Topics in the theories of means and discrete dynamical systems

Master's thesis submitted by

LARS VAN DEN BERG

Supervisor: PROF. DR. GUNTHER CORNELISSEN

Second reader: DR. VALENTIJN KAREMAKER

Department of Mathematics
Faculty of Science
Utrecht University

21st August 2020

Abstract

In Part I, we study means, and prove new results about them. In particular, we define and study the translation means, we prove new results about Gauss composition of means, and we define and study families of means.

In Part II, we study discrete dynamical system. We enhance the proof of a theorem about random discrete dynamical system. Further, we study generalisations of dynamical systems related to the Collatz conjecture, in our analysis of which we use the translation mean.

Acknowledgements

I heartily thank Gunther Cornelissen for being an excellent and knowlegdable teacher and supervisor, for his inspiration and patience, and for the motivating and enjoyable conversations we had. I heartily thank Valentijn Karemaker for taking the effort of being the second reader of this text.

I heartily thank my loved ones, friends, family, family-in-law, for their affection and support, and for being good for me in various ways during the hectic time of researching.

Some of them helped me directly with making progress on this thesis. Berend Ringeling and Harry Smit proofread parts of it, gave me professional advise, and were motivating mentors, especially in the final seven weeks of writing. Berend resolved my headaches by sending me the paper [DMP05], which provided the first example of a non-quasi-arithmetic mean that I knew of (see Theorem 3.8.8). Rob van den Berg also helped me with my struggles on non-quasi-arithmetic Gauss compositions, and he introduced me to the theory of Poisson approximations, relevant in the context of Theorem 6.1.5. Claudia Wieners and Isabel Droste helped me so much with their involvement in the process and keeping me motivated; I want to thank in particular them for their love and care during my research. Claudia learnt me about usage in scientific practice of the harmonic and geometric means (see Example 1.2.5). Isabel explained me how she uses the power mean PM_{-2} in her research.

My friends and former fellow students Evelien van Dijk and Joost Franssen formed together with Berend and me a "thesis talking group" in the fall semester of 2018, which was very motivating and enjoyable, and which helped to reduce the loneliness of doing a master thesis.

I heartily thank my colleagues at Finaps, for their flexibility and coulance, their understanding that I was sometimes a bit overworked, and their interest in my progress on this text. I thank in particular the managing directors Marlies van Sonsbeek and Lonneke Makhaija, and my team members Jochem van Eldijk, Tim Pijl, Pauline ten Brinke and Andrew Hagens.

Contents

Global definitions and notational conventions

Basics

1. Let $A \subseteq \mathbb{R}$ and let $t \in \mathbb{R}$. We write $A_{\leq t} := \{a \in A : a \leq t\}$.
   Analogously for $A_{\geq t}$ and $A_{<t}$ and $A_{>t}$. Further, we write $A_{\neq t} := A \setminus \{t\}$.
2. We write $\mathbb{N} := \mathbb{Z}_{\geq 1}$, and $\mathbb{N}_0 := \mathbb{Z}_{\geq 0}$.
3. Let X be a set and let $f, g : X \to \mathbb{R}$ be functions. We write $f \leq g$ to denote that the inequality holds *pointwise*, that is, $f(x) \leq g(x)$ for all $x \in X$.
   Analogously for $f \geq g$ and $f < g$ and $f > g$.

Analytical

1. Let $m, n \in \mathbb{N}$, let A be an open subset of $\mathbb{R}^m$, let $f : A \to \mathbb{R}^n$ be a function.
   We say that f is *smooth* if f is of differentiability class C^∞.

Let Y be a subspace of a topological space X, let $f, g : Y \to \mathbb{C}$ be functions,
let $p \in X$ be a limit point of Y.

1. We write "$f(x) \sim g(x)$ as $x \to p$" to denote that that $\lim_{x \to p} f(x)/g(x) = 1$.
2. We write "$f(x) = O(g(x))$ as $x \to p$" to denote that there exists $M \in \mathbb{R}_{>0}$ and an open neighbourhood U of p such that g is nonzero on U, and $|f/g| \leq M$ on U.
3. We write "$f(x) = o(g(x))$ as $x \to p$" to denote that $\lim_{x \to p} f(x)/g(x) = 0$.

Probabilistic

1. We denote the probability of the event A by $\mathbf{P}(A)$.
2. We denote the expectation of the random variable X by $\mathbf{E}(X)$.

Miscellaneous

1. We denote the tuples $(0, 0, \ldots, 0), (1, 1, \ldots, 1) \in \mathbb{Z}^n$ by $\mathbf{0}$ and $\mathbf{1}$ respectively, for all $n \geq 1$.
2. Let A be a set, let $n \in \mathbb{N}$, let $\mathbf{x} = (x_1, \ldots, x_n) \in A^n$. For $1 \leq k \leq n$,
   we write $\pi_k(\mathbf{x})$ to denote the kth entry of $\mathbf{x}$, that is, $\pi_k(\mathbf{x}) = x_k$.
3. We write $\overline{\mathbb{R}}$ for the extended real line $\mathbb{R} \cup \{-\infty, \infty\}$.
   Topologically, $\overline{\mathbb{R}}$ is the two-point compactification of $\mathbb{R}$.

Non-mathematical

A result (or definition) is tagged with (*) or (†) to signify that I consider it interesting in its own right, and that I expect that:

(*) it is previously unknown (or undefined), at least in its generality;

(†) it is known by experts, but there is no previously published proof (or definition) of it, at least in its generality.

Introduction

This text consists of two parts: Part I is about means, Part II is about discrete dynamical systems. The two parts are relatively independent from each other, but there are some connections. Gauss composition of means, which we consider in Chapter 3, is obtained by iterating a transformation, and iteration of transformations is the subject of discrete dynamical systems. We do however not use any explicit theory of discrete dynamical systems in our study of Gauss composition. A more important connection between the two parts is that we define a type of means, the *translation means*, that arise naturally when we study dynamical systems that are related to the Collatz conjecture. We use properties of the translation means to easily prove some basic properties of those dynamical systems.

The chapters about means in Part I build upon each other. They contain some classical results, but also quite some new results about means, especially about Gauss composition of means, and about "families of means" (which is a definition of our own). The translation means are also a definition of our own, and we prove properties of them.

Chapter 1 is an introduction to means and their basic properties. One of the most fundamental operations on means, conjugation, is introduced at the and of Chapter 1 and further studied in Chapter 2, where we study in particular the power means and the translation means. Chapter 3 deals with another very interesting operation on means, namely Gauss composition. Chapter 4 studies the way in which means in the same "family" are related with each other.

The chapters 5,6,7 about discrete dynamical systems rely less heavily on each other than the chapters about means. Chapter 5 deals with basic definitions and properties of discrete dynamical systems, and we illustrate them with some interesting examples. Chapter 6 studies stochastic properties of random discrete dynamical systems on finite sets. Chapter 7 is about very specific discrete dynamical systems on the natural numbers: namely, systems that are related to the Collatz conjecture.

Main new results in this text (in order of appearance)

1. Results about how the "compressing properties" of means (see Definition 1.1.14.7 and Lemma 1.1.28) are preserved by conjugation and composition of means: Theorem 2.1.3.4, Proposition 2.2.1, Corollary 2.2.2; for the power and translation means: Theorem 2.6.6.5 and Theorem 2.7.4.4; for composition: Theorem 3.1.1.

2. A proof of the continuity of the "full power mean map": Theorem 2.6.6.6

3. Properties of the translation means: Theorem 2.7.4. They imply a new proof of the classical inequality of pythagorean means: Corollary 2.7.6.

4. Gauss composition commutes with conjugation: Theorem 3.5.3. Under certain conditions, Gauss composition commutes with pointwise limit: Theorem 3.6.3; this implies the "continuity theorem" 3.6.5 for Gauss composition.

5. Properties of Gauss composites of power means and Gauss composites of translation means: Theorem 3.7.4. Uncountably many of them are not quasi-arithmetic: Theorem 3.8.6.

6. Power means and translation means form families of means: Corollary 4.3.7

7. Combinatorial and analytical properties of families of means: Theorem 4.4.1, Theorem 4.6.1, and Theorem 4.7.3.

8. A new proof of the result that if $\mathbf{W} \in \mathbb{R}_{>0}^{n\times n}$ is a stochastic matrix, then the rows of the stochastic matrix $\lim_{m\to\infty} \mathbf{W}^m$ are equal to each other: Theorem 4.7.5.

9. A formula for Gauss composition of means of the same family: Theorem 4.7.7 and Corollary 4.7.8.

10. More accurate error terms in some asymptotic formulas about random finite flow graphs: Theorem 6.1.5.

11. Relations between (on the one hand) flows and cycles of generalised Collatz transformations, and (on the other hand) the arithmetic and translation means: Theorem 7.5.4. The theorem implies a new proof of Corollary 7.5.7, that "the average exponent in a cycle of a generalised Collatz transformation is close to a logarithm".

Part I

Topics in the theory of means

Chapter 1

An introduction to means

The Pythagorean means – the arithmetic mean, the harmonic mean and the geometric mean – are well-known. Less known is that there is a general notion of what a "mean" is, and a well-established literature about means, for example the overview works [Bul03], [HLP34], [Haj13], and research papers on specific topics like [MP15] and [DMP05].

In this introductory chapter, we define what a mean is, and what additional properties it may or may not have; we give concrete examples to illustrate the new notions, and prove basic results. In the last part, we define conjugation of means, which is the subject of study of the next chapter.

1.1 Means: basic definitions, properties, and examples

Definition 1.1.1. Let $A \subseteq \mathbb{R}$, let $n \in \mathbb{N}$. We denote the minimum and maximum functions $A^n \to A$ by $\mathrm{Min}_{A,n}$ and $\mathrm{Max}_{A,n}$. That is, for $\mathbf{x} = (x_1, \ldots, x_n) \in A^n$,

$$\mathrm{Min}_{A,n}(\mathbf{x}) := \min\{x_1, \ldots, x_n\}, \qquad \mathrm{Max}_{A,n}(\mathbf{x}) := \max\{x_1, \ldots, x_n\}.$$

When A or n is clear from the context, we may hide A or n from the notation.

Definition 1.1.2 (Mean)**.** Let $A \subseteq \mathbb{R}$, let $n \in \mathbb{N}$, let $M : A^n \to \mathbb{R}$ be a function. We say that M is a *mean (on A^n)* if

$$\mathrm{Min}_{A,n} \leq M \leq \mathrm{Max}_{A,n}.$$

If moreover M maps A^n into A, we say that M is an *internal mean (on A^n)*.

***Remark* 1.1.3.** If A is an interval of $\mathbb{R}$, then every mean on A^n is an internal mean.

In definitions of means that I read in the literature (for instance in [CFKT14] and [Haj13]), means are restricted to A^n where A is an interval, and/or where $n = 2$. In particular, there is no distinction between means and internal means.

The definition of means in [HLP34] is different; it is closer to what we call "families of means" in Chapter 4.

Example 1.1.4. $(x_1, \ldots, x_n) \mapsto (x_1 + \cdots + x_n)/n$ defines an internal mean on $\mathbb{Q}^n$. $(x_1, \ldots, x_n) \mapsto (x_1 \cdots x_n)^{1/n}$ defines a mean on $\mathbb{Q}^n_{>0}$, and an internal mean on $\mathbb{R}^n_{>0}$.

Example 1.1.5. Let $M : \mathbb{Q}^n_{>0} \to \mathbb{Q}_{>0}$ map $(x_1, \dots, x_n)$ to $(x_1 \cdots x_n)^{1/n}$ if $x_1 \cdots x_n$ is an nth power in $\mathbb{Q}$, and to $(x_1 + \cdots + x_n)/n$ otherwise. Clearly, M is an internal mean.

Before we study more examples, we introduce in §1.1.1 some additional properties that a mean may or may not have, so that we can compare means by comparing their properties. Before we introduce those properties, we make a few simple definitions and fix some notations.

Definition 1.1.6 (Diagonal)**.** Let $A \subseteq B$ be any two sets, let $n \in \mathbb{N}$.

1. We write $\mathrm{diag}_{A,n}$ or simply diag_n for the function $A \to A^n : a \mapsto (a, \dots, a)$.
2. We write $\mathrm{diag}(A^n)$ for the image of $\mathrm{diag}_{A,n}$.
3. Let $M : A^n \to B$ be a function. We say that M is *constant-preserving* if $M \circ \mathrm{diag}_{A,n} = \mathrm{id}_A$.

Definition 1.1.7 (Vector comparison)**.** Let $\mathbf{x} = (x_1, \dots, x_n), \mathbf{y} = (y_1, \dots, y_n) \in \mathbb{R}^n$.
We write $\mathbf{x} \leq \mathbf{y}$ to denote that $x_i \leq y_i$ for all i.
We write $\mathbf{x} < \mathbf{y}$ to denote that $\mathbf{x} \leq \mathbf{y}$, and that $x_i < y_i$ for *at least one* i.

Definition 1.1.8 (Monotonic multivariable functions)**.** Let $f : A^n \to \mathbb{R}$ be a function, for some $A \subseteq \mathbb{R}$ and $n \in \mathbb{N}$. Then

$$f \text{ is called } \left\{ \begin{array}{l} \textit{(monotonically) increasing} \\ \textit{strictly increasing} \\ \textit{(monotonically) decreasing} \\ \textit{strictly decreasing} \end{array} \right. \quad \text{if} \quad \left\{ \begin{array}{l} \mathbf{x} \leq \mathbf{y} \implies f(\mathbf{x}) \leq f(\mathbf{y}) \\ \mathbf{x} < \mathbf{y} \implies f(\mathbf{x}) < f(\mathbf{y}) \\ \mathbf{x} \leq \mathbf{y} \implies f(\mathbf{x}) \geq f(\mathbf{y}) \\ \mathbf{x} < \mathbf{y} \implies f(\mathbf{x}) > f(\mathbf{y}). \end{array} \right.$$

We say that f *is monotonic* to indicate that f is either increasing or decreasing, and f *is strictly monotonic* to indicate that f is either strictly increasing or strictly decreasing.

Fact 1.1.9. *Let $A \subseteq \mathbb{R}$, let $n \in \mathbb{N}$, let $M : A^n \to \mathbb{R}$ be a function.*

1. *Suppose M is a mean. Then M is constant-preserving.*
2. *Suppose M is constant-preserving and increasing. Then M is a mean.*

Proof. 1. For $a \in A$, we have $a = \mathrm{Min}(\mathrm{diag}_n(a)) \leq M(\mathrm{diag}_n(a)) \leq \mathrm{Max}(\mathrm{diag}_n(a)) = a$.
2. Let $\mathbf{x} = (x_1, \dots, x_n) \in A^n$, say $x_i = \mathrm{Min}(\mathbf{x})$ and $x_j = \mathrm{Max}(\mathbf{x})$. Then

$$\mathrm{Min}(\mathbf{x}) = M(\mathrm{diag}_n(x_i)) \leq M(\mathbf{x}) \leq M(\mathrm{diag}_n(x_j)) = \mathrm{Max}(\mathbf{x}). \qquad \square$$

Definition 1.1.10. Let $n \in \mathbb{N}$, let $U \subseteq \mathbb{R}^n$, let $f : U \to \mathbb{R}$ be a function, let F be a subfield of $\mathbb{R}$. We say that f is *F-rational* if f is the quotient of two polynomial functions on U with coefficients in F, such that the denominator nowhere vanishes on U.

Definition 1.1.11. Let A be a set, let $n \in \mathbb{N}$, let $\mathbf{x} = (x_1, \dots, x_n) \in A^n$.

1. Let $i \in \mathbb{N}_{\leq n}$, let $t \in A$. We write $\mathbf{x}_{[i,t]}$ for the element of A^n that we get by replacing the ith variable of $\mathbf{x}$ by t. That is,

$$\mathbf{x}_{[i,t]} = (x_1, \dots, x_{i-1}, t, x_{i+1}, \dots, x_n) \in A^n.$$

2. Let $m \in \mathbb{N}$, and let $f : \mathbb{N}_{\leq m} \to \mathbb{N}_{\leq n}$ be a function. We write
$$\mathbf{x}_{\circ f} := (x_{f(1)}, x_{f(2)}, \dots, x_{f(m)}) \in A^m.$$

3. We write S_n for the permutation group of $\mathbb{N}_{\leq n}$.

***Remark* 1.1.12.** In the context of part 2 of Definition 1.1.11, let $g : \mathbb{N}_{\leq k} \to \mathbb{N}_{\leq m}$ be another function. Then $(\mathbf{x}_{\circ f})_{\circ g} = \mathbf{x}_{\circ (f \circ g)} \in A^k$. Further, $\mathbf{x}_{\circ \,\mathrm{id}} = \mathbf{x}$. In particular, the map $S_n \times A^n \to A^n : (\sigma, \mathbf{x}) \mapsto \mathbf{x}_{\circ \sigma}$ is a right group action of S_n on A^n.

1.1.1 Properties that a mean may or may not have

The definition of means that we gave, is very general, so that there is not much interesting to prove about them in full generality. Therefore, we define some properties, so that we can prove more interesting things about means with certain properties.

***Remark* 1.1.13.** Any property of a mean can also be "considered locally": if M is a mean on A^n, and $B \subseteq A$, we say that "M has property X on B" to denote that the mean $M|_{B^n}$ on B^n has property X.

Definition 1.1.14. Let $A \subseteq \mathbb{R}$, let $n \in \mathbb{N}$, let $M : A^n \to \mathbb{R}$ be a mean.

1. M is called ***increasing*** or ***strictly increasing*** respectively if it is so in the sense of Definition 1.1.8.

2. M is called ***continuous*** if M is a continuous function on A^n.
   M is called ***smooth*** if A is open in $\mathbb{R}$, and M is a smooth function on A^n.

3. For a subfield F of $\mathbb{R}$, we say that M is ***F-rational*** if M is an F-rational function. In the case that $F = \mathbb{R}$, we simply say that M is *rational*.

4. M is called ***scale-invariant*** if A is closed under multiplication by $A_{>0}$, and
$$M(\lambda \mathbf{x}) = \lambda M(\mathbf{x}) \qquad \forall(\lambda \in A_{>0}) \forall(\mathbf{x} \in A^n).$$

5. M is called ***symmetric*** if
$$M(\mathbf{x}_{\circ \sigma}) = M(\mathbf{x}) \qquad \forall(\sigma \in S_n) \forall(\mathbf{x} \in A^n).$$

6. M is called ***strict*** if for all $\mathbf{x} \in A^n \setminus \mathrm{diag}(A^n)$, we have $\mathrm{Min}(\mathbf{x}) < M(\mathbf{x}) < \mathrm{Max}(\mathbf{x})$.

7. (†) Let $(c, C) \in \mathbb{R}^2$. We say that M is ***$[c, C]$-compressing*** if $0 < c \leq C$, and M is strict, and for all $\mathbf{x} \in A^n \setminus \mathrm{diag}(A^n)$
$$\frac{\mathrm{Max}(\mathbf{x}) - M(\mathbf{x})}{M(\mathbf{x}) - \mathrm{Min}(\mathbf{x})} \in [c, C].$$
   When $C = c^{-1}$, we simply write $[c]$-compressing instead of $[c, c^{-1}]$-compressing.

   We say that M is ***compressing*** if for all $(a, b) \in A^2$ with $a < b$, there exists (c, C) such that M is $[c, C]$-compressing on $A \cap [a, b]$.

8. (†) Let $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$, say $\mathbf{a} = (a_1, \ldots, a_n)$. Suppose that $A \subseteq \mathbb{R}_{>0}$.

   We say that M is $(\cdot, \mathbf{a})$-*power-asymptotic* if A is unbounded, and for all $\mathbf{x} \in A^n$ there exists $c \in \mathbb{R}_{>0}$ such that $M(\mathbf{x}_{[i,t]}) \sim ct^{a_i}$ as $t \to \infty$.

   We say that M is $(\mathbf{a}, \cdot)$-*power-asymptotic* if 0 is a limit point of A, and for all $\mathbf{x} \in A^n$ there exists $c \in \mathbb{R}_{>0}$ such that $M(\mathbf{x}_{[i,t]}) \sim ct^{a_i}$ as $t \to 0$.

   We say that M is $(\mathbf{a}, \mathbf{b})$***-power-asymptotic*** if it is both $(\cdot, \mathbf{b})$-power-asymptotic and $(\mathbf{a}, \cdot)$-power-asymptotic. If $\mathbf{a} = \mathrm{diag}_n(a)$ for some $a \in \mathbb{R}$, we simply say that M is $(a, \mathbf{b})$-*power-asymptotic*, likewise for $(\cdot, a)$ et cetera.

***Remark* 1.1.15** (Remarks about Definition 1.1.14)**.**

1. The properties listed in 1, 2, 4, 5 and 6 of the definition are fairly standard in the literature on means, they are for instance used in [Bul03] and [HLP34]. The other parts are, as far as I know, not standard.

   In some sources, for instance [Bul03] and [HLP34], the word 'homogeneous' is used instead of 'scale-invariant'.

2. We might have defined scale-invariance by requiring that there exists $p \in \mathbb{R}$ such that $M(\lambda\mathbf{x}) = \lambda^p M(\mathbf{x})$, for all $\lambda, \mathbf{x}$, which would have been closer to the common usage of the word scale-invariance, but the fact that M is constant-preserving would imply (except in the trivial case that $A \subseteq \{0\}$) that $p = 1$.

3. If M is $(\mathbf{a}, \mathbf{b})$-power-asymptotic, then $\mathbf{a} \in [0,1]^n$ and $\mathbf{b} \in [0,1]^n$. Otherwise, the condition $\mathrm{Min}(\mathbf{x}) \leq M(\mathbf{x}) \leq \mathrm{Max}(\mathbf{x})$ would be violated for sufficiently small $\mathbf{x}$ or for sufficiently large $\mathbf{x}$.

4. We investigate the "compressing" property and explain its name in §1.1.4.

***Remark* 1.1.16** (Means on A^n where $n = 1$)**.** Let $A \subseteq \mathbb{R}$. There exists only one mean on A, namely id_A. It is strictly increasing, continuous, smooth if A is open, $\mathbb{Q}$-rational, scale-invariant, symmetric, compressing (a vacuous truth, since $A = \mathrm{diag}(A)$), and $(1,1)$-power-asymptotic. There is nothing more to say about the case $n = 1$. We will therefore tacitly exclude this case in the statements of several results.

1.1.2 Examples and facts about means and their properties

Fact 1.1.17 (Properties of Min and Max.)**.** *Let $n \in \mathbb{N}_{\geq 2}$. The means* Min *and* Max *on $\mathbb{R}^n$ are increasing, continuous, scale-invariant and symmetric, but not strictly increasing, not smooth, not rational, not compressing, and not power-asymptotic.*

The same holds for the means Min *and* Max *on $\mathbb{R}^n_{>0}$, except that they are power-asymptotic:* Min *is $(1,0)$-power-asymptotic,* Max *is $(0,1)$-power-asymptotic.*

Proof. All statements are clear, except perhaps for the "power-asymptotic on $\mathbb{R}^n_{>0}$" statements, which we show for Min on $\mathbb{R}^n_{>0}$, it follows analogously for Max. Clearly, for all sufficiently small t we have $\mathrm{Min}(\mathbf{x}_{[i,t]}) = t = 1t^1$. On the other hand, for all sufficiently large t, there is $d > 0$ such that $\mathrm{Min}(\mathbf{x}_{[i,t]}) = d = dt^0$. □

Definition 1.1.18. Let $n \in \mathbb{N}_{\geq 2}$. The *arithmetic mean*, the *geometric mean* and the *harmonic mean* are

$$\begin{array}{lll} \mathrm{AM}_n: & \mathbb{R}^n \to \mathbb{R}: & (x_1,\dots,x_n) \mapsto (x_1+\cdots+x_n)/n, \\ \mathrm{GM}_n: & \mathbb{R}^n_{>0} \to \mathbb{R}_{>0}: & (x_1,\dots,x_n) \mapsto (x_1\cdots x_n)^{1/n}, \\ \mathrm{HM}_n: & \mathbb{R}^n_{>0} \to \mathbb{R}_{>0}: & (x_1,\dots,x_n) \mapsto \left((x_1^{-1}+\cdots+x_n^{-1})/n\right)^{-1}. \end{array}$$

***Remark* 1.1.19.** The means in Definition 1.1.18 are called the *pythagorean means.* We consider them, and their "weighted" generalisations, more extensively in §1.2.

It is easily seen that AM_n, GM_n and HM_n are strictly increasing, smooth, scale-invariant, symmetric, strict means. By Lemma 1.1.26 below, they are compressing. Moreover, AM_n is $\mathbb{Q}$-rational and $(0,1)$-power-asymptotic, HM_n is $\mathbb{Q}$-rational and $(1,0)$-power-asymptotic, and GM_n is $(\frac{1}{n},\frac{1}{n})$-power-asymptotic. All these statements are encompassed and generalised by Proposition 1.2.6. We study further generalisations in §2.6 and §2.7: the *power means* and *translation means.*

The following fact (its proof is trivial) will be used extensively when we consider conjugation of means (in §1.3 and below), because in many cases we can only "conjugate" a mean after restricting it to a subset of its domain.

Fact 1.1.20 (Restricting means to subsets)**.** *Let $B \subseteq A \subseteq \mathbb{R}$, let $n \in \mathbb{N}$, let M be a mean on A^n. For any property in Definition 1.1.14, if M has the property, then the mean $M|_{B^n}$ has the property as well, – provided, in the case of 'smooth', 'scale-invariant' and 'power-asymptotic', that B satisfies the conditions stated for A.*

***Remark* 1.1.21.** In the context of Fact 1.1.20, suppose $B = A \cap I$, where I is an interval of $\mathbb{R}$, and suppose M is an *internal* mean. Then M maps $(A\cap I)^n$ into $A \cap I$, so M restricts to an *internal* mean on $(A \cap I)^n$.

The remainder of this section up to §1.1.4 consists almost entirely of some examples of means, to illustrate how diverse and peculiar they can be.

Example 1.1.22. The internal mean on $\mathbb{Q}^n_{>0}$ from Example 1.1.5 does not satisfy any of the properties listed in Definition 1.1.14, except that it is symmetric and compressing (and hence strict). This is easily seen, for instance by noting that the subset of nth powers is dense in $\mathbb{Q}_{>0}$, and that $M(\mathrm{diag}_n(1)_{[1,t]})$ equals $t^{1/n}$ if t is an nth power, and otherwise equals $t/n + (n-1)/n$.

Example 1.1.23. Let p_n denote the nth prime number, so $(p_1, p_2, \dots) = (2, 3, \dots)$.

We define $M : \mathbb{R}^2_{>0} \to \mathbb{R}^2_{>0}$ as follows. For $x, y \in \mathbb{R}^2_{>0}$ with $x \leq y$, let k be the number of primes in the interval $[x, y]$; if $k \geq 1$, we denote the smallest prime in that interval by p_{j+1}. We define

$$M(x,y) = M(y,x) = \begin{cases} \mathrm{GM}_k(p_{j+1},\dots,p_{j+k}) & \text{if } k \geq 2, \\ \mathrm{GM}_2(x,y) & \text{if } k \in \{0,1\}. \end{cases}$$

Clearly, M is a symmetric, discontinuous mean on $\mathbb{R}^2$, and not strictly increasing. It is however not hard to see that M is increasing. Further, M is compressing; this is not hard to see, using that on a fixed interval $[a,b]$, we need to consider only finitely many geometric means GM_k, and that each geometric mean GM_k is compressing.

Moreover, it turns out that M is $(\cdot,1)$-power-asymptotic. Namely, in [SV11] it is showed that

$$\lim_{n\to\infty} P_n = e^{-1}, \qquad \text{where } P_n := \frac{(p_1\cdots p_n)^{1/n}}{p_n}.$$

The proof is fairly short; the authors proceed by writing

$$\log(P_n) = \frac{\theta(p_n)-p_n}{n} + \frac{p_n}{n} - \log(p_n), \tag{1.1}$$

where $\theta(x) := \sum_{p_i\le x} p_i$ is known as Chebyshev's function; they use a well-established upper bound for $|\theta(x)-x|$, and the prime number theorem, to show that the first term in (1.1) approaches 0 as $k\to\infty$; and they show that the sum of the other two terms approaches -1, which may be seen as a strong variation on the prime number theorem.

If $x_1, x_2, \dots$ is a sequence in $\mathbb{R}_{>0}$, such that $\log(x_n) = o(n)$ as $n\to\infty$, then it is easily seen, by considering the logarithm of the quotient, that for any fixed $j\ge 0$,

$$\lim_{n\to\infty} \frac{\mathrm{GM}_n(x_1,\dots,x_n)}{\mathrm{GM}_{n-j}(x_{j+1},\dots,x_n)} = 1.$$

The sequence of prime numbers satisfies $\log(p_n) = o(n)$ as $n\to\infty$ (for instance by the prime number theorem). Thus, writing $\pi(y)$ for the number of primes smaller or equal to y, we have for any fixed x,

$$M(x,y) \sim \mathrm{GM}_{\pi(y)}(p_1,p_2,\dots,p_{\pi(y)}) = p_{\pi(y)}P_{\pi(y)} \sim yP_{\pi(y)} \sim y/e \quad \text{as } y\to\infty.$$

We used that $p_{\pi(y)}\sim y$ as $y\to\infty$, which follows because $p_{\pi(y)} \le y \le p_{\pi(y)+1}$, and because by the prime number theorem, $p_k/p_{k+1}\to 1$ as $k\to\infty$. Thus, $M(x,y)\sim y/e$ as x is fixed and $y\to\infty$; in particular, M is $(\cdot,1)$-power-asymptotic. ♪

1.1.3 Means on A^n where A is not an interval

As we remarked in Remark 1.1.3, means are usually only considered on A^n where A is an interval of $\mathbb{R}$. Actually, all the concrete examples of means that we encounter in §1.2 and below, are like that; typically, A equals $\mathbb{R}^n_{>0}$ for some $n\ge 2$. In this subsection, we consider some aspects of some means that could not be encountered for means on A^n where A is an interval of $\mathbb{R}$.

Fact 1.1.24. *Let $A\subseteq\mathbb{R}$, let $n\in\mathbb{N}_{\ge 2}$. Suppose there is an interval $I\subseteq\mathbb{R}$ such that $|A\cap I|\in\mathbb{N}_{\ge 2}$. There exists no strictly increasing internal mean on A^n.* ♩

Proof. Let $a,b\in A\cap I$ such that $a<b$, so $k := |A\cap(a,b]|\in\mathbb{N}$. There clearly exists a sequence $\mathbf{x}_0,\mathbf{x}_1,\dots,\mathbf{x}_{nk}$ in A^n such that $(a,\dots,a)=\mathbf{x}_0<\mathbf{x}_1<\dots<\mathbf{x}_{nk}=(b,\dots,b)$. Let M be an increasing internal mean on A^n. The sequence $a=M(\mathbf{x}_0),M(\mathbf{x}_1),\dots,M(\mathbf{x}_{nk})=b$ strictly increases at at most k out of nk steps, so M is not strictly increasing. □

Example 1.1.25. This example serves to show that in Fact 1.1.24, the requirement that $|A\cap I|\in\mathbb{N}_{\ge 2}$ can't be weakened by replacing $\mathbb{N}_{\ge 2}$ by $\mathbb{N}$. In fact, this is an example of a set A that has two non-degenerate intervals among its connected components, such that there is a strictly increasing, continuous, internal mean on A^2.

Let $A = (a,b) \cup \{m\} \cup (c,d)$, for some $a,b,m,c,d \in \mathbb{R}$ with $a < b < m < c < d$. Let $M : A^2 \to A$ be given by $M(m,m) = m$, and

$$M(y,x) = M(x,y) = \begin{cases} \dfrac{x+y}{2} & \text{if } (x,y) \in (a,b)^2 \cup (c,d)^2 \\ \dfrac{x+b}{2} + \dfrac{b-x}{2}\dfrac{y-c}{d-c} & \text{if } (x,y) \in (a,b) \times (c,d) \\ \dfrac{x+b}{2} & \text{if } (x,y) \in (a,b) \times \{m\} \\ \dfrac{y+c}{2} & \text{if } (x,y) \in \{m\} \times (c,d). \end{cases}$$

It is straightforwardly verified that M is a strictly increasing, continuous mean.

Remark about our choice of M: Note that for M to be a strictly increasing, continuous mean, it has to map $(a,b) \times (c,d)$ entirely into either (a,b) or (c,d), by continuity. Out of the two, we chose (a,b). If we fix $x = b - \varepsilon$ for some small $\varepsilon > 0$, and let y increase from c to d, then $M(x,y)$ has only room to increase from $M(b-\varepsilon, m)$ to b. That is the reason for letting our definition of M be so asymmetric in $((a,b),(c,d))$. ♪

1.1.4 About the compressing property

Intuitively, "the more compressing a mean is, the less skewed it can be towards either of the extremes Min and Max".

It is easily seen that for example the geometric mean is not $[c,C]$-compressing on the whole of $\mathbb{R}_{>0}$, for any c, C. That is the reason that we defined the compressing property as a "local" property, that is, a property that has to hold (with possibly different parameters) on each closed interval.

Quite some work in Chapter 2 and §3.1 is devoted to studying the compressing property (and the power asymptotic property) of means that are constructed out of other means. For readers who are not interested in this, we note that everything mentioning "compressing" and "power-asymptotic" in this text, can be skipped at reading without making the text less intelligible.

Our only true application of the compressing property is in Theorem 3.2.4, where the compressing property provides in concrete situations an upper bound for the speed of convergence of a tuple of means to their Gauss composite.

In the case of continuous means on intervals or closed sets, the 'compressing' property is equivalent to the 'strictness' property, as the next lemma shows.

Lemma 1.1.26. *Let A be an interval or a closed subset of $\mathbb{R}$.*
Let M be a continuous, strict mean on A^n. Then M is compressing. ♩

Proof. Let $(a,b) \in A^2$ with $a < b$. Because $(A \cap [a,b])^n$ is compact, and the function

$$f : \quad (A \cap [a,b])^n \to \mathbb{R} : \quad \mathbf{x} \mapsto \frac{\mathrm{Max}(\mathbf{x}) - M(\mathbf{x})}{M(\mathbf{x}) - \mathrm{Min}(\mathbf{x})}$$

is continuous, f attains a minimum $c \in \mathbb{R}$ and a maximum $C \in \mathbb{R}$. We have $0 < c$, because M is strict. Hence, M is $[c, C]$-compressing on $A \cap [a, b]$. $\square$

Next, we introduce an alternative formulation of the "compressing" property, (we call it "alt-compressing"), that involves a different but related fraction. It depends on the context which formulation is more suitable: in the proofs about conjugation of means (Theorem 2.1.3 for example), it is convenient to use the "compressing" formulation, while in the proof about composition of means (Theorem 3.1.1) it is convenient to use the "alt-compressing" formulation. Both formulations are easily translated into each other, as we show in Lemma 1.1.28 and Remark 1.1.29.

Definition 1.1.27. Let $A \subseteq \mathbb{R}$, let $n \in \mathbb{N}$, let $M : A^n \to \mathbb{R}$ be a mean.

Let $(r, R) \in \mathbb{R}^2$ with $0 \le r \le R \le 1$. We say that M is $[r, R]$-***alt-compressing*** if for all $\mathbf{x} \in A^n \setminus \operatorname{diag}(A^n)$

$$\frac{\operatorname{Max}(\mathbf{x}) - M(\mathbf{x})}{\operatorname{Max}(\mathbf{x}) - \operatorname{Min}(\mathbf{x})} \in [r, R].$$

If $R = 1 - r$, we simply write $[r]$-alt-compressing instead of $[r, 1 - r]$-alt-compressing.

Lemma 1.1.28. *Let $f : \mathbb{R}_{>0} \to (0, 1)$ be the bijection given by $f(t) = \dfrac{t}{1+t}$.*
Let $A \subseteq \mathbb{R}$, let $n \in \mathbb{N}$, let M be a mean on A^n, let $c, C, r, R \in \mathbb{R}$ such that $0 < c \le C$ and that $0 < r \le R < 1$.

1. *M is $[c, C]$-compressing $\iff$ M is $[f(c), f(C)]$-alt-compressing.*
2. *M is $[c]$-compressing $\iff$ M is $[f(c)]$-alt-compressing.*

Proof. Let $\mathbf{x} \in A^n \setminus \operatorname{diag}(A^n)$. If M is not a strict mean, then all four statements in 1–2 are false, and we are done. So we suppose that M is strict, and write

$$Q_1 := \frac{\operatorname{Max}(\mathbf{x}) - M(\mathbf{x})}{M(\mathbf{x}) - \operatorname{Min}(\mathbf{x})}, \qquad Q_2 := \frac{\operatorname{Max}(\mathbf{x}) - M(\mathbf{x})}{\operatorname{Max}(\mathbf{x}) - \operatorname{Min}(\mathbf{x})}.$$

Part 1 follows by noting that $Q_2 = f(Q_1)$, and that f is increasing. Part 2 follows from part 1 by noting that $f(c^{-1}) = 1 - f(c)$. $\square$

***Remark* 1.1.29.** For using "$\Longleftarrow$" in Lemma 1.1.28 in concrete situations, it is convenient to note that the inverse of f is given by $f^{-1}(t) = \dfrac{t}{1-t}$.

Corollary 1.1.30. *Let $A \subseteq \mathbb{R}$, let $n \in \mathbb{N}$, let M be a mean on A^n.*
The following are equivalent:

1. *M is compressing.*
2. *For all $(a, b) \in A^2$ with $a < b$, there exists $(c, C) \in \mathbb{R}^2$ with $0 < c \le C$ such that M is $[c, C]$-compressing on $A \cap [a, b]$.*
3. *For all $(a, b) \in A^2$ with $a < b$, there exists $(r, R) \in \mathbb{R}^2$ with $0 < r \le R < 1$ such that M is $[r, R]$-alt-compressing on $A \cap [a, b]$.*

Lemma 1.1.31. *Let $A \subseteq \mathbb{R}$, let $n \in \mathbb{N}$, let $\mathbf{x} \in A^n$, let $a := \mathrm{Min}(\mathbf{x}) \leq \mathrm{Max}(\mathbf{x}) =: b$, and suppose that M is an $[r, R]$-alt-compressing mean on $(A \cap [a, b])^n$. Then $M(\mathbf{x})$ is contained in the interval $[Ra + (1-R)b, ra + (1-R)b]$ of length $(R-r)(b-a)$.*

Proof. This follows straightforwardly by rewriting the relation $r \leq \frac{b-M(\mathbf{x})}{b-a} \leq R$. □

***Remark* 1.1.32.** We can phrase Corollary 1.1.31 very informally like: "All bunches of n numbers that stretch from a to b, so over a distance of $b-a$, are compressed by M into an interval of length $(R-r)(b-a)$." This clarifies the name 'compressing'.

1.2 Weighted pythagorean means

The three so-called *classical pythagorean means* are the (unweighted) arithmetic, geometric, and harmonic mean. We introduced them in Definition 1.1.18, ande in Definition 1.2.3 we introduce their "weighted" generalisations. These are probably the only means that people outside of mathematics normally call means, and that are used in practice – see Example 1.2.5.

An excellent and comprehensive source about the pythagorean means is [Bul03, Chapter II]. All the results about them that we mention here, except about the compressing and power-asymptotic properties, are contained in that chapter.

The next definition, which essentially is a convention on notation, will be used throughout this text, because it hides redundant symbols from a formula, and thus enhances the essence of the formula.

Definition 1.2.1 (Vector operation notation)**.** Let $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$, and $\mathbf{y} = (y_1, \ldots, y_n) \in \mathbb{R}^n$, and $t \in \mathbb{R}$.

1. We write $\mathbf{x}+\mathbf{y}, \mathbf{xy}, \frac{\mathbf{x}}{\mathbf{y}}, \mathbf{x}^{\mathbf{y}} \in \mathbb{R}^n$ for the result of the corresponding entrywise operation, provided that all of $\frac{x_i}{y_i}$ and $x_i^{y_i}$ are defined in the case of $\frac{\mathbf{x}}{\mathbf{y}}$ and $\mathbf{x}^{\mathbf{y}}$ respectively. Moreover, when $\mathbf{x} = \mathrm{diag}_n(t)$, we sometimes substitute t for $\mathbf{x}$ in such an expression; analogously when $\mathbf{y} = \mathrm{diag}_n(t)$. Thus, for instance, $t^{\mathbf{x}} + t$ means the same as $\mathrm{diag}_n(t)^{\mathbf{x}} + \mathrm{diag}_n(t)$, which means the same as $(t^{x_1} + t, t^{x_2} + t, \ldots, t^{x_n} + t)$.

2. We write $\sum \mathbf{x} := \sum_{i=1}^n x_i$, and $\prod \mathbf{x} := \prod_{i=1}^n x_i$.

3. We follow the convention that exponentiation has priority over multiplication, which has priority over summation. Thus, $\sum \mathbf{xy}$ means $\sum(\mathbf{xy})$ and not $(\sum \mathbf{x})\mathbf{y}$, while $\prod \mathbf{x}^{\mathbf{y}}$ means $\prod(\mathbf{x}^{\mathbf{y}})$ and not $(\prod \mathbf{x})^{\mathbf{y}}$. Moreover, with $\prod \mathbf{xy}$ we mean $\prod(\mathbf{xy})$ and not $(\prod \mathbf{x})\mathbf{y}$.

Example 1.2.2. $\sum t\mathbf{y}\mathbf{x}^{-1}$ means $\sum_{i=1}^n t y_i x_i^{-1}$, and $\prod \mathbf{x}^{\mathbf{y}/t}$ means $\prod_{i=1}^n x_i^{y_i/t}$, and $\sum \mathbf{y}(e^{t\mathbf{x}} - 1)$ means $\sum_{i=1}^n y_i(e^{tx_i} - 1)$.

Definition 1.2.3 (AM, GM, HM)**.** Let $\mathbf{w} \in \mathbb{R}^n_{>0}$, for some $n \in \mathbb{N}$.

The *harmonic mean* $\mathrm{HM}_{\mathbf{w}}$, the *geometric mean* $\mathrm{GM}_{\mathbf{w}}$, and the *arithmetic mean* $\mathrm{AM}_{\mathbf{w}}$, *weighted by* $\mathbf{w}$, are functions $A^n \to A$, where $A := \mathbb{R}_{>0}$ in the case of HM and GM, while $A := \mathbb{R}$ in the case of AM. Namely, for $\mathbf{x} \in A^n$, we define

$$\mathrm{HM}_{\mathbf{w}}(\mathbf{x}) = \frac{\sum \mathbf{w}}{\sum \mathbf{w}\mathbf{x}^{-1}}, \quad \mathrm{GM}_{\mathbf{w}}(\mathbf{x}) = \Big(\prod \mathbf{x}^{\mathbf{w}}\Big)^{1/\sum \mathbf{w}}, \quad \mathrm{AM}_{\mathbf{w}}(\mathbf{x}) = \frac{\sum \mathbf{w}\mathbf{x}}{\sum \mathbf{w}}.$$

In the case that $\mathbf{w} \in \operatorname{diag}(\mathbb{R}^n_{>0})$, we normally write $\mathrm{HM}_n, \mathrm{GM}_n, \mathrm{AM}_n$ instead of $\mathrm{HM}_{\mathbf{w}}, \mathrm{GM}_{\mathbf{w}}, \mathrm{AM}_{\mathbf{w}}$, in accordance with Definition 1.1.18. ⸾

***Remark* 1.2.4.** Clearly, it are only the ratios between the "weights" (i.e. the entries of $\mathbf{w}$) that matter, in the sense that for any $\lambda \in \mathbb{R}_{>0}$, we have $\mathrm{HM}_{\lambda\mathbf{w}} = \mathrm{HM}_{\mathbf{w}}$ and $\mathrm{GM}_{\lambda\mathbf{w}} = \mathrm{GM}_{\mathbf{w}}$ and $\mathrm{AM}_{\lambda\mathbf{w}} = \mathrm{AM}_{\mathbf{w}}$. In particular, we can without loss of generality restrict ourselves to the case that the weights are "normalised", i.e. $\sum \mathbf{w} = 1$. But for the sake of clarity and ease of several statements and proofs, for instance in §2.6–§2.7 and in Chapter 4, it is better to allow $\mathbf{w}$ to be any tuple in $\mathbb{R}^n_{>0}$.

Example 1.2.5 (Usage of the weighted pythagorean means)**.**

1. In Newtonian physics, let $\mathcal{B}_k$ be a body of mass w_k and with center of gravity $\mathbf{x}_k \in \mathbb{R}^3$, for $k = 1, \ldots, n$. Let $\mathbf{w} = (w_1, \ldots, w_n)$, and let $(\mathbf{x}'_1, \mathbf{x}'_2, \mathbf{x}'_3) \in \mathbb{R}^{3\times n}$ be the transpose of the matrix $(\mathbf{x}_1, \ldots, \mathbf{x}_n) \in \mathbb{R}^{n\times 3}$. Then the center of gravity of $\mathcal{B}_1, \ldots, \mathcal{B}_k$ together, regarded as one body, is

$$\big(\mathrm{AM}_{\mathbf{w}}(\mathbf{x}'_1), \mathrm{AM}_{\mathbf{w}}(\mathbf{x}'_2), \mathrm{AM}_{\mathbf{w}}(\mathbf{x}'_3)\big).$$

2. In economics,[1] let $\mathcal{P}$ be a producible good, let $\mathcal{F}$ be a $\mathcal{P}$-producing factory. We define the *output* of $\mathcal{F}$ as its yearly produced quantity of $\mathcal{P}$, the *labour* of $\mathcal{F}$ as the yearly total amount of time spent by employees to produce $\mathcal{P}$, and the *productivity* of $\mathcal{F}$ as output dived by labour.

