus we have a contradiction, and can conclude that $p(0) = 1$. Suppose $p(a) = 0$ for some complex number a . Then $p(a^2) = p(a^4) = \dots = 0$. However, a polynomial can only have a finite number of roots. So a must be a root of unity. Therefore, $|a| = 1$. Similarly, the reader can prove that $|a + 1| = 1$. Use this to show that $a^2 + a + 1 = 0$, and therefore that

$$a = \frac{-1}{2} \pm \frac{\sqrt{3}}{2} i.$$

Now there is an important difference between this solution and that we found above. In this case, we have two cube roots of unity, whereas the two values found above were not.

Now the polynomial $p(x)$ has real coefficients. Therefore, if $x - a$ is a factor of $p(x)$ then so is $x - \bar{a}$. So we can conclude that $(x - a)(x - \bar{a}) = x^2 + x + 1$ is a factor. Indeed, any nonconstant factor of $p(x)$ with real coefficients must be of the form $(x^2 + x + 1)^n$.

It follows that $p(x)$ can be written as $p(x) = c(x^2 + x + 1)^n$, where c is some real number. However $p(0) = 1$. Therefore $c = 1$. It can be checked that n can be any positive integer.

11. Make the substitution $g(x) = f(x) - \frac{1}{2}$, and write the functional equation in terms of $g(x)$.
- Show that $g(x + 2a) = g(x)$.
 - Make the substitution $h(x) = 4[g(x)]^2 - \frac{1}{2}$, and show that $h(x + a) = -h(x)$. Show that $h(x)$ can be defined arbitrarily on the interval $[0, a)$ subject to the restriction that $|h(x)| \leq \frac{1}{2}$, and then can be extended to all real x .

12. Define

$$g(x) = f(x) - x f(1) + x f(0) - f(0)$$

and show that $g(x)$ has the same property as $f(x)$. It is sufficient to show that $g(x) = 0$ for all $x \in [0, 1]$.

Now $g(0) = g(1) = 0$. Let M and m be the maximum and minimum values of g on the interval $[0, 1]$. By the extreme value theorem, the sets

$$\{x \in [0, 1] : g(x) = M\}$$

and

$$\{x \in [0, 1] : g(x) = m\}$$

are both nonempty. As f is a continuous function, both of these sets must be closed, and therefore have maximum and minimum values. Let

$$y = \max\{x \in [0, 1] : g(x) = M\}.$$

If $0 < y < 1$, there would exist an $h > 0$ for which

$$M = \frac{g(y-h) + g(y+h)}{2}.$$

As M is the maximum of g , this would imply that $g(y+h) = g(y-h) = M$, contradicting the fact that y is the maximum value at which g attains M . So y must equal 0 or 1. From this, it follows that $g(x) \leq 0$ for all $0 \leq x \leq 1$. Similarly, it can be shown that $g(x) \geq 0$.

13. Let $c = g(0)$. Then $f(x+c) = 2x+5$. So $f(z) = 2(z-c)+5$. Solving for $g(y)$ in the original equation, we get $g(y) = y/2+c$. Therefore, $g[x+f(y)] = (x+2y+5)/2$.

14. Define

$$S = \{x : f(x) = x\}.$$

Use the continuity of f to show that S is a closed set. From condition (b), we know that S is not empty.

Let us suppose that $S \neq \mathbf{R}$. We prove a contradiction. As S is closed, we can let x_0 be a boundary point of S . Suppose that $y \notin S$. Then there exists some real number $r \neq 0$ such that $f(y) = y+r$. Use property (a) and the fact that f is one-to-one to prove by induction that

$$f(y+nr) = (y+nr) + r.$$

Now suppose $x \notin S$. Let $\epsilon = |f(x) - x|$. As f is continuous, there exists a $0 < \delta \leq \epsilon/4$ such that $|z - x| < \delta$ implies that $|f(z) - f(x)| < \epsilon/4$. Additionally, the continuity of f implies that there exists an $\eta > 0$ such that $\eta < \delta$, for which $|w - x_0| < \eta$ implies that $|f(w) - f(x_0)| < \delta$.

In particular, let us choose the real number y such that $|y - x_0| < \eta$, and such that $f(y) \neq y$. (Don't forget that x_0 is a boundary point of S . Therefore, we can get arbitrarily close to x_0 with points that are not in S .) Then

$$\begin{aligned} 0 < |f(y) - y| &\leq |f(y) - f(x_0)| + |f(x_0) - y| \\ &= |f(y) - f(x_0)| + |x_0 - y| \\ &< \delta + \eta \\ &< 2\delta. \end{aligned}$$

Now let $r = f(y) - y$. Because $0 < |r| < 2\delta$, there exists an integer n such that $|(y + nr) - x| < \delta$. But $f(y + nr) = (y + nr) + r$. Therefore

$$\begin{aligned} \epsilon &= |f(x) - x| \\ &\leq |f(x) - f(y + nr)| + |f(y + nr) - x| \\ &< \frac{\epsilon}{4} + |(y + nr) + r - x| \\ &\leq \frac{\epsilon}{4} + |(y + nr) - x| + |r| \\ &< \frac{\epsilon}{4} + \delta + 2\delta \\ &< \frac{\epsilon}{4} + \frac{\epsilon}{4} + \frac{\epsilon}{2} \\ &= \epsilon. \end{aligned}$$

This is a contradiction, which we obtained by assuming that $S \neq \mathbf{R}$.

7.3 Hints for Chapter 2

1. Just follow the proof of Theorem 2.3. You do not need to worry about negative values this time.
2. The steps given in the problem are big hints in themselves.
3. First, suppose that $g(0) = 0$. Using the equation $f(y) = g(0)h(y)$ it will follow that $f(y)$ vanishes for all y . However, the equation $g(x)h(y) = 0$ for all x and y implies that either g vanishes for all real numbers or h does. Similarly, if $h(0) = 0$, the same conclusion follows. Suppose then, that $g(0)$ and $h(0)$ are both nonzero. Set $g(0) = a$ and $h(0) = b$. Prove that $g(x) = b^{-1}f(x)$ and $h(x) = a^{-1}f(x)$. Then prove that

$$f_0(x) = \frac{f(x)}{ab}$$

is continuous, and satisfies Cauchy's exponential equation.

4. Let $f(0) = b$, and define $f_0(x) = f(x) - b$. We need to show that f_0 satisfies Cauchy's equation. First of all, set $y_i = 0$ for all i , and set $x_1 = 0$, $x_2 = x$, and $x_3 = -x$, to show that $f_0(-x) = -f_0(x)$. Secondly, set $y_i = 0$ for all i and write $x_3 = -(x_1 + x_2)$ to obtain the equation $f_0(x_1) + f_0(x_2) + f_0(-x_1 - x_2) = 0$. Then apply the first result.
5. The proof of this result carries through in much the same way as the original proof of Theorem 2.3, with the usual addition replaced by addition modulo one. Because $f(0 + 0) = 2f(0) \pmod{1}$ it follows that $f(0) = 0$

(mod 1). From this we also get $f(1) = 0 \pmod{1}$. From this fact, we can show that

$$2^k f\left(\frac{1}{2^k}\right) = 0$$

mod 1, which is equivalent to

$$f\left(\frac{1}{2^k}\right) = \frac{n}{2^k}$$

mod 1, again. It must be established that the value of n is independent of the choice of k in this formula. From that point on the proof is similar. We extend this to show that

$$f\left(\frac{m}{2^k}\right) = \frac{nm}{2^k} \pmod{1}$$

and use the fact that the dyadic rationals are dense.

6. Replace addition by addition modulo one in both the domain and the range of f . By combining the methods of Problems 4 and 5, it can be shown that $f(x) = nx + b \pmod{1}$.

7. Consider the equation

$$f[(tu)x, (tu)y] = f[t(ux), t(uy)].$$

Apply the modified Euler equation (2.66) once to the left-hand side and twice to the right-hand side to show that the function h satisfies the equation $h(tu) = h(t)h(u)$.

8. Suppose $b > a$. Define $\epsilon = (b - a)/n$, where n is a positive integer. Then define $x_0 = a$, and $x_i = a + i\epsilon$, for $i = 1, \dots, n$. Then

$$\begin{aligned} |f(a) - f(b)| &\leq \sum_{i=1}^n |f(x_i) - f(x_{i-1})| \\ &\leq \sum_{i=1}^n n^{-2} K (b - a)^2 \\ &= n^{-1} K (b - a)^2. \end{aligned}$$

Letting $n \rightarrow \infty$, we have $n^{-1} K (b - a)^2 \rightarrow 0$.

9. No. The function $f(x) = |x|$ is subadditive. (This is the triangle inequality.) In addition $|x| \geq x$. However, $|x| \neq x$ when x is negative.

10. Note first that if $f(x) = 0$ for any rational x , then f must vanish on all rational values. Henceforth we assume that $f(x) \neq 0$ for all x . Let $f(0) = a$. Prove by induction that

$$f[n f(x)] = f(0) [f(x)]^n$$

for all integers n . Note that

$$f(-a) = f[-f(0)] = f(0) [f(0)]^{-1} = 1.$$

Therefore we have

$$f(x+1) = f[x + f(-a)] = f(x) f(-a) = f(x).$$

It follows that $f(m) = a$ for all integers m . Now suppose that m/n is in the range of f for some integers m and n . So there exists an x such that $f(x) = m/n$. We can write

$$\begin{aligned} a &= f(m) \\ &= f\left[n \frac{m}{n}\right] \\ &= f[n f(x)] \\ &= a \left(\frac{m}{n}\right)^n. \end{aligned}$$

As $a \neq 0$ we can cancel to obtain $m/n = \pm 1$.

Can $m/n = -1$? Suppose that -1 is in the range of f . Then $f(x-1) = f(x)(-1)$. But this contradicts the statement that $f(x+1) = f(x)$ for all x .

11. The steps given in the question are hints enough.
12. The zero function is one solution to this equation. Suppose that there exists a real number y_0 such that $f(y_0) \neq 0$. Then

$$f(x) f(y_0) = f\left(\sqrt{x^2 + y_0^2}\right) = f(-x) f(y_0).$$

This implies that $f(x) = f(|x|)$ for all real x . If we define $g(x) = f(\sqrt{x})$ for $x \geq 0$, then g satisfies Cauchy's exponential equation $g(x+y) = g(x)g(y)$ for $x, y \geq 0$.

Now if $g(0) = 0$ it can be seen that $g(x) = 0$ for all $x \geq 0$. Suppose that $g(0) \neq 0$. Then the proof that $g(x) > 0$ for all $x \geq 0$ is similar but not identical to the proof for Cauchy's exponential equation because g is restricted in its domain of definition to nonnegative x . Note that if $g(x) = 0$, then $g(x/2) = 0$. So if there were an $x \geq 0$ for which $g(x) = 0$, then it would be possible to construct a sequence $x, x/2, x/4, \dots$ converging

to zero such that g vanishes on all these values. However, g is continuous. So this would imply that $g(0) = 0$, a possibility that we have already eliminated. It follows that $g(x) \neq 0$ for all nonnegative x . Using $g(x) = [g(x/2)]^2$, we observe that $g(x) > 0$. The rest of the argument parallels the original argument for Cauchy's exponential equation. We conclude that $f(x) = [f(1)]^{x^2}$.

13. Note that $f(x) = 0$ for all x is a solution to the equation. Let $y = 0$. Then $[f(x)]^2 = [f(x)]^2 [f(0)]^2$. If $f(x) \neq 0$ for some value of x , then $f(0) = \pm 1$. But f satisfies the equation if and only if $-f$ does. So we can reduce to the case where $f(0) = 1$. Putting $x = 0$, we can conclude that $f(y) = f(-y)$ for all y . Putting $x = y$, we can conclude that $f(2x) = [f(x)]^4$. From this equation, we see that if $f(x) = 0$ for any x , then $f(x/2) = 0$ as well. Using an argument similar to the previous problem, we see that $f(x) \neq 0$ for all x . Because $f(x) = [f(x/2)]^4$ it follows that $f(x) > 0$ for all x .

By induction, it can be established that

$$f(nx) = [f(x)]^{n^2}.$$

Letting $x = 1/n$ we get $f(1) = [f(1/n)]^{n^2}$, and thereby

$$f(1/n) = [f(1)]^{1/n^2}.$$

So

$$f\left(\frac{m}{n}\right) = f[m(1/n)] = [f(1/n)]^{m^2} = [f(1)]^{m^2/n^2}.$$

Apply continuity to show that $f(x) = [f(1)]^{x^2}$. Again, if $f(x)$ is a solution, then so is $-f(x)$.

14. a) Write $u = f(x)$ and $v = f(y)$. The equation becomes

$$(gf^{-1})\left(\frac{u+v}{2}\right) = \frac{(gf^{-1})(u) + (gf^{-1})(v)}{2}$$

which is Jensen's equation. The result follows by solving Jensen's equation.

- b) Let $g(x) = f(x+t)$. Then the equation can be rewritten as

$$g^{-1}\left[\frac{g(x) + g(y)}{2}\right] = f^{-1}\left[\frac{f(x) + f(y)}{2}\right].$$

Applying the result of part (a), we see that

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

for all real x . This is an example of Vincze's equation. So we must have

$$f(x) = cx + d \quad \text{or} \quad f(x) = cs^x + d.$$

The first of these gives rise to the usual arithmetic mean. The second mean is

$$\log_s \left[\frac{s^x + s^y}{2} \right]$$

which is called the exponential mean.

- c) Use logarithms to transform this property into the translative property of part (b). The two homogeneous means are the geometric mean and the root mean power

$$\left[\frac{x^p + y^p}{2} \right]^{1/p}.$$

15. Let $x = y = 0$ in the equation to obtain $f(0) = 0$. Next, we differentiate both sides of the equation, once with respect to x and once with respect to y . The resulting equation is

$$0 = f''(x+y)f(x-y) - f(x+y)f''(x-y).$$

Let $x = (u+v)/2$ and $y = (u-v)/2$. We obtain

$$f''(u)f(v) = f(u)f''(v).$$

The zero function solves this equation. Looking for nontrivial solutions, we suppose there exists a value v_0 such that $f(v_0) \neq 0$. Then the equation becomes $f''(u) = cf(u)$, where $c = f''(v_0)/f(v_0)$. This is a standard differential equation with the initial condition $f(0) = 0$. If $c > 0$, we have $f(u) = A \sinh(\sqrt{c}u)$. If $c = 0$, we have $f(u) = Au$. Finally, if $c < 0$, we have $f(u) = A \sin(\sqrt{-c}u)$.

16. *Step 1.* Prove that $f_1(0) = 0$, using rule (b).
Step 2. Prove that $f_1(x) = x$, using rule (a) and Step 1.
Step 3. Prove that $f_n(c, 0, \dots, 0, -c) = 0$. To show this, use the symmetry property of f_n to switch c and $-c$, and use rule (b).
Step 4. Prove the required result using induction. From Step 2 it is clear that the result is true for $n = 1$. Suppose it is true for $n = k$. Rule (c) says that $f_{k+1}(x_1, \dots, x_{k+1})$ depends only on x_{k+1} and $x_1 + \dots + x_k$. Let $a_1 + \dots + a_{k+1} = 0$. Then using the observation above, we get

$$\begin{aligned} f_{k+1}(a_1, \dots, a_{k+1}) &= f_{k+1}(-a_{k+1}, 0, \dots, 0, a_{k+1}) \\ &= 0. \end{aligned}$$

Finally, let x_1, \dots, x_{k+1} be arbitrary with $(x_1 + \dots + x_{k+1})/(k+1) = u$. Then

$$\begin{aligned} f_{k+1}(x_1, \dots, x_{k+1}) &= f(x_1 - u, \dots, x_{k+1} - u) + u \\ &= 0 + u \\ &= u. \end{aligned}$$

7.4 Hints for Chapter 3

1. For any x , the sequence $2^{-n}x$, $n = 1, 2, \dots$, converges to zero. Conclude from the functional equation that $f(x) = f(2^{-n}x)$ and use the continuity of f to show that $f(x) = f(0)$ for all x . Hence f is a constant function.
2. The function $f(x) = x + \sin(\pi x)$ will work.
3. First show that $\alpha(x) = \cos x$ is a contractive mapping. Use this to show that $\alpha^n(x) = \cos^n x$ converges to the unique solution to the equation $\cos x_0 = x_0$. As x_0 is in the closure of the orbit of every x , it will follow from the continuity of f that $f(x) = f(x_0)$. So f must be a constant function.
4. Prove that $\Gamma(x+1) = x\Gamma(x)$.
5. Suppose that f is neither strictly increasing nor strictly decreasing. We prove a contradiction. Under this assumption, it must be possible to find a set of real numbers $x < y < z$ such that either $\max[f(x), f(z)] < f(y)$ or else $\min[f(x), f(z)] > f(y)$. It is sufficient to examine the first of these cases as the second is similar. We can also assume that $f(z) > f(x)$ as the other case, namely $f(z) < f(x)$ follows by reversing the roles of x and z . Because $f(z)$ lies in the open interval $(f(x), f(y))$, it follows by the intermediate value theorem that there exists a real number w between x and y such that $f(w) = f(z)$. However, z is greater than x and y . So $w \neq z$. This contradicts the fact that f is a one-to-one function.
6. This is an example of Schröder's equation with $\alpha(x) = sx/(1+x)$. The Koenigs algorithm gives $f(x) = x/(1+x-s)$ as a nontrivial solution. The equation $\alpha(x) = x$ has solutions $x = 0$ and $x = s-1$. We may define f for all x such that $x \neq s-1$.
7. The function $g(x) = x^{-1}$ satisfies Abel's equation $g[\alpha(x)] = g(x) + 1$, where $\alpha(x) = x/(x+1)$. So, using equation (3.38), we deduce that $f(x) = x/(cx+1)$ is a solution to the given commutativity equation for all c .
8. We start with the functions f and g defined on $[1, \infty)$. On the interval $(0, 1)$, define f and g to be $1/f(1/x)$ and $1/g(1/x)$, respectively. Next, set $f(0) = 0$ and $g(0) = 0$. To extend f and g to negative numbers, we define $f(-x) = f(x)$ and $g(-x) = g(x)$ on $(-\infty, 0)$. Defined on the entire real line in this way, the functions f and g can be seen to satisfy the required equations.