   For $k = 1, \ldots, n$, let $\mathcal{F}_k$ be a $\mathcal{P}$-producing factory with output o_k, labour l_k, and productivity $p_k = o_k/l_k$. Then the productivity of $\mathcal{F}_1, \ldots, \mathcal{F}_n$ together, regarded as one factory, is

$$\mathrm{HM}_{(o_1,\ldots,o_n)}(p_1, \ldots, p_n).$$

3. Let $x(t)$ be a positive quantity that changes over the time $t \in \mathbb{R}$. (For example, x could represent a population size or a production rate). For $t, t' \in \mathbb{R}$ with $t < t'$, we define *the average growth rate of x between t and t'* as $g(t, t') := (x(t')/x(t))^{1/(t'-t)}$.

   Let $t_0 < t_1 < \ldots < t_n$, let $d_k := t_k - t_{k-1}$, and let $g_k := g(t_{k-1}, t_k)$. Then

$$g(t_0, t_n) = \mathrm{GM}_{(d_1,\ldots,d_n)}(g_1, \ldots, g_n).$$

[1] I learned about part 2 and 3 of Example 1.2.5 from my wife, Claudia Wieners, who encountered related examples in her interdisciplinary research. In both examples, some scientists were inclined to apply the arithmetic mean instead of the harmonic or geometric mean. I have the impression that inappropriate usage of the arithmetic mean is not uncommon.

4. When averaging quantities of different orders of magnitude, if we are really interested in the average order of magnitude, it is appropriate to use the (neutrally weighted) geometric mean. The arithmetic mean is insensitive for changes of the relatively small quantities, even as they change by orders of magnitude. ♪

Proposition 1.2.6 (Properties of $\mathrm{AM}, \mathrm{GM}, \mathrm{HM}$)**.** *Let $n \in \mathbb{N}_{\geq 2}$, let $\mathbf{w} = (w_1, \ldots, w_n) \in \mathbb{R}^n_{>0}$. We assume without loss of generality that $\sum \mathbf{w} = 1$.*

1. $\mathrm{AM}_{\mathbf{w}}, \mathrm{GM}_{\mathbf{w}}, \mathrm{HM}_{\mathbf{w}}$ *are internal means. They are strictly increasing, smooth, scale-invariant and compressing. They are symmetric if* $\mathbf{w} \in \mathrm{diag}(\mathbb{R}^n_{>0})$.

2. $\mathrm{AM}_{\mathbf{w}}$ *is rational, and on* $\mathbb{R}^n_{>0}$ *it is* $(0,1)$*-power-asymptotic.*
It is $[\mathrm{Min}(\mathbf{w})]$*-alt-compressing, and* $[\frac{\mathrm{Min}(\mathbf{w})}{1-\mathrm{Min}(\mathbf{w})}]$*-compressing.*

$\mathrm{HM}_{\mathbf{w}}$ *is rational and* $(1,0)$*-power-asymptotic.*

$\mathrm{GM}_{\mathbf{w}}$ *is* $(\mathbf{w}, \mathbf{w})$*-power-asymptotic.*

3. $\mathrm{AM}_{\mathbf{w}}(\mathbf{x}+\mathbf{y}) = \mathrm{AM}_{\mathbf{w}}(\mathbf{x}) + \mathrm{AM}_{\mathbf{w}}(\mathbf{y})$ *for all* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$,
$\mathrm{GM}_{\mathbf{w}}(\mathbf{x}\mathbf{y}) = \mathrm{GM}_{\mathbf{w}}(\mathbf{x})\mathrm{GM}_{\mathbf{w}}(\mathbf{y})$ *for all* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n_{>0}$. ♩

Proof. 1. All statements are easily verified; the compressing property follows by Lemma 1.1.26, because we clearly deal with strict, continuous means on intervals.

2. The 'rational' statements are clear. That $\mathrm{GM}_{\mathbf{w}}$ is $(\mathbf{a}, \mathbf{a})$-power-asymptotic, follows because $\mathrm{GM}_{\mathbf{w}}(\mathbf{x}_{[i,t]}) = ct^{w_i}$ for some c not dependent of t. That $\mathrm{AM}_{\mathbf{w}}$ on $\mathbb{R}^n_{>0}$ is $(0,1)$-power-asymptotic, follows because $\mathrm{AM}_{\mathbf{w}}(\mathbf{x}_{[i,t]}) \to c$ as $t \to 0$, for some $c \in \mathbb{R}_{>0}$ independent of t, while $\mathrm{AM}_{\mathbf{w}}(\mathbf{x}_{[i,t]}) \sim tw_i$ as $t \to \infty$. That $\mathrm{HM}_{\mathbf{w}}$ is $(1,0)$-power-asymptotic, follows analogously (and it follows in a more general context from the last statement of Corollary 2.2.4).

The "compressing" statement follows from the "alt-compressing" statement by Lemma 1.1.28. It remains to show the "alt-compressing" statement. Let $\mathbf{x} \in \mathbb{R}^n$ be such that $\mathrm{Min}(\mathbf{x}) =: x < y := \mathrm{Max}(\mathbf{x})$. Say x is at position i in $\mathbf{x}$, and y at position j. Let $\mathbf{x}_- \in \mathbb{R}^n$ be the vector with x in every entry except for y in the jth entry, and let $\mathbf{x}_+ \in \mathbb{R}^n$ be the vector with y in every entry except for x in the ith entry. Then $\mathbf{x}_- \leq \mathbf{x} \leq \mathbf{x}_+$, hence

$$x + w_j(y-x) = \mathrm{AM}_{\mathbf{w}}(\mathbf{x}_-) \leq \mathrm{AM}_{\mathbf{w}}(\mathbf{x}) \leq \mathrm{AM}_{\mathbf{w}}(\mathbf{x}_+) = y + w_j(x-y).$$

Therefore,

$$\frac{\mathrm{AM}(\mathbf{x}) - \mathrm{Min}(\mathbf{x})}{\mathrm{Max}(\mathbf{x}) - \mathrm{Min}(\mathbf{x})} = \frac{\mathrm{AM}(\mathbf{x}) - x}{y-x} \in \Big[w_j, 1-w_j\Big] \subseteq \Big[\mathrm{Min}(\mathbf{w}), 1 - \mathrm{Min}(\mathbf{w})\Big].$$

3. Clear. □

***Remark* 1.2.7.** Below, we show more properties of the weighted pythagorean means:

1. Concrete compression formulas for $\mathrm{GM}_{\mathbf{w}}$ and $\mathrm{HM}_{\mathbf{w}}$. (Corollary 2.2.3.)

2. The classical inequalities $\mathrm{HM}_{\mathbf{w}} \leq \mathrm{GM}_{\mathbf{w}} \leq \mathrm{AM}_{\mathbf{w}}$ on $\mathbb{R}^n_{>0}$, which are strict inequalities on the complement of $\mathrm{diag}(\mathbb{R}^n_{>0})$. (Corollary 2.6.8)

3. "Conjugation relations" between the means. (Fact 1.3.4 and Corollary 1.3.11.)

4. The properties 1.2.6.3 are unique for $\mathrm{AM}_{\mathbf{w}}$ and $\mathrm{GM}_{\mathbf{w}}$ among the continuous means on $\mathbb{R}^n_{>0}$. (Corollary 1.3.12.)

We could have given the proofs at this point, but to make the exposition clearer, we postpone the proofs until we have developed more tools and theory.

1.3 Conjugation and inversion duality

Conjugation is the most basic operation on means. It is extensively (although in less generality) treated in [Bul03, Chapter IV] and in [HLP34, Chapter III]. We first define conjugation of general functions $A^n \to A$, and show in Proposition 1.3.7 basic general properties of conjugation. Then we apply this to means on A^n.

The following definition, or rather a convention on notation, is used throughout this text. The reader might be puzzled why we need names for such common functions, but we do need a name for any function when we want to write down conjugation with that function (as in Definition 1.3.6).

Definition 1.3.1. Let $a \in \mathbb{R}_{\neq 0}$ and let $b \in \mathbb{R}$. We use the following notation for the following bijections:

$$\begin{array}{llllll} \mathrm{pow}_a & : & \mathbb{R}_{>0} \to \mathbb{R}_{>0} & : & x \mapsto x^a \\ \mathrm{mul}_a & : & \mathbb{R} \to \mathbb{R} & : & x \mapsto ax \\ \mathrm{tra}_b & : & \mathbb{R} \to \mathbb{R} & : & x \mapsto b + x \\ \mathrm{inv}_a & : & \mathbb{R}_{\neq 0} \to \mathbb{R}_{\neq 0} & : & x \mapsto ax^{-1} \\ \mathrm{ainv}_b & : & \mathbb{R} \to \mathbb{R} & : & x \mapsto b - x. \end{array}$$

To ease the notation, we use the same notations for the restrictions to subsets; for instance, for any $A \subseteq \mathbb{R}_{\neq 0}$, we have bijections $\mathrm{inv}_a : A \to \mathrm{inv}_a(A)$. Further, we write 'inv' instead of 'inv_1', and 'ainv' instead of 'ainv_0', and 'id' instead of 'mul_1'.

Definition 1.3.2. Let $g : A \to B$ be a function, where A, B are any two sets, and let $n \in \mathbb{N}$. We define *the induced entrywise function* $\langle g \rangle_n$ by

$$A^n \to B^n : \quad \langle g \rangle_n(x_1, \ldots, x_n) = (g(x_1), \ldots, g(x_n)).$$

To ease the notation, we normally write $\langle g \rangle$ instead of $\langle g \rangle_n$.

Example 1.3.3. Let $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n_{>0}$. Then $\langle \exp \rangle \mathbf{x} = (e^{x_1}, \ldots, e^{x_n})$, and $\langle \mathrm{inv} \rangle \mathbf{x} = (x_1^{-1}, \ldots, x_n^{-1})$, and $\langle \mathrm{pow}_p \rangle \mathbf{x} = (x_1^p, \ldots, x_n^p)$.

Fact 1.3.4. *Let $n \in \mathbb{N}$, let $\mathbf{w} \in \mathbb{R}^n_{>0}$. On $\mathbb{R}^n_{>0}$, we have the following relations between the Pythagorean means:*

$$\begin{aligned}
\exp \circ \mathrm{AM}_{\mathbf{w}} \circ \langle \log \rangle &= \mathrm{GM}_{\mathbf{w}}, \\
\mathrm{inv} \circ \mathrm{AM}_{\mathbf{w}} \circ \langle \mathrm{inv} \rangle &= \mathrm{HM}_{\mathbf{w}}, \\
\mathrm{inv} \circ \mathrm{GM}_{\mathbf{w}} \circ \langle \mathrm{inv} \rangle &= \mathrm{GM}_{\mathbf{w}}.
\end{aligned}$$

Proof. The first equality follows by

$$\exp\Big(\frac{\sum \mathbf{w}\langle\log\rangle\mathbf{x}}{\sum \mathbf{w}}\Big) = \big(\exp(\sum \mathbf{w}\langle\log\rangle\mathbf{x})\big)^{1/\sum \mathbf{w}} = \big(\prod\langle\exp\rangle\mathbf{w}\langle\log\rangle\mathbf{x}\big)^{1/\sum \mathbf{w}} = \big(\prod \mathbf{x}^{\mathbf{w}}\big)^{1/\sum \mathbf{w}}.$$

The other equalities follow by even simpler calculations. □

We use the following simple corollary in Theorem 7.5.4, where we analyse "Collatz-like" iterations.

Corollary 1.3.5. *Let $(x_1, x_2, x_3, \ldots)$ be a sequence in $\mathbb{R}_{>0}$, let $\mathbf{x}_n := (x_1, \ldots, x_n)$. Suppose that $\lim_{n\to\infty} x_n = \infty$. Then $\lim_{n\to\infty} \mathrm{HM}_n(\mathbf{x}_n) = \infty$.*

Proof. We have $\lim_{n\to\infty} x_n^{-1} = 0$, so we clearly have $\lim_{n\to\infty} \mathrm{AM}_n(\mathbf{x}_n^{-1}) = 0$. The result follows by noting that $\mathrm{HM}_n(\mathbf{x}_n) = (\mathrm{AM}_n(\mathbf{x}_n^{-1}))^{-1}$ (by Fact 1.3.4). □

1.3.1 Conjugation and inversion duality operators

The content of this subsection is a bit abstract. We apply it almost exclusively to the case that f is an internal mean and g is monotonic, from §1.3.2 onwards. Because the results are just as true without those assumptions on f and g, we present them in their generality, to make the exposition hopefully clearer.

Definition 1.3.6 (Conjugation, inversion, inversion duality)**.** Let A, B be sets, let $g : A \to B$ be a bijection, let $n \in \mathbb{N}$.

1. The *"conjugation with g" operator*, denoted by $[g]$, maps each function $f : A^n \to A$ to a function $f^{[g]} : B^n \to B$, namely

   $$f^{[g]} := g \circ f \circ \langle g^{-1} \rangle.$$

   We informally say that $f^{[g]}$ "is a conjugate of f".

2. Suppose $A \subseteq \mathbb{R}$, let $i : A \to i(A)$ be one of the following bijections:
   $i = \mathrm{inv}_a$ for some $a \in \mathbb{R}_{>0}$ (assuming that $0 \notin A$), or
   $i = \mathrm{ainv}_b$ for or some $b \in \mathbb{R}$. We say that i is an *inversion of A*, and that $f^{[\mathrm{inv}]}$ is the *i-inversion dual of f*.

   We give special attention to the inversion $i = \mathrm{inv}$; we call $[\mathrm{inv}]$ the *inversion duality* operator, and we say that $f^{[\mathrm{inv}]}$ is the *inversion dual of f*.

Proposition 1.3.7 ((†) Properties of conjugation)**.** *Let $g : A \to B$ and $h : B \to C$ be bijections, for any sets A, B, C, let $f, F, f_1, f_2, \ldots : A^n \to A$ be functions, for any $n \in \mathbb{N}$.*

1. $(f^{[g]})^{[h]} = f^{[h \circ g]}$, *and* $f^{[\mathrm{id}]} = f$.

2. *Let* $S(A)$ *be the permutation group of* A, *let* $A^{(A^n)}$ *be the set of functions* $A^n \to A$. *The map* $S(A) \times A^{(A^n)} \to A^{(A^n)} : (g, f) \mapsto f^{[g]}$ *is a group action of* $S(A)$ *on* $A^{(A^n)}$.

For the remaining items, we assume that $A, B, C \subseteq \mathbb{R}$.

3. *Suppose* f *is continuous. If* g *is a homeomorphism, then* $f^{[g]}$ *is continuous.* *In particular, if* A *is an interval or an open subset of* $\mathbb{R}$, *and* g *is continuous, then* $f^{[g]}$ *is continuous.*

4. *Suppose* f *is smooth. If* g *and* g^{-1} *are smooth, then* $f^{[g]}$ *is smooth.* *In particular, if* g *is smooth, and the derivative of* g *is nowhere zero on* A, *then* $f^{[g]}$ *is smooth.*

5. *Let '*$\prec$*' denote either '*$\leq$*' or '*$<$*'. Suppose that* $f \prec F$.
   If g *is strictly increasing, then* $f^{[g]} \prec F^{[g]}$.
   If g *is strictly decreasing, then* $F^{[g]} \prec f^{[g]}$.

6. *Suppose that* g *is monotonic. If* f *is monotonically/strictly increasing/decreasing respectively (any of 4 combinations), then* $f^{[g]}$ *is so as well.*

7. *Suppose that* g *is continuous, and that the pointwise limit* $f := \lim_{k\to\infty} f_k$ *exists and maps* A^n *into* A. *Then the pointwise limit* $\lim_{k\to\infty} f_k^{[g]}$ *exists and equals* $f^{[g]} : B^n \to B$. ⅃

Proof. Part 1 follows easily from the definitions, part 2 follows from part 1.

3. The first statement is clear. Thus, for the other statement, it suffices to show that g^{-1} is continuous.

First case: A is an interval of $\mathbb{R}$. It is easily seen that g, being a continuous injection on A, is strictly monotonic, and therefore is an open map. Hence, g^{-1} is continuous.

Second case: A is an open subset of $\mathbb{R}$. Then A is the disjoint union of open intervals of $\mathbb{R}$, say $A = \bigsqcup_{j\in J} I_j$. On each open interval I_j, we know from the first case that g^{-1} is continuous on the open interval $g(I_j)$. Therefore, g^{-1} is continuous on $B = \bigsqcup_{j\in J} g(I_j)$.

4. The first statement is clear. Thus, for the other statement, it suffices to show that g^{-1} is smooth. It is implicit by smoothness of g that A is open in $\mathbb{R}$. From the proof of the second case of part 3, we know that $g(A) = B$ is open in $\mathbb{R}$. Because the derivative of g is nowhere zero, it follows from the inverse function theorem that g^{-1} is smooth in a neighbourhood of any point of B, that is, g^{-1} is smooth.

5. Suppose that $f < F$, and that g is strictly decreasing. (The other three cases follow analogously.) For all $\mathbf{x} \in A^n$, we have

$$f^{[g]}(\mathbf{x}) = g(f(\langle g^{-1}\rangle \mathbf{x})) > g(F(\langle g^{-1}\rangle \mathbf{x})) = F^{[g]}(\mathbf{x}),$$

where ">" follows because "$\mathbf{y} < \mathbf{z} \implies g(\mathbf{y}) > g(\mathbf{z})$". Hence, $f^{[g]} > F^{[g]}$.

6. Suppose that g is strictly decreasing, and that f is strictly increasing. (The other seven cases follow analogously.) Then also g^{-1} is strictly decreasing. For any $\mathbf{x}, \mathbf{y} \in A^n$,

$$\mathbf{x} < \mathbf{y} \implies \langle g^{-1}\rangle\mathbf{x} > \langle g^{-1}\rangle\mathbf{y} \implies f(\langle g^{-1}\rangle\mathbf{x}) > f(\langle g^{-1}\rangle\mathbf{y}) \implies g(f(\langle g^{-1}\rangle\mathbf{x})) < g(f(\langle g^{-1}\rangle\mathbf{y})).$$

So if $\mathbf{x} < \mathbf{y}$, then $f^{[g]}(\mathbf{x}) < f^{[g]}(\mathbf{y})$.

7. For all $\mathbf{x} \in B^n$, we have by assumption

$$f_n(\langle g^{-1}\rangle\mathbf{x}) \to f(\langle g^{-1}\rangle\mathbf{x}) \quad \text{as } n \to \infty.$$

Because g is continuous and because $f(\langle g^{-1}\rangle\mathbf{x}) \in A$ and $f_n(\langle g^{-1}\rangle\mathbf{x}) \in A$ for all n, it follows that

$$g(f_n(\langle g^{-1}\rangle\mathbf{x})) \to g(f(\langle g^{-1}\rangle\mathbf{x})) \quad \text{as } n \to \infty,$$

that is, $f_n^{[g]}(\mathbf{x}) \to f^{[g]}(\mathbf{x})$. □

Corollary 1.3.8 (Properties of inversion duality)**.** *Let $A \subseteq \mathbb{R}$, let $i : A \to i(A)$ be an inversion of A (as in Definition 1.3.6.2). Let $n \in \mathbb{N}$, let $f, g : A^n \to A$ be functions.*

1. $(f^{[i]})^{[i]} = f$.
2. *If f is continuous, then $f^{[i]}$ is continuous. If f is smooth, then $f^{[i]}$ is smooth.*
3. *If $f \le g$, then $f^{[i]} \ge g^{[i]}$.*
   If $f < g$, then $f^{[i]} > g^{[i]}$.
4. *If f is strictly/monotonically increasing/decreasing, then $f^{[i]}$ is so as well.* ♩

Proof. 1, 2, 3, 4 follow from 1, {3, 4}, 5, 6 respectively of Proposition 1.3.7, using that i is a strictly decreasing, smooth bijection, and an involution. □

1.3.2 Conjugation and inversion duality of means

After the "abstract" §1.3.1, we apply the results to means. In this final subsection of Chapter 1, we only treat some basics, because conjugation of means is the topic of Chapter 2.

Lemma 1.3.9 (The conjugates of the means Min and Max)**.**
Let $A, B \subseteq \mathbb{R}$, and let $g : A \to B$ be a monotonic bijection. Let $n \in \mathbb{N}$.
If g is increasing, then $\mathrm{Min}_n^{[g]} = \mathrm{Min}_n$ and $\mathrm{Max}_n^{[g]} = \mathrm{Max}_n$ on B^n.
If g is decreasing, then $\mathrm{Min}_n^{[g]} = \mathrm{Max}_n$ and $\mathrm{Max}_n^{[g]} = \mathrm{Min}_n$ on B^n. ♩

Proof. Let $\mathbf{x} = (x_1, \dots, x_n) \in B^n$. If g is decreasing, then g^{-1} is decreasing, hence $\mathrm{Min}\big(g^{-1}(x_1), \dots, g^{-1}(x_n)\big) = g^{-1}\big(\mathrm{Max}(x_1, \dots, x_n)\big)$. That is, $\mathrm{Min}^{[g]}(\mathbf{x}) = \mathrm{Max}(\mathbf{x})$. The other three statements follow analogously. □

Lemma 1.3.10 (Conjugation of means)**.** *Let $A, B \subseteq \mathbb{R}$, let $n \in \mathbb{N}$,*
let M be an internal mean on A^n. Let $g : A \to B$ be monotonic bijection.
Then $M^{[g]}$ is an internal mean on B^n. ♩

Proof. By Proposition 1.3.7.5, we have
$\mathrm{Min}_n^{[g]} \leq M^{[g]} \leq \mathrm{Max}_n^{[g]}$ in case that g is increasing, and
$\mathrm{Max}_n^{[g]} \leq M^{[g]} \leq \mathrm{Min}_n^{[g]}$ in case that g is decreasing.
By Lemma 1.3.9, it follows that $\mathrm{Min}_n \leq M^{[g]} \leq \mathrm{Max}_n$. □

We study in the next chapter how the properties of M and $M^{[g]}$ relate to each other. We close this chapter by reformulating the relations between the Pythagorean means in terms of conjugation and inversion duality.

Corollary 1.3.11 (Relations between pythagorean means; inversion duals)**.**
Let $n \in \mathbb{N}$, let $\mathbf{w} \in \mathbb{R}^n_{>0}$.

1. $\mathrm{GM}_{\mathbf{w}} = \mathrm{AM}_{\mathbf{w}}^{[\exp]}$ *on* $\mathbb{R}^n_{>0}$, *and* $\mathrm{GM}_{\mathbf{w}}^{[\log]} = \mathrm{AM}_{\mathbf{w}}$ *on* $\mathbb{R}^n$.

2. $\mathrm{GM}_{\mathbf{w}}^{[\mathrm{inv}]} = \mathrm{GM}_{\mathbf{w}}$ *on* $\mathbb{R}^n_{>0}$, *and* $\mathrm{HM}_{\mathbf{w}}^{[\mathrm{inv}]} = \mathrm{AM}_{\mathbf{w}}$ *on* $\mathbb{R}^n_{>0}$.

3. $\mathrm{Min}_n^{[i]} = \mathrm{Max}_n$ *on* $i(A)^n$, *for any* $A \subseteq \mathbb{R}$ *and any inversion* i *of* A. ⌟

Proof. 1 and 2 follow from Fact 1.3.4, while 3 follows from Fact 1.3.9, using that i is strictly decreasing. □

Corollary 1.3.12 (Uniqueness of additivity of AM and multiplicativity of GM)**.**
Let $n \in \mathbb{N}_{\geq 2}$, let M be a continuous internal mean on $\mathbb{R}^n_{>0}$.

1. *If* $M(\mathbf{x}+\mathbf{y}) = M(\mathbf{x}) + M(\mathbf{y})$ *for all* $\mathbf{x}, \mathbf{y}$, *then* $M = \mathrm{AM}_{\mathbf{w}}$ *for some* $\mathbf{w} \in \mathbb{R}^n_{>0}$.

2. *If* $M(\mathbf{x}\mathbf{y}) = M(\mathbf{x})M(\mathbf{y})$ *for all* $\mathbf{x}, \mathbf{y}$, *then* $M = \mathrm{GM}_{\mathbf{w}}$ *for some* $\mathbf{w} \in \mathbb{R}^n_{>0}$. ⌟

Proof. 1. We have $M(\lambda\mathbf{x}) = \lambda M(\mathbf{x})$ for all $\lambda \in \mathbb{Q}$ by additivity, hence for all $\lambda \in \mathbb{R}$ by continuity. So M is a linear function, so $M(\mathbf{x}) = \sum \mathbf{w}\mathbf{x}$ for some $\mathbf{w} \in \mathbb{R}^n_{>0}$. From $M(\mathrm{diag}_n(1)) = 1$, it follows that $\sum \mathbf{w} = 1$. Hence, $M(\mathbf{x}) = \mathrm{AM}_{\mathbf{w}}$.

2. By Lemma 1.3.10 and Proposition 1.3.7, $M^{[\log]}$ is a continuous internal mean on $\mathbb{R}^n$. Clearly, $M^{[\log]}$ is additive, so by part 1, we have $M^{[\log]} = \mathrm{AM}_{\mathbf{w}}$ for some $\mathbf{w} \in \mathbb{R}^n_{>0}$. Hence, $M = \mathrm{GM}_{\mathbf{w}}$. □

Chapter 2

Conjugation of means

At the end of the previous chapter, we considered properties of conjugation of functions $A^n \to A$ in general, and defined conjugation of means. In this chapter, we study aspects of conjugation of means. The core of this chapter is Theorem 2.1.3, in which it is shown how the properties of a conjugate $M^{[f]}$ of a mean M relate to properties of M itself. The most involved part is to show how the compressing properties of M and $M^{[f]}$ relate; the Lemma's 2.1.1 and 2.1.2 contain the preliminary work that we need for that purpose. The sections 2.2–2.5 are concerned with corollaries and examples of Theorem 2.1.3.

The sections 2.6 and 2.7 deal with very concrete means: power means and translation means, which can be viewed as means that interpolate (and extrapolate) the pythagorean means. Power means (except for Min and Max) and translation means are conjugates of the arithmetic mean, and we use Theorem 2.1.3 for proving properties about them.

The power means are well-known; in contrast, the translation mean has, as far as we know, not been defined previously. However, in §2.6–2.7 we will see that several aspects of power means and translation means are similar to each other. We use translation means in Chapter 7 in the context of the Collatz conjecture.

2.1 Reconstructing properties of conjugate means

The aim of this section is to state and prove Theorem 2.1.3. To prove the "compressing" statement, we need the next two lemma's.

Lemma 2.1.1 (*). *Let $A \subseteq \mathbb{R}$, let $n \in \mathbb{N}$, let M be an internal mean on A^n. Let $a, b, c, C, r, R \in \mathbb{R}$, with $a \neq 0$ and $0 \leq r \leq R \leq 1$ and $0 < c \leq C$.*

Let $f : A \to f(A)$ be the monotonic bijection given by $f(x) = ax + b$.

1. *Suppose M is $[r, R]$-alt-compressing.*
   If f is increasing, then $M^{[f]}$ is $[r, R]$-alt-compressing.
   If f is decreasing, then $M^{[f]}$ is $[1 - R, 1 - r]$-alt-compressing.
2. *Suppose M is $[c, C]$-compressing.*
   If f is increasing, then $M^{[f]}$ is $[c, C]$-compressing.
   If f is decreasing, then $M^{[f]}$ is $[C^{-1}, c^{-1}]$-compressing. ♩

Proof. **1.** Let $\mathbf{x} \in f(A)$. We write $Q := \dfrac{\mathrm{Max}(\mathbf{x}) - M^{[f]}(\mathbf{x})}{\mathrm{Max}(\mathbf{x}) - \mathrm{Min}(\mathbf{x})}$, and $x := \mathrm{Min}(\langle f^{-1}\rangle \mathbf{x})$, $y := M(\langle f^{-1}\rangle \mathbf{x})$, $z := \mathrm{Max}(\langle f^{-1}\rangle \mathbf{x})$.

If f is increasing, then by Lemma 1.3.9,

$$Q = \frac{\mathrm{Max}^{[f]}(\mathbf{x}) - M^{[f]}(\mathbf{x})}{\mathrm{Max}^{[f]}(\mathbf{x}) - \mathrm{Min}^{[f]}(\mathbf{x})} = \frac{f(z) - f(y)}{f(z) - f(x)} = \frac{z - y}{z - x} \in [r, R].$$

If f is decreasing, then by Lemma 1.3.9,

$$Q = \frac{\mathrm{Min}^{[f]}(\mathbf{x}) - M^{[f]}(\mathbf{x})}{\mathrm{Min}^{[f]}(\mathbf{x}) - \mathrm{Max}^{[f]}(\mathbf{x})} = \frac{f(x) - f(y)}{f(x) - f(z)} = \frac{x - y}{x - z} = 1 - \frac{z - y}{z - x} \in [1 - R, 1 - r].$$

2. This follows from part 1 and Lemma 1.1.28. □

Lemma 2.1.2 (†)**.** *Let $A \subseteq \mathbb{R}$, and let $f : A \to \mathbb{R}$ be a strictly monotonic function. Suppose there exist $\varepsilon, E \in \mathbb{R}$ with $0 < \varepsilon \leq E$, such that for all $a, b \in A$ with $a \neq b$, we have $\left|\dfrac{f(b) - f(a)}{b - a}\right| \in [\varepsilon, E]$. Let $x, y, z \in A$ with $x < y < z$ such that $\dfrac{z - y}{y - x} \in [c, C]$ for some $c, C \in \mathbb{R}$ with $0 < c \leq C$. Then*

$$\frac{f(z) - f(y)}{f(y) - f(x)} \in [\varepsilon c E^{-1}, E C \varepsilon^{-1}].$$ ♩

Proof. We have

$$\frac{f(z) - f(y)}{f(y) - f(x)} = \frac{f(z) - f(y)}{z - y} \cdot \frac{z - y}{y - x} \cdot \frac{y - x}{f(y) - f(x)}. \tag{2.1}$$

The second factor on the right-hand side of (2.1) is in $[c, C]$.
The first factor is in $[\varepsilon, E]$ if f is increasing, and in $[-E, -\varepsilon]$ if f is decreasing.
The third factor is in $[E^{-1}, \varepsilon^{-1}]$ if f is increasing, and in $[-\varepsilon^{-1}, -E^{-1}]$ if f is decreasing.
Hence, the product of the first and the third factor is in $[\varepsilon E^{-1}, E\varepsilon^{-1}]$.
Therefore, the product of the three factors is in $[\varepsilon c E^{-1}, E C \varepsilon^{-1}]$. □

Theorem 2.1.3 (Properties of conjugation of internal means)**.** *Let $A, B \subseteq \mathbb{R}$, let $n \in \mathbb{N}$, let M be an internal mean on A^n, and let $f : A \to B$ be a monotonic bijection.*

1. *$M^{[f]}$ is an* ***internal mean*** *on B^n.*

2. *If M is symmetric, then $M^{[f]}$ is* ***symmetric****.*
   If M is increasing, then $M^{[f]}$ is ***increasing****.*
   If M is strictly increasing, then $M^{[f]}$ is ***strictly increasing****.*

3. *If M and f and f^{-1} are continuous, then $M^{[f]}$ is* ***continuous****.*
   If M and f smooth, and the derivative of f is nowhere zero, then $M^{[f]}$ is ***smooth****.*

4. *(*) Suppose that M is $[c,C]$-compressing, and that there exist $\varepsilon, E \in \mathbb{R}$ with $0 < \varepsilon \leq E$, such that for all $x, y \in A$ with $x \neq y$, we have $\left|\dfrac{f(y)-f(x)}{y-x}\right| \in [\varepsilon, E]$.*
   If f is increasing, then $M^{[f]}$ is $[\varepsilon c E^{-1}, EC\varepsilon^{-1}]$-compressing.
   *If f is decreasing, then $M^{[f]}$ is $[\varepsilon C^{-1}E^{-1}, Ec^{-1}\varepsilon^{-1}]$-**compressing**.*

5. *Suppose that M is scale-invariant, and that there exists $p \in \mathbb{R}_{\neq 0}$ such that $\{b^{1/p} : b \in B_{>0}\} \subseteq A_{>0}$, and $f(\lambda x) = \lambda^p f(x)$ for all $x \in A$ and $\lambda \in A_{>0}$.*
   *Then $M^{[f]}$ is **scale-invariant**.*

6. *(†) Suppose that $A = \mathbb{R}_{>0} = B$, that M is $(\mathbf{a}, \mathbf{b})$-power-asymptotic, and that f is continuous and multiplicative.*
   If f is increasing, then $M^{[f]}$ is $(\mathbf{a}, \mathbf{b})$-power-asymptotic.
   *If f is decreasing, then $M^{[f]}$ is $(\mathbf{b}, \mathbf{a})$-**power-asymptotic**.* ⌟

Proof. The statements are trivially true for $n = 1$. Suppose that $n \geq 2$.

1. This is the content of Lemma 1.3.10.

2. The first statement is clear, the others follow from part 6 of Proposition 1.3.7.

3. This follows from part 3 and 4 of Proposition 1.3.7.

4. *Case 1: f is increasing.* Let $\mathbf{x} \in B^n \setminus \operatorname{diag}(B^n)$. Because f is injective, we have $\langle f^{-1}\rangle \mathbf{x} \in A^n \setminus \operatorname{diag}(A^n)$. Hence,

$$\frac{\operatorname{Max}(\langle f^{-1}\rangle \mathbf{x}) - M(\langle f^{-1}\rangle \mathbf{x})}{M(\langle f^{-1}\rangle \mathbf{x}) - \operatorname{Min}(\langle f^{-1}\rangle \mathbf{x})} \in [c, C].$$

Applying Lemma 2.1.2 and then Lemma 1.3.9, we get

$$\frac{\operatorname{Max}(\mathbf{x}) - M^{[f]}(\mathbf{x})}{M^{[f]}(\mathbf{x}) - \operatorname{Min}(\mathbf{x})} \in [\varepsilon c E^{-1}, EC\varepsilon^{-1}].$$

Case 2: f is decreasing. Then $\operatorname{ainv} \circ f$ is increasing, so by Case 1 it follows that $M^{[\operatorname{ainv}\circ f]}$ is $[\varepsilon c E^{-1}, EC\varepsilon^{-1}]$-compressing. Because $f = \operatorname{ainv} \circ \operatorname{ainv} \circ f$, we have

$$M^{[f]} = (M^{[\operatorname{ainv}\circ f]})^{[\operatorname{ainv}]},$$

hence it follows from Lemma 2.1.1.1 that $M^{[f]}$ is $[\varepsilon C^{-1}E^{-1}, Ec^{-1}\varepsilon^{-1}]$-compressing.

5. Let $\lambda \in B_{>0}$ and $x \in B$. Because $\lambda^{1/p} \in A_{>0}$, we have $f(\lambda^{1/p} f^{-1}(x)) = \lambda x$. Hence B is closed under multiplication by $B_{>0}$. Moreover, $f^{-1}(\lambda x) = \lambda^{1/p} f^{-1}(x)$. Therefore, for $\mathbf{x} \in B^n$,

$$f\big(M\big(\langle f^{-1}\rangle \lambda \mathbf{x}\big)\big) = f\big(M\big(\lambda^{1/p}\langle f^{-1}\rangle \mathbf{x}\big)\big) = f\big(\lambda^{1/p} M\big(\langle f^{-1}\rangle \mathbf{x}\big)\big) = \lambda f\big(M\big(\langle f^{-1}\rangle \mathbf{x}\big)\big),$$

that is, $M^{[f]}(\lambda \mathbf{x}) = \lambda M^{[f]}(\mathbf{x})$.

6. We write $\mathbf{a} = (a_1, \ldots, a_n) \in \mathbb{R}^n$ and $\mathbf{b} = (b_1, \ldots, b_n) \in \mathbb{R}^n$.
Suppose that f is decreasing; the case that f is increasing follows analogously. Because $f : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ is a strictly decreasing bijection, we have

$$\lim_{t\to 0} f^{-1}(t) = \infty \quad \text{and} \quad \lim_{t\to\infty} f^{-1}(t) = 0 \tag{2.2}$$

Let $\mathbf{x} \in \mathbb{R}^n_{>0}$, and let $i \in \mathbb{N}_{\leq n}$. Then $\langle f^{-1}\rangle(\mathbf{x}_{[i,t]}) = (\langle f^{-1}\rangle \mathbf{x})_{[i,f^{-1}(t)]}$. Because M is $(\mathbf{a},\mathbf{b})$-power-asymptotic, and because of (2.2), there exist $c, c' \in \mathbb{R}_{>0}$ such that

$$\lim_{t\to 0} \frac{M\big((\langle f^{-1}\rangle \mathbf{x})_{[i,f^{-1}(t)]}\big)}{c(f^{-1}(t))^{b_i}} = 1 = \lim_{t\to\infty} \frac{M\big((\langle f^{-1}\rangle \mathbf{x})_{[i,f^{-1}(t)]}\big)}{c'(f^{-1}(t))^{a_i}}. \tag{2.3}$$

We apply f to all three expressions in (2.3), and we simplify the resulting outer two expressions: because f is continuous, we can "interchange f and the limit operator", and because f is multiplicative, we can "interchange f and the quotient operator". Then we simplify the resulting numerators and denominators; in the denominators, we use that f is multiplicative, and that $f(x^y) = f(x)^y$ for any $x, y \in \mathbb{R}_{>0}$, which follows by first considering $y \in \mathbb{Q}$, and then using continuity for $y \in \mathbb{R}$. Thus, after these simplifications, we get

$$\lim_{t\to 0} \frac{M^{[f]}(\mathbf{x}_{[i,t]})}{f(c)t^{b_i}} = f(1) = \lim_{t\to\infty} \frac{M^{[f]}(\mathbf{x}_{[i,t]})}{f(c')t^{a_i}}.$$

Because $f(1), f(c), f(c') \in \mathbb{R}_{>0}$, it follows that, $M^{[f]}$ is $(\mathbf{b},\mathbf{a})$-asymptotic. □

2.2 Consequences of the theorem

The first consequences of Theorem 2.1.3 that we describe, are concrete formulas for how the compressing properties of specific means M relate to the compressing properties of conjugates of M. We start generally in Proposition 2.2.1, and end concretely in the corollaries 2.2.3 and 2.2.4.

Proposition 2.2.1 ((*) Reconstruction of the compressing parameters)**.**

Let A, B be open subsets of $\mathbb{R}$, let $n \in \mathbb{N}$, let M be an internal mean on A^n.

Let $g : A \to B$ be a continuously differentiable, monotonic bijection that is either convex or concave.

Let $a, b \in A$ with $a \leq b$. Suppose that $M^{[g]}$ is $[c, C]$-compressing on $B \cap [g(a), g(b)]$ (if g is increasing) or $B \cap [g(b), g(a)]$ (if g is decreasing). Let $q := \dfrac{g'(a)}{g'(b)}$. Then

$$M \text{ is on } A \cap [a,b] \begin{cases} [qc, q^{-1}C]\text{-compressing} & \text{if } g \text{ is increasing and convex} \\ [q^{-1}c, qC]\text{-compressing} & \text{if } g \text{ is increasing and concave} \\ [q^{-1}C^{-1}, qc^{-1}]\text{-compressing} & \text{if } g \text{ is decreasing and convex} \\ [qC^{-1}, q^{-1}c^{-1}]\text{-compressing} & \text{if } g \text{ is decreasing and concave.} \end{cases}$$ ⅃

Proof. We write $f := g^{-1}$, so we have $M = (M^{[g]})^{[f]}$; and we write $Q(x, y) := \dfrac{f(y) - f(x)}{y - x}$ for $x \neq y$. Note that the derivative of g is nowhere zero on A, since otherwise it would follow by concavity or convexity that g can't be strictly monotonic; thus, in particular, q and q^{-1} are well-defined.

Case 1: g is increasing. Then f is increasing, and $M^{[g]}$ is $[c, C]$-compressing on $B \cap [g(a), g(b)]$. First, suppose that g is convex. Then f is concave, hence Q is a decreasing function on A^2. So

$$\frac{1}{g'(b)} = f'(g(b)) = \lim_{x\to b} Q(x,b) \le Q(x,y) \le \lim_{x\to a} Q(x,a) = f'(g(a)) = \frac{1}{g'(a)}. \tag{2.4}$$

Hence, Theorem 2.1.3.4 applies with $\varepsilon = 1/g'(b)$ and $E = 1/g'(a)$, and it yields that M is $[qc, q^{-1}C]$-compressing. Next, suppose that g is concave. Then f is convex, hence Q is an increasing function on A^2. Hence, by the same formulas as in (2.4) but with '$\le$' replaced by '$\ge$', it follows that Theorem 2.1.3.4 applies with $\varepsilon = 1/g'(a)$ and $E = 1/g'(b)$. Thus, M is $[q^{-1}c, qC]$-compressing.

Case 2: g is decreasing. Then $M^{[g]}$ is $[c, C]$-compressing on $B \cap [g(b), g(a)]$, and the bijection $G := \mathrm{ainv} \circ g : A \to -B$ is increasing. Because $M^{[G]} = (M^{[g]})^{[\mathrm{ainv}]}$, it follows from Lemma 2.1.1 that $M^{[G]}$ is $[C^{-1}, c^{-1}]$-compressing on $B \cap [G(b), G(a)]$. Thus, Case 1 applies with G substituted for g. The results follow by noting that g is convex if and only if G is concave, and that $G'(a)/G'(b) = g'(a)/g'(b)$. □

We want to stress that we don't claim that the compressing parameters in Proposition 2.2.1 and in the corollaries below are optimal parameters; they can only be seen as upper and lower bounds for the optimal parameters.

We use the next corollary for determining compressing parameters of the power means and the translation means, in §2.6–2.7.

Corollary 2.2.2 ((*) Compressing parameters for quasi-arithmetic means)**.**
Let A, B be open subsets of $\mathbb{R}$, let $n \in \mathbb{N}_{\ge 2}$, let M be an internal mean on A^n.

Suppose that $M^{[g]} = \mathrm{AM}_{\mathbf{w}}$ on B^n, for some $\mathbf{w} \in \mathbb{R}^n_{>0}$ with $\sum \mathbf{w} = 1$, and for some continuously differentiable, monotonic bijection $g : A \to B$ that is either convex or concave. Then for any $a, b \in A$ with $a \le b$, we have that M on $A \cap [a, b]$ is

$$\begin{cases} [qc]\text{-compressing} & \text{if } g \text{ is increasing and convex, or decreasing and concave;} \\ [q^{-1}c]\text{-compressing} & \text{if } g \text{ is increasing and concave, or decreasing and convex;} \end{cases}$$

where $q := \dfrac{g'(a)}{g'(b)}$ *and* $c := \dfrac{\mathrm{Min}(\mathbf{w})}{1 - \mathrm{Min}(\mathbf{w})}$. ⅃

Proof. This follows from Proposition 2.2.1 by using that $\mathrm{AM}_{\mathbf{w}}$ is $[c]$-compressing, which we know from Proposition 1.2.6. □

Corollary 2.2.3 ((*) Compressing parameters for Pythagorean means)**.**
Let $n \in \mathbb{N}_{\ge 2}$, let $\mathbf{w} \in \mathbb{R}^n_{>0}$ with $\sum \mathbf{w} = 1$, let $c := \mathrm{Min}(\mathbf{w})/(1 - \mathrm{Min}(\mathbf{w}))$.
For all $a, b \in \mathbb{R}_{>0}$ with $a \le b$,
$\mathrm{GM}_{\mathbf{w}}$ *is $[ca/b]$-compressing on $[a, b]$, and*
$\mathrm{HM}_{\mathbf{w}}$ *is $[ca^2/b^2]$-compressing on $[a, b]$.* ⅃

Proof. This follows by applying Corollary 2.2.2, using that $\mathrm{GM}_{\mathbf{w}}^{[\log]} = \mathrm{AM}_{\mathbf{w}} = \mathrm{HM}_{\mathbf{w}}^{[\mathrm{inv}]}$, and that $\log : \mathbb{R}_{>0} \to \mathbb{R}$ is increasing and concave,
while $\mathrm{inv} : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ is decreasing and convex. □

Corollary 2.2.4 ((*) Properties of inversion duals of internal means)**.**
Let $A \subseteq \mathbb{R}_{>0}$, let $n \in \mathbb{N}$, and let M be an internal mean on A^n.
Let i be an inversion of A, as in Definition 1.3.6.

Then $M^{[i]}$ is an internal mean on $i(A)^n$. If M is symmetric, increasing, strictly increasing, continuous, or smooth, respectively, then $M^{[i]}$ is so as well.

Further, let $a, b \in i(A)$. If M is $[c, C]$-compressing on $A \cap [i(b), i(a)]$, then

$$M^{[i]} \text{ is on } i(A) \cap [a,b] \begin{cases} [C^{-1}, c^{-1}]\text{-compressing} & \text{if } i = \operatorname{ainv}_t \text{ for some } t \in \mathbb{R}; \\ [\frac{a^2}{b^2}C^{-1}, \frac{b^2}{a^2}c^{-1}]\text{-compressing} & \text{if } i = \operatorname{inv}_t \text{ for some } t \in \mathbb{R}_{>0}. \end{cases}$$

Further, specifically for "$i = \operatorname{inv}$": If M is scale-invariant, then $M^{[\mathrm{inv}]}$ is scale-invariant. If M is $(\mathbf{a}, \mathbf{b})$-power-asymptotic, then $M^{[\mathrm{inv}]}$ is $(\mathbf{b}, \mathbf{a})$-power-asymptotic. ⌟

Proof. The "compressing" statement follows from Corollary 2.2.1, using that i is decreasing and convex, and that $(M^{[i]})^{[i]} = M$. The other statements follow by applying Theorem 2.1.3 with $f := i : A \to i(A)$, and specifically with $f = \operatorname{inv}$ in the case of the "scale-invariant" and the "power-asymptotic" statements. □

2.3 Examples of means constructed by conjugation

In this short section, we give two miscellaneous examples of means that are obtained by conjugation of other means, to illustrate some aspects of Theorem 2.1.3: the first example concerns an unusual kind of scale-invariant mean, the second example and the question that follows it are concerned with F-rational means.