9. The constancy of f can be proved for all a and b such that b/a is irrational and positive. Take any real values x and x_0 , and suppose that $f(x) > f(x_0)$. Then there exists a $\delta > 0$ such that $f(x) > f(t)$ for all t such that $|t - x_0| < \delta$. Next, the reader should prove that there exist positive integers n and m such that

$$|x - x_0 - ma + nb| < \delta.$$

So $f(x) > f(x - ma + nb)$. However, from our functional inequality, we have

$$f(x - ma + nb) \geq f(x + nb) \geq f(x)$$

which is a contradiction. Similarly, $f(x) < f(x_0)$ also leads to a contradiction. We conclude that $f(x) = f(x_0)$ for all x . That is, f is a constant function.

The reader should note that we have not used the continuity of f at any place other than x_0 . Thus it is sufficient to assume that there exists a point x_0 at which f is continuous.

10. This problem is a bit of a mathematical joke, and is much easier than it looks (in retrospect). Let $g(x) = a$ for all x . Then our equation becomes $f(a) = a$. But this is true for all real numbers a . So $f(x) = x$ for all x .
11. Let $\alpha(x) = (x - 1)/x$. To get some insight into the nature of this functional equation, we examine the splinter of α . We get $\alpha^2(x) = -1/(x - 1)$ and $\alpha^3(x) = x$. The fact that the splinter of x is finite suggests that the functional equation can be solved by simple substitution methods. Replacing x by $\alpha(x)$ in the functional equation, we get $f[\alpha(x)] + f[\alpha^2(x)] = 2 - 1/x$. Replacing x by $\alpha^2(x)$ in the original functional equation gives us $f[\alpha^2(x)] + f(x) = (x - 2)/(x - 1)$. Adding this last equation to our original functional equation and subtracting the intermediate equation, we get

$$f(x) = \frac{x^3 - x^2 - 1}{2x(x - 1)}.$$

12. Such a function does not exist. For suppose f satisfies the conditions of the problem. Let c be the smallest integer multiple of $1/4$ such that $f(x) \leq c$ for all x . Use (c) to show that $f(2) = 2$ and conclude that $c \geq 2$. Also, by definition, there exists some real number x such that $f(x) \geq c - 1/4$. So

$$c \geq f\left(x + \frac{1}{x^2}\right) = f(x) + \left[f\left(\frac{1}{x}\right)\right]^2 \geq c - \frac{1}{4} + \left[f\left(\frac{1}{x}\right)\right]^2.$$

Therefore $f(1/x) \geq -1/2$. Additionally,

$$c \geq f\left(\frac{1}{x} + x^2\right) = f\left(\frac{1}{x}\right) + [f(x)]^2 \geq -\frac{1}{2} + \left(c - \frac{1}{4}\right)^2.$$

This last inequality can be checked to be incompatible with the inequality $c \geq 2$.

13. As in problem 11, above, we see that if we let $\alpha(z) = 1 - z$, then the splinter $\alpha^n(z)$ is a finite set because $\alpha^2(z) = z$. Replace z by $1 - z$ in the functional equation to get $f(1 - z) + (1 - z)f(z) = 2 - z$. Eliminate $f(1 - z)$ from the two functional equations to obtain

$$(1 - z + z^2)f(z) = 1 - z + z^2.$$

It is tempting to leap to the conclusion that $f(z) = 1$ for all z . However, the exceptional values are those values of z for which $1 - z + z^2 = 0$ as we have not assumed that f is continuous. Thus $f(z) = 1$ for all z except possibly $z_0 = e^{\pi i/3}$ and $z_1 = e^{-\pi i/3}$. We have $z_0 + z_1 = 1$ and $z_0 z_1 = 1$. Write $f(z_0) = 1 + w_0$ and $f(z_1) = 1 + w_1$. Plug these expressions back into the original functional equation to obtain $w_0 + z_0 w_1 = 0$. Thus w_0 is arbitrary, and w_1 can be defined in terms of w_0 .

14. Let x be any real number. Define $a_{k,l} = f(x + k/6 + l/7)$. Then

$$a_{k,l+1} + a_{k+1,l} = a_{k,l} + a_{k+1,l+1}.$$

For any positive integer m , we can sum these equations over values of k , where $0 \leq k \leq m - 1$, to obtain

$$a_{0,l+1} + a_{m,l} = a_{0,l} + a_{m,l+1}.$$

Adding these equations over $0 \leq l \leq n - 1$ for each positive integer n yields

$$a_{0,n} + a_{m,0} = a_{0,0} + a_{m,n}.$$

In particular, we can set $m = 6$ and $n = 7$ to obtain $2f(x + 1) = f(x) + f(x + 2)$. So the sequence $f(x + n)$ is an arithmetic progression. However, by assumption (a), this sequence is bounded. So $f(x + 1) = f(x)$ for all x .

15. A power function of the form $f(x) = x^p$ almost works, because the functional equation can be reduced to $1^p + 2^p + 3^p = 0$. If x is negative, however, the expression x^p is ambiguous or multi-valued. This problem can be eliminated by considering a function of the form $f(x) = |x|^p$. Unfortunately, there is no real number p that satisfies the equation $1^p + 2^p + 3^p = 0$. But if we could find a complex number p satisfying this equation, then we could define $f(x) = \Re(|x|^p)$, where \Re denotes the function which gives the real part of any complex number. To ensure that f is continuous, we need to have $\Re(p) \geq 0$.

Let $p = a + b\sqrt{-1}$. Then to satisfy the equation $1^p + 2^p + 3^p = 0$ we need to have

$$2^a \sin(b \log 2) + 3^a \sin(b \log 3) = 0$$

and

$$2^a \cos(b \log 2) + 3^a \cos(b \log 3) = -1.$$

When $a = 0$, the first equation is satisfied by $b = \pi / \log(3/2)$. However, in this case, the left-hand side of the second equation is greater than -1 . On the other hand, as a approaches infinity, then the first equation is solved by a value of b which approaches the solution to $\sin(b \log 3) = 0$. A value of b which solves this is $b = 3\pi / \log 3$. For this choice of b the left-hand side of the second equation goes to $-\infty$ which is less than -1 . To complete the proof, the reader will need to argue carefully that somewhere between these extremes lies a solution to both equations.

The reader with a more practical bent may find such an existence argument a bit unsatisfying. Those of such a persuasion may be satisfied to know that

$$p \approx .9405921833 + 8.230709548 \sqrt{-1}$$

will work. A plot of the function $f(x) = \Re(|x|^p)$ can be found in Figure 7.1.

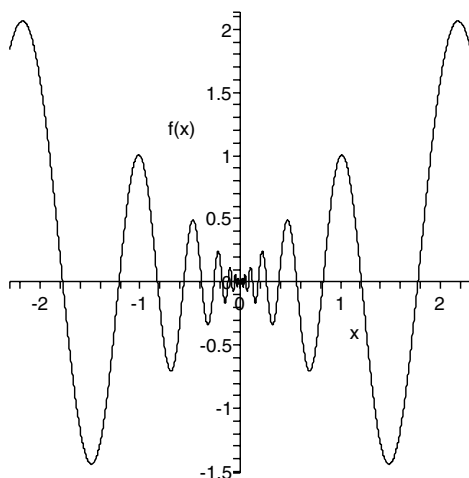


Fig. 7.1. A function satisfying the required functional equation

16. Suppose we could partition the nonzero real numbers into n sets A_1, A_2, \dots, A_n such that
- a) there exists a function $g_i : A_i \rightarrow A_{i+1}$ which is one-to-one and onto A_{i+1} for all $i = 1, \dots, n - 1$;

b) if $x \in A_i$ then $x^{-1} \in A_i$.

We could then define $f(x) = g_i(x)$ for $x \in A_i$, and $f[g_{n-1}g_{n-2} \cdots g_1(x)] = x^{-1}$ for all $x \in A_1$. As every element $y \in A_n$ can be represented uniquely as $g_{n-1}g_{n-2} \cdots g_1(x)$ for $x \in A_1$, it follows that f is well defined on the nonzero real numbers, and that $f^n(x) = x^{-1}$.

17. The function $f(x) = e^x$ satisfies these conditions. We need to prove that it is the only one. We can prove using induction and condition (a) that $f(x) = [f(2^{-n}x)]^{2^n}$. So $(1 + 2^{-n}x)^{2^n} \leq f(x)$ by condition (c). Similarly, we have $f(-x) \geq (1 - 2^{-n}x)^{2^n}$. Therefore,

$$(1 + 2^{-n}x)^{2^n} \leq f(x) = [f(-x)]^{-1} \leq (1 + 2^{-n}x)^{2^n} (1 - 4^{-n}x^2)^{-(2^n)}.$$

Taking limits as $n \rightarrow \infty$ gives us $e^x \leq f(x) \leq e^x$, which gives us the required result.

18. The function does not exist. Consider a sequence defined by $a_1 = 1$, $a_2 = f(1)$, and $a_n = f(a_{n-1})$ for all $n \geq 2$. The functional equation can be used to prove that the sequence $\{a_n\}$ contains all positive integers. To see this, note first that the number 1 is obviously in the sequence. If $a_n = x$, then

$$a_{n+f(x)} = f^{f(a_n)}(a_n) = f^{f(x)}(x) = x + 1.$$

So by induction the sequence must hit every positive integer. Also all members of the sequence are distinct. (If $a_n = a_m$ where $n \neq m$ then the sequence cycles and therefore the range of the sequence is finite. But this would contradict the result just proved, namely that the sequence contains all positive integers.) Finally, $\{a_n\}$ is an increasing sequence: If $x = a_k$, then $x + 1 = a_{k+f(x)}$. So $x + 1$ appears later in the sequence than x for all x . So the sequence must be $a_n = n$ for all n . Therefore $f(n) = n + 1$, which does not work.

19. We have $f[f(1000)] = 1/999$. So the range of f contains 999 and $1/999$. By the intermediate value theorem, there exists an a such that $f(a) = 500$. So $f(500) = f[f(a)] = 1/500$.
20. The function f must be constant or be the identity function. Letting $x = 0$ we see that $f(y) = f[f(y)]$. So $f(x) = x$ on the range of f . Let a and b be the minimum and maximum values, respectively, of f . (The minimum or maximum could be infinite.) As f is continuous, the intermediate value theorem tells us that the interval (a, b) lies in the range of f . So f is the identity on (a, b) .
If $a = b$ then f is constant. If $a = -\infty$ and $b = +\infty$, then $f(x) = x$. Suppose that a is finite and $b > a$. The continuity of f implies that $f(a) = \lim_{x \rightarrow a^+} f(x) = a$. So in this case f must be the identity on $[a, b)$. So $f(x) = x$ on some interval $[a, a + 2d]$ for $d > 0$. Let $0 < t < d$. Then

$f(a-t) = a+s$ for some $s > d$. (If $s \leq d$ then $a = f(t+a-t) = f(t+a+s) = a+s+t$, which is a contradiction.) However, the statements $f(a-t) > a+d$ for all t in $(0, a)$ and $f(a) = a$ contradict the intermediate value theorem.

The case where b is finite and $a < b$ is similar.

21. Try $f(x) = x - [x]$.
22. The function f must be constant. Suppose not. Then there exist real numbers x and y such that y lies between x and $x + f(x)$ and such that $f(x) \neq f(y)$. Then there exists a straight line L defined by the equation $x + ny = c$, where n is some positive integer, that separates the point $(y, f(y))$ from the points $(x, f(x))$ and $(x + f(x), f(x))$ on the graph of the function f . By the intermediate value theorem, there exist two distinct points A and B with x -coordinates a and b respectively that lie on the graph of the function and on the line L . By induction we can establish that $f(x) = f[x + nf(x)]$. So $f(c) = f[a + nf(a)] = f(a)$. Similarly, $f(c) = f[b + nf(b)] = f(b)$. So $f(a) = f(b)$. Thus A and B have the same y -coordinate. But this means that $a + nf(a) = b + nf(b)$ expressed in terms of L . So $a = b$ which is a contradiction, because A and B are distinct.
23. In fact, there are infinitely many expressions for $u(x)$ in terms of f and g . Here are three:

$$\begin{aligned} u(x) &= g(x) - f(x+3) + f(x+1) + f(x-1) - f(x-3) \\ &= -g(x+2) + f(x+5) - f(x+3) + f(x+1) + f(x-1) \\ &= g(x+4) - f(x+7) + f(x+5) - f(x+3) + f(x+1). \end{aligned}$$

Can you find a way of generating all such formulas?

24. The unique solution is $f(x) = 1/x$, a result which is known as Dubikajtis' theorem. To prove this, consider a set \mathbf{Q} which is defined as the smallest set of real numbers such that
- $1 \in \mathbf{Q}$;
 - if $x \in \mathbf{Q}$ then $x + 1 \in \mathbf{Q}$;
 - if $x \in \mathbf{Q}$ then $x^{-1} \in \mathbf{Q}$.

One way to define what we mean by *smallest* here is to define \mathbf{Q} to be the intersection of all subsets of \mathbf{R} which have these three properties. The first step in the proof is to show that \mathbf{Q} is the set of all positive rational numbers. To demonstrate this, we first prove that \mathbf{Q} contains all positive integers. This follows easily from the fact that \mathbf{Q} is closed under translations by 1. Because $2 \in \mathbf{Q}$, it follows from the closure of \mathbf{Q} under reciprocals that $1/2 \in \mathbf{Q}$. Any positive rational of the form $(2m+1)/2$ can be written as $m + 1/2$, which must be in \mathbf{Q} because of (b) above and

the fact that $1/2 \in \mathbf{Q}$. Having established that any rational of the form $n/j \in \mathbf{Q}$ for all $1 \leq j \leq k$ we proceed to show that a rational of the form $n/(k+1) \in \mathbf{Q}$ by writing $n/(k+1) = m + [(k+1)/p]^{-1}$, where p is a positive integer less than or equal to k . By induction, we see that \mathbf{Q} contains all the positive rationals, and therefore must be this set.

The next step is to show that $f(1) = 1$. To prove this, we first show by induction that $f(n+1) = c/(1+nc)$, where c is defined to be $f(1)$. Therefore $f[c/(1+nc)] = f^2(n+1) = n+1$. So

$$\begin{aligned} f\left[\frac{1+(n+1)c}{1+nc}\right] &= f\left(1 + \frac{c}{1+nc}\right) \\ &= f\left(\frac{c}{1+nc}\right) \frac{1}{1+f\left(\frac{c}{1+nc}\right)} \\ &= \frac{n+1}{n+2}. \end{aligned}$$

Taking limits on both sides of this equation as $n \rightarrow \infty$, and using the continuity of f gives us $f(1) = 1$.

The next stage of the proof is to show that $f(x) = 1/x$ on the set \mathbf{Q} . Clearly, $f(x) = 1/x$ when $x = 1$. If $f(x) = 1/x$, then

$$f(x+1) = \frac{f(x)}{1+f(x)} = \frac{x^{-1}}{1+x^{-1}} = \frac{1}{x+1}$$

and

$$f(x^{-1}) = f^2(x) = x = (x^{-1})^{-1}.$$

So $f(x) = x^{-1}$ on \mathbf{Q} .

Using the continuity of f we extend the result from \mathbf{Q} to all of $(0, \infty)$.

25. If we let $y = x$, we get $f(2x) + f[2f(x)] = f(2f[x+f(x)])$. Next, let us replace x with $f(x)$ in this equation, and subtract the equation above from the resulting equation. We get

$$f[2f^2(x)] - f(2x) = f(2f[f(x) + f^2(x)]) - f(2f[x+f(x)]).$$

Suppose $f^2(x) > x$. Then the left-hand side is less than 0. So

$$f[f(x) + f^2(x)] > f[x+f(x)].$$

This implies that $f(x) + f^2(x) < x + f(x)$, which is a contradiction. The case $f^2(x) < x$ is disposed of similarly.