Example 2.3.1. Let $p > 0$ and $a > 0$, and let

$$f: \quad \mathbb{R} \to \mathbb{R}: \quad x \mapsto \begin{cases} (x/a)^{1/p} & \text{if } x \geq 0 \\ -(-x)^{1/p} & \text{if } x \leq 0. \end{cases}$$

So f is an increasing, continuous bijection. Moreover, $f(\lambda x) = \lambda^{1/p} f(x)$ for all $x \in \mathbb{R}$ and $\lambda \in \mathbb{R}_{>0}$. Thus, if M is a scale-invariant mean on $\mathbb{R}^n$, then Theorem 2.1.3 tells us that $M^{[f]}$ is a scale-invariant mean on $\mathbb{R}^n$. If M is symmetric, increasing, strictly increasing, or continuous, respectively, then $M^{[f]}$ is so as well. However, f is not in general smooth; and, as $(x, y) \to (0, 0)$, the difference quotient $(f(y) - f(x))/(y - x)$ approaches zero in the case that $1/p > 1$, and it approaches infinity in the case that $1/p < 1$. So Theorem 2.1.3 tells us nothing about smoothness or compressingness of $M^{[f]}$.

For a concrete example: let $\mathbf{w} \in \mathbb{R}^n_{>0}$ for some $n \geq 1$; then $\mathrm{AM}^{[f]}_{\mathbf{w}}$ is a strictly increasing, continuous, scale-invariant mean on $\mathbb{R}^n$. We have for instance

$$\mathrm{AM}^{[f]}_3(2, 3, -5) = f\Big(\frac{a(2^p + 3^p) - 5^p}{3}\Big) = \begin{cases} \Big(\dfrac{2^p + 3^p - a^{-1}5^p}{3}\Big)^{1/p} & \text{if } a(2^p + 3^p) \geq 5^p \\ -\Big(\dfrac{-a(2^p + 3^p) + 5^p}{3}\Big)^{1/p} & \text{else.} \end{cases}$$

Thus, informally speaking, a serves as the "weight factor for positive numbers relative to negative numbers": "As $a \to \infty$, the influence of negative numbers vanishes, while as $a \to 0$, the influence of positive numbers vanishes; at $a = 1$, the influences are balanced."

Example 2.3.2. Let F be a subfield of $\mathbb{R}$. For any matrix $\mathbf{M} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in F^{2\times 2}$ with $\det(M) \neq 0$, the bijection

$$f_{\mathbf{M}}: \quad \mathbb{R}_{\neq -d/c} \to \mathbb{R}_{\neq a/c}: \quad x \mapsto \frac{ax+b}{cx+d}$$

restricts to a monotonic injection on $\mathbb{R}_{>-d/c}$, which is decreasing and convex and with image $\mathbb{R}_{>a/c}$ if $\det(M) > 0$, while it is increasing and concave and with image $\mathbb{R}_{<a/c}$ if $\det(M) < 0$. Thus, is M is an internal mean on $\mathbb{R}^n_{>-d/c}$, then $M^{[f_{\mathbf{M}}]}$ is an internal mean on $\mathbb{R}^n_{>a/c}$ or $\mathbb{R}^n_{<a/c}$. If M is compressing, then Corollary 2.2.1 easily yields formulas for the compressingness of $M^{[f_{\mathbf{M}}]}$, using that $f'_{\mathbf{M}}(y)/f'_{\mathbf{M}}(x) = (\frac{cx+d}{cy+d})^2$ for any $x, y \in \mathbb{R}_{>-d/c}$. Moreover, because $f_{\mathbf{M}}^{-1} = f_{\mathbf{M}^{-1}}$ is F-rational, it follows that if M is F-rational, then $M^{[f_{\mathbf{M}}]}$ is F-rational.

I did not attempt to solve the following question:

***Question* 2.3.3.** Let F be a subfield of $\mathbb{R}$, let I be an interval of $\mathbb{R}$, let $n \in \mathbb{N}$, let M be an F-rational mean on I. Does it follow that there exists $f_{\mathbf{M}}$ as in Example 2.3.2, and $\mathbf{w} \in F^n_{>0}$, such that $M = \mathrm{AM}_{\mathbf{w}}^{[f_{\mathbf{M}}]}$, where $\mathrm{AM}_{\mathbf{w}}$ is restricted to some interval J for which $f_{\mathbf{M}} : J \to I$ is a monotonic bijection?

2.4 Quasi-arithmetic means

This section introduces an important subset of means constructed by conjugation: *quasi-arithmetic* means, which are continuous conjugates of (weighted) arithmetic means on intervals. They are probably the most comprehensively studied kind of means. The power means and translation means that we study in §2.6–2.7 are the main examples.

Definition 2.4.1. Let $n \in \mathbb{N}$, let A be a non-degenerate interval of $\mathbb{R}$.
A mean M on A^n is called *quasi-arithmetic* if there exists $\mathbf{w} \in \mathbb{R}^n_{>0}$,
and a non-degenerate interval B of $\mathbb{R}$, and a continuous monotonic bijection $g : B \to A$, such that $M = \mathrm{AM}_{\mathbf{w}}^{[g]}$.

***Remark* 2.4.2.** Some sources, like [MP15] and [HLP34], add to the definition of the quasi-arithmetic property the condition that $\mathbf{w} \in \mathrm{diag}_n(\mathbb{R}^n_{>0})$. This is equivalent (by Fact 2.4.3.3) to the condition that M is symmetric.

Fact 2.4.3 (Basic facts about quasi-arithmetic (and slightly more general) means)**.**
Let $n \in \mathbb{N}$, let $B \subseteq \mathbb{R}$ and $C \subseteq A \subseteq \mathbb{R}$,
let $g : B \to A$ be a continuous monotonic bijection. Let $M = \mathrm{AM}_{\mathbf{w}}^{[g]}$.

1. *M is a strictly increasing, continuous, compressing internal mean on A^n.*

2. *M restricted to C^n is a quasi-arithmetic mean on C^n, provided that C is a non-degenerate interval of $\mathbb{R}$.*

3. *M is symmetric if and only if $\mathbf{w} \in \operatorname{diag}_n(\mathbb{R}^n_{>0})$.* ♩

Proof. 1. Follows from Theorem 2.1.3.

2. The restriction of g to $D := g^{-1}(C)$ is a continuous monotonic bijection $D \to C$, and $M|_{C^n} = \mathrm{AM}_{\mathbf{w}}^{[g|_D]}$.

3. Because $M = \mathrm{AM}_{\mathbf{w}}^{[g]}$ and $M^{[g^{-1}]} = \mathrm{AM}_{\mathbf{w}}$, it follows from Theorem 2.1.3.2 that M is symmetric if and only if $\mathrm{AM}_{\mathbf{w}}$ is symmetric. Clearly, because there are $x, y \in B$ with $x \neq y$, the mean $\mathrm{AM}_{\mathbf{w}}$ on B^n is symmetric if and only if $\mathbf{w} \in \operatorname{diag}_n(\mathbb{R}^n_{>0})$. □

Example 2.4.4 (Examples of means that are or are not quasi-arithmetic)**.**

1. All the concrete examples of continuous means that we studied up to this point, except for Min and Max and Examples 1.1.25, are quasi-arithmetic.

2. In §2.6 and §2.7 we study power means and translation means. All of them, except for Min and Max (extreme cases of power means), are quasi-arithmetic means.

3. I expect that the *composite power means* and the *composite translation means* that we define in Definition 3.7.1, are in general not quasi-arithmetic; I say 'in general', because these means depend on multiple real parameters, and there are parameter combinations for which the mean is quasi-arithmetic. My precise expectation is Conjecture 3.8.9. We proof for infinitely many of those means that they are not quasi-arithmetic (Theorems 3.8.6 and 3.8.8).

4. In [Haj13, §9], it is shown that the so-called logarithmic mean on $\mathbb{R}^2_{>0}$, defined by $L(a, b) = \dfrac{a - b}{\log(a) - \log(b)}$, is not quasi-arithmetic. ♪

***Remark* 2.4.5** (Results and literature about quasi-arithmetic means)**.**
Much is known about quasi-arithmetic means; probably because they are ubiquitous on the one hand, but very concrete on the other hand.

The means that are considered in the classical book [HLP34] are almost exclusively quasi-arithmetic means.

Chapter IV of the encyclopaedic book [Bul03] contains many results about quasi-arithmetic means that we don't mention; it mainly focusses on (complicated) inequalities.

Theorem 1 on pp. 287–291 of the encyclopaedic book [AD89] gives a complete characterisation of two-dimensional symmetric quasi-arithmetic means M in terms of a simple "bisymmetry" functional equation for M. This has been generalised in recent years to means in all dimensions n and that are not necessarily symmetric: 𝄿

Theorem 2.4.6 (Characterisation of quasi-arithmetic means)**.**
Let A be a non-degenerate interval of $\mathbb{R}$, let $n \geq 2$, let $M : A^n \to A$ be a function. The conditions 1 and 2 are equivalent:

1. *M is a quasi-arithmetic mean on A^n.*

2. *M is constant-preserving, strictly increasing, and continuous.*
   Moreover, M satisfies the bisymmetry functional equation*:*

$$M\big(M(\mathbf{x}_1),\dots,M(\mathbf{x}_n)\big) = M\big(M(\mathbf{x}'_1),\dots,M(\mathbf{x}'_n)\big) \qquad \forall(\mathbf{x}_1,\dots,\mathbf{x}_n) \in A^{n\times n},$$

   where $(\mathbf{x}'_1,\dots,\mathbf{x}'_n) \in A^{n\times n}$, denotes the transpose of $(\mathbf{x}_1,\dots,\mathbf{x}_n)$. ⅃

Reference. Theorem C and the sentence directly below it in [MP15]. □

***Remark* 2.4.7** (Continuation of 2.4.5)**.** Theorem 5 in [MP15] – the statement of it involves Gauss composition, the subject of study of our next chapter – is a further generalisation of Theorem 2.4.6, although only for the case of symmetric means.

Apart from Theorem 2.4.6, the most important general result about quasi-arithmetic means is probably Theorem 2.4.8, which gives a necessary and sufficient relation between f and g for the equality $\mathrm{AM}_{\mathbf{w}}^{[f]} = \mathrm{AM}_{\mathbf{w}}^{[g]}$ to hold. We don't use Theorem 2.4.6 in this text, but we do use Theorem 2.4.8, especially in Chapter 3 to show that certain means are not quasi-arithmetic.

Theorem 2.4.8. *Let $n \in \mathbb{N}_{\geq 2}$, let A, B, I be non-degenerate intervals of $\mathbb{R}$, let $f : A \to I$ and $g : B \to I$ be continuous monotonic bijections. We have*

$$\mathrm{AM}_n^{[f]} = \mathrm{AM}_n^{[g]} \text{ on } I^n \iff f = g \circ \mathrm{tra}_b \circ \mathrm{mul}_a \text{ for some } a, b \in \mathbb{R}.$$ ⅃

(Reference to) proof. "$\Longleftarrow$" is clear. For the other implication, suppose that $\mathrm{AM}_n^{[f]} = \mathrm{AM}_n^{[g]}$. In the notation of [MP15, Theorem A], this is written as $A_{f^{-1}} = A_{g^{-1}}$. By the implication "(ii) $\implies$ (iii)" in that Theorem A, it follows that there are $a, b \in \mathbb{R}$ such that $g^{-1}(x) = af^{-1}(x) + b$ for all $x \in I$; that is, $g^{-1} = \mathrm{tra}_b \circ \mathrm{mul}_a \circ f^{-1}$. Hence, $f = g \circ \mathrm{tra}_b \circ \mathrm{mul}_a$.

Right after the formulation of Theorem A, there are references to its three independent original sources. □

2.5 The conjugation group action

We don't really apply the material of this section anywhere else in this text, except for proving analogous statements, like that we have a conjugation action on families of means (Corollary 4.3.9). So the reader may skip over this section and the text remains as intelligible. However, we find the observation that conjugation defines group actions too interesting and obvious to neglect it; peculiarly, I found no mention of group actions in other text on means, perhaps because of the strong emphasis in most of the literature about means on analytical rather than algebraic aspects.

Corollary 2.5.1 (Conjugation action on means)**.** *Let $n \in \mathbb{N}_{\geq 2}$, let $A \subseteq \mathbb{R}$, let G be the group of monotonic bijections $A \to A$, with composition of functions as operation. Let $\mathbb{M}$ be the set of internal means on A^n.*

1. *G acts on $\mathbb{M}$ by the map $(f, M) \mapsto M^{[f]}$.*

2. *The subsets* $\{$*symmetric means*$\}$ *and* $\{$*increasing means*$\}$ *and* $\{$*strictly increasing means*$\}$ *of* $\mathbb{M}$ *are invariant under* G.

3. *The subset* $\{$*continuous means*$\}$ *of* $\mathbb{M}$ *is invariant under the subgroup* $G_{cont} := \{f \in G : f \text{ is a homeomorphism}\}$ *of* G.

4. *The subsets* $\{$*smooth means*$\}$ *and* $\{$*compressing means*$\}$ *of* $\mathbb{M}$ *are invariant under the subgroup* $G_{smoo} := \{f \in G : f \text{ is smooth and } f' \text{ has no zeros on } A\}$ *of* G.

Proof. The statements follow directly from Theorem 2.1.3 and Proposition 1.3.7.2. □

Example 2.5.2. Let $n \in \mathbb{N}_{\geq 2}$. This example serves to show that, in the context of Corollary 2.5.1, the action of G_{cont} on $\{$continuous internal means on $A^n\}$ is not free.

1. The means Min_n and Max_n on A^n together form one orbit, and their stabiliser is the subgroup of index 2 in G_{cont} consisting of the increasing continuous bijections $A \to A$. This follows from Fact 1.3.9.

2. If M is a scale-invariant internal mean on A^n, and A is closed under multiplication, then the stabiliser of M clearly contains as a subgroup $\{\text{mul}_a : a \in A_{>0}\}$.

3. The stabiliser of $\text{AM}_{\mathbf{w}}$ on $\mathbb{R}^n$ clearly contains $H := \{\text{tra}_b \circ \text{mul}_a : \ b \in \mathbb{R}, \ a \in \mathbb{R}_{\neq 0}\}$ as a subgroup, for all $\mathbf{w} \in \mathbb{R}^n_{>0}$. If $\mathbf{w} \in \text{diag}(\mathbb{R}^n_{>0})$, then the stabiliser equals H; this follows by Theorem 2.4.8 (taking $A = I = B = \mathbb{R}$ and $g = \text{id}$).

4. The stabiliser of $\text{GM}_{\mathbf{w}}$ on $\mathbb{R}^n_{>0}$ contains $H' := \{\text{mul}_b \circ \text{pow}_a : \ b \in \mathbb{R}_{>0}, \ a \in \mathbb{R}_{\neq 0}\}$ as a subgroup. This follows by the facts that H stabilises $\text{AM}_{\mathbf{w}}$, and
$$\text{GM}_{\mathbf{w}} = \text{AM}_{\mathbf{w}}^{[\exp]}, \qquad \exp \circ \text{tra}_b = \text{mul}_{e^b} \circ \exp, \qquad \exp \circ \text{mul}_a = \text{pow}_a \circ \exp .$$

5. In Proposition 3.8.6 we construct, for all $n \geq 3$, a "continuum" of increasing, continuous, symmetric internal means M_p on $\mathbb{R}^n_{>0}$ —parametrised by $p \in \mathbb{R}$ such that the map $p \mapsto M_p$ is injective— such that the means M_p are all in the same orbit of the action by G_{cont}, and whose stabilisers all contain $\{\text{id}, \text{inv}\}$ as a subgroup.

***Question* 2.5.3.** Let $A \subseteq \mathbb{R}$, let $n \in \mathbb{N}_{\geq 2}$. Let G_{cont} and G_{smoo} be as in Corollary 2.5.1.

1. What is the cardinality of the set of orbits of the action of G_{cont} on $\{$symmetric, strictly increasing, continuous internal means on $A^n\}$?

2. What is the cardinality of the set of orbits of the action of G_{smoo} on $\{$symmetric, strictly increasing, smooth internal means on $A^n\}$?

***Remark* 2.5.4.** I did not attempt to solve Question 2.5.3. In the Theorems 3.8.6 and 3.8.8, we show that for all $n \geq 2$ there are symmetric, strictly increasing, compressing, scale-invariant internal means on $\mathbb{R}^n_{>0}$ that are not quasi-arithmetic. It follows for the first question in 2.5.3, in the case that $A = \mathbb{R}_{>0}$, that the cardinality is at least two.

2.6 Power means

Power means, which we define in Definition 2.6.2, are probably the best-known and most used kind of means, after the pythagorean means (which are examples of power means). They are correspondingly well-studied; several properties of them that we mention, and many more that we don't mention (mainly about inequalities), can be found for example in [HLP34, Chapters II & III & V] and [Bul03, Chapter III].

We did however not find the "full continuity and smoothness" statements (Theorem 2.6.6.6) in the literature. We make Definition 2.6.1 to conceptualise the "full continuity" statement, because we study more examples with similar "full continuity" properties: namely, Theorem 2.7.4 about translation means, Theorems 3.6.5 and 3.7.4.2 about Gauss compositions, and Theorem 4.6.1 about families of continuous means, provide other examples of *continuums of means.*

Definition 2.6.1 ((†) Notation $\mathbb{M}_{A^n}$, and Continuum of means)**.** Let $A \subseteq \mathbb{R}$, let $n \in \mathbb{N}$.

1. We write $\mathbb{M}_{A^n}$ to denote the set of means on A^n.
2. Let let T be a topological space, and let

$$M: \quad T \to \mathbb{M}_{A^n}: \quad \tau \mapsto M_\tau$$

be a function. If the map

$$T \times A^n \to A: \quad (\tau, \mathbf{x}) \mapsto M_\tau(\mathbf{x})$$

is continuous, with respect to the product topology on $T \times A^n$, then we say that the map M is a *continuum of means (on A^n).* ≀

2.6.1 About the definition of power means

Definition 2.6.2. Let $p \in \overline{\mathbb{R}}$, let $n \in \mathbb{N}$, let $\mathbf{w} \in \mathbb{R}^n_{>0}$. The *power mean* $\mathrm{PM}_{p,\mathbf{w}}$ is the internal mean on $\mathbb{R}^n_{>0}$ given by

$$\mathrm{PM}_{p,\mathbf{w}} = \begin{cases} \mathrm{Max}_n & \text{if } p = \infty \\ \mathrm{AM}_{\mathbf{w}}^{[\mathrm{pow}_{1/p}]} & \text{if } p \in \mathbb{R}_{\neq 0} \\ \mathrm{GM}_{\mathbf{w}} & \text{if } p = 0 \\ \mathrm{Min}_n & \text{if } p = -\infty. \end{cases}$$

In the case that $\mathbf{w} \in \mathrm{diag}(\mathbb{R}^n_{>0})$, we simply write $\mathrm{PM}_{p,n}$ instead of $\mathrm{PM}_{p,\mathbf{w}}$. ≀

***Remark* 2.6.3.** More explicitly, for $p \in \mathbb{R}_{\neq 0}$ and $\mathbf{x} \in \mathbb{R}^n_{>0}$ we have

$$\mathrm{PM}_{p,\mathbf{w}}(\mathbf{x}) = \left(\frac{\sum \mathbf{w}\mathbf{x}^p}{\sum \mathbf{w}} \right)^{1/p}. \tag{2.5}$$

Clearly, $\mathrm{PM}_{p,\mathbf{w}} = \mathrm{PM}_{p,\lambda\mathbf{w}}$ for any $\lambda \in \mathbb{R}_{>0}$, so we may assume without loss of generality that $\sum \mathbf{w} = 1$.

Power means encompass the pythagorean means, because

$$\mathrm{HM}_{\mathbf{w}} = \mathrm{PM}_{-1,\mathbf{w}}, \qquad \mathrm{GM}_{\mathbf{w}} = \mathrm{PM}_{0,\mathbf{w}}, \qquad \mathrm{AM}_{\mathbf{w}} = \mathrm{PM}_{1,\mathbf{w}}. \tag{2.6}$$

The reason for the particular definition of the power mean in the case that $p \in \{0, -\infty, \infty\}$, becomes clear from part 3 of the following theorem: the definition is such that $(p, \mathbf{w}) \mapsto \mathrm{PM}_{p,\mathbf{w}}$ is a continuum of means, for every n.

Thus, informally speaking, we can see the map $p \mapsto \mathrm{PM}_{p,\mathbf{w}}$ as a continuous interpolation between the means $\mathrm{Min}_n, \mathrm{HM}_{\mathbf{w}}, \mathrm{GM}_{\mathbf{w}}, \mathrm{AM}_{\mathbf{w}}, \mathrm{Max}_n$ as p increases from $-\infty$ to ∞; moreover, this map is increasing in p, by Theorem 2.6.6.7.

2.6.2 Jensen's inequality in the context of the arithmetic mean

In the proof of the classical result that the map $p \mapsto \mathrm{PM}_{p,\mathbf{w}}(\mathbf{x})$ is increasing (Theorem 2.6.6), we use the classical inequality called Jensen's inequality, or actually a direct corollary of it (2.6.5).

Theorem 2.6.4 (Jensen's inequality)**.** *Let A be an interval of $\mathbb{R}$, let $f : A \to \mathbb{R}$ be a convex function. Let $n \in \mathbb{N}_{\geq 2}$, let $\mathbf{w} \in \mathbb{R}^n_{>0}$, let $\mathbf{x} \in A^n$. Then*

$$f\left(\frac{\sum \mathbf{w}\mathbf{x}}{\sum \mathbf{w}}\right) \leq \frac{\sum \mathbf{w}\langle f \rangle \mathbf{x}}{\sum \mathbf{w}} \tag{2.7}$$

If f is strictly convex and $\mathbf{x} \notin \mathrm{diag}(A^n)$, then the inequality in (2.7) is strict.

Theorem 2.6.4 is a direct reformulation of [Bul03, §1.4, Theorem 12]. That location contains a nice exposition about Jensen's inequality and related inequalities.

Corollary 2.6.5. *Let A be an interval of $\mathbb{R}$, let $f : A \to A$ be a strictly convex bijection. Let $n \in \mathbb{N}_{\geq 2}$, let $\mathbf{w} \in \mathbb{R}^n_{>0}$. For all $\mathbf{x} \in A^n \setminus \mathrm{diag}(A^n)$, we have*

$$\mathrm{AM}^{[f]}_{\mathbf{w}}(\mathbf{x}) < \mathrm{AM}_{\mathbf{w}}(\mathbf{x}).$$

Proof. This follows directly by substituting $\langle f^{-1} \rangle \mathbf{x}$ for $\mathbf{x}$ in (2.7). □

2.6.3 Properties of power means

The next theorem constitutes the core of §2.6. It consists of well-known and unknown properties of power means. The well-known properties (without † or ∗) are mentioned in some way in [HLP34, Chapters II & III & V] and [Bul03, Chapter III]. Also well-known is that the map $p \mapsto \mathrm{PM}_{p,\mathbf{w}}(\mathbf{x})$ is continuous; but the "full continuity" statement in part 6 is, to my knowledge, not yet treated in the literature.

Theorem 2.6.6 (Power means)**.** *Let $n \in \mathbb{N}_{\geq 2}$, let $p \in \overline{\mathbb{R}}$, let $\mathbf{w} \in \mathbb{R}^n_{>0}$, let $\mathbf{x} \in \mathbb{R}^n_{>0}$.*

1. *"Duality": $\mathrm{PM}^{[\mathrm{inv}]}_{p,\mathbf{w}} = \mathrm{PM}_{-p,\mathbf{w}}$.*

2. *"Kinship": $\mathrm{PM}^{[\mathrm{pow}_q]}_{p,\mathbf{w}} = \mathrm{PM}_{p/q,\mathbf{w}}$, for all $q \in \mathbb{R}_{\neq 0}$.*

3. *"Mean": $\mathrm{PM}_{p,\mathbf{w}}$ is a continuous, increasing, scale-invariant internal mean on $\mathbb{R}^n_{>0}$. It is symmetric if $\mathbf{w} \in \mathrm{diag}(\mathbb{R}^n_{>0})$. It is strictly increasing and smooth if $p \notin \{\infty, -\infty\}$.*

4. *(†) "Power-asymptotic": Let $\sum \mathbf{w} = 1$. Then $\mathrm{PM}_{p,\mathbf{w}}$ is*
   - *$(0,1)$-power-asymptotic if $p > 0$;*
   - *$(\mathbf{w},\mathbf{w})$-power-asymptotic if $p = 0$;*
   - *$(1,0)$-power-asymptotic if $p < 0$.*

5. *(*) "Compressing": Let $\sum \mathbf{w} = 1$. Let $a, b \in \mathbb{R}_{>0}$ with $a < b$, let $c := \frac{\mathrm{Min}(\mathbf{w})}{1-\mathrm{Min}(\mathbf{w})}$. Then $\mathrm{PM}_{p,\mathbf{w}}$ on the interval $[a,b]$ is*
   - *$[\,(a/b)^{p-1}c\,]$-compressing if $1 \le p < \infty$;*
   - *$[\,(a/b)^{1-p}c\,]$-compressing if $-\infty < p \le 1$.*

6. *(†) "Continuous and smooth": The function*
$$\psi : \quad \overline{\mathbb{R}} \times \mathbb{R}^n_{>0} \times \mathbb{R}^n_{>0} \to \mathbb{R}_{>0} : \quad (p, \mathbf{w}, \mathbf{x}) \mapsto \mathrm{PM}_{p,\mathbf{w}}(\mathbf{x})$$
is continuous; that is, the map $\overline{\mathbb{R}} \times \mathbb{R}^n_{>0} \to \mathbb{M}_{\mathbb{R}^n_{>0}} : (p, \mathbf{w}) \mapsto \mathrm{PM}_{p,\mathbf{w}}$ is a continuum of means.

   The restriction of ψ to $(\mathbb{R}_{\neq 0}) \times \mathbb{R}^n_{>0} \times \mathbb{R}^n_{>0}$ is smooth.

7. *"Increasing in p": The function*
$$\varphi : \quad \overline{\mathbb{R}} \to \mathbb{R}_{>0} : \quad p \mapsto \mathrm{PM}_{p,\mathbf{w}}(\mathbf{x})$$
is continuous and increasing, and it is strictly increasing if $\mathbf{x} \notin \mathrm{diag}(\mathbb{R}^n_{>0})$. ♩

Proof.

1. ("Duality".) For $p \in \{0, \pm\infty\}$, this is contained in Corollary 1.3.11. For $p \in \mathbb{R}_{\neq 0}$, note that $\mathrm{inv} \circ \mathrm{pow}_{1/p} = \mathrm{pow}_{-1/p}$, hence $(\mathrm{AM}_{\mathbf{w}}^{[\mathrm{pow}_{1/p}]})^{[\mathrm{inv}]} = \mathrm{AM}_{\mathbf{w}}^{[\mathrm{pow}_{-1/p}]}$.

2. ("Kinship"). For $p \notin \{0, \pm\infty\}$, this follows because $\mathrm{pow}_q \circ \mathrm{pow}_{1/p} = \mathrm{pow}_{q/p}$. For $p \in \{\pm\infty\}$ it follows from Lemma 1.3.9, using that pow_q is increasing if $q > 0$ and decreasing if $q < 0$. For $p = 0$, it follows from the calculation
$$\mathrm{GM}_{\mathbf{w}}^{[\mathrm{pow}_q]} = (\mathrm{AM}_{\mathbf{w}}^{[\exp]})^{[\mathrm{pow}_q]} = (\mathrm{AM}_{\mathbf{w}}^{[\mathrm{mul}_q]})^{[\exp]} = \mathrm{AM}_{\mathbf{w}}^{[\exp]} = \mathrm{GM}_{\mathbf{w}},$$
where the second equality follows because $\mathrm{pow}_q \circ \exp = \exp \circ \mathrm{mul}_q$, and the third equality follows because $\mathrm{AM}_{\mathbf{w}}$ is scale-invariant.

3. ("Mean"). For $p \in \{\pm\infty\}$, the statements are contained in Fact 1.1.17, and for $p \in \{0, 1\}$ in Proposition 1.2.6.1. For $p \in \mathbb{R}_{\neq 0}$, the statements follow from Theorem 2.1.3, applied with $M := \mathrm{AM}_{\mathbf{w}}$ and $f := \mathrm{pow}_{1/p} : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$. We use in particular that $f(\lambda x) = \lambda^{1/p} f(x)$ for all $\lambda, x \in \mathbb{R}_{>0}$.

4. ("Power-asymptotic".) This is contained in Proposition 1.2.6.2 for $p = 0$, and it follows from Theorem 2.1.3.6 for $p \neq 0$, using in particular that f is multiplicative.

5. ("Compressing".) For $p = 0$, this is contained in Corollary 2.2.3. For $p \in \mathbb{R}_{\neq 0}$, it follows from Corollary 2.2.2, applied with $g := \mathrm{pow}_p : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$. We use that g is increasing and convex if $p \geq 1$, increasing and concave if $0 < p \leq 1$, and decreasing and convex if $p < 0$.

6. ("Continuous and smooth".) It is clear that ψ is smooth on $(\mathbb{R}_{\neq 0}) \times \mathbb{R}^n_{>0} \times \mathbb{R}^n_{>0}$, because there it is given by (2.5). Hence, to show continuity on $\overline{\mathbb{R}} \times \mathbb{R}^n_{>0} \times \mathbb{R}^n_{>0}$, it suffices to show that

$$\lim_{(p,\mathbf{w},\mathbf{x}) \to (p_0,\mathbf{w}_0,\mathbf{x}_0)} \mathrm{PM}_{p,\mathbf{w}}(\mathbf{x}) = \mathrm{PM}_{p_0,\mathbf{w}_0}(\mathbf{x}_0) \tag{2.8}$$

for all $(p_0, \mathbf{w}_0, \mathbf{x}_0) \in \{0, -\infty, \infty\} \times \mathbb{R}^n_{>0} \times \mathbb{R}^n_{>0}$.

Case 1: $p_0 \in \{\pm\infty\}$. It suffices to show (2.8) for $p_0 = \infty$, because then it follows by continuity of inv that $\mathrm{PM}^{[\mathrm{inv}]}_{p,\mathbf{w}}(\mathbf{x}) \to \mathrm{PM}^{[\mathrm{inv}]}_{\infty,\mathbf{w}_0}(\mathbf{x}_0)$ as $(p, \mathbf{w}, \mathbf{x}) \to (p_0, \mathbf{w}_0, \mathbf{x}_0)$, that is, $\mathrm{PM}_{-p,\mathbf{w}}(\mathbf{x}) \to \mathrm{PM}_{-\infty,\mathbf{w}_0}(\mathbf{x}_0)$.

By multiplying (2.5) with $\frac{\mathrm{Max}(\mathbf{x})}{\mathrm{Max}(\mathbf{x})}$, we get

$$\mathrm{PM}_{p,\mathbf{w}}(\mathbf{x}) = \begin{cases} \mathrm{Max}(\mathbf{x})\Big(\dfrac{\sum \mathbf{w}(\frac{\mathbf{x}}{\mathrm{Max}(\mathbf{x})})^p}{\sum \mathbf{w}}\Big)^{1/p} & \text{if } p \in \mathbb{R}_{>0} \\ \mathrm{Max}(\mathbf{x}) & \text{if } p = \infty. \end{cases} \tag{2.9}$$

As we want to study the behaviour as $(p, \mathbf{w}, \mathbf{x}) \to (\infty, \mathbf{w}_0, \mathbf{x}_0)$, we may assume without loss of generality that $p > 0$, and that there exists $c > 0$ such that $\mathrm{Min}(\mathbf{w})/\sum \mathbf{w} > c$ for all $\mathbf{w}$. Thus, the expression between big brackets in (2.9) is bounded from above by 1, and bounded from below by c. Hence, that expression raised to the power $1/p$ approaches 1 as $p \to \infty$. Hence, $\mathrm{PM}_{p,\mathbf{w}}(\mathbf{x}) \to \mathrm{Max}(\mathbf{x}_0)$ as $(p, \mathbf{w}, \mathbf{x}) \to (\infty, \mathbf{w}_0, \mathbf{x}_0)$.

Case 2: $p_0 = 0$. It suffices to show that

$$\lim_{(p,\mathbf{w},\mathbf{x}) \to (0,\mathbf{w}_0,\mathbf{x}_0)} \mathrm{PM}^{[\log]}_{p,\mathbf{w}}(\mathbf{x}) = \mathrm{AM}_{\mathbf{w}_0}(\mathbf{x}_0), \tag{2.10}$$

because then it would follow by continuity of log and exp that $\mathrm{PM}_{p,\mathbf{w}}(\mathbf{x}) \to \mathrm{GM}_{\mathbf{w}_0}(\mathbf{x}_0)$.

If $p \in \mathbb{R}_{\neq 0}$, we have

$$\mathrm{PM}^{[\log]}_{p,\mathbf{w}}(\mathbf{x}) = \frac{1}{p}\log\Big(\frac{\sum \mathbf{w}(\langle\exp\rangle\mathbf{x})^p}{\sum \mathbf{w}}\Big) = \frac{1}{p}\log\Big(\frac{\sum \mathbf{w}e^{p\mathbf{x}}}{\sum \mathbf{w}}\Big) = \frac{1}{p}\log\Big(1 + \frac{\sum \mathbf{w}(e^{p\mathbf{x}} - 1)}{\sum \mathbf{w}}\Big).$$

For $x \in \mathbb{R}$, we have $e^{px} - 1 = px + O(p^2x^2)$ as $px \to 0$, hence

$$\sum \mathbf{w}(e^{p\mathbf{x}} - 1) = \big(p\sum \mathbf{w}\mathbf{x}\big) + O(p^2\mathrm{Max}(\mathbf{x})^2)\sum \mathbf{w} \qquad \text{as } (p, \mathbf{x}) \to (0, \mathbf{x}_0).$$

Thus, if $p \in \mathbb{R}_{\neq 0}$, we have

$$\mathrm{PM}^{[\log]}_{p,\mathbf{w}}(\mathbf{x}) = \frac{1}{p}\log(1 + pf), \quad \text{where } f := \frac{\sum \mathbf{w}\mathbf{x}}{\sum \mathbf{w}} + O(p\,\mathrm{Max}(\mathbf{x})^2) \text{ as } (p, \mathbf{x}) \to (0, \mathbf{x}_0).$$

Moreover, for $x \in \mathbb{R}$, we have $\log(1+x) = x + O(x^2)$. Thus, we have

$$\mathrm{PM}^{[\log]}_{p,\mathbf{w}}(\mathbf{x}) = \begin{cases} f + O(pf^2) & \text{if } p \in \mathbb{R}_{\neq 0} \\ \mathrm{AM}_{\mathbf{w}}(\mathbf{x}) & \text{if } p = 0. \end{cases}$$

As $(p, \mathbf{w}, \mathbf{x}) \to (0, \mathbf{w}_0, \mathbf{x}_0)$, we clearly have $f \to \mathrm{AM}_{\mathbf{w}_0}(\mathbf{x}_0)$ and $pf^2 \to 0$, and therefore $\mathrm{PM}^{[\log]}_{p,\mathbf{w}}(\mathbf{x}) \to \mathrm{AM}_{\mathbf{w}_0}(\mathbf{x}_0)$. This completes the proof of (2.10).

7. ("Increasing in p".)[1] If $\mathbf{x} \in \mathrm{diag}(\mathbb{R}^n_{>0})$ then φ is constant, so we suppose that $\mathbf{x} \notin \mathrm{diag}(\mathbb{R}^n_{>0})$. Let $(p, q) \in \overline{\mathbb{R}}^2$ such that $p < q$. First, suppose $(p, q) \in \mathbb{R}^2_{>0}$. Then

$$\mathrm{AM}_{\mathbf{w}}(\mathbf{x})^{[\mathrm{pow}_{q/p}]} < \mathrm{AM}_{\mathbf{w}}(\mathbf{x})$$

by Lemma 2.6.5, because $\mathrm{pow}_{q/p}$ is convex. Hence,

$$\varphi(p) = \mathrm{AM}_{\mathbf{w}}(\mathbf{x})^{[\mathrm{pow}_{1/p}]} = (\mathrm{AM}_{\mathbf{w}}(\mathbf{x})^{[\mathrm{pow}_{q/p}]})^{[\mathrm{pow}_{1/q}]} < \mathrm{AM}_{\mathbf{w}}(\mathbf{x})^{[\mathrm{pow}_{1/q}]} = \varphi(q);$$

the inequality follows by Proposition 1.3.7.5 (applied with $A := \mathbb{R}^n_{>0} \backslash \mathrm{diag}(\mathbb{R}^n_{>0})$), because $\mathrm{pow}_{1/q}$ is strictly increasing. We conclude that φ is strictly increasing on $\mathbb{R}_{>0}$.

Next, suppose $(p, q) \in \mathbb{R}^2_{<0}$. Then $(-q, -p) \in \mathbb{R}^2_{>0}$ and $-q < -p$, hence

$$\mathrm{PM}_{-q,\mathbf{w}}(\mathbf{x}) < \mathrm{PM}_{-p,\mathbf{w}}(\mathbf{x})$$

by what we just showed. Because inv is strictly decreasing, it follows by Proposition 1.3.7.6 (applied with $A := \mathbb{R}^n_{>0} \setminus \mathrm{diag}(\mathbb{R}^n_{>0})$) that

$$\varphi(q) = \mathrm{PM}^{[\mathrm{inv}]}_{-q,\mathbf{w}}(\mathbf{x}) > \mathrm{PM}^{[\mathrm{inv}]}_{-p,\mathbf{w}}(\mathbf{x}) = \varphi(p).$$

Thus, φ is strictly increasing on $\mathbb{R}_{\neq 0}$.
Because φ is continuous on $\overline{\mathbb{R}}$, by part 3, it follows that φ is strictly increasing on $\overline{\mathbb{R}}$. ☐

Example 2.6.7. Let $x_1, \ldots, x_n, w_1, \ldots, w_n \in \mathbb{R}_{>0}$ such that $\sum w_i = 1$.
Part 3 of Theorem 2.6.6 tells us that

$$\lim_{p \to 0} (w_1 x_1^p + \cdots + w_n x_n^p)^{1/p} = x_1^{w_1} \cdots x_n^{w_n}.$$ ♪

Corollary 2.6.8. *Let $n \in \mathbb{N}$, let $\mathbf{w} \in \mathbb{R}^n_{>0}$. For all $\mathbf{x} \in \mathbb{R}^n_{>0}$, we have*

$$\mathrm{HM}_{\mathbf{w}}(\mathbf{x}) \leq \mathrm{GM}_{\mathbf{w}}(\mathbf{x}) \leq \mathrm{AM}_{\mathbf{w}}(\mathbf{x}),$$

and the inequalities are strict if and only if $\mathbf{x} \notin \mathrm{diag}(\mathbb{R}^n_{>0})$. 𝅗𝅥

Proof. This follows from Theorem 2.6.6.6, by the identities (2.6). ☐

[1] We based our proof of this on the proof in `https://en.wikipedia.org/wiki/Generalized_mean#Proof_of_power_means_inequality`.

***Remark* 2.6.9.** Corollary 2.6.8 is a classical result. In the Handbook [Bul03, §2.2.1], we read: "Although the inequality between the arithmetic and geometric means in its simplest form [i.e. with $\mathbf{w} = (1,1)$] was probably known in antiquity [by a geometric interpretation], the general result for weighted means of n numbers seems to have first appeared in print in the nineteenth century, in the notes of Cauchy's course (...), although certain results of Newton are related to this inequality".

The classical book [HLP34] states in §2.5: "This theorem is so fundamental that we propose to give a number of proofs, of varying degrees of simplicity and generality."

We give a new proof in Corollary 2.7.6.

2.6.4 The uniqueness of power means

The next theorem, which tells us that the power means without Min and Max are the only symmetric, scale-invariant, quasi-arithmetic means on $\mathbb{R}^n_{>0}$, will be useful in §3.8 for proving that certain means are not quasi-arithmetic.

Theorem 2.6.10. *Let $n \in \mathbb{N}$, let M be a symmetric, scale-invariant, quasi-arithmetic mean on $\mathbb{R}^n_{>0}$. There exists $p \in \mathbb{R}$ such that $M = \mathrm{PM}_{p,n}$.*

Proof. This proof is based on the proof of [HLP34, §3.3, Theorem 84].

Because M is symmetric and quasi-arithmetic, there exist by Fact 2.4.3.3 a non-degenerate interval B of $\mathbb{R}$, and a continuous monotonic bijection $f : B \to \mathbb{R}_{>0}$, such that $M = \mathrm{AM}_n^{[f]}$.

For $b \in \mathbb{R}$, we have $\mathrm{AM}_n = \mathrm{AM}_n^{[\mathrm{tra}_b]}$ on $\mathbb{R}^n$, hence we clearly have $M = \mathrm{AM}_n^{[f \circ \mathrm{tra}_b]}$, where $f \circ \mathrm{tra}_b$ maps $B - b \to \mathbb{R}_{>0}$. Since $f \circ \mathrm{tra}_b$ is continuous monotonic bijection and $B - b$ is a non-degenerate interval of $\mathbb{R}$, and since there exists b such that $(f \circ \mathrm{tra}_b)(0) = 1$, we can assume without loss of generality that $f(0) = 1$.

Because the mean $M = \mathrm{AM}_n^{[f]}$ is scale-invariant, it follows that for all $y \in \mathbb{R}_{>0}$, we have $\mathrm{AM}_n^{[f]} = (\mathrm{AM}_n^{[f]})^{[\mathrm{mul}_y]} = \mathrm{AM}_n^{[\mathrm{mul}_y \circ f]}$ on $\mathbb{R}^n_{>0}$. Thus, Theorem 2.4.8 tells us that there exist numbers $a(y), b(y) \in \mathbb{R}$ such that

$$f^{-1} \circ \mathrm{mul}_y \circ f = \mathrm{tra}_{b(y)} \circ \mathrm{mul}_{a(y)}.$$

Thus, writing $g := f^{-1} : \mathbb{R}_{>0} \to B$, we have

$$g \circ \mathrm{mul}_y = \mathrm{tra}_{b(y)} \circ \mathrm{mul}_{a(y)} \circ g.$$

Hence,

$$g(xy) = a(y)g(x) + b(y) \qquad \forall x, y \in \mathbb{R}_{>0}. \tag{2.11}$$

Moreover, $g(1) = 0$ since $f(0) = 1$, so by (2.11) we have $g(y) = b(y)$. Hence,

$$g(xy) = a(y)g(x) + g(y) \qquad \forall x, y \in \mathbb{R}_{>0}. \tag{2.12}$$

Because $g(yx) = g(xy)$, it follows that

$$a(y)g(x) + g(y) = a(x)g(y) + g(x) \qquad \forall x, y \in \mathbb{R}_{>0}. \tag{2.13}$$

If $x \neq 1$ and $y \neq 1$, then $g(x)$ and $g(y)$ are nonzero, so by (2.13) we have

$$\frac{a(y)-1}{g(y)} = \frac{a(x)-1}{g(x)}.$$

That is, the function $\mathbb{R}_{>0} \setminus \{1\} \to \mathbb{R} : y \mapsto \frac{a(y)-1}{g(y)}$ is constant, say equal to $c \in \mathbb{R}$. So $a(y) = cg(y) + 1$ for all $y > 0$ with $y \neq 1$. Substituting this in (2.12), it follows that

$$g(xy) = cg(x)g(y) + g(x) + g(y) \qquad \forall x, y \in \mathbb{R}_{>0}, \tag{2.14}$$

since (2.14) clearly holds in the case that $y = 1$, that is, $g(y) = 0$.

Case 1: $c = 0$. Then $g(xy) = g(x) + g(y)$. As g is continuous, it follows by elementary analysis that $g = \mathrm{mul}_s \circ \log$ for some $s \in \mathbb{R}$. Since g is injective, we have $s \neq 0$, so $f = \exp \circ \mathrm{mul}_{s^{-1}}$. Hence,

$$M = \mathrm{AM}_n^{[f]} = (\mathrm{AM}_n^{[\mathrm{mul}_{s^{-1}}]})^{[\exp]} = \mathrm{AM}_n^{[\exp]} = \mathrm{GM}_n = \mathrm{PM}_{0,n}.$$

Case 2: $c \neq 0$. Let $h := \mathrm{tra}_1 \circ \mathrm{mul}_c \circ g : \ \mathbb{R}_{>0} \to B'$, where $B' = cB + 1$. Then

$$h(xy) = cg(xy) + 1 = c\big((cg(x)+1)g(y) + g(x)\big) + 1 = (cg(x)+1)(cg(y)+1) = h(x)h(y).$$

Because h is continuous and not constant, it follows by elementary analysis that $h = \mathrm{pow}_p$ for some $p \in \mathbb{R}_{\neq 0}$. Thus,

$$M = \mathrm{AM}_n^{[f]} = \mathrm{AM}_n^{[h^{-1} \circ \mathrm{tra}_1 \circ \mathrm{mul}_c]} = \mathrm{AM}_n^{[h^{-1}]} = \mathrm{AM}_n^{[\mathrm{pow}_{1/p}]} = \mathrm{PM}_{p,n}. \qquad \square$$

2.7 Translation means

This section was written as much as possible in synchronisation with the previous section, to stress the analogies between the power mean map $(p, \mathbf{w}) \mapsto \mathrm{PM}_{p,\mathbf{w}}$ and the translation mean map $(p, \mathbf{w}) \mapsto \mathrm{TM}_{p,\mathbf{w}}$ (both are maps on $\overline{\mathbb{R}} \times \mathbb{R}^n_{>0}$, for all n). In particular, both maps are continuums of means, and they interpolate (or extrapolate) between the pythagorean means, and for fixed $\mathbf{w}$ the maps are increasing, and we have the inversion duality $p \mapsto -p$. We will also see clear differences between the power means and translation means in Theorem 2.7.4: the power asymptotic properties are different, and a translation mean is not scale-invariant unless it is a pythagorean mean (instead, we have the identities in part 7 of the theorem).

Whereas the power means are widely known, the translation means have not been defined previously (in this generality), as far as I know. We give a first application of basic properties of translation means in Theorem 7.5.4, about generalised Collatz dynamics.

2.7.1 About the definition of the translation means

Definition 2.7.1 (*)**.** Let $b \in \overline{\mathbb{R}}$, let $n \in \mathbb{N}$, let $\mathbf{w} \in \mathbb{R}^n_{>0}$. The *translation mean* $\mathrm{TM}_{b,\mathbf{w}}$ is the internal mean on $\mathbb{R}^n_{>0}$ given by

$$\mathrm{TM}_{b,\mathbf{w}} = \begin{cases} \mathrm{AM}_{\mathbf{w}} & \text{if } b = \infty \\ \mathrm{GM}_{\mathbf{w}}^{[\mathrm{tra}_{-b}]} & \text{if } b \in \mathbb{R}_{\geq 0} \\ \mathrm{GM}_{\mathbf{w}}^{[\mathrm{inv} \circ \mathrm{tra}_b]} & \text{if } b \in \mathbb{R}_{\leq 0} \\ \mathrm{HM}_{\mathbf{w}} & \text{if } b = -\infty, \end{cases}$$

where $\mathrm{GM}_{\mathbf{w}}$ denotes the geometric mean restricted to $\mathbb{R}^n_{>|b|}$, in case that $b \in \mathbb{R}$.

In the case that $\mathbf{w} \in \mathrm{diag}(\mathbb{R}^n_{>0})$, we simply write $\mathrm{TM}_{b,n}$ instead of $\mathrm{TM}_{b,\mathbf{w}}$.

***Remark* 2.7.2.**

1. We have overlapping definitions for $b = 0$, but because $\mathrm{GM}_{\mathbf{w}}^{[\mathrm{inv}]} = \mathrm{GM}_{\mathbf{w}}$, it is well-defined.

2. We spell out more explicitly what the definition means, in the case that $b \in \mathbb{R}$.

   Case 1: $b \in \mathbb{R}_{\geq 0}$. Then $\mathrm{GM}_{\mathbf{w}}$ is a mean on $\mathbb{R}^n_{>b}$, and we have a monotonic, continuous bijection $\mathrm{tra}_{-b} : \mathbb{R}_{>b} \to \mathbb{R}_{>0} : x \mapsto x - b$. Hence (by Theorem 2.1.3), $\mathrm{GM}_{\mathbf{w}}^{[\mathrm{tra}_{-b}]}$ is a mean on $\mathbb{R}^n_{>0}$. For $\mathbf{x} \in \mathbb{R}^n_{>0}$, it is given by

   $$\mathrm{TM}_{b,\mathbf{w}}(\mathbf{x}) = \mathrm{GM}_{\mathbf{w}}(\mathbf{x} + b) - b = \left(\prod (\mathbf{x} + b)^{\mathbf{w}} \right)^{1/\sum \mathbf{w}} - b. \tag{2.15}$$

   Case 2: $b \in \mathbb{R}_{\leq 0}$. Then $-b \in \mathbb{R}_{\leq 0}$, and

   $$(\mathrm{TM}_{-b,\mathbf{w}})^{[\mathrm{inv}]} = (\mathrm{GM}_{\mathbf{w}}^{[\mathrm{tra}_b]})^{[\mathrm{inv}]} = \mathrm{GM}_{\mathbf{w}}^{[\mathrm{inv} \circ \mathrm{tra}_b]} = \mathrm{TM}_{b,\mathbf{w}}.$$

   Hence, for $\mathbf{x} \in \mathbb{R}^n_{>0}$, we have

   $$\mathrm{TM}_{b,\mathbf{w}}(\mathbf{x}) = (\mathrm{GM}_{\mathbf{w}}(\mathbf{x}^{-1} - b) + b)^{-1} = \left(\left(\prod (\mathbf{x}^{-1} - b)^{\mathbf{w}} \right)^{1/\sum \mathbf{w}} + b \right)^{-1}. \tag{2.16}$$

3. The reason for the particular definition of the translation mean in the case that $b \in \{-\infty, \infty\}$, becomes clear from part 5 of Theorem 2.7.4: the definition is such that $(b, \mathbf{w}) \mapsto \mathrm{TM}_{b,\mathbf{w}}$ is a continuum of means.

   Thus, informally speaking, we can see the map $b \mapsto \mathrm{TM}_{b,\mathbf{w}}$ as a continuous interpolation between $\mathrm{HM}_{\mathbf{w}}, \mathrm{GM}_{\mathbf{w}}, \mathrm{AM}_{\mathbf{w}}$, as b increases from $-\infty$ to ∞. Moreover, this map is increasing in b, by part 6 of the theorem. To prove that "increasing" property, we use the following lemma.

2.7.2 Properties of the translation means

Lemma 2.7.3 (*). *Let $n \in \mathbb{N}$, let $\mathbf{w} \in \mathbb{R}^n_{>0}$, let $\mathbf{x} \in \mathbb{R}^n_{>0}$. For all $b \in \mathbb{R}_{>0}$, we have*

$$\frac{d}{db}\mathrm{GM}_{\mathbf{w}}(\mathbf{x}+b) = \frac{\mathrm{GM}_{\mathbf{w}}(\mathbf{x}+b)}{\mathrm{HM}_{\mathbf{w}}(\mathbf{x}+b)}.$$

In particular, if $\mathbf{x} \notin \mathrm{diag}(\mathbb{R}^n_{>0})$, then

$$\frac{d}{db}\mathrm{GM}_{\mathbf{w}}(\mathbf{x}+b) > 1, \tag{2.17}$$

$$\lim_{b\to\infty}\frac{d}{db}\mathrm{GM}_{\mathbf{w}}(\mathbf{x}+b) = 1. \tag{2.18}$$

Proof. Because $\mathrm{GM}_{\mathbf{w}} = \mathrm{AM}^{[\exp]}_{\mathbf{w}}$, we have $\mathrm{GM}_{\mathbf{w}}(\mathbf{x}+b) = \exp\left(\sum \mathbf{w}\langle\log\rangle(\mathbf{x}+b)\right)$. Hence, using the chain rule,

$$\frac{d}{db}\mathrm{GM}_{\mathbf{w}}(\mathbf{x}+b) = \mathrm{GM}_{\mathbf{w}}(\mathbf{x}+b)\sum\frac{\mathbf{w}}{\mathbf{x}+b} = \frac{\mathrm{GM}_{\mathbf{w}}(\mathbf{x}+b)}{\mathrm{HM}_{\mathbf{w}}(\mathbf{x}+b)}.$$

(2.17) follows by Corollary 2.6.8, and (2.18) follows from

$$\mathrm{GM}_{\mathbf{w}}(\mathbf{x}+b) \sim b \sim \mathrm{HM}_{\mathbf{w}}(\mathbf{x}+b) \quad \text{as } b\to\infty,$$

which is easily verified because $\mathrm{GM}_{\mathbf{w}}$ and $\mathrm{HM}_{\mathbf{w}}$ are scale-invariant. □

The next theorem, consisting of (unknown) properties of translation means, constitutes the core of §2.7.

Theorem 2.7.4 ((*) Translation means).
Let $n \in \mathbb{N}_{\geq 2}$, let $b \in \overline{\mathbb{R}}$, let $\mathbf{w} \in \mathbb{R}^n_{>0}$, let $\mathbf{x} \in \mathbb{R}^n_{>0}$.

1. *"Duality": $\mathrm{TM}^{[\mathrm{inv}]}_{b,\mathbf{w}} = \mathrm{TM}_{-b,\mathbf{w}}$.*

2. *"Mean": $\mathrm{TM}_{b,\mathbf{w}}$ is a strictly increasing, smooth internal mean on $\mathbb{R}^n_{>0}$. It is symmetric if $\mathbf{w} \in \mathrm{diag}(\mathbb{R}^n_{>0})$.*

3. *"Power-asymptotic": Suppose $\sum \mathbf{w} = 1$. Then $\mathrm{TM}_{b,\mathbf{w}}$ is*

    - *$(0,1)$-power-asymptotic if $b = \infty$;*
    - *$(0,\mathbf{w})$-power-asymptotic if $b \in \mathbb{R}_{>0}$;*
    - *$(\mathbf{w},\mathbf{w})$-power-asymptotic if $b = 0$;*
    - *$(\mathbf{w},0)$-power-asymptotic if $b \in \mathbb{R}_{<0}$;*
    - *$(1,0)$-power-asymptotic if $b = -\infty$.*

4. *"Compressing": Suppose $\sum \mathbf{w} = 1$. Let $\alpha, \beta \in \mathbb{R}_{>0}$ with $\alpha < \beta$, let $c := \frac{\mathrm{Min}(\mathbf{w})}{1-\mathrm{Min}(\mathbf{w})}$. Then $\mathrm{TM}_{b,\mathbf{w}}$ on the interval $[\alpha,\beta]$ is*

- $[c]$*-compressing if* $b = \infty$*;*
- $\left[\dfrac{\alpha+b}{\beta+b}c\right]$*-compressing if* $0 \le b < \infty$*;*
- $\left[\dfrac{\alpha(b\alpha-1)}{\beta(b\beta-1)}c\right]$*-compressing if* $-\infty < b \le 0$*;*
- $\left[\dfrac{\alpha^2}{\beta^2}c\right]$*-compressing if* $b = -\infty$*.*

5. *"Continuous and smooth": The function*

$$\psi : \quad \overline{\mathbb{R}} \times \mathbb{R}^n_{>0} \times \mathbb{R}^n_{>0} \to \mathbb{R}_{>0} : \quad (b, \mathbf{w}, \mathbf{x}) \mapsto \mathrm{TM}_{b,\mathbf{w}}(\mathbf{x})$$

is continuous, that is, the map $\overline{\mathbb{R}} \times \mathbb{R}^n_{>0} \to \mathbb{M}_{\mathbb{R}^n_{>0}} \ : \ (b, \mathbf{w}) \mapsto \mathrm{TM}_{b,\mathbf{w}}$ *is a continuum of means.*

The restriction of ψ *to* $(\mathbb{R}_{\neq 0}) \times \mathbb{R}^n_{>0} \times \mathbb{R}^n_{>0}$ *is smooth.*

6. *"Increasing in b": The function*

$$\varphi : \quad \overline{\mathbb{R}} \to \mathbb{R}_{>0} : \quad b \mapsto \mathrm{TM}_{b,\mathbf{w}}(\mathbf{x})$$

is continuous and increasing, and it is strictly increasing if $\mathbf{x} \notin \mathrm{diag}(\mathbb{R}^n_{>0})$*.*

7. *"Scaling": For all* $\lambda \in \mathbb{R}_{>0}$*, we have*

$$\begin{aligned} \lambda\, \mathrm{TM}_{b,\mathbf{w}}(\mathbf{x}) &= \mathrm{TM}_{\lambda b,\mathbf{w}}(\lambda \mathbf{x}) && \text{if } b \in \overline{\mathbb{R}}_{\ge 0}, \\ \lambda\, \mathrm{TM}_{b,\mathbf{w}}(\mathbf{x}) &= \mathrm{TM}_{\lambda^{-1} b,\mathbf{w}}(\lambda \mathbf{x}) && \text{if } b \in \overline{\mathbb{R}}_{\le 0}. \end{aligned}$$

In particular, if $b \notin \{0, \infty, -\infty\}$*, then* $\mathrm{TM}_{b,\mathbf{w}}$ *is not scale-invariant.* ⅃

Proof.

1. ("Duality".) We noted this already in Remark 2.7.2 for $b \ge 0$, and from that it follows for $b \le 0$ because [inv] is an involution.

2 and 3. ("Mean" and "Power-asymptotic".) For $b \in \{0, \pm\infty\}$, the statements are contained in Proposition 1.2.6. For $b \in \mathbb{R}_{>0}$, the statements, except the "power-asymptotic" statements, follow from Theorem 2.1.3, applied with $M := \mathrm{GM}_{\mathbf{w}}$ and $f := \mathrm{tra}_{-b} : \mathbb{R}_{>b} \to \mathbb{R}_{>0}$. For $b \in \mathbb{R}_{<0}$, the statements follow from those for $b \in \mathbb{R}_{>0}$ by Corollary 2.2.4, using that $\mathrm{TM}_{b,\mathbf{w}} = \mathrm{TM}^{[\mathrm{inv}]}_{-b,\mathbf{w}}$.

It remains to prove the "power-asymptotic" statements in the case that $b \in \mathbb{R}_{>0}$. Let $i \in \mathbb{N}_{\le n}$, and write

$$\mathrm{TM}_{b,\mathbf{w}}(\mathbf{x}) = \mathrm{GM}_{\mathbf{w}}(\mathbf{x} + b) - b = (x_i + b)^{w_i} C - b,$$

where $C > 0$ does not depend on x_i. Note that as $x_i \to 0$, then $\mathrm{GM}_{\mathbf{w}}(\mathbf{x} + b)$ approaches a constant that is strictly larger than b. Hence $\mathrm{TM}_{b,\mathbf{w}}(\mathbf{x}) \sim c x_i^0$ as $x_i \to 0$, for some $c > 0$. Further, it is clear that $\mathrm{TM}_{b,\mathbf{w}}(\mathbf{x}) \sim C x_i^{w_i}$ as $x_i \to \infty$. We conclude that $\mathrm{TM}_{b,\mathbf{w}}$ is $(0, \mathbf{w})$-power-asymptotic.

4. ("Compressing".) For $b \in \{0, \pm\infty\}$, the statements are contained in Proposition 1.2.6 and Corollary 2.2.3. For $b \in \mathbb{R}_{\geq 0}$, we have $\mathrm{TM}_{b,\mathbf{w}}^{[\log \circ \mathrm{tra}_b]} = \mathrm{AM}_{\mathbf{w}}$, thus the statement follows by applying Corollary 2.2.2 with the increasing and concave function $g := \log \circ \mathrm{tra}_b$. For $b \in \mathbb{R}_{\leq 0}$, we have $\mathrm{TM}_{b,\mathbf{w}}^{[\log \circ \mathrm{tra}_{-b} \circ \mathrm{inv}]} = \mathrm{AM}_{\mathbf{w}}$, hence we can apply Corollary 2.2.2 with $g := \log \circ \mathrm{tra}_{-b} \circ \mathrm{inv}$, that is, with the bijection

$$g: \quad \mathbb{R}_{>0} \to \mathbb{R}_{>\log(-b)}: \quad g(x) = \log(x^{-1} - b).$$

We have $g'(x) = \big(x(bx-1)\big)^{-1}$, hence g is decreasing and convex. Thus, the statement follows from Corollary 2.2.2, by noting that $g'(\beta)/g'(\alpha) = (\alpha(b\alpha - 1))/(\beta(b\beta - 1))$.

5. ("Continuous and smooth".) For any subset I of $\overline{\mathbb{R}}$, we write $D_I := I \times \mathbb{R}^n_{>0} \times \mathbb{R}^n_{>0}$, and we denote $D_{\overline{\mathbb{R}}}$ simply by D. On $D_{(-\infty,0)}$ and on $D_{(0,\infty)}$, the function ψ is given by (2.16) and (2.15) respectively, so ψ is clearly smooth on $D_{\mathbb{R}_{\neq 0}}$. Moreover, ψ is clearly continuous on $D_{(-\infty,0]}$ and also on $D_{[0,\infty)}$. Because the intersection of $D_{(-\infty,0]}$ and $D_{[0,\infty)}$ is a closed subset of D, it follows that ψ is continuous on their union $D_{\mathbb{R}}$.

Hence, to show that ψ is continuous on D, it suffices to show that

$$\lim_{(b,\mathbf{w},\mathbf{x}) \to (b_0,\mathbf{w}_0,\mathbf{x}_0)} \mathrm{TM}_{b,\mathbf{w}}(\mathbf{x}) = \mathrm{TM}_{b_0,\mathbf{w}_0}(\mathbf{x}_0) \tag{2.19}$$

for all $(b_0, \mathbf{w}_0, \mathbf{x}_0) \in \{-\infty, \infty\} \times \mathbb{R}^n_{>0} \times \mathbb{R}^n_{>0}$. It suffices to show it for $b_0 = \infty$, because then it follows by continuity of inv that $\mathrm{TM}_{b,\mathbf{w}}^{[\mathrm{inv}]}(\mathbf{x}) \to \mathrm{TM}_{\infty,\mathbf{w}_0}^{[\mathrm{inv}]}(\mathbf{x}_0)$ as $(b, \mathbf{w}, \mathbf{x}) \to (b_0, \mathbf{w}_0, \mathbf{x}_0)$, that is, $\mathrm{TM}_{-b,\mathbf{w}}(\mathbf{x}) \to \mathrm{TM}_{-\infty,\mathbf{w}_0}(\mathbf{x}_0)$.

For $b \in \mathbb{R}_{>0}$, we have

$$\begin{aligned} \mathrm{TM}_{b,\mathbf{w}}(\mathbf{x}) = \mathrm{GM}_{\mathbf{w}}(\mathbf{x} + b) - b = b\big(\mathrm{GM}_{\mathbf{w}}(\tfrac{\mathbf{x}}{b} + 1) - 1\big); \\ \mathrm{GM}_{\mathbf{w}}(\tfrac{\mathbf{x}}{b} + 1) = \exp\big(\mathrm{AM}_{\mathbf{w}}\langle\log\rangle(\tfrac{\mathbf{x}}{b} + 1)\big). \end{aligned} \tag{2.20}$$

We analyse this chain of expressions from inside out, in order to obtain by backsubstitution an expression for $\mathrm{TM}_{b,\mathbf{w}}(\mathbf{x})$ that reveals its behaviour as $b \to \infty$.

Because $\log(x+1) = x + O(x^2)$ for $x \in \mathbb{R}_{>0}$, we have

$$\mathrm{AM}_{\mathbf{w}}\langle\log\rangle(\tfrac{\mathbf{x}}{b} + 1) = \mathrm{AM}_{\mathbf{w}}(\tfrac{\mathbf{x}}{b}) + O(\tfrac{\mathrm{Max}(\mathbf{x})^2}{b^2}) = b^{-1} f, \quad \text{where } f := \mathrm{AM}_{\mathbf{w}}(\mathbf{x}) + O(\tfrac{\mathrm{Max}(\mathbf{x})^2}{b}).$$

Because $\exp(x) = 1 + x + O(x^2)$ as $x \to 0$, and because $b^{-1} f \to 0$ as $b \to \infty$, we have

$$\mathrm{GM}_{\mathbf{w}}(\tfrac{\mathbf{x}}{b} + 1) = \exp(b^{-1} f) = 1 + b^{-1} f + O(b^{-2} f^2) \quad \text{as } b \to \infty.$$

Therefore,

$$\mathrm{TM}_{b,\mathbf{w}}(\mathbf{x}) = b\big(\mathrm{GM}_{\mathbf{w}}(\tfrac{\mathbf{x}}{b} + 1) - 1\big) = f + O(b^{-1} f^2) \quad \text{as } b \to \infty.$$

Because $f \to \mathrm{AM}_{\mathbf{w}_0}(\mathbf{x}_0)$ and $b^{-1} f^2 \to 0$ as $(b, \mathbf{w}, \mathbf{x}) \to (\infty, \mathbf{w}_0, \mathbf{x}_0)$, it follows that $\mathrm{TM}_{b,\mathbf{w}}(\mathbf{x}) \to \mathrm{AM}_{\mathbf{w}_0}(\mathbf{x}_0)$, as desired.

6. ("Increasing in b".) Continuity follows from part 3, and φ is constant if $\mathbf{x} \in \operatorname{diag}(\mathbb{R}^n_{>0})$. Suppose that $\mathbf{x} \notin \operatorname{diag}(\mathbb{R}^n_{>0})$. For $b \in \mathbb{R}_{>0}$, we have $\varphi(b) = \mathrm{GM}_{\mathbf{w}}(\mathbf{x}+b) - b$, so it follows from (2.17) that φ is strictly increasing on $\mathbb{R}_{>0}$.

Let $b_1, b_2 \in \mathbb{R}$ with $b_1 < b_2 < 0$. Then $0 < -b_2 < -b_1$, hence

$$\mathrm{TM}_{-b_2,\mathbf{w}}(\mathbf{x}) = \varphi(-b_2) < \varphi(-b_1) = \mathrm{TM}_{-b_1,\mathbf{w}}(\mathbf{x}).$$

Because inv is strictly decreasing, it follows by Proposition 1.3.7.6 (applied with $A := \mathbb{R}^n_{>0} \setminus \operatorname{diag}(\mathbb{R}^n_{>0})$) that

$$\varphi(b_2) = \mathrm{TM}^{[\mathrm{inv}]}_{-b_2,\mathbf{w}}(\mathbf{x}) > \mathrm{TM}^{[\mathrm{inv}]}_{-b_1,\mathbf{w}}(\mathbf{x}) = \varphi(b_1).$$

We conclude that φ is strictly increasing on $\mathbb{R}_{\neq 0}$. By continuity, it follows that φ is strictly increasing on $\overline{\mathbb{R}}$.

7. ("Scaling".) For $b \in \{0, \infty, -\infty\}$, the equalities follow by scale-invariance of $\mathrm{TM}_{b,\mathbf{w}}$. For $b \in \mathbb{R}_{>0}$, the equality follows from scale-invariance of $\mathrm{GM}_{\mathbf{w}}$; namely,

$$\lambda\,\mathrm{TM}_{b,\mathbf{w}}(\mathbf{x}) = \lambda\big(\mathrm{GM}_{\mathbf{w}}(\mathbf{x}+b) - b\big) = \mathrm{GM}_{\mathbf{w}}(\lambda\mathbf{x} + \lambda b) - \lambda b = \mathrm{TM}_{\lambda b,\mathbf{w}}(\lambda\mathbf{x}).$$

For $b \in \mathbb{R}_{<0}$, the equality follows from the duality $\mathrm{TM}_{b,\mathbf{w}} = \mathrm{TM}^{[\mathrm{inv}]}_{-b,\mathbf{w}}$; namely,

$$\lambda\,\mathrm{TM}_{b,\mathbf{w}}(\mathbf{x}) = \big(\lambda^{-1}\mathrm{TM}_{-b,\mathbf{w}}(\mathbf{x}^{-1})\big)^{-1} = \big(\mathrm{TM}_{-\lambda^{-1}b,\mathbf{w}}((\lambda\mathbf{x})^{-1})\big)^{-1} = \mathrm{TM}_{\lambda^{-1}b,\mathbf{w}}(\lambda\mathbf{x}).$$

The statement about scale-invariance follows from part 4; for instance, if $\lambda > 1$ and $b \in \mathbb{R}_{>0}$ and $\mathbf{x} \notin \operatorname{diag}(\mathbb{R}^n_{>0})$, then $\lambda\,\mathrm{TM}_{b,\mathbf{w}}(\mathbf{x}) = \mathrm{TM}_{\lambda b,\mathbf{w}}(\lambda\mathbf{x}) > \mathrm{TM}_{b,\mathbf{w}}(\lambda\mathbf{x})$. ☐

Example 2.7.5. Let $x_1, \ldots, x_n, w_1, \ldots, w_n \in \mathbb{R}_{>0}$ such that $\sum w_i = 1$. Part 3 of Theorem 2.7.4 tells us that

$$\lim_{b\to\infty} (x_1+b)^{w_1} \cdots (x_n+b)^{w_n} - b = w_1x_1 + \cdots + w_nx_n.$$

Corollary 2.7.6. *Let $b, c \in \overline{\mathbb{R}}$ with $b \leq 0 \leq c$, let $n \in \mathbb{N}$, let $\mathbf{w} \in \mathbb{R}^n_{>0}$. Then*

$$\mathrm{HM}_{\mathbf{w}} \leq \mathrm{TM}_{b,\mathbf{w}} \leq \mathrm{GM}_{\mathbf{w}} \leq \mathrm{TM}_{c,\mathbf{w}} \leq \mathrm{AM}_{\mathbf{w}}.$$

Proof. This follows directly by Theorem 2.7.4.6 and by the definitions of $\mathrm{TM}_{-\infty,\mathbf{w}}$ and $\mathrm{TM}_{0,\mathbf{w}}$ and $\mathrm{TM}_{\infty,\mathbf{w}}$. ☐

***Remark* 2.7.7.** In particular, Corollary 2.7.6 yields a new proof of the classical inequalities between the Pythagorean means (Corollary 2.6.8).

Chapter 3

Gauss composition of means

Gauss composition of means is, like conjugation of means that we studied in the previous chapter, a way to make new means out of old means. Its definition is however more involved, its properties are more obscure, and the results about it appear to be "deeper". This is already the case for one of the simplest non-trivial examples: the arithmetic-geometric mean (AGM), which was first studied by C.F. Gauss.

3.1 Reconstructing properties of composite means

Before we state and proof our basic theorem (3.2.4) about Gauss composition, we need a more general theorem (3.1.1) about "normal" composition of means. Gauss composition is obtained as the limit of "normal" composition.

Theorem 3.1.1 is not only fundamental for Gauss composition, but also for families of means, that we study in Chapter 4. However, the most involved part of the proof is in the "compressing" property, which we nowhere use in its full strength (we only use that the composite is compressing, not the explicit parameters).

Theorem 3.1.1 ((†) Composition of means)**.** *Let $A \subseteq \mathbb{R}$, let $m, n \in \mathbb{N}$.*
Let M be a mean on A^n.
For each $k \in \mathbb{N}_{\leq n}$, let $n_k \in \mathbb{N}$, let M_k be an internal mean on A^{n_k}, and let $f_k : \mathbb{N}_{\leq n_k} \to \mathbb{N}_{\leq m}$ be a function. Let $\mathcal{M}$ be the map

$$\mathcal{M}: \quad A^m \to \mathbb{R}: \quad \mathbf{x} \mapsto M\big(M_1(\mathbf{x}_{\circ f_1}), \ldots, M_n(\mathbf{x}_{\circ f_n})\big).$$

1. *$\mathcal{M}$ is a* **mean** *on A^n.*

2. *If all of the means $M, M_1, \ldots, M_n$ are* **increasing**, **continuous**, **smooth**, *or* **scale-invariant**, *respectively, then $\mathcal{M}$ is so as well.*
   If $M, M_1, \ldots, M_n$ are F-rational for some subfield F of $\mathbb{R}$, then $\mathcal{M}$ is F-**rational**.

3. *Let $B \subseteq \mathbb{R}$, let $g : A \to B$ be a bijection. Suppose M is an internal mean. Then for all $\mathbf{x} \in B^n$,*

$$\mathcal{M}^{[g]}(\mathbf{x}) = M^{[g]}\big(M_1^{[g]}(\mathbf{x}_{\circ f_1}), \ldots, M_n^{[g]}(\mathbf{x}_{\circ f_n})\big).$$

4. *Suppose that* $M_1, \dots, M_n$ *are symmetric, and that* $f_k \in S_n$ *for all* $k \in \mathbb{N}_{\le n}$. *Then* $\mathcal{M}$ *is* **symmetric**.

5. *Suppose that* M *is strictly increasing, and that there is a subset* $K \subseteq \mathbb{N}_{\le n}$ *such that* $\bigcup_{k\in K} f_k(\mathbb{N}_{\le n_k}) = \mathbb{N}_{\le n}$, *and such that* M_k *is strictly increasing for each* $k \in K$. *Then* $\mathcal{M}$ *is* **strictly increasing**.

6. *(*) Let* $r, R, s, S \in \mathbb{R}$ *such that* $0 \le r \le R \le 1$ *and* $0 \le s \le S \le 1$. *Suppose that* M *is* $[r, R]$*-alt-compressing, and that there is a subset* $K \subseteq \mathbb{N}_{\le n}$ *such that* $\bigcup_{k\in K} f_k(\mathbb{N}_{\le n_k}) = \mathbb{N}_{\le n}$, *and such that* M_k *is* $[s, S]$*-alt-compressing for each* $k \in K$. *Then* $\mathcal{M}$ *is* $[rs, 1-(1-R)(1-S)]$*-alt-***compressing**.

7. *Suppose that* M *is scale-invariant and continuous, that* M_k *is* $(\cdot, a)$*-power-asymptotic, and that* $f_k \in S_n$, *for all* $k \in \mathbb{N}_{\le n}$. *Then* $\mathcal{M}$ *is* $(\cdot, a)$-**power-asymptotic**.

   Analogously with '·' and 'a' interchanged.

8. *Suppose that all* f_k *are injective, and that their images form are a partition of* $\mathbb{N}_{\le n}$; *so in particular,* $\sum n_k = n$. *Suppose that* M *is* $(\cdot, b)$*-power-asymptotic, and* $M_1, \dots, M_n$ *are are* $(\cdot, a)$*-power-asymptotic. If* $a > 0$ *or* M *is continuous, then* $\mathcal{M}$ *is* $(\cdot, ab)$-**power-asymptotic**.

   Analogously with '·' and 'number' interchanged. ♩

The proof of the theorem is below the next example, which serves to make the statement of the theorem more concrete.

Example 3.1.2. Let $p, \beta \in \overline{\mathbb{R}}$, and let $\mathcal{M} : \mathbb{R}^3_{>0} \to \mathbb{R}_{>0}$ be given by

$$\mathcal{M}(x, y, z) = \mathrm{GM}_2\big(\mathrm{PM}_{p,4}(x, x, x, y), \mathrm{TM}_{\beta,2}(y, z)\big).$$

In the notation of Theorem 3.1.1, we have $n = 2$ and $m = 3$, and $f_1 : \mathbb{N}_{\le 4} \to \mathbb{N}_{\le 3}$ maps 1, 2 and 3 to 1 and maps 4 to 2, while $f_2 : \mathbb{N}_{\le 2} \to \mathbb{N}_{\le 3}$ maps 1 to 2 and 2 to 3.

Theorem 3.1.1 tells us that $\mathcal{M}$ is an increasing, continuous and scale-invariant mean. Moreover, it tells us that if $p \notin \{\pm\infty\}$, then $\mathcal{M}$ is strictly increasing and smooth.

Regarding the compressing property, suppose for concreteness that $p \ge 1$ and that $\beta \ge 0$. Let $a, b \in \mathbb{R}_{>0}$ with $a < b$. From Theorem 2.6.6 and Theorem 2.7.4 and Lemma 1.1.28, it follows that on the interval $[a, b]$, the mean $\mathrm{PM}_{p,4}$ is $[a^{p-1}/(a^{p-1} + 3b^{p-1})]$-alt-compressing, and $\mathrm{TM}_{\beta,2}$ is $[(a+\beta)/(a+b+2\beta)]$-alt-compressing, and GM_2 is $[a/(a+b)]$-alt-compressing. Hence, writing

$$s := \mathrm{Min}\big(\frac{a^{p-1}}{a^{p-1} + 3b^{p-1}}, \frac{a+\beta}{a+b+2\beta}\big),$$

it follows from Theorem 3.1.1 that $\mathcal{M}$ is $[sa/(a+b)]$-alt-compressing.

Theorem 3.1.1 tells us nothing about asymptotic properties of $\mathcal{M}$. Moreover, it is straightforwardly verified that $\mathcal{M}$ is *not* symmetric, for instance by comparing $\mathcal{M}(1, y, y)$ with $\mathcal{M}(y, 1, y)$. ♪

Proof of Theorem 3.1.1.

1 & 2 & 3. Clear.

4. ("symmetric".) Clear, because $M_k(\mathbf{x}_{\circ f_k}) = M_k(\mathbf{x})$ for all k and all $\mathbf{x}$.

5. ("strictly increasing".) Let $\mathbf{x} = (x_i)_i \in A^n$ and $\mathbf{y} = (y_i)_i \in A^n$ such that $\mathbf{x} < \mathbf{y}$, say $x_j < y_j$. There exists $k \in K$ such that j is in the image of f_k. Because M_k is strictly increasing, we have $M_k(\mathbf{x}_{\circ f_k}) < M_k(\mathbf{y}_{\circ f_k})$. Because all of the M_i are increasing, and M is strictly increasing, it follows that $\mathcal{M}(\mathbf{x}) < \mathcal{M}(\mathbf{y})$.

6. ("alt-compressing".) Let $\mathbf{x} \in A^n \setminus \operatorname{diag}(A^n)$. We write

$$\mathbf{y} := \big(M_1(\mathbf{x}_{\circ f_1}), \dots, M_n(\mathbf{x}_{\circ f_n})\big) \in A^n,$$

so we have $\mathcal{M}(\mathbf{x}) = M(\mathbf{y})$. Further, we write (for each $k \in \mathbb{N}_{\leq n}$)

$$\begin{aligned} c_{\mathbf{x}} &:= \operatorname{Min}(\mathbf{x}) & d_{\mathbf{x}} &:= \operatorname{Max}(\mathbf{x}) \\ c_{\mathbf{y}} &:= \operatorname{Min}(\mathbf{y}) & d_{\mathbf{y}} &:= \operatorname{Max}(\mathbf{y}) \qquad Q := \frac{d_{\mathbf{x}} - \mathcal{M}(\mathbf{x})}{d_{\mathbf{x}} - c_{\mathbf{x}}}. \\ c_k &:= \operatorname{Min}(\mathbf{x}_{\circ f_k}) & d_k &:= \operatorname{Max}(\mathbf{x}_{\circ f_k}) \end{aligned}$$

Our aim is to prove that $Q \geq rs$. Supposing we proved that, we complete the proof by duality. Namely, we have then that $\mathcal{M}$ is $[rs, 1]$-alt-compressing. By part 3, we have

$$\mathcal{M}^{[\text{ainv}]}(\mathbf{x}) = M^{[\text{ainv}]}\big(M_1^{[\text{ainv}]}(\mathbf{x}_{\circ f_1}), \dots, M_n^{[\text{ainv}]}(\mathbf{x}_{\circ f_n})\big).$$

Lemma 2.1.1 tells us that $M_k^{[\text{ainv}]}$ is $[1-S, 1-s]$-alt-compressing for each $k \in K$, and that $M^{[\text{ainv}]}$ is $[1-R, 1-r]$-alt-compressing. Hence, by what we are supposing to have proved, it follows that $\mathcal{M}^{[\text{ainv}]}$ is $[(1-R)(1-S), 1]$-alt-compressing. Applying Lemma 2.1.1 again, we get that $\mathcal{M}$ is $[0, 1-(1-R)(1-S)]$-alt-compressing. Thus, we get the desired result that $\mathcal{M}$ is $[rs, 1-(1-R)(1-S)]$-alt-compressing.

Thus, it remains to show that $Q \geq rs$.

Because $\bigcup_{k \in K} f_k(\mathbb{N}_{\leq n_k}) = \mathbb{N}_{\leq n}$, there exists $k \in K$ such that the number $c_{\mathbf{x}}$ is an entry of the tuple $\mathbf{x}_{\circ f_k}$. So we have

$$c_{\mathbf{x}} = c_k \leq M_k(\mathbf{x}_{\circ f_k}) \leq d_k \leq d_{\mathbf{x}}. \tag{3.1}$$

It is easily seen that $M_k(\mathbf{x}_{\circ f_i}) < d_{\mathbf{x}}$, because otherwise it would follow, since M_k is strict, that each inequality in (3.1) is an equality, violating that $c_{\mathbf{x}} < d_{\mathbf{x}}$. So we may write

$$Q = Q_2 Q_1, \qquad \text{where} \quad Q_2 = \frac{d_{\mathbf{x}} - M(\mathbf{y})}{d_{\mathbf{x}} - M_k(\mathbf{x}_{\circ f_k})}, \quad \text{and} \quad Q_1 = \frac{d_{\mathbf{x}} - M_k(\mathbf{x}_{\circ f_k})}{d_{\mathbf{x}} - c_{\mathbf{x}}}. \tag{3.2}$$

We show that $Q_1 \geq s$ and $Q_2 \geq r$, which implies the desired result that $Q \geq rs$.

If $c_k = d_k$, then by (3.1) we have $Q_1 = 1 \geq s$. If $c_k < d_k$, we use (3.1) to find that[1]

$$Q_1 \geq \frac{d_k - M_k(\mathbf{x}_{\circ f_k})}{d_k - c_k} \geq s.$$

[1] In this derivation it is essential that $c_{\mathbf{x}} = c_k$. If there is no such k, then it is easily seen that $\mathcal{M}$ is not even a strict mean, provided that $|A| \geq 2$.

If $M_k(\mathbf{x}_{\circ f_k}) \geq M(\mathbf{y})$, then $Q_2 \geq 1 \geq r$. If $M_k(\mathbf{x}_{\circ f_k}) < M(\mathbf{y})$, we use the inequalities

$$c_{\mathbf{y}} \leq M_k(\mathbf{x}_{\circ f_k}) < M(\mathbf{y}) \leq d_{\mathbf{y}} \leq d_{\mathbf{x}}$$

to find that

$$Q_2 \geq \frac{d_{\mathbf{y}} - M(\mathbf{y})}{d_{\mathbf{y}} - M_k(\mathbf{x}_{\circ f_k})} \geq \frac{d_{\mathbf{y}} - M(\mathbf{y})}{d_{\mathbf{y}} - c_{\mathbf{y}}} \geq r.$$

7. ("power-asymptotic 1".) Let $i \in \mathbb{N}_{\leq n}$. For each k, there exists $c_k > 0$ such that $M_k(\mathbf{x}_{\circ f_k})/x_i^a \to c_k$ as $x_i \to \infty$ as the other $x_{i'}$ are fixed. Hence, by scale-invariance and continuity of M, we have

$$\frac{\mathcal{M}(\mathbf{x})}{x_i^a} = M\Big(\frac{M_1(\mathbf{x}_{\circ f_1})}{x_i^a}, \ldots, \frac{M_n(\mathbf{x}_{\circ f_n})}{x_i^a}\Big) \to M(c_1, \ldots, c_n) \quad \text{as } x_i \to \infty.$$

Thus, $\mathcal{M}$ is $(\cdot, a)$-power-asymptotic. The analogous statement follows by the same proof, with $x_i \to \infty$ replaced by $x_i \to 0$.

8. ("power-asymptotic 2".) Let $i \in \mathbb{N}_{\leq n}$ and $\mathbf{x} \in A^n$. For $t \in A$, we write as a shorthand $\mathbf{x}' := \mathbf{x}_{[i,t]}$, and

$$\begin{aligned} \mathbf{y} &:= (y_1, \ldots, y_n) := \big(M_1(\mathbf{x}_{\circ f_1}), \ldots, M_n(\mathbf{x}_{\circ f_n})\big), \\ \mathbf{y}' &:= (y_1', \ldots, y_n') := \big(M_1(\mathbf{x}'_{\circ f_1}), \ldots, M_n(\mathbf{x}'_{\circ f_n})\big). \end{aligned}$$

Let k be the unique index such that i is in the image of f_k. Because f_k is injective, the variable t occurs only once in $\mathbf{x}_{\circ f_k}$. Because M_k is $(\cdot, a)$-power-asymptotic, it follows that there exists a function $t \mapsto c_t$ such that $y_k' = c_t t^a$, and $c_t \to c$ as $t \to \infty$ for some constant $c > 0$. Moreover, $y_j' = y_j$ for all $j \neq k$. Hence, we have $\mathbf{y}' = \mathbf{y}_{[k, c_t t^a]}$.

Case 1: $a > 0$. Because M is $(\cdot, b)$-power-asymptotic, there exists $d > 0$ such that

$$\mathcal{M}(\mathbf{x}') = M(\mathbf{y}') = M(\mathbf{y}_{[k, c_t t^a]}) = (d + o(1))(c_t t^a)^b \sim d c^b t^{ab} \quad \text{as } t \to \infty,$$

using that $c_t t^a \to \infty$ as $t \to \infty$. Hence, $\mathcal{M}$ is $(\cdot, ab)$-power-asymptotic, as desired.

Case 2: $a \leq 0$ and M is continuous. We know by Remark 1.1.14.3 that $a \geq 0$, so $a = 0$. Hence $\mathbf{y}'$ approaches a constant vector as $t \to \infty$, hence $\mathcal{M}(\mathbf{x}') = M(\mathbf{y}')$ approaches a constant as $t \to \infty$, because M is continuous. Hence $\mathcal{M}$ is $(\cdot, 0)$-power-asymptotic, as desired.

The analogous statement follows analogously, with $t \to \infty$ replaced by $t \to 0$. $\square$

3.2 Constructing Gauss composition

Now that we proved Theorem 3.1.1 about "normal" composition of means, we only need a lemma about pointwise limits of means before we can study Gauss composition, which is the pointwise limit of normal composition.

3.2.1 Preliminary work

Lemma 3.2.1 (Limits of means)**.** *Let $A \subseteq \mathbb{R}$, let $n \in \mathbb{N}$, let M_k be a mean on A^n, for each $k \in \mathbb{N}$. Suppose that the pointwise limit $M := \lim_{k\to\infty} M_k$ exists on A^n.*

1. *M is a mean.*
2. *If each M_k is increasing, scale-invariant, or symmetric, respectively, then M is so as well.*
3. *Let $r, R \in \mathbb{R}$ such that $0 \le r \le R \le 1$. If each M_k is $[r,R]$-alt-compressing, then M is $[r,R]$-alt-compressing.*
4. *Suppose that A is an interval. Let $B \subseteq \mathbb{R}$, let $g : A \to B$ be a continuous monotonic bijection. Then $M_k^{[g]}$ is an internal mean on B^n for all k, and the pointwise limit $\lim_{k\to\infty} M_k^{[g]}$ exists as an internal mean on B^n, and is equal to $M^{[g]}$. That is,*

$$\lim_{k\to\infty} M_k^{[g]} = (\lim_{k\to\infty} M_k)^{[g]} \quad : \quad B^n \to B. \qquad ♩$$

Proof. **1 & 2 & 3.** Clear.

4. Because A is an interval, M and M_k are internal means on A^n. Lemma 1.3.10 tells us that $M^{[g]}$ and $M_k^{[g]}$ are internal means on B^n. Proposition 1.3.7.7 tells us that $\lim_{k\to\infty} M_k^{[g]} = M^{[g]}$. □

Definition 3.2.2 (Multidimensional mean transformations)**.** Let $A \subseteq \mathbb{R}$, let $n \in \mathbb{N}$.

1. We write $\mathbb{M}_{\mathrm{int}(A^n)}$ for the set of internal means on A^n, and $\mathbb{M}_{\mathrm{comp}(A^n)}$ for the set of *compressing* internal means on A^n.
2. Let $\mathbf{M} = (M_1, \ldots, M_n) \in \mathbb{M}^n_{\mathrm{int}(A^n)}$
   By "*the means of* $\mathbf{M}$" we mean the means $M_1, \ldots, M_n$ on A^n.
   We regard $\mathbf{M}$ as a *transformation* of A^n, namely

$$\mathbf{M} : \quad A^n \to A^n : \quad \mathbf{x} \mapsto \big(M_1(\mathbf{x}), \ldots, M_n(\mathbf{x})\big). \qquad 𝄽$$

Example 3.2.3. The transformation $(\mathrm{AM}_2, \mathrm{HM}_2) \in \mathbb{M}^2_{\mathrm{comp}(\mathbb{R}^2_{>0})}$ maps (a,b) to $(\frac{a+b}{2}, \frac{2ab}{a+b})$.

The transformation $(\mathrm{AM}_{(1,2,3)}, \mathrm{AM}_{(4,5,6)}, \mathrm{AM}_{(7,8,9)}) \in \mathbb{M}^3_{\mathrm{comp}(\mathbb{R}^3)}$ maps (a,b,c) to $(\frac{a+2b+3c}{6}, \frac{4a+5b+6c}{15}, \frac{7a+8b+9c}{24})$. ♪

3.2.2 Statement of the basic theorem of Gauss composition

Theorem 3.2.4 (Gauss composition of compressing means)**.** *Let $A \subseteq \mathbb{R}$, let $n \in \mathbb{N}$, let $\mathbf{M} = (M_1, \ldots, M_n) \in \mathbb{M}^n_{\mathrm{comp}(A^n)}$. For all $m \in \mathbb{N}_0$ and $k \in \mathbb{N}_{\le n}$, let $\mathcal{M}_{k,m}$ be the map*

$$\mathcal{M}_{k,m} := M_k \circ \mathbf{M}^{\circ m} : \quad A^n \to A.$$

1. *Let $a, b \in A$ with $a < b$. There exist $r, R \in \mathbb{R}$ with $0 < r \leq R \leq 1$ such that each mean M_k is $[r, R]$-alt-compressing on $A \cap [a, b]$.*

2. *(*) Let a, b, r, R be as in part 1. Let $\mathbf{x} \in (A \cap [a, b])^n$. We write as a shorthand $\mathbf{x}_m := \mathbf{M}^{\circ m}(\mathbf{x})$ for $m \geq 0$. For all m and k, we have*

$$\mathrm{Max}(\mathbf{x}_m) - \mathrm{Min}(\mathbf{x}_m) \leq (R - r)^m (b - a), \qquad \textit{and} \tag{3.3}$$
$$\mathrm{Min}(\mathbf{x}_m) \leq \mathrm{Min}(\mathbf{x}_{m+1}) \leq \mathcal{M}_{k,m}(\mathbf{x}) \leq \mathrm{Max}(\mathbf{x}_{m+1}) \leq \mathrm{Max}(\mathbf{x}_m). \tag{3.4}$$

   *In particular, for all k and m, the map $\mathcal{M}_{k,m}$ is an **internal mean** on A^n.*

3. *For all $\mathbf{x} \in A^n$, there exists a unique number $\mathcal{M}(\mathbf{x}) \in \mathbb{R}$ such that $\lim_{m \to \infty} \mathbf{M}^{\circ m}(\mathbf{x}) = \mathrm{diag}_n(\mathcal{M}(\mathbf{x}))$.*

   *The map $\mathcal{M} : A^n \to \mathbb{R}$ is a **mean**. It satisfies $\mathcal{M} \circ \mathbf{M}^{\circ m} = \mathcal{M}$, for all $m \geq 0$.*

4. *(†) Let $a, b \in A$ with $a < b$, and let $k \in \mathbb{N}_{\leq n}$.*
   The sequence of means $(\mathcal{M}_{k,m})_{m \in \mathbb{N}}$ converges on $(A \cap [a, b])^n$ uniformly to $\mathcal{M}$.