7.5 Hints for Chapter 4

1. Let

$$p(x) = |f(x)| - |g(x)| + h(x).$$

The critical values of the function $p(x)$ are at $x = -1$ and $x = 0$. Because the function is linear between these critical values we can write

$$p(x) = a|x + 1| + b|x| + cx + d.$$

Solve for the constants a , b , c , and d .

2. a) First note that if x is a root of $f(x)$ then so is $x^2 + x$. Similarly, if $x + 1$ is a root of $f(x)$ then so is $x^2 + x$. The latter statement is equivalent to saying that if x is a root of $f(x)$ then

$$(x - 1)^2 + (x - 1) = x^2 - x$$

is also a root. However, if $x \neq 0$ is any root of $f(x)$ in the complex plane, then

$$\max(|x^2 + x|, |x^2 - x|) > |x|.$$

Thus, if there were a nonzero root, we would be able to generate an infinite sequence of distinct roots. This could only happen if $f(x)$ were a constant polynomial with $f(x) \equiv 0$ or $f(x) \equiv 1$.

The only other possibility is that $f(x)$ is not constant, and 0 is the only root. In this case, $f(x) = x^k$ for some k .

- b) Examining the degree of each polynomial, we observe that

$$\deg f \cdot \deg g = \deg f + \deg g.$$

This is satisfied in nonnegative integers if both degrees are zero or both degrees are two. In the first case, we must have

$$f(x) = c, \quad g(x) = 1,$$

where the constant c is arbitrary.

When the degrees of both polynomials are two, we must have some point x in the complex plane such that $g(x) = 0$. Then

$$\begin{aligned} f(0) &= f(g(x)) \\ &= f(x)g(x) \\ &= 0. \end{aligned}$$

This implies that $f(x)$ can be written in the form $f(x) = cx(x + a)$ for some constant a . It is easily checked that

$$\begin{aligned} f(x) &= c(x^2 + ax) \\ g(x) &= x^2 + ax - a \end{aligned}$$

is the general solution where a and c are arbitrary.

3. Set $x = 0$ and $y = 0$. This is a case in which $xy + yz + zx = 0$. So the equation reduces to

$$f(0) + f(-z) + f(z) = 2f(z).$$

In this equation, the special case $z = 0$ yields the formula $3f(0) = 2f(0)$, which implies that $f(0) = 0$. Therefore

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

which is equivalent to $f(z) = f(-z)$. This holds for all z . The polynomial f is even. Therefore, we can write $f(x) = g(x^2)$, for some polynomial g with real coefficients. So

$$g((x-y)^2) + g((y-z)^2) + g((z-x)^2) = 2g((x+y+z)^2)$$

for all x, y , and z such that $xy + yz + zx = 0$.

For any such triple x, y, z , let $r = x - y$, $s = y - z$ and $t = z - x$. Then

$$\begin{aligned} (x+y+z)^2 &= x^2 + y^2 + z^2 \\ &= (x^2 + y^2 + z^2) - (xy + yz + zx) \\ &= \frac{1}{2}(r^2 + s^2 + t^2) \\ &= \frac{1}{2}[r^2 + s^2 + (-r-s)^2] \\ &= r^2 + s^2 + rs. \end{aligned}$$

So our functional equation in g reduces to

$$g(r^2) + g(s^2) + g((r+s)^2) = 2g(r^2 + rs + s^2).$$

The special case in this equation where $r = s$ yields

$$2g(r^2) + g(4r^2) = 2g(3r^2).$$

The next step is to show that this equation must be satisfied for all real r . So far we have only verified this equation for values of r such that $r = x - y$, $r = y - z$, and such that $xy + yz + zx = 0$. However, it is possible to show that any real number r has such a representation. The values

$$x = r + \frac{|r|}{\sqrt{3}}, \quad y = \frac{|r|}{\sqrt{3}}, \quad z = \frac{|r|}{\sqrt{3}} - r$$

work. So, letting $u = r^2$, we deduce that

$$2g(u) + g(4u) = 2g(3u)$$

for all $u \geq 0$. Let $g(u) = a_0 + a_1 u + \cdots + a_n u^n$. A simple argument based on matching the coefficients on each side of the equation now shows that

$a_j = 0$ for all $j \geq 3$. In addition, $g(0) = f(0) = 0$. So $a_0 = 0$. Therefore, we can write

$$g(u) = a_1 u + a_2 u^2.$$

It follows that we can write

$$f(r) = g(r^2) = a_1 r^2 + a_2 r^4.$$

It can be checked that every function of this form satisfies the original functional equation.

4. Consider $f(1)$. We are given that $f(f(1)) = 3$, and that $f(n)$ is strictly increasing in n . If $f(1) = 1$, then $f(f(1)) = 1$, which is a contradiction. So $f(1) \geq 2$. If $f(1) > 2$, then $f(f(n)) > 3$ which is also a contradiction, because $f(f(1)) = 3 \cdot 1 = 3$. So $f(1) = 2$. Next, we see that $f(2) = f(f(1)) = 3$. Continuing we get

$$f(3) = f(f(2)) = 3 \cdot 2 = 6.$$

Using reasoning of this kind, it is left to the reader to verify by induction that for any $k = 0, 1, 2, \dots$, we must have

$$\begin{aligned} f(3^k) &= 2 \cdot 3^k \text{ and} \\ f(2 \cdot 3^k) &= 3^{k+1}. \end{aligned}$$

What about the numbers between these values? Suppose $3^k < n < 2 \cdot 3^k$. Because $f(n)$ is strictly increasing,

$$2 \cdot 3^k < f(n) < 3^{k+1}.$$

There are $3^k - 1$ numbers strictly between $2 \cdot 3^k$ and 3^{k+1} . These are the possible values of $f(n)$. There are also $3^k - 1$ numbers n strictly between 3^k and $2 \cdot 3^k$. Because $f(n)$ is strictly increasing, it follows that

$$\begin{aligned} f(3^k + j) &= f(3^k) + j \\ &= 2 \cdot 3^k + j \end{aligned}$$

for $j = 0, 1, 2, \dots, 3^k$.
So

$$\begin{aligned} f(2 \cdot 3^k + j) &= f(f(3^k + j)) \\ &= 3 \cdot (3^k + j) \\ &= 3^{k+1} + 3j \end{aligned}$$

for all $j = 0, 1, 2, \dots, 3^k$.

We leave it to the reader to evaluate $f(2001)$.

5. a) Set $x = 0$. We get

$$f(f(y)) = (f(0))^2 + y.$$

So it is natural to define $f(0) = a$ so that $f(f(y)) = a^2 + y$. The function $h(y) = a^2 + y$ on the right-hand side is onto. That is, as y ranges over all reals, so does $h(y)$. Therefore $f(y)$ must be onto as well.

Because f is onto, there must exist a number b such that $f(b) = 0$. Setting $x = b$ in the original functional equation yields

$$f(f(y)) = y.$$

Thus f is an involution. So

$$\begin{aligned} (f(x))^2 + y &= f(x f(x) + f(y)) \\ &= f[f(f(x)) f(x) + f(y)], \text{ replacing } x \text{ with } f(f(x)) \\ &= f[f(x) f(f(x)) + f(y)], \text{ commuting factors} \\ &= (f(f(x)))^2 + y \\ &= x^2 + y. \end{aligned}$$

So $(f(x))^2 = x^2$ for all x .

Looking at the original functional equation, we see that the solution $f(x) = x$ works. Let us look for other solutions. If we have some other solution, then there must exist some value $r \neq 0$ such that $f(r) = -r$. So for any y ,

$$\begin{aligned} r^2 + y &= (f(r))^2 + y \\ &= f(r f(r) + f(y)). \end{aligned}$$

Let us square this last expression to see what we get.

$$\begin{aligned} [f(r f(r) + f(y))]^2 &= [r f(r) + f(y)]^2 \\ &= [-r^2 + f(y)]^2. \end{aligned}$$

Putting these two chains of identities together, we have

$$[r^2 + y]^2 = [-r^2 + f(y)]^2.$$

But this can only hold if $f(y) = -y$. It can be checked that this is a solution to the original functional equation.

We conclude that the only two solutions are $f(x) = x$ for all x and $f(x) = -x$ for all x .

- b) We begin by looking for simple polynomial solutions to the equation. If $f(x)$ is a polynomial, then each side of the equation is a polynomial in both x and y . Matching the degree in the variable x in both sides, we conclude that $[\deg f]^2 = 2[\deg f]$. This is solved by $\deg f = 0$ or

deg $f = 2$. Upon further inspection, we observe that $f(x) = 0$ works in the former case, and $f(x) = x^2$ works in the latter case.

Are there any other solutions? Suppose there exists some real number $c \neq 0$ such that $f(c) \neq c^2$. We show that $f(x) = 0$ for all x . Let us make the substitution

$$y = \frac{x^2 - f(x)}{2}$$

in our original functional equation. Then

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

Therefore

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

So for every x , either $f(x) = 0$ or $f(x) = x^2$. In particular, $f(0) = 0$. Additionally, $f(x) = f(-x)$ for all real x .

So if there are any solutions other than the two polynomial solutions, they can only be of the form where $f(x) = 0$ for some values of $x \neq 0$, and $f(x) = x^2$ for the other remaining nonzero values. Set $x = 0$ in the original functional equation. Then

$$f(-y) = f(y)$$

for all y . Suppose that there is some value $c \neq 0$ such that $f(c) = 0$. Now set $x = c$ in the original functional equation. Then

$$f(y) = f(c^2 - y).$$

Putting these last equations together, we have

$$f(y) = f(-y) = f(c^2 + y).$$

Because $c \neq 0$, we can conclude that f is periodic, with period at most c^2 .

Next, we substitute $y = c^2$ into the original equation, and use the periodicity that we have just found to conclude that

$$\begin{aligned} f(f(x)) &= f(f(x) + c^2) \\ &= f(x^2 - c^2) + 4f(x)c^2 \\ &= f(x^2) + 4c^2f(x). \end{aligned}$$

However, if we substitute $y = 0$ in the original functional equation, we get

$$f(f(x)) = f(x^2).$$

If we put these last two equations together, we see that $c^2f(x) = 0$ for all x . But by assumption $c \neq 0$. So $f(x) = 0$ for all x .

We conclude that there are only two solutions: $f(x) = 0$ for all x , and $f(x) = x^2$ for all x .

6. Consider the case where $f(x) = x$ for all x . This is a strictly monotonic function. Then $x = x g(y)$ for all x . It follows that one solution is $g(y) = 1$ for all y .

Are there any other solutions? Suppose f is any strictly monotonic function and g a corresponding solution. Set $y = 0$. Then

$$f(x) = f(x)g(0) + f(0)$$

for all x . In order for this to hold, we must have $g(0) = 1$ and $f(0) = 0$. Because f is strictly monotonic, we can conclude that $f(x) \neq 0$ for any value of $x \neq 0$.

Note that the left-hand side of the functional equation is symmetric in x and y . We can exploit that symmetry to conclude that

$$\begin{aligned} f(x)g(y) + f(y) &= f(x+y) \\ &= f(y)g(x) + f(x) \end{aligned}$$

for all x and y . Equivalently, we may write

$$\frac{g(y) - 1}{f(y)} = \frac{g(x) - 1}{f(x)}$$

for all $x \neq 0$ and all $y \neq 0$. This can only hold if the left-hand side and the right-hand side are in fact constants. So there is some constant k such that

$$\frac{g(x) - 1}{f(x)} = k$$

for all $x \neq 0$. Therefore, any solution for $g(x)$ can be written in terms of $f(x)$ as $g(x) = kf(x) + 1$ for $x \neq 0$. Because we already know that $f(0) = 0$ and $g(0) = 1$, this equation must also work for $x = 0$.

Let us consider the possible choices for $g(x)$ that this equation imposes. If $k = 0$, then $g(x) = 1$ for all x . But this is a solution that we have already found.

Suppose $k \neq 0$. Then

$$\begin{aligned} g(x+y) &= kf(x+y) + 1 \\ &= kf(x)g(y) + kf(y) + 1 \\ &= (g(x) - 1)g(y) + g(y) \\ &= g(x)g(y). \end{aligned}$$

So $g(x)$ is a strictly monotonic function satisfying Cauchy's exponential equation. It follows that $g(x) = a^x$ for some $a > 0$ and that $f(x) = k^{-1}(a^x - 1)$. This works. Thus we have all the solutions.

7. Consider the function $g(x) = f(x+1) - 1$. The function g is nondecreasing and satisfies

$$g(0) = 0, \quad g(-1) = -1.$$

Let $r < 0$ and $s > 0$. Then we can write $r = x - 1$ and $s = y - 1$ for $x < 1$ and $y > 1$. Therefore

$$\begin{aligned} g(-r s) &= g(-(x-1)(y-1)) \\ &= f(1 - (x-1)(y-1)) - 1 \\ &= f(x+y-xy) - 1 \\ &= f(x) + f(y) - f(x)f(y) - 1 \\ &= -(f(x)-1)(f(y)-1) \\ &= -g(r)g(s). \end{aligned}$$

This is equivalent to

$$g(rs) = -g(-r)g(s)$$

for all $r, s > 0$. The condition that the nondecreasing function g satisfies this multiplicative property with $g(0) = 0$ and $g(-1) = -1$ is equivalent to the original problem. So it suffices to find all such g .

We first consider the case where $g(1) = 0$. Then for any $s > 0$ we have

$$\begin{aligned} g(s) &= g(s1) \\ &= -g(-s)g(1) \\ &= -g(-s) \cdot 0 \\ &= 0. \end{aligned}$$

So any nondecreasing function g such that $g(-1) = -1$ and $g(s) = 0$ for $s \geq 0$ will satisfy the functional equation for g .

Next, we consider the case where $g(1) \neq 0$. Note that for all $x > 0$,

$$\begin{aligned} g(x) &= g(x1) \\ &= -g(-x)g(1). \end{aligned}$$

Define $h(x) = g(x)/g(1)$. Then for all $x, y > 0$,

$$\begin{aligned} h(xy) &= g(xy)/g(1) \\ &= -g(-x)g(y)/g(1) \\ &= [g(x)/g(1)]g(y)/g(1) \\ &= h(x)h(y). \end{aligned}$$

Moreover, $h(0) = 0$, and $h(1) = 1$. Because h is a monotone function satisfying these properties, we conclude that there exists some $k > 0$

such that $h(x) = x^k$ for all $x > 0$. Thus $g(x) = cx^k$, where $c = g(1)$. Furthermore for $x > 0$,

$$\begin{aligned} g(-x) &= -\frac{g(x)}{g(1)} \\ &= -\frac{cx^k}{c} \\ &= -x^k. \end{aligned}$$

Thus our full family of solutions for f is in the first case,

$$f(x) = \begin{cases} 1 & \text{for } x \geq 1, \\ 0 & \text{for } x = 0, \\ \text{any} & \text{for } x < 0. \end{cases}$$

(Here f is constrained additionally to be nondecreasing.) The full family of functions in the second case is

$$f(x) = \begin{cases} c(x-1)^k + 1 & \text{for } x > 1, \\ 1 & \text{for } x = 1, \\ -(1-x)^k + 1 & \text{for } x < 1. \end{cases}$$

Here, $c > 0$ and $k \geq 0$.

8. Yes, the function exists. In fact the function is unique up to a sign change. Let x be a positive rational number. Write $x = a/b$, where a and b are positive integers that are relatively prime. We apply the Euclidean algorithm to the pair (a, b) to construct a sequence of remainders r_0, r_1, \dots, r_n as follows. Set $r_0 = a$, $r_1 = b$. Then

$$r_{j-1} = q_j r_j + r_{j+1}$$

for $j = 1, \dots, n$, where finally $r_{n+1} = 0$. Because the value of n is a function of x , we can write $n(x)$ for this. For each $x \in \mathbf{Q}$, define

$$s(x) = \begin{cases} (-1)^{n(x)} & \text{for } x > 0, \\ 1 & \text{for } x = 0, \\ (-1)^{n(-x)+1} & \text{for } x < 0. \end{cases}$$

For example, if $x = 5/4$, then

$$r_0 = 5, \quad r_1 = 4, \quad r_2 = 1, \quad r_3 = 0$$

so that $n(5/4) = 2$. Thus $s(5/4) = (-1)^2 = 1$.

Then $s(x)$ satisfies the required conditions. For suppose $x + y = 0$ where x and y are distinct rational numbers. Then one is positive and the other negative, say $x > 0$ and $y < 0$. Then

$$\begin{aligned}
 s(x) &= (-1)^{n(x)} \\
 s(y) &= (-1)^{n(x)+1} \\
 &= -s(x) \\
 &= -[s(x)]^{-1}.
 \end{aligned}$$

Thus $s(x)s(y) = -1$.

Next, suppose that $x + y = 1$. Because x and y are distinct, neither is equal to $1/2$. Moreover, one of them is positive, say x . Write $x = a/b$ as above. Then $y = (b - a)/b$. There are two cases to consider: $y < 0$ and $y > 0$.