5. *Suppose that all the M_k are 'good', where 'good' is an arbitrary but fixed property out of **increasing**, **strictly increasing**, **continuous**, **smooth**, F**-rational**, **scale-invariant**, **symmetric**, $(\cdot, a)$-**power-asymptotic**, $(a, \cdot)$-**power-asymptotic** (where $a \in [0, 1]$ and F is a subfield of $\mathbb{R}$). Then $\mathcal{M}_{k,m}$ is good, for all k and m.*

6. *Suppose that all the M_k are 'good', where 'good' is an arbitrary but fixed property out of **increasing**, **continuous**, **scale-invariant**, **symmetric**. Then $\mathcal{M}$ is good.*

7. *(*) Let a, b, r, R be as in part 1.*
   The means $\mathcal{M}$ and $\mathcal{M}_{k,m}$ are $[r, R]$-alt-compressing on $A \cap [a, b]$.
   *In particular, $\mathcal{M}$ and $\mathcal{M}_{k,m}$ are **compressing**.*

8. *For any continuous function $f : A^n \to \mathbb{R}$, it holds that*

$$f = f \circ \mathbf{M} \iff f = f \circ \mathrm{diag}_n \circ \mathcal{M}.$$

   Hence, if M is a continuous mean on A^n, then

$$M = M \circ \mathbf{M} \iff M = \mathcal{M}. \qquad ⅃$$

The proof of the theorem is below Example 3.2.10. We could have placed the proof here, but we prefer to first clarify the meaning of the theorem.

***Remark* 3.2.5.** In the context of part 8 of the theorem, the equation $f = f \circ \mathbf{M}$ (where $f \in \mathbb{R}^{(A^n)}$ is the variable) is called the **functional equation of $\mathcal{M}$**. Part 8 tells us that if f is a continuous solution of the functional equation of $\mathcal{M}$, then $f(\mathbf{x}) = f(\mu, \mu, \ldots, \mu)$, were $\mu := \mathcal{M}(\mathbf{x})$.

***Remark* 3.2.6.** In the notation of the theorem, we have for all $m \geq 0$ that

$$\mathbf{M}^{\circ(m+1)} = (\mathcal{M}_{1,m}, \ldots, \mathcal{M}_{n,m}) \in \mathbb{M}^n_{\mathrm{comp}(A^n)}.$$

Definition 3.2.7 (Gauss composition of compressing means)**.** Let $A \subseteq \mathbb{R}$, let $n \in \mathbb{N}$, let $\mathbf{M} = (M_1, \ldots, M_n) \in \mathbb{M}^n_{\mathrm{comp}(A^n)}$.

1. The mean $\mathcal{M} : A^n \to \mathbb{R}$ that we constructed in Theorem 3.2.4, that is,

$$\mathcal{M}(\mathbf{x}) = \pi_k\big(\lim_{m\to\infty} \mathbf{M}^{\circ m}(\mathbf{x})\big) \qquad \forall k \in \mathbb{N}_{\leq n},$$

is called the *Gauss composition of* $\mathbf{M}$.

2. The *Gauss composition map* $\bigotimes : \mathbb{M}^n_{\mathrm{comp}(A^n)} \to \mathbb{M}_{\mathrm{comp}(A^n)}$ is the function that maps $\mathbf{M} \in \mathbb{M}^n_{\mathrm{comp}(A^n)}$ to the Gauss composition of $\mathbf{M}$.

As alternative notations, we write $\bigotimes \mathbf{M} =: \bigotimes_{k=1}^n M_k =: M_1 \otimes \cdots \otimes M_n$.

***Remark* 3.2.8.** The statement and proof of Theorem 3.2.4 are our own, but we saw (afterwards) that the parts 1,3,5,6,8 they overlap with [Mat99, Lemma 1 & Theorem 1 & Theorem 2]. Matkowski states his results in [Mat99] only in the case that A is an interval. On another aspect, it is stronger: it supposes that only one of the means $M_1, \ldots, M_n$ is strict, instead of all of them. (When A is an interval, being strict is equivalent to being compressing, see Lemma 1.1.26.)

3.2.3 First example of Gauss composition: the AGM

Definition 3.2.9. The *arithmetic-geometric mean on* $\mathbb{R}^2_{>0}$ is $\mathrm{AGM} := \mathrm{AM}_2 \otimes \mathrm{GM}_2$.

Example 3.2.10. By Theorem 3.2.4, AGM is an increasing, continuous, scale-invariant, symmetric, compressing mean on $\mathbb{R}^2_{>0}$. It is the unique continuous mean on $\mathbb{R}^2_{>0}$ that satisfies the functional equation

$$\mathrm{AGM}(a, b) = \mathrm{AGM}(\tfrac{a+b}{2}, \sqrt{ab}).$$

For any $(a, b) \in \mathbb{R}^2_{>0}$, we can compute $\mathrm{AGM}(a, b)$ as the common limit of a_m and b_m as $m \to \infty$, where $(a_m, b_m) := (\mathrm{AM}_2, \mathrm{GM}_2)^{\circ m}(a, b)$. Explicitly, the recursion is given by

$$(a_0, b_0) = (a, b), \qquad (a_{m+1}, b_{m+1}) = (\tfrac{1}{2}(a_m + b_m), \sqrt{a_m b_m}) \quad \text{for } m \geq 0.$$

Clearly, if $a_0 \geq b_0$, then $a_m \geq a_{m+1} \geq b_{m+1} \geq b_m$ for all m, because $\mathrm{AM}_2 \geq \mathrm{GM}_2$. Hence,

$$a_m \geq \lim_{m\to\infty} a_m = \mathrm{AGM}(a, b) = \lim_{m\to\infty} b_m \geq b_m \qquad \text{if } a_0 \geq b_0. \tag{3.5}$$

By (3.3), $|a_m - b_m|$ converges at least exponentially, but the convergence seems to be much faster; I did not attempt to formally state or prove that, but I expect that a tight upper bound for $|a_m - b_m|$ can be derived by the observation that when a and b are close together, say $|a - b| = 2\varepsilon$, then $\sqrt{ab}$ is very close to $(a + b)/2 =: M$, namely $|\sqrt{ab} - M| = |\sqrt{M^2 - \varepsilon^2} - M| \approx \varepsilon^2/(2M)$.

The following example is worked out in more detail in [Cox84, Example 2]. Starting with $(a_0, b_0) = (\sqrt{2}, 1)$, we have $|a_4 - b_4| < 10^{-19}$. Using that $a_4 \geq \mathrm{AGM}(\sqrt{2}, 1) \geq b_4$, we obtain the decimal expansion $\mathrm{AGM}(\sqrt{2}, 1) = 1.1981402347355922074\ldots$ in which all 20 decimals are correct. Gauss did this computation by hand.

***Remark* 3.2.11.** 'Gauss composition of means' is named after C.F. Gauss, who initiated the subject by intensely studying the AGM and revealing deep insights about it, especially by extending the domain of the AGM to certain sets of complex numbers, and by studying the AGM in relation with elliptic integrals and modular functions. As far as I know, the only "Gauss composition of means" that Gauss studied himself, is the AGM (and probably the related GHM and AHM). For a nice and modern overview to the work of Gauss on the AGM, we refer to [Cox84]. We return to this subject in §3.3.

The term "Gauss composition" is used in [DMP05] and [CFKT14] and [MP15], for example, but not in [Mat99].

3.2.4 Proof of the basic theorem of Gauss composition

Proof of Theorem 3.2.4.

1. By Corollary 1.1.30, for each $k \in \mathbb{N}_{\leq n}$ there exist r_k, R_k with $0 < r_k \leq R_k < 1$ such that M_k is $[r_k, R_k]$-alt-compressing on $A \cap [a, b]$. So we can take $r := \mathrm{Min}(r_1, \ldots, r_n)$ and $R := \mathrm{Max}(R_1, \ldots, R_n)$.

2. (3.4) follows easily by noting that $\mathbf{x}_{m+1} = \big(M_1(\mathbf{x}_m), \ldots, M_n(\mathbf{x}_m)\big)$.

(3.3) follows by induction on m. For $m = 0$ it is trivial. Suppose it holds for some $m \geq 0$, so we have $b' - a' \leq (R - r)^m (b - a)$, where $a' := \mathrm{Min}(\mathbf{x}_m) \leq \mathrm{Max}(\mathbf{x}_m) =: b'$. Lemma 1.1.31 tells us that there is an interval I of length $(R - r)(b' - a')$ such that $M_k(\mathbf{x}_m) \in I$, for all k. It follows that $\mathbf{x}_{m+1} \in I^n$, in particular we have

$$\mathrm{Max}(\mathbf{x}_{m+1}) - \mathrm{Min}(\mathbf{x}_{m+1}) \leq (R - r)(b' - a') \leq (R - r)^{m+1}(b - a),$$

as desired.

3. The first statement follows because $\lim_{m\to\infty} \mathrm{Max}(\mathbf{x}_m) - \mathrm{Min}(\mathbf{x}_m) = 0$, by (3.3). That $\mathcal{M}$ is a mean follows from (3.4). Clearly, for any $k \in \mathbb{N}_{\leq n}$ and $m \geq 0$, we have

$$\mathcal{M}(\mathbf{x}) = \pi_k\big(\lim_{i\to\infty} \mathbf{M}^{\circ i}(\mathbf{x})\big) = \pi_k\big(\lim_{i\to\infty} \mathbf{M}^{\circ(i+m)}(\mathbf{x})\big) = \mathcal{M}(\mathbf{M}^{\circ m}(\mathbf{x})).$$

4. Let r, R be as in part 1, corresponding to the given a, b. Part 2 tells us that for all $\mathbf{x} \in (A \cap [a, b])^n$, the numbers $\mathcal{M}_{k,m}(\mathbf{x})$ and $\mathcal{M}(\mathbf{x})$ are both contained in the interval $[\mathrm{Min}(\mathbf{x}_m), \mathrm{Max}(\mathbf{x}_m)]$, hence $|\mathcal{M}_{k,m}(\mathbf{x}) - \mathcal{M}(\mathbf{x})| \leq (R - r)^m (b - a)$.

5. Note that for all $m \geq 1$, we have $\mathcal{M}_{k,m} = \mathcal{M}_{k,m-1} \circ \mathbf{M}$, so

$$\mathcal{M}_{k,m}(\mathbf{x}) = \mathcal{M}_{k,m-1}\big(M_1(\mathbf{x}), \ldots, M_n(\mathbf{x})\big) \qquad \text{for all } \mathbf{x} \in A^n. \tag{3.6}$$

We show by induction on m that $\mathcal{M}_{k,m}$ is good. For $m = 0$ it is tautologically true. If $\mathcal{M}_{k,m-1}$ is good, for some $m \geq 1$, then it follows from (3.6) and Theorem 3.1.1 that $\mathcal{M}_{k,m}$ is good.

6. Part 5 tells us that all the means $\mathcal{M}_{k,m}$ are good. Moreover, Lemma 3.2.1 tells us that $\mathcal{M}$ is good, except in the case that 'good' means 'continuous'. Thus, suppose that $\mathcal{M}_{k,m}$ are continuous, for all m and k. By part 4 and the uniform limit theorem, it follows that $\mathcal{M}$ is continuous on $(A \cap [a,b])^n$, for all $a, b \in A$ with $a < b$. Hence, it follows that $\mathcal{M}$ is continuous on A^n.

7. That $\mathcal{M}_{k,m}$ is $[r,R]$-alt-compressing on $A \cap [a,b]$, is tautologically true for $m = 0$, and for $m \geq 1$ it follows easily from the fact (derived from (3.4)) that

$$\mathrm{Min}\big(M_1(\mathbf{x}), \ldots, M_n(\mathbf{x})\big) \leq \mathcal{M}_{k,m}(\mathbf{x}) \leq \mathrm{Max}\big(M_1(\mathbf{x}), \ldots, M_n(\mathbf{x})\big),$$

and from the fact that each M_k is $[r,R]$-alt-compressing on $A \cap [a,b]$.

Lemma 3.2 tells us that $\mathcal{M}$ is $[r,R]$-alt-compressing on $A \cap [a,b]$.

8. The implication '$\Longleftarrow$' follows directly from the identity $\mathcal{M} \circ \mathbf{M} = \mathcal{M}$, which we showed in part 3. Conversely, suppose that $f = f \circ \mathbf{M}$. Then $f = f \circ \mathbf{M}^{\circ m}$ for all $m \geq 0$. Because f is continuous, we have $f(\mathbf{x}) = f(\lim_{m\to\infty} \mathbf{M}^{\circ m}(\mathbf{x})) = f(\mathrm{diag}_n(\mathcal{M}(\mathbf{x}))$. $\square$

3.2.5 Implications and questions

Corollary 3.2.12. *Let $A \subseteq \mathbb{R}$, let $n \in \mathbb{N}$, let $\mathbf{M} \in \mathbb{M}^n_{\mathrm{comp}(A^n)}$. Suppose that the means of $\mathbf{M}$ are continuous. For all $m \geq 1$, we have $\bigotimes \mathbf{M}^{\circ m} = \bigotimes \mathbf{M}$.*

Proof. We write $\mathcal{M} := \bigotimes \mathbf{M}$ and $\mathcal{M}_m := \bigotimes \mathbf{M}^{\circ m}$. We have $\mathcal{M} = \mathcal{M} \circ \mathbf{M}^{\circ m}$, by part 3 (of Theorem 3.2.4). So $\mathcal{M}$ satisfies the functional equation of $\mathcal{M}_m$. Moreover, $\mathcal{M}$ is a continuous mean, by part 3 and 6. Hence, by part 8, $\mathcal{M} = \mathcal{M}_m$. $\square$

***Remark* 3.2.13.** By part 5 of Theorem 3.2.4, many properties of the means of $\mathbf{M}$ are inherited by the means of $\mathbf{M}^{\circ m}$, for all $m \geq 1$. But in part 6 of the theorem, we only mention for some of those properties that they are inherited by $\mathcal{M} := \bigotimes \mathbf{M}$. The reality is that I don't know if the other properties are inherited by $\mathcal{M}$, except for F-rationality. Namely, we will see in Proposition 3.9.1 that $\mathrm{AM}_2 \otimes \mathrm{HM}_2 = \mathrm{GM}_2$. Clearly, AM_2 and HM_2 are $\mathbb{Q}$-rational, but GM_2 is not even $\mathbb{R}$-rational.

For the remaining properties, I know that they are not inherited by pointwise limits of means in general. For example, if $p_1, p_2 \ldots$ is a sequence in $\mathbb{R}_{>0}$ such that $\lim_{m\to\infty} p_m = \infty$, then every power mean PM_{n,p_m} is strictly increasing and smooth, but the pointwise limit $\lim_{m\to\infty} \mathrm{PM}_{n,p_m}$ equals Max_n, which is neither strictly increasing nor smooth. Moreover, every translation mean TM_{n,p_m} is $(0, \frac{1}{n})$-power-asymptotic, but the pointwise limit $\lim_{m\to\infty} \mathrm{TM}_{n,p_m}$ equals AM_n, which is $(0,1)$-power-asymptotic.

However, these are examples where the "limit is taken in an extreme direction". The limit involved in Gauss composition seems to be much "nicer"; intuitively (see Remark 3.4.4), $\mathcal{M}$ is "in between means that are all good", and by Theorem 3.2.4.4, "the means $\mathcal{M}_{k,m}$ converge uniformly and exponentially to $\mathcal{M}$ on bounded intervals of A". I expect that the limiting process is nice enough to preserve the remaining properties, in the sense of the following conjecture.

***Conjecture* 3.2.14** (†?)**.** Let $A \subseteq \mathbb{R}$, let $n \in \mathbb{N}$, let $\mathbf{a} \in [0,1]^n$, let $M_1, \ldots, M_n \in \mathbb{M}_{\mathrm{comp}(A^n)}$. Suppose that each M_k is 'good', where 'good' is an arbitrary but fixed property out of **smooth**, **strictly increasing**, $(\cdot, \mathbf{a})$**-power-asymptotic**, $(\mathbf{a}, \cdot)$**-power-asymptotic**. Then $\bigotimes_{k=1}^n M_k$ is good.

3.3 The Arithmetic-Geometric Mean

The arithmetic-geometric mean is undoubtedly the best-studied example of Gauss composition; already by Gauss and contemporaries of him, much was known about this mean (see [Cox84, Introduction]). The main motivation for Gauss to consider the AGM, appears to have been the following theorem.

Theorem 3.3.1 (Integral equation for the AGM)**.** *For all* $(a,b) \in \mathbb{R}^2_{>0}$*, we have*

$$\mathrm{AGM}(a,b) = \begin{cases} \dfrac{\pi}{2I(a,b)} & \text{if } a \geq b, \\ \dfrac{\pi}{2I(b,a)} & \text{if } a \leq b, \end{cases} \qquad \text{where } I(a,b) = \int_0^{\frac{\pi}{2}} \left\|(a\cos\varphi,\ b\sin\varphi)\right\|_2^{-1} d\varphi.$$

(Reference to) proof. Let $(a,b) \in \mathbb{R}^2_{>0}$. By Theorem 3.2.4.6, the mean AGM is symmetric. Therefore, it suffices to show that if $a \geq b$, then $\mathrm{AGM}(a,b) = (\pi/2)I(a,b)^{-1}$.

Thus, suppose $a \geq b$. It can be verified by nontrivial but elementary calculations, see [Cox84, Thm. 1.1], that $I(a,b) = I(\frac{a+b}{2}, \sqrt{ab})$. In other words, $I(a,b) = I((\mathrm{AM}_2, \mathrm{GM}_2)(a,b))$. Moreover, by Corollary 2.6.8 we have $\mathrm{AM}_2(a,b) \geq \mathrm{GM}_2(a,b)$. Thus, it follows by induction on m that $I(a,b) = I((\mathrm{AM}_2, \mathrm{GM}_2)^{\circ m}(a,b))$ for all $m \geq 0$. Because I is a continuous (even smooth) function of $(a,b) \in \mathbb{R}^2_{>0}$, it follows by letting $m \to \infty$ that $I(a,b) = I(\mu,\mu)$, where $\mu := \mathrm{AGM}(a,b)$. Hence,

$$I(a,b) = I(\mu,\mu) = \int_0^{\pi/2} \mu^{-1} d\varphi = (\pi/2)\mu^{-1}. \qquad \square$$

***Remark* 3.3.2.** The integral $I(a,b)$ can be written in terms of a so-called complete elliptic integral of the first kind; see [Cox84, p279]. On page 280–281 of that review article, it is explained that Gauss did his computation of $\mathrm{AGM}(\sqrt{2},1)$, that we mentioned in Example 3.2.10, to compute the arc length of a lemniscate of Bernoulli, namely the length ℓ of the algebraic curve in $\mathbb{R}^2$ defined by $(x^2+y^2)^2 = 2(x^2-y^2)$; it turns out that $\ell = 4I(\sqrt{2},1) = (2\pi)/\mathrm{AGM}(\sqrt{2},1)$. Gauss wrote about this in his diary: "the demonstration of this fact will surely open an entirely new field of analysis." Quoting [Cox84, p281]: "It was in trying to understand the real meaning of this equality that several streams of Gauss' thought came together and produced the exceptionally rich mathematics which we will explore in §2."

In [Cox84, §2], the AGM is extended to a multi-valued function on $\mathbb{C}^2$ – multi-valued because at each iteration of $(a,b) \mapsto (\frac{a+b}{2}, \sqrt{ab})$, there are two possible choices for $\sqrt{ab}$, thus leading to an uncountable set of sequences, all of them converging in $\mathbb{C}^2$. However, for only countably many of those sequences, the limit is non-zero. Gauss showed that for $|a| \geq |b|$, except in the trivial case that $a = \pm b$ or $b = 0$, the set of non-zero values of $\mathrm{AGM}(a,b)$ equals

$$\left\{(d\mu^{-1} + ic\lambda^{-1})^{-1} \ : \ d \in 4\mathbb{Z}+1,\ c \in 4\mathbb{Z},\ \gcd(d,c) = 1\right\},$$

where μ is the "simplest" value of $\mathrm{AGM}(a,b)$, and λ is the "simplest" value of $\mathrm{AGM}(a+b, a-b)$; see [Cox84, Theorem 2.2], and above the theorem for the definition of the "simplest" value. The proof of the theorem is deep; quoting Cox, the proof "makes extensive use of theta functions and modular functions of level four", and "we will finally discover the 'entirely new field of analysis' predicted by Gauss".

It is beyond the scope of this text to delve into these matters.

As a concrete application of Theorem 3.3.1, we mention that it leads to sequences that converge very rapidly to the number π.

Theorem 3.3.3. *Let* $\mathbf{M} := (\mathrm{AM}_2, \mathrm{GM}_2) \in \mathbb{M}^2_{\mathrm{comp}(\mathbb{R}^2_{>0})}$. *For* $m \in \mathbb{N}$, *let*

$$\Delta_m := f \circ \mathbf{M}^{\circ m} : \quad \mathbb{R}^2_{>0} \to \mathbb{R}, \qquad \text{where } f(x,y) := x^2 - y^2.$$

For all $(a,b) \in \mathbb{R}^2_{>0}$ *such that* $a^2 + b^2 = 1$, *we have*

$$\pi = \frac{4 \cdot \mathrm{AGM}(1,a) \cdot \mathrm{AGM}(1,b)}{1 - \sum_{m=1}^{\infty} 2^m \big(\Delta_m(1,a) + \Delta_m(1,b)\big)}.$$

Proof. This is Theorem 1a of [Sal76, §5]. It follows by combining Theorem 3.3.1 with a certain identity of elliptic integrals, called Legendre's relation. □

***Remark* 3.3.4.** As we mentioned in Example 3.2.10, for $(x,y) \in \mathbb{R}^2_{>0}$, the sequence $\mathbf{M}^{\circ m}(x,y)$ converges very rapidly to $\mathrm{AGM}(x,y)$. Thus, Theorem 3.3.3 yields a very efficient algorithm for computing π. It is most efficient when $a = b$, that is, $a = \sqrt{1/2} = b$. Writing out the corresponding formula explicitly leads to

Corollary 3.3.5. *Let the sequence* $(a_m, b_m)_{m\in\mathbb{N}}$ *be recursively defined by* $(a_0, b_0) = (1, (\frac{1}{2})^{\frac{1}{2}})$, *and by* $(a_{m+1}, b_{m+1}) = \big(\frac{1}{2}(a_m + b_m),\ (a_m b_m)^{\frac{1}{2}}\big)$ *for* $m \geq 0$. *Then*

$$\pi = \frac{4(\lim_{m\to\infty} a_m)^2}{1 - \sum_{m=1}^{\infty} 2^{m+1}\big(a_m^2 - b_m^2\big)}.$$

3.4 Properties of the Gauss composition map

We studied properties of the means $\bigotimes \mathbf{M}$ in Theorem 3.2.4, but in this section we study properties of the map $\bigotimes$ itself.

Proposition 3.4.1 (†?). *Let* A *be a subset of* $\mathbb{R}$, *let* $n \in \mathbb{N}$, *let* $\mathbb{M} := \mathbb{M}_{\mathrm{comp}(A^n)}$.
Let $M \in \mathbb{M}$, *let* $\mathbf{M} = (M_1, \dots, M_n) \in \mathbb{M}^n$ *and* $\mathbf{N} = (N_1, \dots, N_n) \in \mathbb{M}^n$.
Let $\bigotimes : \mathbb{M}^n \to \mathbb{M}$ *be the Gauss composition map.*

1. *"Betweenness": If* $f, g : A^n \to \mathbb{R}$ *are functions such that* $f \leq M_k \leq g$ *for all* k, *then* $f \leq \bigotimes \mathbf{M} \leq g$.
2. *"Increasing": Suppose that* $M_k \leq N_k$ *for all* k, *and that either all the means of* $\mathbf{M}$ *or all the means of* $\mathbf{N}$ *are increasing. Then* $\bigotimes \mathbf{M} \leq \bigotimes \mathbf{N}$.
3. *"Constant-preserving":* $\bigotimes \circ \operatorname{diag}_{\mathbb{M},n} = \mathrm{id}_{\mathbb{M}}$.

4. *"Symmetric": Suppose that the means of* $\mathbf{M}$ *are symmetric. Then for all* $\sigma \in S_n$*, we have* $\bigotimes \mathbf{M}_{\circ\sigma} = \bigotimes \mathbf{M}$*.*

Proof. **1.** Let $\mathbf{x} \in A^n$. Using that $\mathrm{Min} \leq M_k \leq \mathrm{Max}$, it follows by induction on $m \geq 1$ that $\mathrm{diag}_n(f(\mathbf{x})) \leq \mathbf{M}^{\circ m}(\mathbf{x}) \leq \mathrm{diag}_n(g(\mathbf{x}))$, hence the result follows by letting $m \to \infty$.

2. Let $\mathbf{y}, \mathbf{z} \in A^n$ such that $\mathbf{y} \leq \mathbf{z}$; we claim that $\mathbf{M}(\mathbf{y}) \leq \mathbf{N}(\mathbf{z})$. Namely, if each M_k is increasing, then $M_k(\mathbf{y}) \leq M_k(\mathbf{z}) \leq N_k(\mathbf{z})$, while if each N_k is increasing, then $M_k(\mathbf{y}) \leq N_k(\mathbf{y}) \leq N_k(\mathbf{z})$. So in any case, $M_k(\mathbf{y}) \leq N_k(\mathbf{z})$ for all k, that is, $\mathbf{M}(\mathbf{y}) \leq \mathbf{N}(\mathbf{z})$.

Let $\mathbf{x} \in A^n$. From our claim, it follows by induction on $m \geq 0$ that $\mathbf{M}^{\circ m}(\mathbf{x}) \leq \mathbf{N}^{\circ m}(\mathbf{x})$, hence the result follows by letting $m \to \infty$.

3. This follows directly from the definitions since means are constant-preserving, but it also follows directly from part 1, by $M \leq \bigotimes_{k=1}^n M \leq M$.

4. Let $\mathbf{x} \in A^n$ and $\sigma \in S_n$. By symmetry of the means M_k, it follows that $\mathbf{M}_{\circ\sigma}(\mathbf{x}_{\circ\sigma}) = \mathbf{M}_{\circ\sigma}(\mathbf{x}) = (\mathbf{M}(\mathbf{x}))_{\circ\sigma}$. Hence, it follows by induction on $m \geq 1$ that $(\mathbf{M}_{\circ\sigma})^{\circ m}(\mathbf{x}) = (\mathbf{M}^{\circ m}(\mathbf{x}))_{\circ\sigma}$. The result follows by letting $m \to \infty$ and noting that $\mathrm{diag}(A^n)$ is invariant under σ. $\square$

***Conjecture* 3.4.2** (†?)**.** *"Strictly increasing":* Suppose that in addition to the conditions of part 2 ("Increasing") of Proposition 3.4.1, there is a subset D of A^n such that $M_k < N_k$ on D, for at least one $k \in \mathbb{N}_{\leq n}$. Then $\bigotimes \mathbf{M} < \bigotimes \mathbf{N}$ on D.

Conditional proof: Suppose that Conjecture 3.2.14 is true as far as 'strictly increasing' is concerned. Thus, we have that either (case 1) the mean $\mathcal{M} := \bigotimes \mathbf{M}$ is strictly increasing, or (case 2) the mean $\mathcal{N} := \bigotimes \mathbf{N}$ is strictly increasing. Moreover, for all $\mathbf{x} \in D$, we have $M_k(\mathbf{x}) \leq N_k(\mathbf{x})$ for all k, with strict inequality for at least one k, so $\mathbf{M}(\mathbf{x}) < \mathbf{N}(\mathbf{x})$. It follows (using Proposition 3.4.1.3 and the functional equations of $\mathcal{M}$ and $\mathcal{N}$) that

$$\begin{aligned} \mathcal{M}(\mathbf{x}) &= \mathcal{M}(\mathbf{M}(\mathbf{x})) < \mathcal{M}(\mathbf{N}(\mathbf{x})) \leq \mathcal{N}(\mathbf{N}(\mathbf{x})) = \mathcal{N}(\mathbf{x}) \qquad \text{in case 1,} \\ \mathcal{M}(\mathbf{x}) &= \mathcal{M}(\mathbf{M}(\mathbf{x})) \leq \mathcal{N}(\mathbf{M}(\mathbf{x})) < \mathcal{N}(\mathbf{N}(\mathbf{x})) = \mathcal{N}(\mathbf{x}) \qquad \text{in case 2.} \end{aligned}$$ $\square$

***Remark* 3.4.3.** The requirement that the means of $\mathbf{M}$ are symmetric, can't be left away in part 4 of Proposition 3.4.1: in Example 4.7.9, we give an example of two non-symmetric means M and N on $\mathbb{R}^2$ such that $M \otimes N \neq N \otimes M$.

***Remark* 3.4.4.** The properties of the Gauss composition map that we listed in Proposition 3.4.1 and Conjecture 3.4.2 reflect to a certain extend the properties of a (strictly) increasing, (symmetric) mean. We will see in Theorem 3.6.3 that continuity is also reflected by the Gauss composition map. We may intuitively view $\bigotimes$ as "a (partially increasing, symmetric, continuous) mean on $\mathbb{M}^n$", and the Gauss composition of $M_1, \ldots, M_n$ as "the average of the means $M_1, \ldots, M_n$".

***Conjecture* 3.4.5** (*)**.** For all $(i, j) \in \mathbb{N}^2_{\leq n}$, let M_{ij} be a continuous, symmetric, compressing mean on A^n. Then

$$\bigotimes_{j=1}^n \Big(\bigotimes_{i=1}^n M_{ij} \Big) = \bigotimes_{i=1}^n \Big(\bigotimes_{j=1}^n M_{ij} \Big).$$

Remarks about this conjecture. The requirement that the means of $\mathbf{M}$ are symmetric, can't be left away: in Example 4.7.9, we give an example of four non-symmetric means M_{11}, M_{12}, M_{21}, M_{22} on $\mathbb{R}^2$, such that

$$(M_{11} \otimes M_{21}) \otimes (M_{12} \otimes M_{22}) \neq (M_{11} \otimes M_{12}) \otimes (M_{21} \otimes M_{22}).$$

If the conjecture is true, then it follows that $\bigotimes_{i=1}^n \left(\bigotimes_{j=1}^n M_{\sigma(ij)}\right) = \bigotimes_{i=1}^n \left(\bigotimes_{j=1}^n M_{ij}\right)$ for all $\sigma \in S^{n \times n}$, because we also have the symmetries from Proposition 3.4.1.4, and together with the symmetry from this conjecture, they generate the entire symmetry group $S^{n\times n}$.

I expect that the conjecture is true, because Gauss composition of symmetric means seems to be a very symmetric process in all respects; it seems to me that $\bigotimes_{i=1}^n \left(\bigotimes_{j=1}^n M_{ij}\right)$ is in a sense "the average of the means M_{ij}", and that the order should not matter. But I can imagine that my intuition is wrong here. □

Example 3.4.6. Let M, M_1, M_2 be three continuous, symmetric, compressing means on $\mathbb{R}^2_{>0}$, and suppose that M is scale-invariant. By Proposition 3.9.1 below, we have that $M \otimes M^{[\text{inv}]} = \text{GM}_2$. Thus, if Conjecture 3.4.5 is true, then it follows that

$$(M \otimes M_1) \otimes (M^{[\text{inv}]} \otimes M_2) = \text{GM}_2 \otimes (M_1 \otimes M_2).$$

For example, if would follow that for all $p, b, c \in \mathbb{R}$,

$$(\text{PM}_{2,p} \otimes \text{TM}_{2,b}) \otimes (\text{PM}_{2,-p} \otimes \text{TM}_{2,c}) = \text{GM}_2 \otimes (\text{TM}_{2,b} \otimes \text{TM}_{2,c}).$$ ♪

3.5 Gauss composition commutes with Conjugation

We could informally say that Lemma 3.2.1.4 tells us that "under good conditions, conjugation commutes with pointwise limit". In this section, we show that "conjugation commutes with Gauss composition", in the sense of Theorem 3.5.3. In the next section, we show that "under good conditions, pointwise limit commutes with Gauss composition", in the sense of Theorem 3.6.3.

Definition 3.5.1. Let A, B be subsets of $\mathbb{R}$, let $g : A \to B$ be a monotonic bijection, let $n \in \mathbb{N}$. For $\mathbf{M} = (M_1, \ldots, M_n) \in \mathbb{M}^n_{\text{comp}(A^n)}$, we write

$$\mathbf{M}^{[g]} := (M_1^{[g]}, \ldots, M_n^{[g]}) \in \mathbb{M}^n_{\text{comp}(B^n)}.$$ 𝄽

***Remark* 3.5.2.** That $\mathbf{M}^{[g]} \in \mathbb{M}^n_{\text{comp}(B^n)}$ follows from Theorem 2.1.3.

For any monotonic bijection $h : B \to C$ where $C \subseteq \mathbb{R}$, we clearly have $(\mathbf{M}^{[g]})^{[h]} = \mathbf{M}^{[h\circ g]}$ and $\mathbf{M}^{[\text{id}_A]} = \mathbf{M}$, by Proposition 1.3.7.1. 𝄿

Theorem 3.5.3 (†)**.** *Let A be an interval of $\mathbb{R}$, let $n \in \mathbb{N}$, let $\mathbf{M} \in \mathbb{M}^n_{\text{comp}(A^n)}$. Let $B \subseteq \mathbb{R}$ and let $g : A \to B$ be a continuous monotonic bijection. Then $\bigotimes \mathbf{M} \in \mathbb{M}_{\text{comp}(A^n)}$, and*

$$\left(\bigotimes \mathbf{M}\right)^{[g]} = \bigotimes \mathbf{M}^{[g]} \quad \in \mathbb{M}_{\text{comp}(B^n)}.$$ 𝅗𝅥

Proof. We write $\mathcal{M} := \bigotimes \mathbf{M}$ and $\mathcal{M}_g := \bigotimes \mathbf{M}^{[g]}$.

Because A is an interval, $\mathcal{M}$ is an internal mean on A^n, and $\mathcal{M}$ is compressing by Theorem 3.2.4. So $\mathcal{M} \in \mathbb{M}_{\mathrm{comp}(A^n)}$, and therefore $\mathcal{M}^{[g]} \in \mathbb{M}_{\mathrm{comp}(B^n)}$.

Let $k \in \mathbb{N}_{\leq n}$ be arbitrary but fixed. We have, as pointwise limits,

$$\mathcal{M} = \lim_{m\to\infty} M_k \circ \mathbf{M}^{\circ m} \qquad \text{and} \qquad \mathcal{M}_g = \lim_{m\to\infty} M_k^{[g]} \circ (\mathbf{M}^{[g]})^{\circ m}. \tag{3.7}$$

By induction on $m \geq 0$, it follows from Theorem 3.1.1.3 that

$$M_k^{[g]} \circ (\mathbf{M}^{[g]})^{\circ m} = (M_k \circ \mathbf{M}^{\circ m})^{[g]}. \tag{3.8}$$

With (3.7) and (3.8) combined, Lemma 1.3.7.7 tells us that $\mathcal{M}_g = \mathcal{M}^{[g]}$, as desired. □

More elegant proof in the case that the means of $\mathbf{M}$ *are continuous.* Then $\mathcal{M}$ is continuous, hence $\mathcal{M}^{[g]}$ is continuous. Further, we have

$$\mathcal{M}^{[g]} \circ \mathbf{M}^{[g]} = (\mathcal{M} \circ \mathbf{M})^{[g]} = \mathcal{M}^{[g]},$$

the first equality by Theorem 3.1.1.3. Hence, $\mathcal{M}^{[g]}$ satisfies the functional equation of $\mathcal{M}_g$, so Theorem 3.2.4.9 tells us that $\mathcal{M}^{[g]} = \mathcal{M}_g$. □

Example 3.5.4. The *geometric-harmonic mean* is defined as $\mathrm{GM}_2 \otimes \mathrm{HM}_2$ on $\mathbb{R}^2_{>0}$. There is no need to study it separately from the arithmetic-geometric mean, because by Theorem 3.5.3 and Proposition 3.4.1.4, we have $(\mathrm{GM}_2 \otimes \mathrm{HM}_2)^{[\mathrm{inv}]} = \mathrm{AM}_2 \otimes \mathrm{GM}_2$.

3.6 Gauss composition commutes with Limit

We first prove Theorem 3.6.3, that "the Gauss composition map reflects continuity". Then we use that theorem to prove Theorem 3.6.5, that a 1-dimensional continuum of means on A^n gives rise via Gauss composition to an n-dimensional continuum of means on A^n (if the topological space T is regarded as 1-dimensional).

***Remark* 3.6.1.** Let $A \subseteq \mathbb{R}$ and $n \in \mathbb{N}$. Because the elements of $\mathbb{M}^n_{\mathrm{comp}(A^n)}$ are transformations of A^n, we have a notion of pointwise convergence in $\mathbb{M}^n_{\mathrm{comp}(A^n)}$.

Concretely, let $\mathbf{M}_\ell = (M_{1,\ell}, \ldots, M_{n,\ell}) \in \mathbb{M}^n_{\mathrm{comp}(A^n)}$ for each $\ell \in \mathbb{N}$, and let $\mathbf{M} = (M_1, \ldots, M_n) \in \mathbb{M}^n_{\mathrm{comp}(A^n)}$. The condition that $\lim_{\ell\to\infty} \mathbf{M}_\ell = \mathbf{M}$ pointwise, is clearly equivalent to the condition that for all $k \in \mathbb{N}_{\leq n}$, we have $\lim_{\ell\to\infty} M_{k,\ell} = M_k$ pointwise.

Example 3.6.2. For $\ell \in \mathbb{N}$, let $b_\ell, c_\ell \in \overline{\mathbb{R}}$, and let $\mathbf{w}_\ell, \mathbf{v}_\ell \in \mathbb{R}^2_{>0}$, such that

$$b_\ell \to \infty, \quad c_\ell \to 0, \quad \mathbf{w}_\ell \to (1,1), \quad \mathbf{v}_\ell \to (1,1), \qquad \text{as } \ell \to \infty.$$

The theorem on translation means (2.7.4.5) tells us that

$$\lim_{\ell\to\infty} (\mathrm{TM}_{b_\ell,\mathbf{w}_\ell}, \mathrm{TM}_{c_\ell,\mathbf{v}_\ell}) = (\mathrm{AM}_2, \mathrm{GM}_2).$$

Theorem 3.6.3 (*). *Let $A \subseteq \mathbb{R}$ and $n \in \mathbb{N}$. Let $\mathbf{M}_\ell = (M_{1,\ell}, \dots, M_{n,\ell}) \in \mathbb{M}^n_{\mathrm{comp}(A^n)}$ for each $\ell \in \mathbb{N}$, and let $\mathbf{M} = (M_1, \dots, M_n) \in \mathbb{M}^n_{\mathrm{comp}(A^n)}$.*

Suppose that all the means of $\mathbf{M}$ are continuous, and that we have pointwise

$$\lim_{\ell\to\infty} \mathbf{M}_\ell = \mathbf{M}.$$

Suppose further that for each $k \in \mathbb{N}_{\leq n}$ the following holds: If $\mathbf{y}_1, \mathbf{y}_2, \dots$ is a sequence in A^n that converges to some $\mathbf{y} \in A^n$, then

$$\text{either} \qquad \lim_{\ell\to\infty} |M_{k,\ell}(\mathbf{y}_\ell) - M_k(\mathbf{y}_\ell)| = 0 \qquad (\textit{Case 1})$$

$$\text{or} \qquad \lim_{\ell\to\infty} |M_{k,\ell}(\mathbf{y}_\ell) - M_{k,\ell}(\mathbf{y})| = 0. \qquad (\textit{Case 2})$$

Then we have pointwise

$$\lim_{\ell\to\infty} \bigotimes \mathbf{M}_\ell = \bigotimes \mathbf{M}.$$

***Remark* 3.6.4.** Note that Case 1 strengthens the condition that $M_{k,\ell}$ converges pointwise to M_k, while Case 2 can be seen as "asymptotic continuity of $M_{k,\ell}$ as $\ell \to \infty$".

If the convergence of $M_{k,\ell}$ to M_k is locally uniform, then clearly Case 1 is satisfied.

Proof. Let $\mathbf{x} \in A^n$. Our proof consists of three parts:
1. We show: If $\lim_{\ell\to\infty} \mathbf{y}_\ell = \mathbf{y}$ for some $\mathbf{y}, \mathbf{y}_1, \mathbf{y}_2, \dots \in A^n$, then $\lim_{\ell\to\infty} \mathbf{M}_\ell(\mathbf{y}_\ell) = \mathbf{M}(\mathbf{y})$.
2. We show by induction on m that

$$\lim_{\ell\to\infty} \mathbf{M}_\ell^{\circ m}(\mathbf{x}) = \mathbf{M}^{\circ m}(\mathbf{x}) \quad \text{for all } m \in \mathbb{N}. \tag{3.9}$$

3. From (3.9) we derive

$$\lim_{\ell\to\infty} \bigotimes \mathbf{M}_\ell(\mathbf{x}) = \bigotimes \mathbf{M}(\mathbf{x}). \tag{3.10}$$

Part 1. For all $k \in \mathbb{N}_{\leq n}$, we have

$$|M_{k,\ell}(\mathbf{y}_\ell) - M_k(\mathbf{y})| \leq |M_{k,\ell}(\mathbf{y}_\ell) - M_k(\mathbf{y}_\ell)| + |M_k(\mathbf{y}_\ell) - M_k(\mathbf{y})|, \tag{3.11}$$

$$|M_{k,\ell}(\mathbf{y}_\ell) - M_k(\mathbf{y})| \leq |M_{k,\ell}(\mathbf{y}_\ell) - M_{k,\ell}(\mathbf{y})| + |M_{k,\ell}(\mathbf{y}) - M_k(\mathbf{y})|. \tag{3.12}$$

As $\ell \to \infty$, then the rightmost term of (3.11) approaches 0, by continuity of M_k, and the rightmost term of (3.12) approaches 0, by pointwise convergence of $\mathbf{M}_\ell$ to $\mathbf{M}$. Moreover, the middle terms of (3.11) and (3.12) approach 0 in Case 1 and 2 respectively. Hence, we have for all $k \in \mathbb{N}_{\leq n}$ that $\lim_{\ell\to\infty} M_{k,\ell}(\mathbf{y}_\ell) = M_k(\mathbf{y})$. Hence, $\lim_{\ell\to\infty} \mathbf{M}_\ell(\mathbf{y}_\ell) = \mathbf{M}(\mathbf{y})$.

Part 2. For $m = 1$, (3.9) is true by assumption. Suppose that (3.9) is true for a particular $m \geq 1$. Then

$$\lim_{\ell\to\infty} \mathbf{M}_\ell^{\circ(m+1)}(\mathbf{x}) = \lim_{\ell\to\infty} \mathbf{M}_\ell(\mathbf{M}_\ell^{\circ m}(\mathbf{x})) = \mathbf{M}(\mathbf{M}^{\circ m}(\mathbf{x})) = \mathbf{M}^{\circ(m+1)}(\mathbf{x});$$

the middle equality follows from part 1. Thus, (3.9) is true for $m+1$.

Part 3. We have

$$
\begin{aligned}
&|\bigotimes \mathbf{M}_\ell(\mathbf{x}) - \bigotimes \mathbf{M}(\mathbf{x})| = \\
&|\bigotimes \mathbf{M}_\ell(\mathbf{x}) - \mathbf{M}_\ell^{\circ m}(\mathbf{x})| + |\mathbf{M}_\ell^{\circ m}(\mathbf{x}) - \mathbf{M}^{\circ m}(\mathbf{x})| + |\mathbf{M}^{\circ m}(\mathbf{x}) - \bigotimes \mathbf{M}(\mathbf{x})|.
\end{aligned} \tag{3.13}
$$

As $\ell \to \infty$, the first and last term of (3.13) approach 0 by definition of Gauss composition, and the middle term approaches 0 by part 2. This completes the proof of (3.10). □

Theorem 3.6.5 (*)**.** *Let $A \subseteq \mathbb{R}$ and $n \in \mathbb{N}$.*
Let T be a topological space, and let

$$T \to \mathbb{M}_{\mathrm{comp}(A^n)} : \quad \tau \mapsto M_\tau$$

be a continuum of means on A^n (see Definition 2.6.1). Then the map

$$T^n \to \mathbb{M}_{\mathrm{comp}(A^n)} : \quad (\tau_1, \dots, \tau_n) \mapsto \bigotimes_{k=1}^{n} M_{\tau_k}$$

is a continuum of means on A^n. ⅃

Proof. We write

$$G : \; T \times A^n : \; (\tau, \mathbf{x}) \mapsto M_\tau(\mathbf{x}), \qquad \mathcal{G} : \; T^n \times A^n : \; \big((\tau_1, \dots, \tau_n), \mathbf{x}\big) \mapsto \bigotimes_{k=1}^{n} M_{\tau_k}(\mathbf{x}).$$

So we have that G is continuous, and we must show that $\mathcal{G}$ is continuous.