We consider the case $y < 0$ first. In this case,

$$\begin{aligned}
 s(y) &= (-1)^{n(-y)+1} \\
 &= (-1)^{n((a-b)/b)+1}.
 \end{aligned}$$

It is left to the reader to show that

$$n(a/b) = n((a - b)/b)$$

using the fact that $a > b$ for this case. (The two sequences generated by the Euclidean algorithm are the same once they get to $a \bmod b$ and $a - b \bmod b$, respectively. So $s(x)s(y) = -1$.)

The other case to consider is $y > 0$. Then

$$0 < x < \frac{1}{2} < y < 1.$$

It is left to the reader to show that $n(y) = n(x) + 1$. Once again $s(x)s(y) = -1$.

Finally, suppose that $xy = 1$. Then x and y are both positive or both negative. Thus it is sufficient to check that $n(y) = n(x) + 1$, which is left to the reader.

Bibliography

1. Abel, N. H. 1881: Détermination d'une fonction au moyen d'une équation qui ne contient qu'une seule variable. *Oeuvres complètes, vol. 2*, Christiania (Oslo), 36–39.
2. Aczél, J.: 1966, *Lectures on Functional Equations and Their Applications*. New York, Birkhäuser.
3. Aczél, J.: 1984, On history, applications and theory of functional equations. In *Functional Equations: History, Applications and Theory*, J. Aczél (ed.), Mathematics and Its Applications. Boston, Reidel.
4. Aczél, J. and Dhombres, J.: 1985, *Functional Equations Containing Several Variables*. Reading, Mass., Addison-Wesley.
5. d'Alembert, J.: 1769, Mémoire sur les principes de mécanique. *Hist. Acad. Paris 1769*, pp. 278–286.
6. Alexanderson, G. L., Klosinski, L. F., and Larson, L. C.: 1985, *The William Lowell Putnam Mathematical Competition. Problems and Solutions: 1938–1964*. Mathematical Association of America, Washington, DC.
7. Cauchy, A.-L.: 1821, *Cours d'analyse de l'École Polytechnique, Vol. 1. Analyse algébrique*. Chap. V, Paris (Oeuvres, Ser. 2, Vol. 3, Paris 1897, pp. 98–113, 220).
8. Gleason, A. M., Greenwood, R. E. and Kelly, L. M.: 1980, *The William Lowell Putnam Mathematical Competition. Problems and Solutions: 1965–1984*. Mathematical Association of America, Washington, DC.
9. Hamel, G.: 1905, Eine Basis aller Zahlen und die unstetigen Lösungen der Funktionalgleichung $f(x + y) = f(x) + f(y)$. *Math. Ann.* 60, pp. 459–462.
10. Kedlaya, K. S., Poonen, B., and Vakil, R.: 2002, *The William Lowell Putnam Mathematical Competition 1985–2000. Problems, Solutions and Commentary*. Mathematical Association of America, Washington, DC.
11. Koenigs, G.: 1884, Recherches sur les intégrales de certaines équations fonctionnelles. *Ann. Sci. Ec. Norm. Sup. (3)* 1, Supplement, pp. 3–41.
12. Kuczma, M., Choczewski, B. and Ger, R.: 1990, *Iterative Functional Equations*. Cambridge, UK, Cambridge University Press.

13. Lévy, P.: 1928, Fonctions à croissance régulière et itération d'ordre fractionnaire. *Ann. Mat. Pura Appl. (4)* 5, pp. 269–298.
14. Oresme, N.: c. 1352, *Tractatus de configurationibus qualitatum et motuum*. Paris (Ed. transl. and comm. M. Clagett, University of Wisconsin Press, Madison, 1968).
15. Pexider, H. W.: 1903, Notiz über Funktionaltheoreme. *Monatsh. Math. Phys.* 14, pp. 293–301.
16. Vincze, E.: 1962, Eine allgemeinere Methode in der Theorie der Funktionalgleichungen I, *Publ. Math. Debrecen* 9, pp. 149–163.
17. Young, G. S.: 1958, The linear functional equation. *Amer. Math. Monthly* 65, pp. 37–38.

Index

- Abel's equation, 15, 63, 64, 67, 76
 - generalization, 67
- Abelian group, 88
- Acceleration, 3
- Aczél, J., 3, 123
- Alembert, J. d', 9
 - also known as Darenberg, 9
 - also known as Rond, J. le, 9
 - d'Alembert's equation, 10, 17, 18, 45
 - Darenberg changed to d'Alembert, 9
 - law school, 9
 - parents and background, 9
 - placed in orphanage, 9
 - solutions to d'Alembert's equation, 10
 - supported by father financially, 9
- Algebra, 1
- Analysis, 6, 7
- Area, 6
- Aristotle, 2
- Arithmetic f.e., 82
- Arithmetic function, 82
- Arithmetic mean, 51
- Average, 37
- Axiom of Choice
 - nonconstructive existence theorem, 96
- Axiom of choice, 93, 96

- Böttcher's equation, 63, 66
- Babbage, C., 10, 14, 15, 19, 55, 61
 - analytical engine, 11
 - cowcatcher, 11
 - difference engine, 11
 - essay on calculus of functions, 12
 - flat rate postage, 11
 - skeleton keys, 11
- Bastille, 6
- Big Omega, 84
- Binet's formula, 20
- Binomial coefficients, 7
- Binomial theorem, 7, 35, 38
- Black Death, 1, 2
- Bonaparte, N., 7
- British mathematics, 11

- Calculus, 6–8, 17, 18
 - differential equations, 17, 18
 - integral equations, 17
 - Leibniz's methodology, 11
 - Newtonian methodology, 11
- Cambridge Analytical Society, 11
 - renamed Cambridge Philosophical Society, 11
- Canadian Mathematical Society, 16
- Carr, G. S., 22
- Cauchy's equation, 7, 24, 31, 35, 37–40, 43, 95
 - for functions bounded below, 34
 - for monotone function, 35
 - on a circle, 50
 - solution on rationals, 33
 - solution on reals, 34
 - tame solutions, 96
 - two variables, 43
 - two-variable solution, 44

- wild solutions, 96
- Cauchy's exponential equation, 8, 38, 39, 42
- Cauchy's inequality, 42, 50
- Cauchy, A.-L., 6–9, 18, 31, 93
 - Cauchy's equation, 8–10, 18, 31, 35
 - Cauchy's exponential equation, 38
 - early years, 6
 - family, 7
 - founder of analysis, 7
 - full solution to d'Alembert's equation, 10
 - problems with republicans and Napoleonic movements, 7
 - royalist sympathies, 7
 - starting point for theory of f.e., 6
- Challenges and puzzles, 16
- Circular cosine, 10
- Clagett translation, 2
- Classification of problems, 18
- Closure
 - under multiplication, 88
 - under reciprocation, 88
- College of Navarre, College of, 2
- Combinations of forces, 9
- Combinatorial arguments, 8
- Commutative group, 88
- Commutativity, 41
- Commutativity equation, 63, 67, 76
 - solutions for, 67
- Competition, 1, 16–19, 22, 84
 - 1995 Romanian Winter, 80
 - IMO, 17
 - IMO 1972, 19
 - IMO 1997 shortlist, 56, 71, 76
 - IMO 1998, 84
 - IMO 2001, 88
 - IMO 2001 jury, 88
 - IMO 2001 shortlisted problem, 87
 - IMO gold medal problem, 84
 - IMO medals, 84
 - problems, 17
 - Putnam, 17, 18
 - Putnam 1947, 18
 - Putnam 1955, 18
 - Putnam 1956, 18
 - Putnam 1959, 19
 - Putnam 1963, 19
 - Putnam 1966, 22
 - Putnam 1971, 62
 - Putnam 1972, 20
 - Putnam 1979, 70
 - Putnam 1988, 21
 - Putnam 1996, 59
- Completely multiplicative function, 83
- Composition of functions, 69
- Computing, 11
- Conic sections, 4, 6
- Conjugacy equation, 15, 62, 64, 67
 - finding solutions, 64
- Constant of integration, 17
- Continued fractions, 22
- Cosine
 - circular, 10, 48
 - hyperbolic, 10
- Cosine function, 10, 46, 48
 - double angle formula, 46
 - summation formula, 48
- Crux Mathematicorum, 16
- Cyclic function, 61
- d'Alembert's equation, 45, 49
- Degree of polynomial, 18, 79
- Dense set, 47
- Descartes, R., 2
 - graphs of functions, 2
- Dhombres, J., 3, 123
- Differential equations versus f.e., 19
- Differentiation, 8, 9
- Difform intensities, 2
- Direct and inverse calculation, 12
 - inverse more difficult, 12
- Directoire, 7
- Distance, 3
 - signed, 3
- Divergent series, 22
- Dubikajtis' theorem, 112
- Elementary methods without calculus, 18
- Enlightenment, 2
- Equation in two variables, 10
- Equivalence class, 58
- Equivalence relation, 58
 - reflexivity, 58
 - symmetry, 58
 - transitivity, 58
- Euler's equation, 39, 44