Let $\mathbf{z} := \big((\tau_1, \dots, \tau_n), \mathbf{x}\big) \in T^n \times A^n$, and for each $\ell \in \mathbb{N}$, let

$$\mathbf{z}_\ell := \big((\tau_{1,\ell}, \dots, \tau_{1,\ell}), \mathbf{x}_\ell\big) \in T^n \times A^n \quad \text{such that} \quad \lim_{\ell \to \infty} \mathbf{z}_\ell = \mathbf{z}.$$

We have

$$|\mathcal{G}(\mathbf{z}_\ell) - \mathcal{G}(\mathbf{z})| = \left| \bigotimes_{k=1}^{n} M_{\tau_{k,\ell}}(\mathbf{x}_\ell) - \bigotimes_{k=1}^{n} M_{\tau_k}(\mathbf{x}_\ell) \right| + \left| \bigotimes_{k=1}^{n} M_{\tau_k}(\mathbf{x}_\ell) - \bigotimes_{k=1}^{n} M_{\tau_k}(\mathbf{x}) \right|. \tag{3.14}$$

By continuity of G, the means M_τ are continuous, for all $\tau \in T$. Therefore, the mean $\bigotimes_{k=1}^{n} M_{\tau_k}$ is continuous (by Theorem 3.2.4). So the rightmost term in (3.14) converges to 0 as $\ell \to \infty$. Further, from continuity of G it easily follows that the conditions of Theorem 3.6.3 are satisfied, with M_{τ_k} substituted for M_k, and with $M_{\tau_{k,\ell}}$ substituted for $M_{k,\ell}$. It is namely easily seen that continuity of G implies the condition of Case 1.[2] Hence, Theorem 3.6.3 tells us that the middle term in (3.14) converges to 0 as $\ell \to \infty$. Thus, we conclude that

$$\lim_{\ell \to \infty} |\mathcal{G}(\mathbf{z}_\ell) - \mathcal{G}(\mathbf{z})| = 0,$$

so $\mathcal{G}$ is continuous in $\mathbf{z}$. As $\mathbf{z}$ is arbitrary, $\mathcal{G}$ is continuous. □

[2]That is however a detour, because we need Case 1 only to conclude that the left-hand side of (3.11) converges to 0 as $\ell \to \infty$, which is already clear from continuity of G.

3.7 Composite power and translation means

The remaining sections 3.7–3.9 of this chapter are a bit less abstract and more concrete than the previous sections: we mainly apply the theory that we built so far to show some interesting aspects of some concrete Gauss composites. Actually, all the concrete Gauss composites that we consider, are particular cases of the next definition; this is not strange, as the power means and translation means are virtually the only means about which we derived enough "nice" properties to produce Gauss composites of which we know that they have "nice" properties. The aim of this section is to derive those "nice" properties.

Definition 3.7.1 (*)**.** Let $n \in \mathbb{N}$. We define the *composite translation mean map* $\mathcal{TM}_n$ and the *composite power mean map* $\mathcal{PM}_n$ by

$$\mathcal{TM}_n : \overline{\mathbb{R}}^n \times \mathbb{R}^{n\times n}_{>0} \times \mathbb{R}^n_{>0} \to \mathbb{R}_{>0} : \ \big((b_1, \dots, b_n), (\mathbf{w}_1, \dots, \mathbf{w}_n), \mathbf{x}\big) \mapsto \bigotimes\nolimits_{k=1}^n \mathrm{TM}_{b_k,\mathbf{w}_k}(\mathbf{x}),$$

$$\mathcal{PM}_n : \mathbb{R}^n \times \mathbb{R}^{n\times n}_{>0} \times \mathbb{R}^n_{>0} \to \mathbb{R}_{>0} : \ \big((p_1, \dots, p_n), (\mathbf{w}_1, \dots, \mathbf{w}_n), \mathbf{x}\big) \mapsto \bigotimes\nolimits_{k=1}^n \mathrm{PM}_{p_1,\mathbf{w}_k}(\mathbf{x}).$$

For $\mathbf{t} \in \overline{\mathbb{R}}^n \times \mathbb{R}^{n\times n}_{>0}$, we denote by $\mathcal{TM}_{\mathbf{t}}$ the mean on $\mathbb{R}^n_{>0}$ given by $\mathcal{TM}_{\mathbf{t}}(\mathbf{x}) := \mathcal{TM}_n(\mathbf{t}, \mathbf{x})$. For $\mathbf{t} \in \mathbb{R}^n \times \mathbb{R}^{n\times n}_{>0}$, we denote by $\mathcal{PM}_{\mathbf{t}}$ the mean on $\mathbb{R}^n_{>0}$ given by $\mathcal{PM}_{\mathbf{t}}(\mathbf{x}) := \mathcal{PM}_n(\mathbf{t}, \mathbf{x})$.

***Remark* 3.7.2.** Although the map

$$(\overline{\mathbb{R}} \times \mathbb{R}^n_{>0}) \times \mathbb{R}^n_{>0} \to \mathbb{R}_{>0} : \quad ((p, \mathbf{w}), \mathbf{x}) \mapsto \mathrm{PM}_{p,\mathbf{w}}(\mathbf{x})$$

is continuous (by Theorem 2.6.6.6) we can't apply Theorem 3.6.5 to define $\mathcal{PM}_n$ on $(\overline{\mathbb{R}}^n \times \mathbb{R}^{n\times n}_{>0}) \times \mathbb{R}^n_{>0}$, because the means $\mathrm{PM}_{\infty,\mathbf{w}} = \mathrm{Max}_n$ and $\mathrm{PM}_{-\infty,\mathbf{w}} = \mathrm{Min}_n$ are not compressing. 𝄢

Example 3.7.3. Let $\mathbf{W} := (\mathbf{w}_1, \mathbf{w}_2, \mathbf{1}, \mathbf{w}_2, \mathbf{w}_1) \in \mathbb{R}^{5\times 5}_{>0}$. We have

$$\begin{aligned} M_1 &:= \mathcal{TM}_{((-\infty,-2,0,2,\infty),\mathbf{W})} = \mathrm{HM}_{\mathbf{w}_1} \otimes \mathrm{TM}_{-2,\mathbf{w}_2} \otimes \mathrm{GM}_5 \otimes \mathrm{TM}_{2,\mathbf{w}_2} \otimes \mathrm{AM}_{\mathbf{w}_1}, \\ M_2 &:= \mathcal{PM}_{((-1,-2,0,2,1),\mathbf{W})} = \mathrm{HM}_{\mathbf{w}_1} \otimes \mathrm{PM}_{-2,\mathbf{w}_2} \otimes \mathrm{GM}_5 \otimes \mathrm{PM}_{2,\mathbf{w}_2} \otimes \mathrm{AM}_{\mathbf{w}_1}. \end{aligned}$$

By Theorem 3.2.4, M_1 and M_2 are increasing, continuous, compressing means on $\mathbb{R}^n_{>0}$, and M_2 is scale-invariant. If $\mathbf{w}_1 = \mathbf{w}_2 = \mathbf{1}$, then M_1 and M_2 are symmetric.

Moreover, Theorem 3.5.3 tells us that $M_1^{[\mathrm{inv}]} = M_1$ and $M_2^{[\mathrm{inv}]} = M_2$. ♪

Theorem 3.7.4 ((*) Composite power and translation means)**.** *Let $n \in \mathbb{N}$. Let "$\mathcal{XM}, \mathrm{XM}, P$" denote either "$\mathcal{PM}, \mathrm{PM}, \mathbb{R}$" or "$\mathcal{TM}, \mathrm{TM}, \overline{\mathbb{R}}$" (arbitrary but fixed). Let $W := \mathbb{R}^n_{>0}$, let $\mathbf{x} \in \mathbb{R}^n_{>0}$, let $(\mathbf{p}, \mathbf{W}) := \big((p_1, \dots, p_n), (\mathbf{w}_1, \dots, \mathbf{w}_n)\big) \in P^n \times W^n$. So by definition we have*

$$\mathcal{XM}_{(\mathbf{p},\mathbf{W})} = \bigotimes_{k=1}^n \mathrm{XM}_{p_k,\mathbf{w}_k}.$$

1. *"Mean": $\mathcal{XM}_{(\mathbf{p},\mathbf{W})}$ is an increasing, continuous, compressing mean on $\mathbb{R}^n_{>0}$. Moreover, $\mathcal{PM}_{(\mathbf{p},\mathbf{W})}$ is scale-invariant.*

2. *"Continuum of means": The map* $\mathcal{XM}_n : P^n \times W^n \times \mathbb{R}^n_{>0} \to \mathbb{R}_{>0}$ *is continuous. In other words, the map* $P^n \times W^n \to \mathbb{M}_{\mathrm{comp}(\mathbb{R}^n_{>0})} \ : \ (\mathbf{p}, \mathbf{W}) \mapsto \mathcal{XM}_{(\mathbf{p},\mathbf{W})}$ *is a continuum of means.*

3. *"Increasing in* $\mathbf{p}$*": The map* $\varphi : P^n \times \mathbb{R}^n_{>0} \to \mathbb{R}_{>0} : (\mathbf{p}, \mathbf{x}) \mapsto \mathcal{XM}_{(\mathbf{p},\mathbf{W})}(\mathbf{x})$ *is increasing.*

4. *"Diagonally": If* $\mathbf{p} = \mathrm{diag}_n(p)$ *and* $\mathbf{W} = \mathrm{diag}_n(\mathbf{w})$*, then* $\mathcal{XM}_{(\mathbf{p},\mathbf{W})} = \mathrm{XM}_{p,\mathbf{w}}$.

5. *"Scale-invariant in* $\mathbf{W}$*": Let* $(\lambda_1, \dots, \lambda_n) \in \mathbb{R}^n_{>0}$*, and let* $\mathbf{W}' := (\lambda_1 \mathbf{w}_1, \dots, \lambda_n \mathbf{w}_n)$*. Then* $\mathcal{XM}_{(\mathbf{p},\mathbf{W}')} = \mathcal{XM}_{(\mathbf{p},\mathbf{W})}$.

6. *"Symmetry": Suppose* $\mathbf{W} \in \mathrm{diag}(\mathbb{R}^{n\times n}_{>0})$*. Then* $\mathcal{XM}_{(\mathbf{p},\mathbf{W})}$ *is a symmetric mean. Moreover, for all* $\sigma \in S_n$ *we have* $\mathcal{XM}_{(\mathbf{p}\circ\sigma,\mathbf{W})} = \mathcal{XM}_{(\mathbf{p},\mathbf{W})}$.

7. *"Kinship": For any* $q \in \mathbb{R}_{\neq 0}$*, we have* $\mathcal{PM}^{[\mathrm{pow}_q]}_{(\mathbf{p},\mathbf{W})} = \mathcal{PM}_{(q^{-1}\mathbf{p},\mathbf{W})}$.
If $b \in \mathbb{R}$ *such that* $p_k \geq b \geq 0$ *for all* k*, then* $\mathcal{TM}^{[\mathrm{tra}_b]}_{(\mathbf{p},\mathbf{W})} = \mathcal{TM}_{(-b+\mathbf{p},\mathbf{W})}$ *on* $\mathbb{R}^n_{>b}$.

8. *"Duality":* $\mathcal{XM}^{[\mathrm{inv}]}_{(\mathbf{p},\mathbf{W})} = \mathcal{XM}_{(-\mathbf{p},\mathbf{W})}$.

9. *"Scaling of* $\mathcal{TM}$*": Let* $\lambda \in \mathbb{R}_{>0}$*. Let* $\lambda * \mathbf{p} := (\lambda_1 p_1, \dots, \lambda_n p_n)$*, where* $\lambda_k := \lambda$ *if* $p_k \geq 0$ *and* $\lambda_k := \lambda^{-1}$ *if* $p_k < 0$*.*
Then $\lambda \mathcal{TM}_{(\mathbf{p},\mathbf{W})}(\mathbf{x}) = \mathcal{TM}_{(\lambda*\mathbf{p},\mathbf{W})}(\lambda \mathbf{x})$.

10. *"Functional equation":* $\mathcal{XM}_{(\mathbf{p},\mathbf{W})}(\mathbf{x}) = \mathcal{XM}_{(\mathbf{p},\mathbf{W})}\big(\mathrm{XM}_{p_1,\mathbf{w}_1}(\mathbf{x}), \dots, \mathrm{XM}_{p_n,\mathbf{w}_n}(\mathbf{x})\big)$.

Proof.

1. Follows directly from the theorem on Gauss composition (3.2.4), using the corresponding results about power means and translation means (Theorems 2.6.6 and 2.7.4).

2. Follows directly by applying Theorem 3.6.5 to the result that power means and translation means form continuums (Theorems 2.6.6 and 2.7.4).

3. That φ is increasing in $\mathbf{x}$ is covered by part 1. So it suffices to show that φ is increasing in $\mathbf{p}$, that is, that the map $\psi : P^n \to \mathbb{R}_{>0} : \mathbf{p} \mapsto \mathcal{XM}_{(\mathbf{p},\mathbf{W})}(\mathbf{x})$ is increasing, for all $\mathbf{x}$. That ψ is increasing, follows directly from Proposition 3.4.1.2, using (by 2.6.6 and 2.7.4) that if $p_k \leq p'_k$, then $\mathrm{XM}_{p_k,\mathbf{w}_k} \leq \mathrm{XM}_{p'_k,\mathbf{w}_k}$.

4. Follows directly from Proposition 3.4.1.3.

5. Follows directly from the fact that $\mathrm{XM}_{p_k,\,\lambda_k\mathbf{w}_k} = \mathrm{XM}_{p_k,\mathbf{w}_k}$.

6. Each of the means $\mathrm{XM}_{p_k,\mathbf{w}_k}$ is symmetric (by 2.6.6 and 2.7.4), hence the mean $\mathcal{XM}_{(\mathbf{p},\mathbf{W})}$ is symmetric (by 3.2.4). It follows (by 3.4.1) that $\mathcal{XM}_{(\mathbf{p}\circ\sigma,\mathbf{W})} = \mathcal{XM}_{(\mathbf{p},\mathbf{W})}$.

7. We have (by 2.6.6) that $\mathrm{PM}^{[\mathrm{pow}_q]}_{p_k,\mathbf{w}_k} = \mathrm{PM}_{q^{-1}p_k,\,\mathbf{w}_k}$, for all k. Hence (by 3.5.3), $\mathcal{PM}^{[\mathrm{pow}_q]}_{(\mathbf{p},\mathbf{W})} = \mathcal{PM}_{(q^{-1}\mathbf{p},\mathbf{W})}$.

Let $b \in \mathbb{R}$ and suppose that $p_k \geq b \geq 0$ for all k. Because for all $c \geq 0$ we have $\mathrm{TM}_{c,\mathbf{w}_k} = \mathrm{GM}_{\mathbf{w}_k}^{[\mathrm{tra}_{-c}]}$, it follows that $\mathrm{TM}_{p_k,\mathbf{w}_k}^{[\mathrm{tra}_b]} = \mathrm{TM}_{-b+p_k,\mathbf{w}_k}$ on $\mathbb{R}^n_{>b}$. Hence (by 3.5.3), $\mathcal{TM}_{(\mathbf{p},\mathbf{W})}^{[\mathrm{tra}_b]} = \mathcal{TM}_{(-b+\mathbf{p},\mathbf{W})}$.

8. This is the special case $q = -1$ of part 7.

9. By Theorem 2.7.4, we have $\lambda\mathrm{TM}_{(p_k,\mathbf{w}_k)}(\mathbf{x}) = \mathrm{TM}_{(\lambda_k p_k,\mathbf{w}_k)}(\lambda\mathbf{x})$. Let

$$\mathbf{M} := (\mathrm{TM}_{(p_1,\mathbf{w}_1)}, \dots, \mathrm{TM}_{(p_n,\mathbf{w}_n)}), \quad \mathbf{M}_\lambda := (\mathrm{TM}_{(\lambda_1 p_1,\mathbf{w}_1)}, \dots, \mathrm{TM}_{(\lambda_n p_n,\mathbf{w}_n)}),$$

so we have $\mathbf{M} \in \mathbb{M}_{\mathrm{comp}(\mathbb{R}^n_{>0})}$ and $\mathbf{M}_\lambda \in \mathbb{M}_{\mathrm{comp}(\mathbb{R}^n_{>0})}$, and $\lambda\mathbf{M}(\mathbf{x}) = \mathbf{M}_\lambda(\lambda\mathbf{x})$.

It follows by induction on $m \geq 1$ that

$$\lambda\mathbf{M}^{\circ m}(\mathbf{x}) = \mathbf{M}_\lambda^{\circ m}(\lambda\mathbf{x}) : \tag{3.15}$$

we just noted it for $m = 1$, and if it is true for a certain $m \geq 1$, then

$$\lambda\mathbf{M}^{\circ(m+1)}(\mathbf{x}) = \lambda\mathbf{M}\big(\mathbf{M}^{\circ m}(\mathbf{x})\big) = \mathbf{M}_\lambda\big(\lambda\mathbf{M}^{\circ m}(\mathbf{x})\big) = \mathbf{M}_\lambda\big(\mathbf{M}_\lambda^{\circ m}(\lambda\mathbf{x})\big) = \mathbf{M}_\lambda^{\circ(m+1)}(\lambda\mathbf{x}).$$

Letting $m \to \infty$ in (3.15), yields the desired result that $\lambda\mathcal{TM}_{(\mathbf{p},\mathbf{W})}(\mathbf{x}) = \mathcal{TM}_{(\lambda*\mathbf{p},\mathbf{W})}(\lambda\mathbf{x})$.

10. Follows directly from Theorem 3.2.4. □

***Remark* 3.7.5.** *Suppose that Conjecture 3.2.14 is true.*

In addition to part 1 of Theorem 3.7.4, it follows that the means $\mathcal{XM}_{(\mathbf{p},\mathbf{W})}$ are strictly increasing and smooth. Moreover, it follows that Conjecture 3.4.2 is true. Thus, it follows in addition to part 3 of Theorem 3.7.4 that for all $\mathbf{x} \in \mathbb{R}^n_{>0} \setminus \mathrm{diag}(\mathbb{R}^n_{>0})$, the map $\mathbf{p} \mapsto \mathcal{XM}_{(\mathbf{p},\mathbf{W})}(\mathbf{x})$ is strictly increasing. ⫲

***Remark* 3.7.6.** In fact, we not just know that $\mathcal{XM}_{(\mathbf{p},\mathbf{W})}$ is compressing, but for all $a, b \in \mathbb{R}_{>0}$ with $a < b$, we can easily compute $c(a,b)$ such that $\mathcal{XM}_{(\mathbf{p},\mathbf{W})}$ is $[c(a,b)]$-compressing on the interval $[a,b]$. Namely, if $\mathrm{XM}_{p_k,\mathbf{w}_k}$ is $[c_k(a,b)]$-compressing, then (by Theorem 3.2.4 and Lemma 1.1.28) we can take $c(a,b) = \mathrm{Min}(c_1(a,b), \dots, c_n(a,b))$. Explicit expressions for $c_k(a,b)$ are given in Theorem 2.6.6.5 for the power means and in Theorem 2.7.4.4 for the translation means. ⫲

3.8 Gauss composites that are or aren't quasi-arithmetic

In this section, we apply the results about the means $\mathcal{PM}_{(\mathbf{p},\mathbf{W})}$ and $\mathcal{TM}_{(\mathbf{p},\mathbf{W})}$ that we derived in the previous section, to show that some of them are quasi-arithmetic, and that certain others of them are not quasi-arithmetic. The basic technique for this is to test whether a quasi-arithmetic mean could satisfy the functional equation of the other mean; this is made explicit by Proposition 3.8.1. In the case of $\mathcal{PM}_{(\mathbf{p},\mathbf{1})}$, we don't have to test all quasi-arithmetic means but just the power means; we use that in Theorem 3.8.6.

Proposition 3.8.1 (†)**.** *[Functional equation for quasi-arithmetic Gauss composites]*
Let $n \in \mathbb{N}$*, let* $A, B \subseteq \mathbb{R}$*, let* $\mathbf{M} = (M_1, \ldots, M_n) \in \mathbb{M}^n_{\mathrm{comp}(A^n)}$*,*
let $\mathbf{w} = (w_1, \ldots, w_n) \in \mathbb{R}^n_{>0}$*.*
Let $f : A \to B$ *be a monotonic bijection such that* $f^{-1} : B \to A$ *is continuous.*
Suppose that the arithmetic mean $\mathrm{AM}_{\mathbf{w}}$ *maps* B^n *into* B*;*
we regard $\mathrm{AM}_{\mathbf{w}}$ *as an internal mean on* B^n*.*

The following two statements are equivalent:

1. $\bigotimes \mathbf{M} = \mathrm{AM}_{\mathbf{w}}^{[f^{-1}]}$.
2. $\sum_{k=1}^n w_k f(x_k) = \sum_{k=1}^n w_k f(M_k(\mathbf{x}))$, *for all* $\mathbf{x} := (x_1, \ldots, x_n) \in A^n$.

Proof. We show that the statements 1–4 are equivalent, where

3. $\mathrm{AM}_{\mathbf{w}}^{[f^{-1}]} = \mathrm{AM}_{\mathbf{w}}^{[f^{-1}]} \circ \mathbf{M}$.
4. $f^{-1}\Big(\dfrac{\sum_{k=1}^n w_k f(x_k)}{\sum \mathbf{w}}\Big) = f^{-1}\Big(\dfrac{\sum_{k=1}^n w_k f(M_k(\mathbf{x}))}{\sum \mathbf{w}}\Big)$, for all $\mathbf{x} := (x_1, \ldots, x_n) \in A^n$.

"1 $\iff$ *3"* is true by Theorem 3.2.4.8, because $\mathrm{AM}_{\mathbf{w}}^{[f^{-1}]}$ is a continuous mean on A^n, because f^{-1} is continuous.
"3 $\iff$ *4"* is true because 4 is simply an explicit rewording of 3.
"4 $\iff$ *2"* follows because f^{-1} is an injection. □

The next definition is just a restriction of $\mathcal{PM}_n$ that we studied in §3.7, but we give it a name of its own because it turns out that we can prove several interesting aspects about them: Proposition 3.8.4 and Theorem 3.8.6.

Definition 3.8.2 (*)**.** Let $n \in \mathbb{N}$, let $\mathbf{w} \in \mathbb{R}^n_{>0}$. For $p \in \mathbb{R}$, let $\mathbf{p}_n(p) \in \mathbb{R}^n$ be defined as

$$\mathbf{p}_n(p) := \begin{cases} (-p, \ldots, -p, p, \ldots, p) & \text{if } n \text{ is even,} \\ (-p, \ldots, -p, 0, p, \ldots, p) & \text{if } n \text{ is odd,} \end{cases}$$

so that $\mathbf{p}(-p) = -\mathbf{p}(p)$ for all p. We define the *p-symmetric composite power mean*

$$\mathcal{PMS}_{p,\mathbf{w}} := \mathcal{PM}_{(\mathbf{p}_n(p),\ \mathrm{diag}_n(\mathbf{w}))},$$

where the right-hand side is the composite power mean as in Definition 3.7.1.

In the case that $\mathbf{w} = \mathrm{diag}_n(1)$, we simply write $\mathcal{PMS}_{p,n}$ instead of $\mathcal{PMS}_{p,\mathbf{w}}$.

***Remark* 3.8.3.** More explicitly, we have

$$\mathcal{PMS}_{p,\mathbf{w}} = \begin{cases} \mathrm{PM}_{-p,\mathbf{w}} \otimes \cdots \otimes \mathrm{PM}_{-p,\mathbf{w}} \otimes \mathrm{PM}_{p,\mathbf{w}} \otimes \cdots \otimes \mathrm{PM}_{p,\mathbf{w}} & \text{if } n \text{ is even,} \\ \mathrm{PM}_{-p,\mathbf{w}} \otimes \cdots \otimes \mathrm{PM}_{-p,\mathbf{w}} \otimes \mathrm{GM}_{\mathbf{w}} \otimes \mathrm{PM}_{p,\mathbf{w}} \otimes \cdots \otimes \mathrm{PM}_{p,\mathbf{w}} & \text{if } n \text{ is odd.} \end{cases}$$

We study $\mathcal{PMS}_{p,\mathbf{w}}$ for $n = 2$ in Proposition 3.8.4, and for all $n \geq 2$ but with $\mathbf{w}$ restricted to $\mathrm{diag}_n(\mathbb{R}^n_{>0})$ in Theorem 3.8.6.

Proposition 3.8.4. (*).[3] *Let $p \in \mathbb{R}$, let $\mathbf{w} = (v, w) \in \mathbb{R}^2_{>0}$. We have*

$$(\mathrm{PM}_{-p,\mathbf{w}} \otimes \mathrm{PM}_{p,\mathbf{w}} = \mathrm{GM}_{\mathbf{w}}) \iff (v = w \ \text{ or } \ p = 0). \tag{3.16}$$

In particular, we have

$$\mathrm{AM}_{\mathbf{w}} \otimes \mathrm{HM}_{\mathbf{w}} = \mathrm{GM}_{\mathbf{w}} \iff \mathrm{HM}_{\mathbf{w}} \otimes \mathrm{AM}_{\mathbf{w}} = \mathrm{GM}_{\mathbf{w}} \iff \mathbf{w} \in \mathrm{diag}(\mathbb{R}^2_{>0}). \tag{3.17}$$

Proof. Clearly, (3.17) consists of the special cases $p = -1$ and $p = 1$ of (3.16). Further, (3.16) is trivially true for $p = 0$. Thus, it suffices to prove (3.16), supposing that $p \neq 0$.

Because $\mathrm{GM}_{\mathbf{w}} = \mathrm{AM}_{\mathbf{w}}^{[f^{-1}]}$ where $f := \log : \mathbb{R}_{>0} \to \mathbb{R}$, Proposition 3.8.1 tells us that the following statements are equivalent:

$$\mathrm{PM}_{-p,\mathbf{w}} \otimes \mathrm{PM}_{p,\mathbf{w}} = \mathrm{GM}_{\mathbf{w}}. \tag{3.18}$$

$$v \log(x) + w \log(y) = v \log\left((vx^p + wy^p)^{1/p}\right) + w \log\left((vx^{-p} + wy^{-p})^{-1/p}\right) \quad \text{for all } (x, y) \in \mathbb{R}^2_{>0}. \tag{3.19}$$

Applying to both sides of (3.19) the injection $\mathrm{pow}_p \circ \exp$ (it is an injection because $p \neq 0$), it follows that (3.18) is equivalent to

$$x^{vp} y^{wp} = \frac{(vx^p + wy^p)^v}{(vx^{-p} + wy^{-p})^w} \quad \text{for all } (x, y) \in \mathbb{R}^2_{>0}. \tag{3.20}$$

If (3.20) is true for $x = 1 = y$, then $(v + w)^v = (v + w)^w$, hence $v = w$.

Conversely, if $v = w$, then applying the injection $\mathrm{pow}_{1/v}$ to (3.20) results in $x^p y^p = \dfrac{x^p + y^p}{x^{-p} + y^{-p}}$, which is true for all $(x, y) \in \mathbb{R}^2_{>0}$. $\square$

Example 3.8.5. Proposition 3.8.4 tells us that $\mathrm{PM}_{p,2} \otimes \mathrm{PM}_{-p,2} = \mathrm{GM}_2$, for all $p \in \mathbb{R}$. We write out more explicitly what this means. Let $(a, b) \in \mathbb{R}^2_{>0}$, and let the sequence $((a_m, b_m))_{m \in \mathbb{N}_0}$ in $\mathbb{R}^2_{>0}$ be recursively defined by

$$(a_{m+1}, b_{m+1}) = \left(\left(\frac{a_m^p + b_m^p}{2}\right)^{1/p}, \left(\frac{2a_m^p b_m^p}{a_m^p + b_m^p}\right)^{1/p}\right), \qquad (a_0, b_0) = (a, b).$$

Then $\lim_{m\to\infty} a_m = \sqrt{ab} = \lim_{m\to\infty} b_m$.

In particular, taking $p = 1$, we get a sequence $((a_m, b_m))_m$ in $\mathbb{Q}^2_{>0}$ that rapidly converges to $(\sqrt{ab}, \sqrt{ab})$. ♪

The most interesting aspect of the next theorem is (in our eyes) that for $n \geq 3$, it shows the existence of a continuum of distinct means that are not quasi-arithmetic, but that nevertheless have nice properties, and moreover the special property of self-duality.

[3] (*) applies mainly to "$\Longrightarrow$"; the other implication is contained in [Mat99, Proposition 1], but its prove is just the statement that the calculation can be done, whereas we show the calculation. Proposition 2 and 3 in the same paper provide some other examples of pairs of symmetric means M_r, M_{-r} such that $M_r \otimes M_{-r} = \mathrm{GM}_2$.

Theorem 3.8.6 ((*) Properties of $\mathcal{PMS}_{p,n}$).
Let $p \in \mathbb{R}$ and $n \in \mathbb{N}_{\geq 2}$. We write $\mathcal{PMS}_p := \mathcal{PMS}_{p,n}$.

1. *"Continuum of nice means": The map $\mathbb{R} \to \mathbb{M}_{\mathrm{comp}(\mathbb{R}^n_{>0})} : p \mapsto \mathcal{PMS}_p$ is a continuum of increasing, compressing, symmetric, scale-invariant means on $\mathbb{R}^n_{>0}$.*
2. *"Kinship": $\mathcal{PMS}_p^{[\mathrm{pow}_q]} = \mathcal{PMS}_{p/q}$, for all $q \in \mathbb{R}_{\neq 0}$.*
3. *"Self-duality": $\mathcal{PMS}_p^{[\mathrm{inv}]} = \mathcal{PMS}_p$.*
4. *"Generally non-geometric": If $n = 2$ or $p = 0$, then $\mathcal{PMS}_p = \mathrm{GM}_n$. However, if $n \geq 3$ and $p \neq 0$, then $\mathcal{PMS}_p \neq \mathrm{GM}_{\mathbf{w}}$, for all $\mathbf{w} \in \mathbb{R}^n_{>0}$.*
5. *"Generally non-quasi-arithmetic": $\mathcal{PMS}_p$ is quasi-arithmetic iff $n = 2$ or $p = 0$.*
6. *"Injective": Suppose $n \geq 3$. The map $p \mapsto \mathcal{PMS}_p : \mathbb{R} \to \mathbb{M}_{\mathrm{comp}(\mathbb{R}^n_{>0})}$ is injective.* ⌟

Proof. **1 & 2.** Follow directly by Theorem 3.7.4.

3. By part 2 (with $q = -1$) we have $\mathcal{PMS}_p^{[\mathrm{inv}]} = \mathcal{PMS}_{-p}$, and by part 6 ("symmetry") of Theorem 3.7.4, it follows that $\mathcal{PMS}_{-p} = \mathcal{PMS}_p$.

4. If $n = 2$, the statement follows from Proposition 3.8.4. If $p = 0$, we have $\mathcal{PMS}_p = \bigotimes_{k=1}^n \mathrm{GM}_n = \mathrm{GM}_n$.

Suppose $n \geq 3$ and $p \neq 0$. Because $\mathcal{PMS}_p$ is symmetric, while $\mathrm{GM}_{\mathbf{w}}$ is only symmetric for $\mathbf{w} \in \mathrm{diag}(\mathbb{R}^n_{>0})$, it suffices to show that the statement

$$\mathcal{PMS}_p = \mathrm{GM}_n \tag{3.21}$$

is false. By Theorem 3.2.4.8, and by the fact that

$$(\mathrm{PM}_{-p,n}, \mathrm{PM}_{0,n}, \mathrm{PM}_{p,n}) = (\mathrm{HM}_n, \mathrm{GM}_n, \mathrm{AM}_n)^{[\mathrm{pow}_{1/p}]},$$

it follows that the statement (3.21) is equivalent to

$$\mathrm{GM}_n = \mathrm{GM}_n \circ \mathbf{M}, \qquad \text{where} \tag{3.22}$$

$$\mathbf{M} := \mathbf{N}^{[\mathrm{pow}_{1/p}]}, \quad \mathbf{N} := \begin{cases} (\mathrm{HM}_n, \ldots, \mathrm{HM}_n, \mathrm{AM}_n, \ldots, \mathrm{AM}_n) & \text{if } n \text{ is even,} \\ (\mathrm{HM}_n, \ldots, \mathrm{HM}_n, \mathrm{GM}_n, \mathrm{AM}_n, \ldots, \mathrm{AM}_n) & \text{if } n \text{ is odd,} \end{cases}$$

where in each case there are as many occurrences of HM_n as of AM_n. Further, (3.22) is clearly equivalent to $\mathrm{GM}_n^{[\mathrm{pow}_p]} = (\mathrm{GM}_n \circ \mathbf{M})^{[\mathrm{pow}_p]}$. We have that

$$\mathrm{GM}_n^{[\mathrm{pow}_p]} = \mathrm{GM}_n, \qquad \text{and} \qquad (\mathrm{GM}_n \circ \mathbf{M})^{[\mathrm{pow}_p]} = \mathrm{GM}_n^{[\mathrm{pow}_p]} \circ \mathbf{M}^{[\mathrm{pow}_p]} = \mathrm{GM}_n \circ \mathbf{N}.$$

Thus, it follows that (3.21) is equivalent to $\mathrm{GM}_n = \mathrm{GM}_n \circ \mathbf{N}$, that is, to

$$\prod(x_1, \ldots, x_n) = \prod \mathbf{N}(x_1, \ldots, x_n) \qquad \forall (x_1, \ldots, x_n) \in \mathbb{R}^n_{>0}. \tag{3.23}$$

Case 1: n is even. Then (3.23) reduces to

$$x_1 \cdots x_n = \Big(\frac{x_1 + \cdots + x_n}{x_1^{-1} + \cdots + x_n^{-1}}\Big)^{n/2} \qquad \forall (x_1, \ldots, x_n) \in \mathbb{R}^n_{>0}, \tag{3.24}$$

which is false, *because* $n > 2$: for $(x_1, \ldots, x_n) \in \mathbb{Q}^n_{>0}$, the right-hand side of (3.24), in contrast to the left-hand side, is always an $(n/2)$th power in $\mathbb{Q}$, where $n/2 \in \mathbb{N}_{\geq 2}$.

Case 2: n is odd. Then (3.23) reduces to

$$x_1 \cdots x_n = \Big(\frac{x_1 + \cdots + x_n}{x_1^{-1} + \cdots + x_n^{-1}}\Big)^{\frac{n-1}{2}} (x_1 \cdots x_n)^{1/n} \qquad \forall (x_1, \ldots, x_n) \in \mathbb{R}^n_{>0}, \tag{3.25}$$

which is false: for $(x_1, \ldots, x_n) \in \mathbb{Q}^n_{>0}$, the factor $(x_1 \cdots x_n)^{1/n}$ is not always in $\mathbb{Q}$, while the other two factors are in $\mathbb{Q}$.

5. Because $\mathcal{PMS}_p$ is a symmetric and scale-invariant mean on $\mathbb{R}^n_{>0}$ (by part 1), it follows by Theorem 2.6.10 that $\mathcal{PMS}_p$ is quasi-arithmetic iff $\mathcal{PMS}_p = \mathrm{PM}_{q,n}$ for some $q \in \mathbb{R}$. Because $\mathcal{PMS}_p$ is self-dual (by part 3), while the only self-dual symmetric power mean is GM_n, it follows that $\mathcal{PMS}_p$ is quasi-arithmetic iff $\mathcal{PMS}_p = \mathrm{GM}_n$. By part 4, $\mathcal{PMS}_p = \mathrm{GM}_n$ iff $n = 2$ or $p = 0$.

6. Suppose that $\mathcal{PMS}_p = \mathcal{PMS}_{p'}$ for some $p, p' \in \mathbb{R}$ with $p \neq p'$. We aim to arrive at a contradiction. If p or p' equals 0, then it follows from part 5 that $p = 0 = p'$. So we have that p and p' are nonzero. Thus, either $|p/p'| > 1$ or $|p'/p| > 1$; we suppose without loss of generality that $|p/p'| > 1$. Let $q := p/p'$. By part 2, we have $\mathcal{PMS}_p^{[\mathrm{pow}_q]} = \mathcal{PMS}_{p'} = \mathcal{PMS}_p$. So by induction on $m \in \mathbb{N}$, we have

$$\mathcal{PMS}_p = \mathcal{PMS}_p^{[(\mathrm{pow}_q)^{\circ m}]} = \mathcal{PMS}_{p/q^m}.$$

It follows from Theorem 3.7.4.2 that the map

$$\mathbb{R} \times \mathbb{R}^n_{>0} \to \mathbb{R}_{>0} : \quad (p, \mathbf{x}) \mapsto \mathcal{PMS}_p(\mathbf{x})$$

is continuous. In particular, we have pointwise $\lim_{p \to 0} \mathcal{PMS}_p = \mathcal{PMS}_0$.
Because $\lim_{m\to\infty} p/q^m = 0$, it follows that pointwise

$$\mathcal{PMS}_p = \lim_{m\to\infty} \mathcal{PMS}_{p/q^m} = \mathcal{PMS}_0 = \mathrm{GM}_n.$$

By part 4, it follows that $p = 0$: a contradiction. □

Corollary 3.8.7 (*)**.** *Let $n \geq 3$. There are uncountably many means on $\mathbb{R}^n_{>0}$ that are increasing, symmetric, continuous, compressing, scale-invariant, non-quasi-arithmetic, and self-dual.* ⅃

We show in Proposition 3.9.1 that Corollary 3.8.7 is not true for $n = 2$, namely that there exists no such mean on $\mathbb{R}^2_{>0}$. However, dropping the self-dual property, we have a positive result:

Theorem 3.8.8. AGM *is an increasing, symmetric, continuous, compressing, scale-invariant, non-quasi-arithmetic mean on* $\mathbb{R}^2_{>0}$. *It is smooth on* $\mathbb{R}^2_{>0} \setminus \operatorname{diag}(\mathbb{R}^2_{>0})$. ♩

Proof. The statement about smoothness follows from the integral equation of Theorem 3.3.1, by the well-known "Leibniz rule for differentiating an integral"[4], because the integrand is a smooth function of a, b and φ.

The other statements, except for the "non-quasi-arithmetic" statement, follow from Theorem 3.2.4.

It remains to prove that AGM is not quasi-arithmetic. Suppose AGM is quasi-arithmetic. By Theorem 2.6.10, because AGM is symmetric and scale-invariant on $\mathbb{R}^n_{>0}$, it follows that AGM is a power mean $\mathrm{PM}_{p,2}$ for some $p \in \mathbb{R}$. It is easily seen that $\mathrm{AGM} \neq \mathrm{GM}_2$, for instance because $(\mathrm{AM}_2 \otimes \mathrm{GM}_2)(x, y) > \mathrm{GM}_2(x, y)$ if $x \neq y$, which follows from the fact that $\mathrm{AM}_2(x, y) > \mathrm{GM}_2(x, y)$. Hence, $\mathrm{AGM} = \mathrm{AM}_2^{[\mathrm{pow}_{1/p}]}$ for some $p \in \mathbb{R}_{\neq 0}$. From Proposition 3.8.1, there follows a functional equation for the function $f := \mathrm{pow}_p$:

$$f(x) + f(y) = f(\tfrac{x+y}{2}) + f(\sqrt{xy}) \qquad \forall (x, y) \in \mathbb{R}^2_{>0}. \tag{3.26}$$

That is, $x^p + y^p = (\frac{x+y}{2})^p + x^{p/2} y^{p/2}$. Evaluating in $y = 1$, we get

$$x^p + 1 = (\tfrac{x+1}{2})^p + x^{p/2} \qquad \forall (x, y) \in \mathbb{R}^2_{>0}. \tag{3.27}$$

Taking the limit $x \to 0$ if $p > 0$, and the limit $x \to \infty$ if $p > 0$, yields $1 = (1/2)^p$, so $p = 0$. But we noted already that $p \neq 0$. We conclude that AGM is not quasi-arithmetic. □

Alternative proof that AGM *is not quasi-arithmetic.* [DMP05, Theorem 2.1] shows something stronger, namely that the only functions $f : \mathbb{R}_{>0} \to \mathbb{R}$ that satisfy (3.26) are the constant functions. The authors derive this from a more general result about the functional equation $a(x + y) + b(xy) = c(x) + c(y)$ for $(x, y) \in \mathbb{R}^2_{>0}$. □

***Conjecture* 3.8.9** (*)**.** Let $n \in \mathbb{N}$, and let, as in Theorem 3.7.4, "$\mathcal{XM}, \mathrm{XM}, P$" denote either "$\mathcal{PM}, \mathrm{PM}, \mathbb{R}$" or "$\mathcal{TM}, \mathrm{TM}, \overline{\mathbb{R}}$" (arbitrary but fixed). Let $W := \mathbb{R}^n_{>0}$, let $\mathbf{x} \in \mathbb{R}^n_{>0}$, let $(\mathbf{p}, \mathbf{W}) := \big((p_1, \dots, p_n), (\mathbf{w}_1, \dots, \mathbf{w}_n)\big) \in P^n \times W^n$.

$$\mathcal{XM}_{(\mathbf{p},\mathbf{W})} \text{ is quasi-arithmetic} \iff \mathbf{p} \in \operatorname{diag}(P).$$ ♪

Remark about this conjecture. We show the implication "$\Longleftarrow$" in Corollary 4.7.8.

I expect that the special case of "$\Longrightarrow$" where $\mathcal{XM} = \mathcal{PM}$ and $\mathbf{W} = \mathbf{1}$ can be solved using the same strategy as in Theorem 3.8.6: that is, by using that if $\mathcal{PM}_{(\mathbf{p},\mathbf{1})}$ is quasi-arithmetic, then it must be a symmetric power mean $\mathrm{PM}_{q,n}$; and then showing that $\mathrm{PM}_{q,n}$ does not satisfies the functional equation of $\mathcal{PM}_{(\mathbf{p},\mathbf{1})}$. Namely, solving that equation results in an equality of complex analytic functions and hence in an equality of power series; this leads to infinitely many polynomial equations in the finitely many unknowns $q, p_1, \dots, p_n$. □

[4] Here we use only the special case $\frac{d}{dx}\left(\int_a^b f(x,t)\,dt\right) = \int_a^b \frac{\partial}{\partial x} f(x,t)\,dt$ of the rule, where f is a function of class C^1 on $\mathbb{R}^2_{>0}$. See for example [Fla73] for the full statement and a generalisation.

***Conjecture* 3.8.10.** Apart from the implication "$\Longleftarrow$" in Conjecture 3.8.9, there are other known examples of tuples of means (not all equal to each other) whose Gauss composition is quasi-arithmetic; see [Mat99, Proposition 4 & 5]. ♪

3.9 Self-dual means on $\mathbb{R}^2_{>0}$

We concluded in Corollary 3.8.7 that there are uncountably many means on $\mathbb{R}^n_{>0}$ that are symmetric, scale-invariant and self-dual (among several other properties), for all $n \geq 3$. The situation for $n = 2$ is quite different, as we show in

Proposition 3.9.1 (†)**.** *Let M be a symmetric, scale-invariant internal mean on $\mathbb{R}^2_{>0}$.*

1. *If $M^{[\mathrm{inv}]} = M$, then $M = \mathrm{GM}_2$.*
2. *$M \otimes M^{[\mathrm{inv}]} = \mathrm{GM}_2$.* ♩

Proof. **1.** Let $(x, y) \in \mathbb{R}^2_{>0}$. By symmetry and scale-invariance, and by $M^{[\mathrm{inv}]} = M$ respectively, we have

$$xM(1, x^{-1}y) = M(x, y) = yM(1, xy^{-1}), \quad \text{and} \quad M(1, x^{-1}y)M(1, xy^{-1}) = 1.$$

It follows that $M(x, y)^2 = xy$. So $M = \mathrm{GM}_2$.

2. M and $M^{[\mathrm{inv}]}$ are symmetric, scale-invariant internal means on $\mathbb{R}^2_{>0}$. Hence, $\mathcal{M} := M \otimes M^{[\mathrm{inv}]}$ is a symmetric, scale-invariant internal mean on $\mathbb{R}^2_{>0}$. Morevover, $\mathcal{M}^{[\mathrm{inv}]} = M^{[\mathrm{inv}]} \otimes M = \mathcal{M}$. Thus it follows from part 2 that $\mathcal{M} = \mathrm{GM}_2$.

3. Follows from part 2, because $\mathrm{PM}_{2,-b} = \mathrm{PM}^{[\mathrm{inv}]}_{2,b}$ (by Theorem 2.6.6), and because these power means are scale-invariant. □

***Remark* 3.9.2.** Proposition 3.9.1.2 yields another proof of the fact that
$\mathrm{PM}_{2,-b} \otimes \mathrm{PM}_{2,b} = \mathrm{GM}_2$, which we showed in Proposition 3.8.4 by a functional equation.

Thus, contrary to the situation for $n \geq 3$, for $n = 2$ we have that the geometric mean GM_n is the only symmetric, scale-invariant, self-dual mean on $\mathbb{R}^n_{>0}$. However, when we drop the 'scale-invariant' condition, this is no longer true:

Proposition 3.9.3 (*)**.** *Let $b \in \mathbb{R}$, let $\mathcal{M}_b := \mathcal{TM}_{((-b,b),\,(\mathbf{1},\mathbf{1}))} = \mathrm{TM}_{-b,2} \otimes \mathrm{TM}_{b,2}$.*

1. *$\mathcal{M}$ is symmetric, continuous, and self-dual, i.e. $\mathcal{M}^{[\mathrm{inv}]}_b = \mathcal{M}_b$.*
2. *Suppose $b \neq 0$. For all $\mathbf{w} \in \mathbb{R}^2_{>0}$, we have $\mathcal{M}_b \neq \mathrm{GM}_{\mathbf{w}}$.* ♩

Proof. 1. Follows from parts 2, 6 and 8 of Theorem 3.7.4.

2. Suppose that $b > 0$; this is without loss of generality, because $\mathcal{M}_b = \mathcal{M}_{-b}$. Suppose that $\mathcal{M}_b = \mathrm{GM}_{\mathbf{w}}$, for some $\mathbf{w} \in \mathbb{R}^2_{>0}$. Because $\mathcal{M}_b$ is symmetric, it follows that $\mathcal{M}_b = \mathrm{GM}_2$. By squaring the functional equation $\mathrm{GM}_2 = \mathrm{GM}_2 \circ (\mathrm{TM}_{-b,2}, \mathrm{TM}_{b,2})$, we get

$$xy = \frac{\sqrt{(x+b)(y+b)} - b}{\sqrt{(x^{-1}+b)(y^{-1}+b)} - b} \qquad \forall x, y \in \mathbb{R}_{>0}.$$

Substituting $y = 1$ and simplifying, we get

$$\sqrt{1+b}\big(x\sqrt{x^{-1}+b} - \sqrt{x+b}\big) = b(x-1) \qquad \forall x \in \mathbb{R}_{>0}. \tag{3.28}$$

It is easily seen that (3.28) is false: dividing both sides by x and letting $x \to \infty$, the left-hand side approaches $\sqrt{1+b}\sqrt{b}$, while the right-hand side approaches b. $\square$

Chapter 4

Families of means

In the previous three chapters, we saw many results about various kinds of means. The most important concrete examples that we considered, were the power means and translation means; almost all concrete means that we studied were examples of them, or Gauss composites of them.

The theory up to now is in an important sense not entirely satisfactory. We talked for instance about "the" translation means $\mathrm{TM}_{p,\mathbf{w}}$, but in fact we have a distinct mean for every pair $(p, \mathbf{w})$. How are they related to each other? We showed in Theorem 2.7.4 that $(p, \mathbf{w}) \mapsto \mathrm{TM}_{p,\mathbf{w}}$ is a continuum of means. But that is a separate result for each dimension n. How do they relate to each other across different dimensions?

This and similar questions are resolved in the current chapter: the map $(\mathbf{w}, \mathbf{x}) \mapsto \mathrm{TM}_{p,\mathbf{w}}(\mathbf{x})$ forms a *family of means*, for each p separately, but across the dimensions n. Whereas in the previous chapters, the weights $\mathbf{w}$ "were just there", they play a central role in this chapter; in particular, they relate to each other between different dimensions according to the "Coherence" property (in Definition 4.1.2).

The definition of "families of means" is our own. Consequently, virtually all the results in this chapter were formulated and proved by ourselves, and therefore the material is a bit more experimental in nature than that of the previous chapters.

4.1 Definition and first properties of families of means

Definition 4.1.1. We write $\bullet$ for the *concatenation operator on* $\bigsqcup_{n\in\mathbb{N}_0} \mathbb{R}^n$. That is, for $\mathbf{x} = (x_1, \ldots, x_m) \in \mathbb{R}^m$ and $\mathbf{y} = (y_1, \ldots, y_n) \in \mathbb{R}^n$, we have $\mathbf{x} \bullet \mathbf{y} = (x_1, \ldots, x_m, y_1, \ldots, y_n) \in \mathbb{R}^{m+n}$. If $m = 0$, then by convention $\mathbf{x} \bullet \mathbf{y} = \mathbf{y} \bullet \mathbf{x} = \mathbf{y}$. Moreover, for $k \in \mathbb{N}_0$, we write $\mathbf{x}^{\bullet k} := \mathbf{x} \bullet \cdots \bullet \mathbf{x} \in \mathbb{R}^{kn}$ for the concatenation of k copies of $\mathbf{x}$.

Our next definition, of families of means, might seem a bit peculiar, for instance because we only assume that $M_\mathbf{w}$ is a mean when $\mathbf{w} \in \mathbb{R}^2_{>0}$, and we don't assume that $M_\mathbf{w} = M_{\lambda\mathbf{w}}$ for $\lambda > 0$. The reason is that the former property follows from the definition, and the latter property follows from the definition if the means are increasing, as we will see. We want to make as little assumptions as possible, because we think it is more clarifying to show how further properties follow from the definition.

Definition 4.1.2 ((*) Family of means)**.** Let $A \subseteq \mathbb{R}$, let

$$M: \quad \bigsqcup_{n \in \mathbb{N}} \mathbb{R}^n_{>0} \times A^n \to A$$

be a function. For each $n \in \mathbb{N}$ and each $\mathbf{w} \in \mathbb{R}^n_{>0}$, let $M_{\mathbf{w}}$ be the function

$$M_{\mathbf{w}}: \quad A^n \to A: \quad \mathbf{x} \mapsto M(\mathbf{w}; \mathbf{x}).$$

We say that M is a *family of means on A* if it satisfies the following three axioms:

1. *"Means (for $n = 2$)":* $M_{(v,w)}$ is an internal mean on A^2, for all $(v, w) \in \mathbb{R}^2_{>0}$.
2. *"Dominance":* Let $x, y \in A$. Let $(x_k)_{k \in \mathbb{N}}$ be sequence in A with $\lim_{k\to\infty} x_k = x$. Let $(v_k)_{k \in \mathbb{N}}$ and $(w_k)_{k \in \mathbb{N}}$ be sequences in $\mathbb{R}_{>0}$, such that $\lim_{k\to\infty} v_k = \infty$, and $(w_k)_{k \in \mathbb{N}}$ is bounded from above. Then
$$\lim_{k\to\infty} M_{(v_k,w_k)}(x_k, y) = x = \lim_{k\to\infty} M_{(w_k,v_k)}(y, x_k).$$
3. *"Coherence":* Let $k \in \mathbb{N}$. For each $i \in \mathbb{N}_{\leq k}$, let $n_i \in \mathbb{N}$, let $\mathbf{w}_i \in \mathbb{R}^{n_i}_{>0}$ and $\mathbf{x}_i \in A^{n_i}$. Let $\mathbf{x} := \mathbf{x}_1 \bullet \cdots \bullet \mathbf{x}_k \in A^{n_1+\ldots+n_k}$, and let $\mathbf{w} := \mathbf{w}_1 \bullet \cdots \bullet \mathbf{w}_k \in \mathbb{R}^{n_1+\ldots+n_k}_{>0}$. Furhter, let $\mathbf{s} := (\sum \mathbf{w}_1, \ldots, \sum \mathbf{w}_k) \in \mathbb{R}^k_{>0}$. Then
$$M_{\mathbf{w}}(\mathbf{x}) = M_{\mathbf{s}}(M_{\mathbf{w}_1}(\mathbf{x}_1), \ldots, M_{\mathbf{w}_k}(\mathbf{x}_k)).$$ ⌇

Our first concrete example of a family of means is postponed to Proposition 4.2.7.

Definition 4.1.3 ("Goodness" of the means of a family)**.** Let $A \subseteq \mathbb{R}$, let M be a family of means on A. Let "good" denote an arbitrary but fixed property out of
increasing, *strictly increasing*, *continuous*, *smooth*, *F-rational*, *compressing*, *scale-invariant*, where F is a subfield of $\mathbb{R}$.
If for all $\mathbf{w} \in \mathbb{R}^2_{>0}$ the mean $M_{\mathbf{w}}$ is good, then we say that *the means of M are good.* ⌇

Lemma 4.1.4 (Triviality for $n = 1$)**.** *Let $A \subseteq \mathbb{R}$, let M be a family of means on A. For all $w \in \mathbb{R}_{>0}$, we have $M_{(w)} = \mathrm{id}_A$.* ♩

Proof. Let $(w, x) \in \mathbb{R}_{>0} \times A$. We have

$$x = M_{(w,w)}(x, x) = M_{(w,w)}\big(M_{(w)}(x), M_{(w)}(x)\big) = M_{(w)}(x);$$

the outer two equalities follow because $M_{(w,w)}$ is a mean and hence constant-preserving, the inner equality follows from the coherence axiom (with $\mathbf{w}_1 = \mathbf{w}_2 = (w) \in \mathbb{R}^1_{>0}$). □

The following proposition justifies the name "family of means" and the phraseology "Goodness of the means of a family".

Proposition 4.1.5 (*)**.** *Let $A \subseteq \mathbb{R}$, let M be a family of means on A.*

1. *$M_{\mathbf{w}}$ is an internal mean on A^n, for all $n \in \mathbb{N}$ and for all $\mathbf{w} \in \mathbb{R}^n_{>0}$.*

2. *If the means of M are good, where "good" denotes an arbitrary but fixed property out of the ones listed in Definition 4.1.3,*
then $M_{\mathbf{w}}$ is a good mean on A^n, for all $n \in \mathbb{N}$ and for all $\mathbf{w} \in \mathbb{R}^n_{>0}$. ⌟

Proof. The statements for $n = 1$ follow from Lemma 4.1.4, by noting that id_A is a good internal mean on A, for all admitted meanings of "good". For the remainder of the proof, we suppose that $n \geq 2$.

Let $\mathbf{w} = (w_1, \ldots, w_n) \in \mathbb{R}^n_{>0}$, let $\mathbf{x} = (x_1, \ldots, x_n) \in A^n$. We write $\mathbf{w} = \mathbf{w}' \bullet (w_n)$ and $\mathbf{x} = \mathbf{x}' \bullet (x_n)$, where $\mathbf{w}' \in \mathbb{R}^{n-1}_{>0}$ and $\mathbf{x}' \in A^{n-1}$. By Lemma 4.1.4 and the coherence axiom, we have

$$M_{\mathbf{w}}(\mathbf{x}) = M_{(\sum \mathbf{w}', w_n)}\big(M_{\mathbf{w}'}(\mathbf{x}'), x_n\big).$$

We can write this in the notation of Theorem 3.1.1:
let $M := M_{(\sum \mathbf{w}', w_n)}$, let $M_1 := M_{\mathbf{w}'}$, let $M_2 := \mathrm{id}_A$. Then we have

$$M_{\mathbf{w}}(\mathbf{x}) = M\big(M_1(\mathbf{x}_{\circ f_1}), M_2(\mathbf{x}_{\circ f_2})\big), \tag{4.1}$$

where $f_1 : \mathbb{N}_{\leq n-1} \to \mathbb{N}_{\leq n} : k \mapsto k$, and $f_2 : \mathbb{N}_{\leq 1} \to \mathbb{N}_{\leq n} : 1 \mapsto n$.

Let's say that "good" is, in addition to the properties listed in Definition 4.1.3, allowed to mean "internal", so that we can treat part 1 and 2 both at once. Clearly, M_2 is a good mean, and M is a good mean by assumption. Moreover, f_1 and f_2 satisfy the conditions of part 5 and 6 of Theorem 3.1.1. Thus, from Theorem 3.1.1 and (4.1), it follows by induction on $n \geq 2$ that $M_{\mathbf{w}}$ is a good mean, for all $\mathbf{w} \in \mathbb{R}^n_{>0}$. □

Phraseology 4.1.6. Let M be a family of means, let $\mathbf{w} = (w_1, \ldots, w_n) \in \mathbb{R}^n_{>0}$.

1. If $\mathbf{w} = (1)^{\bullet n}$, we normally write M_n instead of $M_{\mathbf{w}}$, and speak of "the neutral mean" M_n.

2. We speak of w_i as "the weight of the ith variable of the mean $M_{\mathbf{w}}$."

3. If the means of M are "good", where "good" is an arbitrary but fixed property out of the list in Definition 4.1.3, then we say that M is a family of "good" means.

   For instance, M is a family of increasing means iff $M_{\mathbf{w}}$ is an increasing mean for all $\mathbf{w}$ (by Proposition 4.1.5.2). ≀

4.2 Basic examples and results about families of means

Throughout this section, let $A \subseteq \mathbb{R}$ and let M be a family of means on A.

4.2.1 Concerning the dominance axiom

Our main reason to formulate the "dominance axiom" with sequences of weights $(v_k)_k$, $(w_k)_k$ such that $\lim_{k\to\infty} v_k = \infty$, instead of $\lim_{k\to\infty} w_k = 0$, is that it immediately implies the following fact; very informally phrased: "averaging with many copies of x, results in the mean being close to x".

Fact 4.2.1. *Let $v \in \mathbb{R}_{>0}$ and $x \in A$, let $n \in \mathbb{N}$, let $\mathbf{w} \in \mathbb{R}^n_{>0}$ and $\mathbf{y} \in A^n$. Then*

$$\lim_{k\to\infty} M_{(v)^{\bullet k}\bullet\mathbf{w}}\big((x)^{\bullet k} \bullet \mathbf{y}\big) = x.$$

Proof. By coherence, we have $M_{(v)^{\bullet k}\bullet\mathbf{w}}\big((x)^{\bullet k} \bullet \mathbf{y}\big) = M_{(kv,\sum\mathbf{w})}\big(M_{(v)^{\bullet k}}((x)^{\bullet k}), M_{\mathbf{w}}(\mathbf{y})\big)$.
We have $M_{(v)^{\bullet k}}((x)^{\bullet k}) = x$, because $M_{(v)^{\bullet k}}$ is a mean (by Proposition 4.1.5).
So $M_{(v)^{\bullet k}\bullet\mathbf{w}}\big((x)^{\bullet k} \bullet \mathbf{y}\big) = M_{(kv,\sum\mathbf{w})}\big(x, M_{\mathbf{w}}(\mathbf{y})\big)$. This approaches x as $k \to \infty$, by the dominance axiom. □

In the case that the means of M are increasing, the dominance axiom implies a stronger variation of itself.

Lemma 4.2.2 (Stronger dominance)**.** *Suppose that the means of M are increasing. Let $x \in A$. Let $(x_k)_{k\in\mathbb{N}}$ and $(y_k)_{k\in\mathbb{N}}$ be sequences in A, such that $\lim_{k\to\infty} x_k = x$, and such that there are $a, b \in A$ such that $a \le y_k \le b$ for all sufficiently large k. Let $(v_k)_{k\in\mathbb{N}}$ and $(w_k)_{k\in\mathbb{N}}$ be sequences in $\mathbb{R}_{>0}$, such that $\lim_{k\to\infty} v_k = \infty$, and such that $(w_k)_{k\in\mathbb{N}}$ is bounded from above. Then*

$$\lim_{k\to\infty} M_{(v_k,w_k)}(x_k, y_k) = x = \lim_{k\to\infty} M_{(w_k,v_k)}(y_k, x_k).$$

Proof. Because the mean $M_{(v_k,w_k)}$ is increasing, we have

$$M_{(v_k,w_k)}(x_k, a) \le M_{(v_k,w_k)}(x_k, y_k) \le M_{(v_k,w_k)}(x_k, b)$$

for all sufficiently large k. Moreover, by the dominance axiom, we have

$$\lim_{k\to\infty} M_{(v_k,w_k)}(x_k, a) = x = \lim_{k\to\infty} M_{(v_k,w_k)}(x_k, b).$$

It follows that $\lim_{k\to\infty} M_{(v_k,w_k)}(x_k, y_k) = x$. The other equality follows analogously. □

4.2.2 Concerning the coherence axiom

Because the coherence axiom in Definition 4.1.2 looks somewhat complicated, we illustrate it with some examples.

Example 4.2.3. Let $\mathbf{x} = (x_1, \ldots, x_6) \in A^6$. The neutral mean of $\mathbf{x}$ can be decomposed in the following ways, for example:

$$\begin{aligned} M_6(x_1, \ldots, x_6) &= M_{(5,1)}(M_5(x_1, \ldots, x_5), x_6) \\ &= M_{(2,1,3)}\big(M_2(x_1, x_2), x_3, M_3(x_4, x_5, x_6)\big) \\ &= M_{(2,2,2)}\big(M_2(x_1, x_2), M_2(x_3, x_4), M_2(x_5, x_6)\big) \\ &= M_{(3,3)}\big(M_3(x_1, x_2, x_3), M_3(x_4, x_5, x_6)\big). \end{aligned}$$

If the means of M are continuous, the last two expressions can be further simplified, because $M_{(2,2,2)} = M_3$ and $M_{(3,3)} = M_2$, as we will show in Theorem 4.6.1.4.

Here is an example of a decomposition of a non-neutral mean of $\mathbf{x}$:

$$M_{(1,\sqrt{2},e,\pi,4,5)}(x_1,\dots,x_6) = M_{(1+\sqrt{2},e,\pi+9)}\big(M_{(1,\sqrt{2})}(x_1,x_2), x_3, M_{(\pi,4,5)}(x_4,x_5,x_6)\big). \quad ♪$$

Some consequences of the coherence axiom are in Lemma 4.2.4; very informally phrased, they tell us: (part 1) "Having some weighted copies of a number x in a mean, amounts to having one copy of x with the sum of the weights"; (part 2) "In a cyclic situation, the mean over multiple cycles is the same as the mean over a single cycle".

Lemma 4.2.4 (†). *Let $A \subseteq \mathbb{R}$, let M be a family of means on A, let $\mathbf{x} = (x_1,\dots,x_n) \in A^n$.*

1. *Let $(k_1,\dots,k_n) \in \mathbb{N}^n$, let $\mathbf{w} = \mathbf{w}_1 \bullet \cdots \bullet \mathbf{w}_n \in \mathbb{R}_{>0}^{k_1+\ldots+k_n}$, where $\mathbf{w}_i \in \mathbb{R}_{>0}^{k_i}$ for each i. Then*

$$M_{\mathbf{w}}\big((x_1)^{\bullet k_1} \bullet \cdots \bullet (x_n)^{\bullet k_n}\big) = M_{(\sum \mathbf{w}_1,\dots,\sum \mathbf{w}_n)}(\mathbf{x}).$$

2. *Let $\mathbf{w} \in \mathbb{R}_{>0}^n$, and let $m \in \mathbb{N}$. Then $M_{\mathbf{w}^{\bullet m}}(\mathbf{x}^{\bullet m}) = M_{\mathbf{w}}(\mathbf{x})$.* ♩

Proof.

1. Writing $\mathbf{s} = (\sum \mathbf{w}_1,\dots,\sum \mathbf{w}_n)$, we have

$$M_{\mathbf{w}}\big((x_1)^{\bullet k_1} \bullet \cdots \bullet (x_n)^{\bullet k_n}\big) = M_{\mathbf{s}}\Big(M_{\mathbf{w}_1}\big((x_1)^{\bullet k_1}\big),\dots,M_{\mathbf{w}_n}\big((x_n)^{\bullet k_n}\big)\Big) = M_{\mathbf{s}}(\mathbf{x}),$$

where the last equality follows because the $M_{\mathbf{w}_i}$ are, like all means, constant-preserving.

2. We have $M_{\mathbf{w}^{\bullet m}}(\mathbf{x}^{\bullet m}) = M_{(\sum \mathbf{w})^{\bullet m}}\Big(\big(M_{\mathbf{w}}(\mathbf{x})\big)^{\bullet m}\Big) = M_{\mathbf{w}}(\mathbf{x})$, where the last equality follows because $M_{(\sum \mathbf{w})^{\bullet m}}$ is constant-preserving. □

4.2.3 (Non-)Examples of families of means

We did not see any concrete examples or non-examples yet of families of means. We fill that gap a bit here, and a lot in §4.3.

Example 4.2.5.
Let M be a function with domain $D := \bigsqcup_{n\in\mathbb{N}} \mathbb{R}_{>0}^n \times \mathbb{R}_{>0}^n$ and codomain $\mathbb{R}_{>0}^n$.

1. Let $M(\mathbf{w},\mathbf{x}) = \mathrm{PM}_{\infty,\mathbf{w}}(\mathbf{x}) = \mathrm{Max}(\mathbf{x})$, for all $(\mathbf{w},\mathbf{x}) \in D$. It is clear that the "means for $n = 2$" axiom and the "coherence" axiom are satisfied. However, the "dominance" axiom is not satisfied, we have for instance $\lim_{k\to\infty} M_{(k,1)}(2,3) = 2 \neq 3$.

2. For all $n \in \mathbb{N}$ and all $\mathbf{w} \in \mathbb{R}_{>0}^n$, let $M_{\mathbf{w}} := \mathrm{GM}_{\mathbf{w}}$ if n is odd, and $M_{\mathbf{w}} := \mathrm{AM}_{\mathbf{w}}$ if n is even. It is easily seen that the "means for $n = 2$" axiom and the "dominance" axiom are satisfied. However, the "coherence" axiom is not satisfied; we have for instance $M_3(x,y,z) = (xyz)^{1/3}$, while
$M_{(2,1)}\big(M_{(1,1)}(x,y), M_1(z)\big) = M_{(2,1)}\big((x+y)/2, z\big) = (x+y+z)/3$.

3. For all $n \in \mathbb{N}$ and all $\mathbf{w} \in \mathbb{R}^n_{>0}$, let $M_{\mathbf{w}} := \mathrm{AM}_{\mathbf{w}}$. Then M is satisfies all three axioms of families of means, as we show in Proposition 4.2.7. ♪

Definition 4.2.6. We define the *arithmetic mean map* AM by

$$\mathrm{AM} : \quad \bigsqcup_{n\in\mathbb{N}} \mathbb{R}^n_{>0} \times \mathbb{R}^n \to \mathbb{R}_{>0} : \quad (\mathbf{w}, \mathbf{x}) \mapsto \mathrm{AM}_{\mathbf{w}}(\mathbf{x}).$$ 𝄽

Proposition 4.2.7 (†)**.** AM *is a family of means on* $\mathbb{R}^n$*.* 𝅗𝅥

Proof. The "means for $n = 2$" axiom is obviously satisfied. The "dominance" axiom is clearly satisfied, because

$$M_{(w_k,v_k)}(y, x_k) = M_{(v_k,w_k)}(x_k, y) = (x_k + (w_k/v_k)y)/(1 + w_k/v_k),$$

which converges to x, provided that $v_k \to \infty$ and $x_k \to x$ and w_k stays bounded.

It remains to prove the "coherence" axiom. Thus, in the notation of the dominance axiom in Definition 4.1.2, let $\mathbf{x} = \mathbf{x}_1 \bullet \cdots \bullet \mathbf{x}_k \in \mathbb{R}^{n_1+\ldots+n_k}$, let $\mathbf{w} = \mathbf{w}_1 \bullet \cdots \bullet \mathbf{w}_k \in \mathbb{R}^{n_1+\ldots+n_k}_{>0}$, and let $\mathbf{s} = (\sum \mathbf{w}_1, \ldots, \sum \mathbf{w}_k) \in \mathbb{R}^k_{>0}$. Then

$$\mathrm{AM}_{\mathbf{s}}\big(\mathrm{AM}_{\mathbf{w}_1}(\mathbf{x}_1), \ldots, \mathrm{AM}_{\mathbf{w}_k}(\mathbf{x}_k)\big) = \frac{1}{\sum \mathbf{s}} \sum_{i=1}^{k} (\textstyle\sum \mathbf{w}_i) \frac{\sum \mathbf{w}_i \mathbf{x}_i}{\sum \mathbf{w}_i} = \frac{1}{\sum \mathbf{w}} \sum_{i=1}^{k} \textstyle\sum \mathbf{w}_i \mathbf{x}_i = \mathrm{AM}_{\mathbf{w}}(\mathbf{x}).$$

□

4.3 Conjugation of families of means

Main results in this section are Theorem 4.3.6 that the conjugate of a family of ("good") means is a family of ("good") means, and its Corollary 4.3.7 that the power means and translation means form families.

Lemma 4.3.1. *Let $B \subseteq A \subseteq \mathbb{R}$, let M be a family of means on A.*
Let $D_B := \bigsqcup_{n\in\mathbb{N}} \mathbb{R}^n_{>0} \times B^n$. Suppose that M maps D_B into B; this condition is in any case satisfied if $B = A \cap I$ for some interval I of $\mathbb{R}$.

1. *$M|_{D_B}$ is a family of means on B.*

2. *Let "good" be an arbitrary but fixed property out of the list in Definition 4.1.3; that is,* { increasing*,* strictly increasing*,* continuous*,* smooth*,* F-rational*,* compressing*,* scale-invariant }*, where F is a subfield of* $\mathbb{R}$*.*

   If the means of M are good, then the means of $M|_{D_B}$ are good, provided that B is open in the case that "good" means "smooth". ♩

Proof. For any $\mathbf{w} \in \mathbb{R}^n_{>0}$, we have $(M|_{D_B})_{\mathbf{w}} = (M_{\mathbf{w}})|_B$. Thus, it is clear that $M|_{D_B}$ satisfies the "means" and "dominance" and "coherence" axioms. That each mean $(M_{\mathbf{w}})|_B$ is good, follows from Fact 1.1.20. □

***Remark* 4.3.2.** In the context of Lemma 4.3.1, we often simply write M instead of $M|_{D_B}$, to avoid overburdening of notation; the context should make it clear on what domain we regard M.

Definition 4.3.3. Let $p \in \overline{\mathbb{R}}$.
We define the *p-power mean map* PM_p and the *p-translation mean map* TM_p by

$$\mathrm{PM}_p : \quad \bigsqcup_{n\in\mathbb{N}} \mathbb{R}^n_{>0} \times \mathbb{R}^n_{>0} \to \mathbb{R}_{>0} : \quad (\mathbf{w}, \mathbf{x}) \mapsto \mathrm{PM}_{p,\mathbf{w}}(\mathbf{x}),$$
$$\mathrm{TM}_p : \quad \bigsqcup_{n\in\mathbb{N}} \mathbb{R}^n_{>0} \times \mathbb{R}^n_{>0} \to \mathbb{R}_{>0} : \quad (\mathbf{w}, \mathbf{x}) \mapsto \mathrm{TM}_{p,\mathbf{w}}(\mathbf{x}).$$

As special cases, we write $\mathrm{GM} := \mathrm{PM}_0 = \mathrm{TM}_0$, and $\mathrm{HM} := \mathrm{PM}_{-1} = \mathrm{TM}_{-\infty}$. ≀

Definition 4.3.4. Let $A, B \subseteq \mathbb{R}$, let $f : A \to B$ be a bijection.
Let $M : \bigsqcup_{n\in\mathbb{N}} \mathbb{R}^n_{>0} \times A^n \to A$ be a function. We write $M^{[f]}$ for the function

$$M^{[f]} : \quad \bigsqcup_{n\in\mathbb{N}} \mathbb{R}^n_{>0} \times B^n \to B : \quad (\mathbf{w}, \mathbf{x}) \mapsto M^{[f]}_{\mathbf{w}}(\mathbf{x}).$$ ≀

Fact 4.3.5. *Let $p \in \mathbb{R}$ and $b \in \overline{\mathbb{R}}$. By Theorem 2.6.6 and Theorem 2.7.4, we have*

$\mathrm{PM}_p = \mathrm{AM}^{[\mathrm{pow}_{1/p}]}$	*if* $p \in \mathbb{R}_{\neq 0}$,	$\mathrm{PM}_p = \mathrm{AM}^{[\exp]}$	*if* $p = 0$.
$\mathrm{TM}_b = \mathrm{AM}^{[\mathrm{tra}_b \circ \exp]}$	*if* $b \in \mathbb{R}_{\geq 0}$,	$\mathrm{TM}_b = \mathrm{AM}^{[\mathrm{inv} \circ \mathrm{tra}_{-b} \circ \exp]}$	*if* $b \in \mathbb{R}_{\leq 0}$.
$\mathrm{TM}_\infty = \mathrm{AM}$		$\mathrm{TM}_{-\infty} = \mathrm{HM}$.	

Theorem 4.3.6 (*)**.** *Let $A, B \subseteq \mathbb{R}$, let $f : A \to B$ be a monotonic homeomorphism. Let M be a family of means on A.*

1. *$M^{[f]}$ is a family of means on B.*
2. *Let "good" be an arbitrary but fixed property out of*
   increasing*,* strictly increasing*,* compressing*,* continuous*.*
   If the means of M are good, then the means of $M^{[f]}$ are good.
3. *If f is smooth and its derivative has no zeros on A,*
   and the means of M are smooth, then the means of $M^{[f]}$ are smooth. ⅃

Proof. **1.** We verify that $M^{[f]}$ satisfies the three axioms of families of means.
"Means": This is satisfied by Lemma 1.3.10.
"Dominance": Let $x, y \in B$, let $(x_k)_{k\in\mathbb{N}}$ be sequence in B with $\lim_{k\to\infty} x_k = x$. Let $(v_k)_{k\in\mathbb{N}}$ and $(w_k)_{k\in\mathbb{N}}$ be sequences in $\mathbb{R}_{>0}$, such that $\lim_{k\to\infty} v_k = \infty$, and $(w_k)_{k\in\mathbb{N}}$ is bounded from above. We have

$$M^{[f]}_{(v_k,w_k)}(x_k, y) = f\Big(M_{(v_k,w_k)}\big(f^{-1}(x_k), f^{-1}(y)\big)\Big). \tag{4.2}$$

Because f^{-1} is continuous, we have $\lim_{k\to\infty} f^{-1}(x_k) = f^{-1}(x)$. Hence, by the dominance property of the family M, it follows that

$$\lim_{k\to\infty} M_{(v_k,w_k)}\big(f^{-1}(x_k), f^{-1}(y)\big) = f^{-1}(x).$$

Because f is continuous, the expression in (4.2) converges to x as $k \to \infty$, as desired. The other equality of the dominance axiom follows analogously.

"Coherence": Let $k \in \mathbb{N}$, let $n_1, \ldots, n_k \in \mathbb{N}$, let $n := n_1 + \ldots + n_k$.
Let $\mathbf{w} := \mathbf{w}_1 \bullet \cdots \bullet \mathbf{w}_k \in \mathbb{R}_{>0}^{n_1+\ldots+n_k}$, let $\mathbf{s} := (\sum \mathbf{w}_1, \ldots, \sum \mathbf{w}_k) \in \mathbb{R}_{>0}^k$.
For $1 \le i \le k$, let $f_i : \mathbb{N}_{\le n_i} \to \mathbb{N}_{\le n}$ be given by $f_i(m) = \sum_{j=1}^{i-1} n_j + m$.
By the coherence property of the family M, we have

$$M_{\mathbf{w}}(\mathbf{x}) = M_{\mathbf{s}}\big(M_{\mathbf{w}_1}(\mathbf{x}_{\circ f_1}), \ldots, M_{\mathbf{w}_k}(\mathbf{x}_{\circ f_k})\big) \qquad \forall \mathbf{x} \in A^n.$$

By Theorem 3.1.1.3, it follows that

$$M_{\mathbf{w}}^{[f]}(\mathbf{x}) = M_{\mathbf{s}}^{[f]}\big(M_{\mathbf{w}_1}^{[f]}(\mathbf{x}_{\circ f_1}), \ldots, M_{\mathbf{w}_k}^{[f]}(\mathbf{x}_{\circ f_k})\big) \qquad \forall \mathbf{x} \in B^n.$$

Hence, $M^{[f]}$ has the coherence property as well.

2. & 3. This follows from Theorem 2.1.3, parts 2, 3 and 4. □

Corollary 4.3.7 (*)**.** PM_p *and* TM_b *are families of means, for all* $p \in \mathbb{R}$ *and* $b \in \overline{\mathbb{R}}$. ⌟

Proof. This follows by combining Theorem 4.3.6, Proposition 4.2.7 and Fact 4.3.5. □

***Remark* 4.3.8.** Corollary 4.3.7 is *not* true for $p \in \{\pm\infty\}$: we saw in Example 4.2.5.1 that Max and Min don't define families of means.

Corollary 4.3.9 ((*))**.** *Let* $A \subseteq \mathbb{R}$.

The group G_{cont}*, consisting of the homeomorphisms* $A \to A$*, acts on the set of families of means on* A*, by the map* $(f, M) \mapsto M^{[f]}$.

The set of families of "good" means on A *is invariant under* G_{cont}*, where "good" is as in Theorem 4.3.6.2* ⌟

4.4 Combinatorial and analytical relations between the weights and the means

The next theorem is the content of this section.

Theorem 4.4.1 (*)**.** *Let* $A \subseteq \mathbb{R}$*, let* M *be a family of means on* A.
Let $n \in \mathbb{N}$*, let* $\mathbf{x} = (x_1, \ldots, x_n) \in A^n$ *and let* $\mathbf{w} = (w_1, \ldots, w_n) \in \mathbb{R}_{>0}^n$.

1. *"Symmetry":* $M(\mathbf{w}^\sigma; \mathbf{x}^\sigma) = M(\mathbf{w}; \mathbf{x})$*, for all* $\sigma \in S_n$.
   In particular, M_n *is a symmetric mean on* A^n.

2. *"Additivity of weights": Suppose that the distinct real numbers that occur as entries of* $\mathbf{x}$ *are* $r_1, \ldots, r_m$*. For* $i \in \mathbb{N}_{\le m}$*, let* s_i *be the sum of the weights of the variables that equal* r_i*. Then*
$$M_{\mathbf{w}}(\mathbf{x}) = M_{(s_1,\ldots,s_m)}(r_1, \ldots, r_m).$$

3. *"Combinability of weights": Let* $\mathbf{w}_1, \ldots, \mathbf{w}_k \in \mathbb{R}_{>0}^n$*, for any* k*. Then*
$$M_{\mathbf{w}_1+\ldots+\mathbf{w}_k}(\mathbf{x}) = M_{(\sum \mathbf{w}_1,\ldots,\sum \mathbf{w}_k)}\big(M_{\mathbf{w}_1}(\mathbf{x}), \ldots, M_{\mathbf{w}_k}(\mathbf{x})\big).$$

In part 4 and 5, we suppose that the means of M are increasing.

4. *"Scale-invariance in the weights": For all $\lambda \in \mathbb{R}_{>0}$, we have $M_{\lambda \mathbf{w}} = M_{\mathbf{w}}$.*

5. *"Continuity in the weights": The map $\mathbb{R}^n_{>0} \to A : \mathbf{w} \mapsto M_{\mathbf{w}}(\mathbf{x})$ is continuous.* ⌟

Proof.

1. (Symmetry.) We first show it in the case that $n = 2$. We write $\mathbf{x} = (x, y)$ and $\mathbf{w} = (v, w)$. Thus, we need to show that

$$M_{(v,w)}(x, y) = M_{(w,v)}(y, x). \tag{4.3}$$

Let $k \in \mathbb{N}$. We have

$$\begin{aligned} M_{(v,w)^{\bullet k} \bullet (v)}\big((x, y)^{\bullet k} \bullet (x)\big) &= M_{(k(v+w),v)}\Big(M_{(v,w)^{\bullet k}}\big((x, y)^{\bullet k}\big), x\Big) \\ &= M_{(k(v+w),v)}\big(M_{(v,w)}(x, y), x\big), \end{aligned} \tag{4.4}$$

where the first equality holds by coherence, and the second equality by part 2 of Lemma 4.2.4. We obviously have $M_{(v,w)^{\bullet k} \bullet (v)}\big((x, y)^{\bullet k} \bullet (x)\big) = M_{(v) \bullet (w,v)^{\bullet k}}\big((x) \bullet (y, x)^{\bullet k}\big)$. By an analogous calculation as in (4.4), we get

$$M_{(v) \bullet (w,v)^{\bullet k}}\big((x) \bullet (y, x)^{\bullet k}\big) = M_{(v,k(v+w))}\big(x, M_{(w,v)}(y, x)\big).$$

Thus, we have

$$M_{(k(v+w),v)}\big(M_{(v,w)}(x, y), x\big) = M_{(v,k(v+w))}\big(x, M_{(w,v)}(y, x)\big).$$

Taking the limit $k \to \infty$ and using the dominance axiom, we get (4.3), as desired.

Now we use that it is true for $n = 2$, in combination with the coherence axiom, to prove that it is true for general n. Let $\tau \in S_n$ be a transposition of two neighbouring numbers k and $k + 1$. Writing $\mathbf{w} =: \mathbf{u} \bullet (w_k, w_{k+1}) \bullet \mathbf{v}$, and $\mathbf{x} =: \mathbf{y} \bullet (x_k, x_{k+1}) \bullet \mathbf{z}$, we have

$$\begin{aligned} M_{\mathbf{w}}(\mathbf{x}) &= M_{\mathbf{u} \bullet (w_k + w_{k+1}) \bullet \mathbf{v}}\big(\mathbf{y} \bullet M_{(w_k, w_{k+1})}(x_k, x_{k+1}) \bullet \mathbf{z}\big) \\ &= M_{\mathbf{u} \bullet (w_{k+1} + w_k) \bullet \mathbf{v}}\big(\mathbf{y} \bullet M_{(w_{k+1}, w_k)}(x_{k+1}, x_k) \bullet \mathbf{z}\big) = M_{\mathbf{w}^\tau}(\mathbf{x}^\tau). \end{aligned}$$

Because $\mathbf{a}^{\rho\sigma} = (\mathbf{a}^\sigma)^\rho$ for all $\mathbf{a} \in \mathbb{R}^n$ and all $\sigma, \rho \in S_n$, and because the group S_n is generated by the transpositions of two neighbouring numbers, it follows that $M_{\mathbf{w}}(\mathbf{x}) = M_{\mathbf{w}^\sigma}(\mathbf{x}^\sigma)$ for all $\sigma \in S_n$, as desired.

2. (Additivity of weights.) Let $\sigma \in S_n$ be a permutation such that

$$\mathbf{x}^\sigma = (r_1)^{\bullet k_1} \bullet \cdots \bullet (r_m)^{\bullet k_m},$$

for some $k_1, \ldots, k_m \in \mathbb{N}$. Write $\mathbf{w}^\sigma = \mathbf{w}_1 \bullet \cdots \bullet \mathbf{w}_m$, where $\mathbf{w}_i \in \mathbb{R}^{k_i}_{>0}$ for each i. Clearly, $\sum \mathbf{w}_i = s_i$. We have $M_{\mathbf{w}}(\mathbf{x}) = M_{\mathbf{w}^\sigma}(\mathbf{x}^\sigma)$ by part 1 (symmetry), and $M_{\mathbf{w}^\sigma}(\mathbf{x}^\sigma) = M_{(s_1,\ldots,s_m)}(r_1, \ldots, r_m)$ by part 1 of Lemma 4.2.4.

3. (Combinability of weights.)
Let $(\mathbf{w}'_1, \dots, \mathbf{w}'_n) \in \mathbb{R}^{n\times k}_{>0}$ be the transpose of $(\mathbf{w}_1, \dots, \mathbf{w}_k) \in \mathbb{R}^{k\times n}_{>0}$. Clearly, $\mathbf{w}_1 + \dots + \mathbf{w}_k = (\sum \mathbf{w}'_1, \dots, \sum \mathbf{w}'_n)$. Thus,

$$\begin{aligned} M_{\mathbf{w}_1+\dots+\mathbf{w}_k}(\mathbf{x}) &= M_{(\sum \mathbf{w}'_1,\dots,\sum \mathbf{w}'_n)}(x_1, \dots, x_n) \\ &= M_{\mathbf{w}'_1\bullet\dots\bullet\mathbf{w}'_n}\big((x_1)^{\bullet k} \bullet \dots \bullet (x_n)^{\bullet k}\big) \\ &= M_{\mathbf{w}_1\bullet\dots\bullet\mathbf{w}_k}\big((\mathbf{x})^{\bullet k}\big) \\ &= M_{(\sum \mathbf{w}_1,\dots,\sum \mathbf{w}_k)}\big(M_{\mathbf{w}_1}(\mathbf{x}), \dots, M_{\mathbf{w}_k}(\mathbf{x})\big), \end{aligned}$$

as desired; the 2nd equality holds by part 1 of Lemma 4.2.4, the 3rd equality by symmetry, and the 4th one by the coherence axiom.

4. (Scale-invariance in the weights.) Let $k \in \mathbb{N}$. Applying the result of part 3 with $\mathbf{w}_i = \mathbf{w}$ for all $i \in \mathbb{N}_{\le k}$, we get that $M_{k\mathbf{w}} = M_{\mathbf{w}}$. It follows that

$$M_{\lambda\mathbf{w}} = M_{\mathbf{w}} \qquad \text{for all } \lambda \in \mathbb{Q}_{>0}.$$

For general $\lambda \in \mathbb{R}_{>0}$, choose a sequence $(q_k)_{k\in\mathbb{N}}$ in $\mathbb{Q}_{>0}$ such that $\lambda - 2^{-k} < q_k < \lambda$. So $\lambda = q_k + \varepsilon_k$, where $0 < \varepsilon_k < 2^{-k}$. We have, by part 3,

$$\begin{aligned} M_{\lambda\mathbf{w}}(\mathbf{x}) = M_{2^k\lambda\mathbf{w}}(\mathbf{x}) &= M_{2^k q_k\mathbf{w} + 2^k\varepsilon_k\mathbf{w}}(\mathbf{x}) \\ &= M_{(2^k q_k \sum \mathbf{w},\ 2^k\varepsilon_k \sum \mathbf{w})}\big(M_{2^k q_k\mathbf{w}}(\mathbf{x}),\ M_{2^k\varepsilon_k\mathbf{w}}(\mathbf{x})\big). \end{aligned} \tag{4.5}$$

Note that $\lim_{k\to\infty} 2^k q_k \sum \mathbf{w} = \infty$, while the sequence $(2^k\varepsilon_k \sum \mathbf{w})_{k\in\mathbb{N}}$ is bounded from above by $\sum \mathbf{w}$. Moreover, we have $\mathrm{Min}(\mathbf{x}) \le M_{2^k\varepsilon_k\mathbf{w}}(\mathbf{x}) \le \mathrm{Max}(\mathbf{x})$, and because $2^k q_k$ is rational, we have $M_{2^k q_k\mathbf{w}}(\mathbf{x}) = M_{\mathbf{w}}(\mathbf{x})$. Hence, by Lemma 4.2.2, the limit of the right-hand side of (4.5) as $k \to \infty$ equals $M_{\mathbf{w}}(\mathbf{x})$. Since the left-hand side does not depend on k, it follows that $M_{\lambda\mathbf{w}}(\mathbf{x}) = M_{\mathbf{w}}(\mathbf{x})$.

5. (Continuity in the weights.) Actually, something stronger is true, namely Lemma 4.4.2. The current result amounts to the special case of the lemma that $\mathbf{x}_k = \mathbf{x}$ for all k. We state the lemma as a separate result, because we need its full strength in the proof of Theorem 4.6.1. □

Lemma 4.4.2. *Let $A \subseteq \mathbb{R}$, let M be a family of increasing means on A.*
Let $n \in \mathbb{N}$, let $\mathbf{x} \in A^n$ and let $\mathbf{w} \in \mathbb{R}^n_{>0}$.
Let $(\mathbf{e}_k, \mathbf{x}_k)_{k\in\mathbb{N}}$ be a sequence in $\mathbb{R}^n \times A^n$, such that $\mathbf{w} + \mathbf{e}_k \in \mathbb{R}^n_{>0}$ for all k, and such that $\lim_{k\to\infty} \mathbf{e}_k = \mathbf{0}$ and $\lim_{k\to\infty} \mathbf{x}_k = \mathbf{x}$.
Moreover, suppose that either the means of M are continuous, or $\mathbf{x}_k = \mathbf{x}$ for all k.
Then $\lim_{k\to\infty} M_{\mathbf{w}+\mathbf{e}_k}(\mathbf{x}_k) = M_{\mathbf{w}}(\mathbf{x})$. ♩

Proof. Step 1: We suppose that $\mathbf{e}_k \in \mathbb{R}^n_{>0}$ for all k. By part 3 and 4 of Theorem 4.4.1, we have

$$M_{\mathbf{w}+\mathbf{e}_k}(\mathbf{x}_k) = M_{(\sum \mathbf{w}, \sum \mathbf{e}_k)}\big(M_{\mathbf{w}}(\mathbf{x}_k), M_{\mathbf{e}_k}(\mathbf{x}_k)\big) = M_{(\frac{\sum \mathbf{w}}{\sum \mathbf{e}_k}, 1)}\big(M_{\mathbf{w}}(\mathbf{x}_k), M_{\mathbf{e}_k}(\mathbf{x}_k)\big). \tag{4.6}$$

We have that $\lim_{k\to\infty}\frac{\sum \mathbf{w}}{\sum \mathbf{e}_k}=\infty$, and that $\lim_{k\to\infty} M_{\mathbf{w}}(\mathbf{x}_k)=\mathbf{x}$ (either by continuity of $M_{\mathbf{w}}$, or because $\mathbf{x}_k=\mathbf{x}$ for all k). Moreover, there are $a,b\in A$ such that
$a\le M_{\mathbf{e}_k}(\mathbf{x}_k)\le b$ for all sufficiently large k. Namely, if $\mathrm{Min}(\mathbf{x}_k)\ge \mathrm{Min}(\mathbf{x})$ for all k, then we can take $a:=\mathrm{Min}(\mathbf{x})$. Otherwise, there exist k such that $\mathrm{Min}(\mathbf{x}_k)<\mathrm{Min}(\mathbf{x})$, and it follows by $\lim_{k\to\infty}\mathbf{x}_k=\mathbf{x}$ that there exists k_0 such that $\mathrm{Min}(\mathbf{x}_{k_0})<\mathrm{Min}(\mathbf{x})$ and $\mathrm{Min}(\mathbf{x}_{k_0})\le \mathrm{Min}(\mathbf{x}_k)$ for all $k>k_0$; thus, we we can take $a:=\mathrm{Min}(\mathbf{x}_{k_0})$. The existence of a suitable b follows analogously.

Hence, Lemma 4.2.2 tells us that (4.6) approaches $M_{\mathbf{w}}(\mathbf{x})$ as $k\to\infty$. This finishes the proof for the case that $(\mathbf{e}_k)_k$ is a sequence in $\mathbb{R}^n_{>0}$.