- generalized, 45
- solution, 44
- solution to generalized, 50
- Europe, 1
- Exhaustion, method of, 6
- Extreme value theorem, 34
- Fibonacci numbers, 20
- Fixed point under iteration, 59
- Fourteenth century, 1
- Fractional part, 58
- France, 7
- French Revolution, 6
- Function
 - analytic, 81
 - antisymmetric, 14
 - arithmetic, 82
 - bounded on interval, 7, 35, 45
 - completely multiplicative, 83, 84
 - complex-valued, 77
 - constant, 4, 9, 58
 - continuous, 7–9, 18, 25–28, 31, 33–35, 37–40, 43–45, 49–51, 58, 59, 69, 70, 75, 76, 78, 101
 - decreasing, 35, 51, 69, 70, 89
 - differentiable, 9
 - even, 47
 - homogeneous, 44
 - homogeneous of degree k , 44
 - increasing, 3, 35–37, 51, 69, 70, 89, 99, 107, 116
 - involution, 13–15, 61
 - linear, 1, 3, 6, 7, 9
 - monotone, 35, 75
 - nondecreasing, 89, 120
 - nowhere continuous, 9
 - nowhere differentiable, 9
 - of two variables, 43
 - one-to-one, 14, 15
 - periodic, 77, 80, 92
 - positive, 39
 - quadratic, 2
 - restricted domain, 18
 - strictly positive, 21
 - symmetric, 13, 52
- Functional equation
 - commutativity equation, 63, 76
 - conjugacy equation, 62, 64
 - existence of solution, 26
 - families of, 57
 - miscellaneous methods, 79
 - nonlinear, 23
 - one variable, 55
 - polynomial, 18, 79, 80
 - simultaneous, 24
 - single variable, 25
 - symmetries in, 91
 - too general, 42
 - two or more variables, 25, 31
 - uniqueness of solution, 26
- Functional equations versus differential equations, 19
- Göttingen University, 93
- Gamma function, 75, 107
- Geometric mean, 51
- Golden ratio, 20
 - as limiting ratio of Fibonacci sequence, 20
- Group, 88
 - commutative, 88
- Group-theoretic methods, 87
- Hamel bases, 35, 93, 95
 - uncountable, 95
 - uniqueness of representation, 94
- Hamel, G., 35, 93
 - 1905 paper of Hamel bases, 93
- Hardy, G. H., 73
- Hardy–Ramanujan number, 73
- Harmonic mean, 51
- Herschel, J., 11
- Herschel, W., 11
- Heuristics for solving f.e., 91
- Heytesbury, William of, 2
- High school mathematics, 16
- High school students, 16
- Hilbert's fourth problem, 93
- Hilbert, D., 93
- Hints to problems, 97
 - Chapter 1, 97
 - Chapter 2, 102
 - Chapter 3, 107
 - Chapter 4, 113
- History, 1, 17, 123
- Hyperbola, 4–6
- Hyperbolic cosine, 10, 46, 47
 - summation formula, 48

- Hyperbolic sine, 48
- Hyperbolic tangent, 48
- Identity, 88
- Identity function, 14
- Indeterminate expression, 18
- Induction, 24, 31, 46, 47, 83, 85
- Ingenuity in solutions, 91
- Integration, 4
- Inverse problem, 12
 - f.e. as, 12
- Italian Renaissance, 16
- Iteration, 58
- Iterative roots, 14, 15, 19, 55, 61, 69
- Iterative square root, 14, 20
- Jensen's equation, 37, 38
 - solution, 38
- Journal of the Indian Mathematical Society, 23
- Klein, F., 93
- Koenigs algorithm, 64, 76, 81
- Lévy algorithm, 66
- Legendre symbol, 84
- Limit, 25, 59, 64, 66
- Linear algebra, 62
- Linear functional equation, 38
 - solution, 38, 51
- Linear functions, 91
- Linearization, 55, 92
- Liouville function, 84
- Lisieux, 2
- Logarithms, 4–6, 8, 39, 91
- Mathematical Association of America, 16
- Mathematical Mayhem, 16
- Mathematics Magazine, 16
- Mean speed theorem, 2
- Mechanics, 2
- Medieval world, 2
- Merton College, 2
- Merton theorem, 2
- Multiplicative group, 88
- Multiplicative identity, 88
- Multiplicative inverse, 88
- Natural philosophy, 2
- Natural sciences, 2
- Nested radical, 21–23, 72
- Nested radicals, 72
- Newton, I., 7
 - derivation of binomial theorem, 7
- Nonconstructive existence, 96
- Nonlinear equation, 23
- Nth order f.e., 13
- Number theory, 84
- Numbers
 - complex, 77, 79
 - dyadic rationals, 47
 - Hardy-Ramanujan number, 73
 - perfect squares, 23
 - positive integers, 82
 - primes, 84, 85
 - rational, 32, 33, 50
 - real, 79
- Orbit, 58
- Oresme, 2
- Oresme, N., 1–4, 6
 - defined functional relationship, 2
 - Grand Master, 2
 - graphs of functions, 2
 - Master of Theology degree, 2
 - Tractatus, 2
- Osborne's rule, 48
- Oxford, 2
- Paris, 2, 7, 9
- Pascal's triangle, 7
- Peacock, G., 11
- Periodic function, 77, 80
- Pexider's equation, 39, 40
 - relation to Cauchy's equation, 40
 - solution, 40
- Physics, 2
- Pi (3.14159...), 22
- Planar region, 6
- Planck, M., 93
- Poincaré's equation, 62, 67
- Polynomial f.e., 79, 80
- Polynomials, 72, 76, 79–81, 91
 - degree, 72, 79, 80
 - roots, 80, 81
- Power function, 39, 82, 83, 91, 109
- Power series, 81, 82
 - equality of coefficients, 82

- Prime factorisation, 83
- Prime factors, 84
 - multiplicity, 84
- Primes, 84
- Problem-solving journals, 16
- Problems, 26, 49, 75, 89
 - for Chapter 1, 26
 - for Chapter 2, 49
 - for Chapter 3, 75
 - for Chapter 4, 89
- Punched cards, 11
- Quadratic equation, 14
- Quasiarithmetic mean, 51
 - arithmetic mean, 51
 - geometric mean, 51
 - harmonic mean, 51
 - homogeneous, 52
 - root mean square, 51
 - translative, 52
- Quasiarithmetic means
 - equivalence, 52
- Ramanujan, S., 21–24, 73
 - birthplace, 22
 - contacts in Indian mathematical community, 23
 - education, 22
 - in *J. Indian Math. Soc.*, 23
- Recreational mathematics, 16
- Recursion, 21, 82
- Recursions, 20
- Reflexivity, 58
- Regularity conditions, 18
- Renaissance, 2
- Representations of groups, 17
- Root mean square, 51
- Rousseau, Mme., 9
- Royal Society of London, 12
- Saint-Jean-le-Rond church, 9
- Saint-Vincent, Gregory of, 4–6
 - Opus geometricum, 4, 6
- Scale factor, 6
- Schröder's equation, 15, 62, 81
 - generalization, 67
 - principal solution, 65
- Schwarz, H., 93
- Second derivative, 17
- Separation of variables, 19
- Sequence, 21, 57
- Sequences, problems involving, 18
- Series, 6, 8, 81, 82
 - convergence, 8, 81
 - divergence, 8
 - rules for multiplying, 8
- Set
 - closed under multiplication, 88
- Shrinking transformation, 6
- Signum function, 15
- Simultaneous f.e., 13, 24
- Sine, 48
 - hyperbolic, 48
- Splinter, 58–61, 64, 65, 92, 108, 109
 - limit of, 59
- Squeeze theorem, 36
- Stretch transformation, 6
- Stretching transformation, 6
- Subadditivity, 50
- Symmetry, 58
- Taiwan, 84
- Tangent, 48
 - hyperbolic, 48
- Transitivity, 58
- Trigonometric functions, 10, 91
- Trigonometric identity, 10, 48
- Uncountable basis, 95
- Undergraduate mathematics, 16
- Uniform difformity, 3
- Uniform functions, 4
- Uniform intensities, 2
- Uniform motion, 3
- Uniform quality, 3
- Uniformly difform function, 4
- Uniformly difform motion, 3
- Uniformly difform quality, 3
- University mathematics, 17
- University of Berlin, 93
- University of Paris, 2
- Vector space of \mathbf{R} over \mathbf{Q} , 95
- Vincze's equation, 40
 - first solution, 41
 - second solution, 42
- Washington, DC, 88
- Woodhouse, R., 11
- Year 1348, 2