Step 2: We drop the assumption that $\mathbf{e}_k\in\mathbb{R}^n_{>0}$.
Let $\delta(\mathbf{e}_k):=\mathrm{Max}\big(\mathbf{w}/(\mathbf{w}+\mathbf{e}_k)\big)+2^{-k}\in\mathbb{R}_{>0}$. Thus, for all $1\le i\le k$, the ith coordinate of the vector $\delta(\mathbf{e}_k)(\mathbf{w}+\mathbf{e}_k)$ is strictly larger than the ith coordinate of $\mathbf{w}$. Hence

$$\delta(\mathbf{e}_k)(\mathbf{w}+\mathbf{e}_k)=\mathbf{w}+\mathbf{e}'_k \qquad \text{for some } \mathbf{e}'_k\in\mathbb{R}^n_{>0}.$$

Moreover, we have $\lim_{k\to\infty}\mathbf{e}_k=0$, and $\lim_{k\to\infty}\delta(\mathbf{e}_k)=1$. Hence, $\lim_{k\to\infty}\mathbf{e}'_k=0$.

By part 4, it follows that

$$M_{\mathbf{w}+\mathbf{e}_k}(\mathbf{x}_k)=M_{\delta(\mathbf{e}_k)(\mathbf{w}+\mathbf{e}_k)}(\mathbf{x}_k)=M_{\mathbf{w}+\mathbf{e}'_k}(\mathbf{x}_k),$$

which approaches $M_{\mathbf{w}}(\mathbf{x})$ as $k\to\infty$, by the result of step 1. This finishes the proof for the general case. □

4.5 Extending M: allowing all but one of the weights to be zero

This short section consists mainly of definitions, which we will use in the next section: there, we show that a continuous family M can be continuously extended to weight vectors $\mathbf{w}$ that contain zeros.

Definition 4.5.1. Let $\mathbf{w}=(w_1,\ldots,w_n)\in\mathbb{R}^n$, for some n.

1. The *support of* $\mathbf{w}$ is the set $\mathrm{supp}(\mathbf{w}):=\{i: w_i\ne 0\}\subseteq\mathbb{N}_{\le n}$.

2. Let $I\subseteq\mathbb{N}_{\le n}$, let $k=|I|$. Say $I=\{i_1,\ldots,i_k\}$, where $i_1<\ldots<i_k$. We define

$$\mathrm{proj}(\mathbf{w},I):=(w_{i_1},\ldots,w_{i_k})\in\mathbb{R}^k.$$

3. Let $k=|\mathrm{supp}(\mathbf{w})|$. We define $\pi_{\mathbf{w}}:\ \mathbb{R}^n_{\ge 0}\setminus\{\mathbf{0}\}\to\mathbb{R}^k_{\ge 0}:\ \mathbf{x}\mapsto\mathrm{proj}(\mathbf{x},\mathrm{supp}(\mathbf{w}))$. 𝄽

Example 4.5.2. Let $\mathbf{w}=(3,4,0,6,7,0,0,1,2,0),\ \ \mathbf{x}=(1,2,3,4,5,6,7,8,9,10)$. Then

$$\mathrm{supp}(\mathbf{w})=\{1,2,4,5,8,9\},\quad \pi_{\mathbf{w}}(\mathbf{w})=(3,4,6,7,1,2),\quad \pi_{\mathbf{w}}(\mathbf{x})=(1,2,4,5,8,9).$$ ♪

Definition 4.5.3 (*)**.** Let $A\subseteq\mathbb{R}$, let M be a family of means on A.

1. We define $\overline{M}: \quad \bigsqcup_{n\in\mathbb{N}} (\mathbb{R}^n_{\geq 0} \setminus \{\mathbf{0}\}) \times A^n \to A: \quad (\mathbf{w};\mathbf{x}) \mapsto M\big(\pi_{\mathbf{w}}(\mathbf{w}); (\pi_{\mathbf{w}}(\mathbf{x})\big).$

   For $\mathbf{w} \in \mathbb{R}^n_{\geq 0} \setminus \{\mathbf{0}\}$, we write $\overline{M}_{\mathbf{w}}$ for the function $A^n \to A : \mathbf{x} \mapsto \overline{M}(\mathbf{w};\mathbf{x})$.

2. We write $D_M, D_{\overline{M}}$ for the domain of $M, \overline{M}$ respectively, and we endow the domains with topologies, namely as subspaces of $\bigsqcup_{n\in\mathbb{N}} \mathbb{R}^{2n}$ with the disjoint union topology.

Example 4.5.4. Let $p \in \mathbb{R} \setminus \{0\}$. Let $\mathbf{w}, \mathbf{x}$ be as in Example 4.5.2. Then

$$\overline{(\mathrm{PM}_p)}_{\mathbf{w}}(\mathbf{x}) = \mathrm{PM}_{p,\pi_{\mathbf{w}}(\mathbf{w})}(\pi_{\mathbf{w}}(\mathbf{x})) = \Big(\frac{3 + 4\cdot 2^p + 6\cdot 4^p + 7\cdot 5^p + 8^p + 2\cdot 9^p}{23}\Big)^{1/p}.$$

Note that this is just the same as $\big((\sum_{i=1}^n w_i x_i^p)/(\sum_{i=1}^n w_i)\big)^{1/p}$, where $\mathbf{w} = (w_1, \ldots, w_n)$ and $\mathbf{x} = (x_1, \ldots, x_n)$. Thus, $\overline{(\mathrm{PM}_p)}$ is given by the same formulas as PM_p.

***Remark* 4.5.5.** Let $A \subseteq \mathbb{R}$, let M be a family of means on A. Clearly, the restriction of $\overline{M}$ to D_M equals M. That is, $\overline{M}$ extends M.

We show in Theorem 4.6.1.1 that if the means of M are continuous and increasing, then M is continuous on D. Because $D_{\overline{M}}$ is contained in the closure of D_M, there is at most one continuous function on $D_{\overline{M}}$ that extends M. We show in part 6 of the theorem that $\overline{M}$ is the unique such function.

4.6 Families of increasing, continuous means

The next theorem is the content of this section.

Theorem 4.6.1 (*).
*Let $A \subseteq \mathbb{R}$, let M be a family of continuous, increasing means on A.
Let $n \in \mathbb{N}$, let $\mathbf{x} = (x_1, \ldots, x_n) \in A^n$ and let $\mathbf{w} = (w_1, \ldots, w_n) \in \mathbb{R}^n_{>0}$.*

1. *"Continuity": The function M is continuous on D_M.*

2. *"Extendability": $\overline{M}$ is the unique continuous extension of M to $D_{\overline{M}}$.*

3. *"Monotony in each weight": Let $\mathbf{w}' \in \mathbb{R}^{n-1}_{>0}$, and let $\mathbf{x} = \mathbf{x}' \bullet (x_n) \in A^n$. The function*

   $$\mu: \quad [0,\infty) \to A: \quad w \mapsto \overline{M}_{\mathbf{w}'\bullet(w)}(\mathbf{x})$$

   is continuous. Writing $y := M_{\mathbf{w}'}(\mathbf{x}')$, we have:

   If $y \leq x_n$, then μ is increasing, with image $[y, x_n)$ or $[y, x_n]$;

   if $y \geq x_n$, then μ is decreasing, with image $(x_n, y]$ or $[x_n, y]$.

   Moreover, if the means of M are strictly increasing, and $y \neq x_n$, then μ is strictly increasing/decreasing, and the image is the half-open interval.

4. *"Interval": A is an interval of $\mathbb{R}$.*

Proof.

1. (Continuity.) Let $(\mathbf{e}_k, \mathbf{x}_k)_{k\in\mathbb{N}}$ be a sequence in $\mathbb{R}^n \times A^n$, such that $\mathbf{w}+\mathbf{e}_k \in \mathbb{R}^n_{>0}$ for all k, and such that $\lim_{k\to\infty} \mathbf{e}_k = \mathbf{0}$ and $\lim_{k\to\infty} \mathbf{x}_k = \mathbf{x}$.

Lemma 4.4.2 tells us that $\lim_{k\to\infty} M_{\mathbf{w}+\mathbf{e}_k}(\mathbf{x}_k) = M_{\mathbf{w}}(\mathbf{x})$.

That is, M is continuous in $(\mathbf{w}, \mathbf{x})$. Thus, M is continuous.

6. (Coherence at vanishing weights.) It is clear that $\overline{M}$ extends M. If there is a continuous extension of M to $D_{\overline{M}}$, then it is unique, because D_M is dense in $D_{\overline{M}}$. So it remains to to prove that $\overline{M}$ is continuous on $D_{\overline{M}}$. Because $D_{\overline{M}}$ has the disjoint subspace topology, it suffices to to prove that $\overline{M}$ is continuous on $(\mathbb{R}^n_{\geq 0} \setminus \{\mathbf{0}\}) \times A^n$, for all $n \in \mathbb{N}$. Because M is continuous on $\mathbb{R}^n_{>0} \times A^n$, which is dense in $(\mathbb{R}^n_{\geq 0} \setminus \{\mathbf{0}\}) \times A^n$, it follows from elementary analysis that it suffices to show that for all $(\mathbf{v}, \mathbf{y}) \in (\mathbb{R}^n_{\geq 0} \setminus (\mathbb{R}^n_{>0} \cup \{\mathbf{0}\})) \times A^n$, we have

$$\lim_{\substack{(\mathbf{w},\mathbf{x})\to(\mathbf{v},\mathbf{y}) \\ (\mathbf{w},\mathbf{x})\in\mathbb{R}^n_{>0}\times A^n}} M(\mathbf{w}, \mathbf{x}) = \overline{M}(\mathbf{v}, \mathbf{y}).$$

Thus, let $\mathbf{y} \in A^n$ and let $\mathbf{v} \in (\mathbb{R}^n_{\geq 0} \setminus (\mathbb{R}^n_{>0} \cup \{\mathbf{0}\}))$, and let $k := |\mathrm{supp}(\mathbf{v})|$, so $1 \leq k \leq n-1$. We have

$$M(\mathbf{w}, \mathbf{x}) - \overline{M}(\mathbf{v}, \mathbf{y}) = \Big(M_{\mathbf{w}}(\mathbf{x}) - M_{\mathbf{w}}(\mathbf{y})\Big) + \Big(M_{\mathbf{w}}(\mathbf{y}) - M_{\pi_{\mathbf{v}}(\mathbf{v})}\big(\pi_{\mathbf{v}}(\mathbf{y})\big)\Big).$$

The first term on the right-hand side approaches zero as $(\mathbf{w}, \mathbf{x}) \to (\mathbf{v}, \mathbf{y})$, by continuity of $M_{\mathbf{w}}$. Hence, it suffices to show that the second term also approaches zero as $(\mathbf{w}, \mathbf{x}) \to (\mathbf{v}, \mathbf{y})$, or equivalently, as $\mathbf{w} \to \mathbf{v}$, since the term does not depend on $\mathbf{x}$. That is, it suffices to show that

$$\lim_{\substack{\mathbf{w}\to\mathbf{v} \\ \mathbf{w}\in\mathbb{R}^n_{>0}}} M_{\mathbf{w}}(\mathbf{y}) = M_{\pi_{\mathbf{v}}(\mathbf{v})}\big(\pi_{\mathbf{v}}(\mathbf{y})\big). \tag{4.7}$$

It is easily seen that for all $\mathbf{z} \in \mathbb{R}^n$ and $\sigma \in S_n$, we have $\pi_{\mathbf{v}^\sigma}(\mathbf{z}^\sigma) = (\pi_{\mathbf{v}}(\mathbf{z}))^\rho$ for some $\rho \in S_k$. Therefore, and by part 1 ("symmetry") of Theorem 4.4.1, the formula (4.7) is equivalent to the formula that we obtain from it by replacing $\mathbf{w}, \mathbf{v}, \mathbf{y}$ by $\mathbf{w}^\sigma, \mathbf{v}^\sigma, \mathbf{y}^\sigma$ respectively. Thus, because $\{\mathbf{w}^\sigma : \mathbf{w} \in \mathbb{R}^n_{>0}\} = \mathbb{R}^n_{>0}$, and $\{\mathbf{y}^\sigma : \mathbf{y} \in A^n\} = A^n$, it suffices to prove (4.7) for the case that all the non-zero entries of $\mathbf{v}$ are at the head, that is,

$$\mathbf{v} = \mathbf{v}_1 \bullet \mathbf{0} \qquad \text{where } \mathbf{v}_1 \in \mathbb{R}^k_{>0}.$$

We further write $\mathbf{w} = \mathbf{w}_1 \bullet \mathbf{w}_2$ and $\mathbf{y} = \mathbf{y}_1 \bullet \mathbf{y}_2$, where $\mathbf{w}_1 \in \mathbb{R}^k_{>0}$ and $\mathbf{y}_1 \in A^k$. We have

$$M_{\mathbf{w}}(\mathbf{y}) = M_{(\sum \mathbf{w}_1, \sum \mathbf{w}_2)}\big(M_{\mathbf{w}_1}(\mathbf{y}_1), M_{\mathbf{w}_2}(\mathbf{y}_2)\big) = M_{\left(\frac{\sum \mathbf{w}_1}{\sum \mathbf{w}_2}, 1\right)}\big(M_{\mathbf{w}_1}(\mathbf{y}_1), M_{\mathbf{w}_2}(\mathbf{y}_2)\big), \tag{4.8}$$

the 1st equality by the coherence property, the 2nd equality by part 4 ("scale-invariance") of Theorem 4.4.1. As $\mathbf{w} \to \mathbf{v}$, then $\mathbf{w}_1 \to \mathbf{v}_1$ and $\mathbf{w}_2 \to \mathbf{0}$, hence $\frac{\sum \mathbf{w}_1}{\sum \mathbf{w}_2} \to \infty$; moreover, $M_{\mathbf{w}_1}(\mathbf{y}_1) \to M_{\mathbf{v}_1}(\mathbf{y}_1)$, by continuity of M, and we have $\mathrm{Min}(\mathbf{y}_2) \leq M_{\mathbf{w}_2}(\mathbf{y}_2) \leq \mathrm{Max}(\mathbf{y}_2)$. Therefore, it follows from Lemma 4.2.2 and (4.8) that

$$\lim_{\mathbf{w}\to\mathbf{v}} M_{\mathbf{w}}(\mathbf{y}) = M_{\mathbf{v}_1}(\mathbf{y}_1).$$

As $M_{\mathbf{v}_1}(\mathbf{y}_1) = M_{\pi_{\mathbf{v}}(\mathbf{v})}\big(\pi_{\mathbf{v}}(\mathbf{y})\big)$, this completes the proof of (4.7).

7. (Monotony in each weight.) For simplicity, we assume that $y \leq x_n$; the other case is proven in an exactly analogous way.

We know that μ is continuous by part 6. Further, we have $\mu(0) = y$, and for any $w \in (0, \infty)$, we have

$$\mu(w) = M_{\mathbf{w}' \bullet (w)}(\mathbf{x}' \bullet (x_n)) = M_{(\sum \mathbf{w}', w)}\big(M_{\mathbf{w}'}(\mathbf{x}'), x_n\big) = M_{(\sum \mathbf{w}', w)}\big(y, x_n\big); \tag{4.9}$$

hence, $\mu(w) \in [y, x_n]$, and $\lim_{w \to \infty} \mu(w) = x_n$ by the dominance axiom. Hence, the image of μ equals $[y, x_n)$ or $[y, x_n]$. Clearly, μ is constant if $y = x_n$; we assume that $y < x_n$. It remains to prove that μ is increasing, and strictly increasing if the means of M are strictly increasing. For any $\varepsilon > 0$, we have

$$\mu(w + \varepsilon) = M_{(\sum \mathbf{w}', w+\varepsilon)}\big(y, x_n\big) = M_{(\sum \mathbf{w}', w, \varepsilon)}\big(y, x_n, x_n\big) = M_{(\sum \mathbf{w}' + w, \varepsilon)}\big(\mu(w), x_n\big); \tag{4.10}$$

the 1st equality by (4.9), the 2nd and 3rd equalities by the coherence axiom. Hence, $\mu(w) \leq \mu(w + \varepsilon) \leq x_n$. So μ is increasing. Suppose that the means of M are strictly increasing. Then we have $\mu(w) < x_n$ by (4.9), and $\mu(w) < \mu(w + \varepsilon) < x_n$ by (4.10). So μ is strictly increasing, and x_n is not in the image.

8. (Interval.) This is a corollary of part 7. Let $x, y \in A$ with $x < y$, and let $z \in \mathbb{R}$ such that $x < z < y$; we must show that $z \in A$. Let

$$\mu : \quad [0, \infty) \to A : \quad w \mapsto M_{(w,1)}(x, y).$$

By part 7, the image of μ is the interval $(x, y]$ of $\mathbb{R}$; in particular, z is in the image of μ, hence $z \in A$. □

4.7 Gauss composition of means within a family

The main results in this final section in the part about means in this text, are Theorem 4.7.3: formulas for the "normal" composition of means within the same family, and Theorem 4.7.7: formula for the Gauss composition of means within the same family. We also give a new proof of Theorem 4.7.5, that the limit of powers of a stochastic matrix has identical rows. And in Example 4.7.9, we apply the theorems to show counterexamples to a result and a conjecture in Chapter 3 in the case that we would remove certain conditions there.

***Remark* 4.7.1.** We regard elements of $\mathbb{R}^n$ by default as row vectors.

Definition 4.7.2. Let $k, n \in \mathbb{N}$, let $\mathbf{w}_k \in \mathbb{R}^n_{>0}$ for $k \in \mathbb{N}_{\leq n}$.
Let $\mathbf{W} := (\mathbf{w}_1, \ldots, \mathbf{w}_n) \in \mathbb{R}^{n \times n}_{>0}$ be the matrix with $\mathbf{w}_k$ as its kth row.

1. We say that $\mathbf{W}$ is a *weight matrix*[1] if $\sum \mathbf{w}_k = 1$ for all $k \in \mathbb{N}_{\leq n}$, that is, if all its rows are normalised.

[1]Another name for "weight matrix" is "stochastic matrix with strictly positive entries", especially in the context of Markov chains.

2. Let M be a family of means on some $A \subseteq \mathbb{R}$. We write

$$M_{\mathbf{W}} := (M_{\mathbf{w}_1}, \ldots, M_{\mathbf{w}_n}) \in \mathbb{M}^n_{\mathrm{int}(A^n)},$$

where $\mathbb{M}_{\mathrm{int}(A^n)}$ is the set of internal means on A^n, as in Definition 3.2.2. As in that definition, we regard $M_{\mathbf{W}}$ as a transformation of A^n.

Theorem 4.7.3 (*). *Let $A \subseteq \mathbb{R}$, let M be a family of means on A. Let $n \in \mathbb{N}$, let $\mathbf{W} \in \mathbb{R}^{n\times n}_{>0}$ be a weight matrix.*

1. *Let $\mathbf{v} \in \mathbb{R}^n_{>0}$. Then* $\quad M_{\mathbf{v}} \circ M_{\mathbf{W}} = M_{\mathbf{vW}} \quad \in \mathbb{M}_{\mathrm{int}(A^n)}$.
2. *Let $\mathbf{V} \in \mathbb{R}^{n\times n}_{>0}$. Then* $\quad M_{\mathbf{V}} \circ M_{\mathbf{W}} = M_{\mathbf{VW}} \quad \in \mathbb{M}^n_{\mathrm{int}(A^n)}$.
3. *If $\mathbf{V}$ is a weight matrix, then $\mathbf{VW}$ is a weight matrix.*
4. *For all $m \in \mathbb{N}$, we have* $\quad M^{\circ m}_{\mathbf{W}} = M_{\mathbf{W}^m}$.

Proof. **1.** Let $\mathbf{x} \in A^n$. We write $\mathbf{v} = (v_1, \ldots, v_n)$. By definition, we have

$$M_{\mathbf{v}} \circ M_{\mathbf{W}}(\mathbf{x}) = M_{(v_1,\ldots,v_n)}\big(M_{\mathbf{w}_1}(\mathbf{x}), \ldots, M_{\mathbf{w}_n}(\mathbf{x})\big).$$

Using that $\sum v_k \mathbf{w}_k = v_k$ for all k, it follows by part 4 and 3 Theorem 4.4.1 that

$$\begin{aligned} M_{(v_1,\ldots,v_n)}\big(M_{\mathbf{w}_1}(\mathbf{x}), \ldots, M_{\mathbf{w}_n}(\mathbf{x})\big) &= M_{(v_1,\ldots,v_n)}\big(M_{v_1\mathbf{w}_1}(\mathbf{x}), \ldots, M_{v_n\mathbf{w}_n}(\mathbf{x})\big) \\ &= M_{(v_1\mathbf{w}_1 + \cdots v_n\mathbf{w}_n)}(\mathbf{x}). \end{aligned}$$

Because $v_1\mathbf{w}_1 + \cdots v_n\mathbf{w}_n = \mathbf{vW}$, we conclude that $M_{\mathbf{v}} \circ M_{\mathbf{W}}(\mathbf{x}) = M_{\mathbf{vW}}(\mathbf{x})$, as desired.

2. We write $\mathbf{V} = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$. It suffices to show that for each $1 \le k \le n$, the kth entry of $M_{\mathbf{V}} \circ M_{\mathbf{W}}$ equals the kth entry of $M_{\mathbf{VW}}$.
The kth entry of $M_{\mathbf{V}} \circ M_{\mathbf{W}}$ is, by definition of $M_{\mathbf{V}}$, equal to $M_{\mathbf{v}_k} \circ M_{\mathbf{W}}$,
The kth row of the matrix $\mathbf{VW}$ is $\mathbf{v}_k\mathbf{W}$, so the kth entry of $M_{\mathbf{VW}}$ equals $M_{\mathbf{v}_k\mathbf{W}}$.
By part 1, we have $M_{\mathbf{v}_k} \circ M_{\mathbf{W}} = M_{\mathbf{v}_k\mathbf{W}}$, which concludes the proof.

3. This follows because a matrix in $\mathbb{R}^{n\times n}_{>0}$ is a weight matrix iff the transpose of $(1)^{\bullet n}$ is an eigenvector with eigenvalue 1.

4. This follows by repeated application of part 2 with $\mathbf{V} = \mathbf{W}$. ∎

Example 4.7.4. Let $\mathbf{W} = (\mathbf{w}_1, \ldots, \mathbf{w}_n) \in \mathbb{R}^{n\times n}_{>0}$ be a weight matrix. By Theorem 4.7.3 with $\mathbf{v} := (1/n)^{\bullet n}$, we have

$$M_n \circ M_{\mathbf{W}} = M_{\mathrm{AM}_n(\mathbf{W})}$$

where $\mathrm{AM}_n(\mathbf{W}) := \frac{1}{n}(\mathbf{w}_1 + \ldots + \mathbf{w}_n)$ is a normalised weight vector.

Very informally phrased: "Averaging neutrally (by M) over n weighted means, amounts to averaging neutrally (by AM) over the n weight vectors."

Theorem 4.7.5. *Let $n \in \mathbb{N}$, let $\mathbf{W} \in \mathbb{R}^{n\times n}_{>0}$ be a weight matrix. The limit $\mathbf{V} := \lim_{m\to\infty} \mathbf{W}^m$ exists in $\mathbb{R}^{n\times n}_{>0}$, and all the rows of $\mathbf{V}$ are identical. Moreover, $\mathbf{V}$ is a weight matrix.*

Proof. This is a well-known result in the theory of Markov chains. See for instance [Ros10, Theorem 4.1]; however, it is stated there without proof.

It is interesting however that a new proof follows easily from the theory of means that we developed so far in this text. Namely, by Theorem 4.7.3, we have

$$\mathrm{AM}_{\mathbf{W}}^{\circ m} = \mathrm{AM}_{\mathbf{W}^m}.$$

Let $1 \leq i \leq n$ and $1 \leq j \leq n$. Let $\mathbf{W}_{m,i,j} \in \mathbb{R}_{>0}$ denote the (i,j)th entry of the matrix $\mathbf{W}^m$, and let $\mathbf{W}_{m,i} := (\mathbf{W}_{m,i,1}, \dots, \mathbf{W}_{m,i,n})$ be its ith row. By definition, we have

$$\mathrm{AM}_{\mathbf{W}^m} = \big(\mathrm{AM}_{\mathbf{W}_{m,1}}, \dots, \mathrm{AM}_{\mathbf{W}_{m,n}}\big).$$

Let $\mathbf{e}_j := \mathbf{0}_{[j,1]} \in \mathbb{R}^n$ be the jth unit vector. By Theorem 4.7.3.3 we have $\sum \mathbf{W}_{m,1} = 1$, hence $\mathrm{AM}_{\mathbf{W}_{m,i}}(\mathbf{e}_j) = \mathbf{W}_{m,i,j}$. Thus,

$$\mathrm{AM}_{\mathbf{W}}^{\circ m}(\mathbf{e}_j) = \mathrm{AM}_{\mathbf{W}^m}(\mathbf{e}_j) = \big(\mathbf{W}_{m,1,j}, \dots, \mathbf{W}_{m,n,j}\big), \tag{4.11}$$

which is the jth column of $\mathbf{W}^m$. By Theorem 3.2.4.3, the vector $\mathrm{AM}_{\mathbf{W}}^{\circ m}(\mathbf{e}_j)$ converges to a vector with identical entries, say $(v_j, \dots, v_j)$, as $m \to \infty$. Hence, by (4.11), it follows that

$$\lim_{m\to\infty} \mathbf{W}^m = (\mathbf{v}, \mathbf{v}, \dots, \mathbf{v}), \qquad \text{where } \mathbf{v} = (v_1, v_2, \dots, v_n).$$

That $\lim_{m\to\infty} \mathbf{W}^m$ is a weight matrix, follows simply because $\mathbf{W}^m$ is a weight matrix for each m, and because the map $\mathbb{R}^n_{>0} \to \mathbb{R}_{>0} : \mathbf{w} \mapsto \sum \mathbf{w}$ is continuous. □

Definition 4.7.6. Let $n \in \mathbb{N}$, let $\mathbf{W} \in \mathbb{R}^{n\times n}_{>0}$ be a weight matrix. We define the vector $\bigotimes \mathbf{W} \in \mathbb{R}^n_{>0}$ by

$$\operatorname{diag}_n(\bigotimes \mathbf{W}) = \lim\nolimits_{m\to\infty} \mathbf{W}^m;$$

its existence is ensured by Theorem 4.7.5. ≀

Theorem 4.7.7 (*)**.** *Let $A \subseteq \mathbb{R}$, let M be a family of compressing means on A. Let $n \in \mathbb{N}$, let $\mathbf{W} \in \mathbb{R}^{n\times n}_{>0}$ be a weight matrix. Then*

$$\bigotimes M_{\mathbf{W}} = M_{\bigotimes \mathbf{W}}.$$ ⅃

Proof. We have

$$\operatorname{diag}_n(\bigotimes M_{\mathbf{W}}) = \lim_{m\to\infty} M_{\mathbf{W}}^{\circ m} = \lim_{m\to\infty} M_{\mathbf{W}^m} = M_{\operatorname{diag}_n(\bigotimes \mathbf{W})} = \operatorname{diag}_n(M_{\bigotimes \mathbf{W}});$$

the 1st equality by definition of $\bigotimes M_{\mathbf{W}}$, the 2nd equality by Theorem 4.7.3, the 3rd equality by Theorem 4.4.1 ("Continuity in the weights"), the 4th equality by the 'row-wise' definition 4.7.2.2 of the operator $M_{\bigotimes \mathbf{W}}$. □

Corollary 4.7.8 ((*) Gauss composition of power/translation means of one parameter)**.** *Let $p \in \mathbb{R}$, let $b \in \overline{\mathbb{R}}$. Let $\mathbf{W} = (\mathbf{w}_1, \dots, \mathbf{w}_n) \in \mathbb{R}^{n\times n}_{>0}$ be a weight matrix. Then we have*

$$\mathcal{PM}_{(\operatorname{diag}_n(p),(\mathbf{w}_1,\dots,\mathbf{w}_n))} = \mathrm{PM}_{p,\bigotimes \mathbf{W}}, \qquad \mathcal{TM}_{(\operatorname{diag}_n(b),(\mathbf{w}_1,\dots,\mathbf{w}_n))} = \mathrm{TM}_{b,\bigotimes \mathbf{W}}.$$ ⅃

Example 4.7.9 (*)**.** For $(i,j) \in \{1,2\}^2$, let $\mathbf{w}_{ij} \in \mathbb{W}$ be given by

$$\begin{pmatrix} \mathbf{w}_{11} & \mathbf{w}_{12} \\ \mathbf{w}_{21} & \mathbf{w}_{22} \end{pmatrix} = \begin{pmatrix} (\frac{1}{3}, \frac{2}{3}) & (\frac{2}{5}, \frac{3}{5}) \\ (\frac{5}{7}, \frac{2}{7}) & (\frac{3}{4}, \frac{1}{4}) \end{pmatrix}$$

Consider the weight matrices $\mathbf{W}_1 := (\mathbf{w}_{11}, \mathbf{w}_{12})$ and $\mathbf{W}_2 := (\mathbf{w}_{21}, \mathbf{w}_{22})$. By matrix diagonalisation, it is easily computed that

$$\lim_{m\to\infty} \mathbf{W}_1^m = \begin{pmatrix} \frac{3}{8} & \frac{5}{8} \\ \frac{3}{8} & \frac{5}{8} \end{pmatrix}, \qquad \lim_{m\to\infty} \mathbf{W}_2^m = \begin{pmatrix} \frac{21}{29} & \frac{8}{29} \\ \frac{21}{29} & \frac{8}{29} \end{pmatrix}.$$

Therefore, we have

$$M_{\mathbf{w}_{11}} \otimes M_{\mathbf{w}_{12}} = \bigotimes M_{\mathbf{W}_1} = M_{(\frac{3}{8},\frac{5}{8})}, \qquad M_{\mathbf{w}_{21}} \otimes M_{\mathbf{w}_{22}} = \bigotimes M_{\mathbf{W}_2} = M_{(\frac{21}{29},\frac{8}{29})}.$$

Further, we similarly compute that

$$\lim_{m\to\infty} \begin{pmatrix} \frac{3}{8} & \frac{5}{8} \\ \frac{21}{29} & \frac{8}{29} \end{pmatrix}^m = \begin{pmatrix} \frac{168}{313} & \frac{145}{313} \\ \frac{168}{313} & \frac{145}{313} \end{pmatrix},$$

hence it follows that

$$(M_{\mathbf{w}_{11}} \otimes M_{\mathbf{w}_{12}}) \otimes (M_{\mathbf{w}_{21}} \otimes M_{\mathbf{w}_{22}}) = M_{(\frac{168}{313},\frac{145}{313})}.$$

By analogous calculation, we find that

$$(M_{\mathbf{w}_{11}} \otimes M_{\mathbf{w}_{21}}) \otimes (M_{\mathbf{w}_{12}} \otimes M_{\mathbf{w}_{22}}) = M_{(\frac{145}{271},\frac{126}{271})}.$$

The numbers $168/313 \approx 0.5367$ and $145/271 \approx 0.5351$, though close together, are not equal. Thus we conclude that there exists examples that satisfy

$$\bigotimes_{j=1}^{n} \Big(\bigotimes_{i=1}^{n} M_{\mathbf{w}_{ij}} \Big) \neq \bigotimes_{i=1}^{n} \Big(\bigotimes_{j=1}^{n} M_{\mathbf{w}_{ij}} \Big).$$

Moreover, we have

$$M_{\mathbf{w}_{11}} \otimes M_{\mathbf{w}_{12}} = M_{(\frac{3}{8},\frac{5}{8})}, \qquad M_{\mathbf{w}_{12}} \otimes M_{\mathbf{w}_{11}} = M_{(\frac{5}{9},\frac{4}{9})}.$$

Hence there exist examples that satisfy $\bigotimes \mathbf{M}_{\circ\sigma} \neq \bigotimes \mathbf{M}$. ♪

Part II

Topics in the theory of discrete dynamical systems

Chapter 5

Discrete dynamical systems and flow graphs

5.1 Iteration

Definition 5.1.1. Let S be a set, let $x \in S$, let $\varphi : S \to S$ be a function.

1. φ is called a *transformation* of S.
2. The pair (S, φ) is called a *discrete dynamical system*.
   We usually refer to the elements of S as points.
3. Let $n \in \mathbb{N}$. The *nth iterate of* φ, denoted by $\varphi^{\circ n} : S \to S$,
   is the n-fold composition $\varphi \circ \cdots \circ \varphi$ if $n \geq 1$, and is id_S if $n = 0$.
4. The *orbit (under* φ*)* of x is the set $\{\varphi^{\circ n}(x) : n \in \mathbb{N}_0\} \subseteq S$.
5. The *flow (under* φ*)* of x is the function $\mathbb{N}_0 \to S : x \mapsto \varphi^{\circ n}(x)$,
   which we usually regard as a sequence $\big(\varphi^{\circ n}(x)\big)_{n \in \mathbb{N}_0}$ in S. ≀

5.1.1 Example: Multi-dimensional recursion

The following example, and the next one which is in fact an example of the example, serves mainly to show how general the previous definition is.

Discrete dynamical systems arise in a wide variety of contexts, in pure mathematics and in applications. A common example arises in modelling real-world systems that change over time, in which case $\varphi^{\circ n}$ represents the state of the system at time n. The definition of discrete dynamical systems may seem too limited to describe many real-world systems, for example when the current state depends on more than one state in the past (this is the case for discretisations of differential equations involving higher order derivatives), or when the behaviour of the system depends on the time parameter. But it turns out that such systems are encompassed by our definition as well, by "storing the time parameter and enough of the past into the transformation", in the following sense.

Example 5.1.2. Let S be a set, let $m \in \mathbb{N}$, let $f : S^m \times \mathbb{N}_0 \to S$ be a function; we informally interpret $\mathbb{N}_0$ as "discrete time", and S as the "state space". Consider the "dynamical system" that is defined in the following way. For $n \in \mathbb{N}$, we write $x_n \in S$ for "the state of the system at time n". Given the first m "initial states" $x_0, \ldots, x_{m-1} \in S$ of the system, the remaining states are recursively given by

$$x_n = f(x_{n-1}, \ldots, x_{n-m}, n) \qquad \forall n \geq m.$$

Thus, the state at time n depends on the m previous states and on the time.

We reformulate this system as a discrete dynamical system in terms of our definition. For $n \geq m-1$, let

$$\mathbf{x}_n := (x_n, x_{n-1}, \ldots, x_{n-(m-1)}, n)^T \in S^m \times \mathbb{N}_0,$$

and let

$$\varphi : \quad S^m \times \mathbb{N}_0 \to S^m \times \mathbb{N}_0 : \quad \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \\ n \end{pmatrix} \mapsto \begin{pmatrix} f(y_1, \ldots, y_m, n+1) \\ y_1 \\ \vdots \\ y_{m-1} \\ n+1 \end{pmatrix}.$$

The discrete dynamical system $(S^m \times \mathbb{N}_0, \varphi)$ mirrors the system defined by f: Namely, for $n \geq m$, we have

$$\varphi(\mathbf{x}_{n-1}) = \varphi \begin{pmatrix} x_{n-1} \\ x_{n-2} \\ \vdots \\ x_{n-m} \\ n-1 \end{pmatrix} = \begin{pmatrix} f(x_{n-1}, \ldots, x_{n-m}, n) \\ x_{n-1} \\ \vdots \\ x_{n-(m-1)} \\ n \end{pmatrix} = \begin{pmatrix} x_n \\ x_{n-1} \\ \vdots \\ x_{n-(m-1)} \\ n \end{pmatrix} = \mathbf{x}_n.$$

Thus, the system defined by f is captured by φ, in each of the m copies of S.

Clearly, if the function f does not depend on the time n, we can proceed analogously by hiding the time dimension; thus, in that case we have $f : S^m \to S$ and $\varphi : S^m \to S^m$.

Example 5.1.3 (Generalised Fibonacci sequences)**.** We consider a concrete case of Example 5.1.2, where (as in the last sentence of that example) the function f does not depend on time. Namely, we let

$$f : \quad \mathbb{C}^2 \to \mathbb{C} : \quad (x, y) \mapsto x + y.$$

Thus, for any $(x_0, x_1) \in \mathbb{C}^2$, we get a sequence $(x_n)_{n \in \mathbb{N}_0}$ in $\mathbb{C}$ that is recursively defined by

$$x_n = f(x_{n-1}, x_{n-2}) = x_{n-1} + x_{n-2} \qquad \forall n \geq 2.$$

In particular, for $(x_0, x_1) = (0, 1)$ we get the well-known Fibonacci sequence.

The corresponding discrete dynamical system, as in Example 5.1.2, is $(\mathbb{C}^2, \varphi)$, where $\varphi : \mathbb{C}^2 \to \mathbb{C}^2$ is the linear map that is given by the matrix $\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$ with respect to the standard basis. Thus, writing $\mathbf{x}_n := (x_n, x_{n-1})^T \in \mathbb{C}^2$ for $n \geq 1$, we have $\varphi(\mathbf{x}_{n-1}) = \mathbf{x}_n$ for all $n \geq 2$.

The eigenvalues of φ are

$$\alpha := \frac{1+\sqrt{5}}{2}, \qquad \beta := \frac{1-\sqrt{5}}{2},$$

and the corresponding eigenvectors

$$\mathbf{v} := \begin{pmatrix} \alpha \\ 1 \end{pmatrix}, \qquad \mathbf{w} := \begin{pmatrix} \beta \\ 1 \end{pmatrix}$$

form a basis of $\mathbb{C}^2$. Thus, writing the "initial values vector" $\mathbf{x}_1$ in terms of the basis of eigenvectors, that is, as

$$\mathbf{x}_1 = a\mathbf{v} + b\mathbf{w} \qquad \text{for some } (a, b) \in \mathbb{C}^2,$$

we get the formula

$$\mathbf{x}_n = a\alpha^{n-1}\mathbf{v} + b\beta^{n-1}\mathbf{w} \qquad \forall n \geq 1,$$

and hence the explicit formula

$$x_n = a\alpha^n + b\beta^n \qquad \forall n \geq 1.$$

Because $|\alpha| > 1$ and $|\beta| < 1$, it follows that $x_n \sim a\alpha^n$ as $n \to \infty$, except when $a = 0$. Thus, if $x_1 \neq \beta x_0$, then the sequence $x_0, x_1, x_2, \ldots$ grows by good approximation as an exponential function with as growth factor the golden ratio, α.

For the initial values $(x_0, x_1) = (0, 1)$, we have $(a, b) = (5^{-1/2}, -5^{-1/2})$. Hence, the explicit formula for the nth Fibonacci number x_n is

$$x_n = \frac{\alpha^n - \beta^n}{\sqrt{5}}.$$

This has been named Binet's Formula, after the French mathematician J.P.M. Binet (1786–1856). He rediscovered the formula, but it was already known to some earlier mathematicians, certainly to Euler and De Moivre; see [Liv03, p. 108]. ♪

5.2 Periodicity

The most essential property of the flow of a point x is that it is either injective, or it has a repeating tail (see Fact 5.2.2.1). In the case of a repeating tail, x is called a *preperiodic* point, as in the next definition.

Definition 5.2.1. Let (S, φ) be a discrete dynamical system, let $x \in S$.

1. x is called *periodic* if there exists $n \geq 1$ such that $\varphi^{\circ n}(x) = x$. The number n is called a *period* of x, and the smallest period of x is called its *primitive period.*

2. x is called a *fixed point* if it has (primitive) period 1.

3. x is called *preperiodic* if $\varphi^{\circ m}(x)$ is periodic for some $m \geq 0$. The smallest such m is called the *preperiodic length* of x. The *eventual primitive period* of x is the primitive period of $\varphi^{\circ m}(x)$.

4. $\mathrm{Per}(S, \varphi, n) := \{x \in S : \varphi^{\circ n}(x) = x\}$ is the set of points with period n.

5. $\mathrm{Per}^*(S, \varphi, n) := \mathrm{Per}(S, \varphi, n) \setminus \bigcup_{k=1}^{n-1} \mathrm{Per}(S, \varphi, k)$ is the set of points with primitive period n.

6. $\mathrm{Per}(S, \varphi) := \bigcup_{n \geq 1} \mathrm{Per}(S, \varphi, n)$ is the set of all periodic points.

We sometimes omit S or φ from the notation if the context makes the meaning clear. ≀

In Fact 5.2.2, we mention two basic facts about periodicity.

Fact 5.2.2. *Let (S, φ) be a discrete dynamical system, let $x \in S$, let $n \in \mathbb{N}$.*

1. *The point x is preperiodic $\iff$ The flow of x is not injective.*

2. *We have* $\mathrm{Per}(S, \varphi, n) = \bigsqcup_{d|n} \mathrm{Per}^*(S, \varphi, d)$. ♩

Proof. **1.** *"$\Longrightarrow$":* Clear. *"$\Longleftarrow$":* Suppose that the flow of x is not injective, say $\varphi^{\circ k}(x) = \varphi^{\circ \ell}(x)$ for some $k, \ell \in \mathbb{N}_0$ with $k < \ell$. Then $\varphi^{\circ k}(x)$ is periodic with period $\ell - k$, so x is preperiodic.

2. Let $x \in \mathrm{Per}(S, \varphi)$, let d_x be the primitive period of x. We have

$$x \in \mathrm{Per}(S, \varphi, n) \iff x \text{ has period } n \iff d_x \text{ divides } n \iff x \in \bigsqcup_{d|n} \mathrm{Per}^*(S, \varphi, d). \quad \square$$

5.2.1 Example: Contractions

In many interesting examples of discrete dynamical systems, the set S has additional structure (like that of a topological space, or a smooth manifold, or an algebraic variety), and that structure is reflected by the map φ. Often, the extra structure is essential for even defining the deeper properties of the dynamical system; an example is the ergodicity property, which can only be defined if S is a measurable space with a finite measure, and φ is a measurable map.

The next proposition, a generalised version of the Banach fixed point theorem, serves as an example in two ways: the dynamical system has additional structure, and some terminology introduced in the previous definitions is illustrated.

Proposition 5.2.3. *Let (X, d) be a metric space and let $\varphi : X \to X$ be a contraction, that is, there exists $c \in [0, 1)$ such that*

$$d\big(\varphi(x), \varphi(y)\big) \leq c\, d\big(x, y\big) \qquad \forall x, y \in X.$$

Then the discrete dynamical system (X, φ) has the following properties:

1. *There are no periodic points except for fixed points.*
2. *There is at most one fixed point.*
3. *If there is a fixed point x^*, then every flow converges to it: $\lim_{n\to\infty} \varphi^{\circ n}(x) = x^*$ for all $x \in X$.*
4. *If X is complete, then there is a fixed point.*

Proof. For any two points $x, y \in X$, and any $n \in \mathbb{N}_1$, we have

$$d(\varphi^{\circ n}(x), \varphi^{\circ n}(y)) \leq c^n d(x, y) \prec d(x, y), \qquad \text{where } \prec \text{ means} \begin{cases} < & \text{if } x \neq y \\ = & \text{if } x = y. \end{cases} \tag{5.1}$$

1. If $p \in X$ has period n, then by applying (5.1) to $(x, y) = (p, \varphi(p))$, the instance of "$\prec$" must be a "$=$". Hence $p = \varphi(p)$, so p is a fixed point.

2. If x^*, x^{**} are fixed points, then by applying (5.1) to $(x, y) = (x^*, x^{**})$ and (for instance) $n = 1$, the instance of "$\prec$" must be a "$=$". Hence, $x^* = x^{**}$.

3. This follows from (5.1) by taking $y := x^*$ and noting that $c^n \to 0$ as $n \to \infty$.

4. Let $x \in X$ be any point and write $x_n := \varphi^{\circ n}(x)$ for all n. Let $m, k \in \mathbb{N}_1$ be natural numbers. By repeatedly applying (5.1) and the triangle inequality, it is easily seen that

$$d(x_{m+k}, x_m) \leq c^m d(x_k, x_0) \leq c^m(c^{k-1} + \ldots + c + 1)d(x_1, x_0) \leq \frac{c^m}{1-c} d(x_1, x_0).$$

Therefore, $(x_n)_n$ is a Cauchy sequence, and by completeness it converges to a limit $x^* \in X$. This is a fixed point, because $\varphi(x^*) = \lim_{n\to\infty} \varphi(x_n) = \lim_{n\to\infty} x_{n+1} = x^*$, where the first equality is valid because φ is continuous (it is even Lipschitz continuous). □

***Remark* 5.2.4.** In the setting of Proposition 5.2.3, all the preperiodic points have eventual period 1, but about the preperiodic lengths that occur, we cannot say anything in general. For example, let $0 < c < 1$ and let $b < 0$, then the function $\varphi(x) := cx + b$ is clearly a contraction on $\mathbb{R}$, and it is easily seen that $(\mathbb{R}, \varphi)$ has no preperiodic points other then the fixed point $b/(1 - c)$ itself. On the other extreme, the function $\psi(x) := \mathrm{Max}(0, \varphi(x))$ is a contraction on $\mathbb{R}_{\geq 0}$, and 0 is the fixed point of $(\mathbb{R}_{\geq 0}, \psi)$. For all $n \geq 0$, the point

$$a_n := \frac{-b}{1-c}\frac{1-c^n}{c^n}$$

satisfies $\varphi^{\circ n}(a_n) = 0$; using that $\varphi(x) < x$ for all x and that $\varphi(x)$ is a strict monotonic increasing function of x, it follows that the interval $(a_n, a_{n+1}]$ consists of preperiodic points for ψ with preperiodic length $n + 1$. Thus, every point of $(\mathbb{R}_{\geq 0}, \psi)$ is preperiodic, and every possible preperiodic length occurs.

5.3 Flow graphs

***Remark* 5.3.1.** Before we discuss flow graphs, we must be clear about our terminology concerning graphs.

1. We allow graphs to have multiple edges, that is, edges that have the same end nodes. Such graphs are commonly called *multigraphs* in the literature, but we just call them graphs.

2. By a *digraph*, we mean a directed graph, that is, a directed multigraph.

3. By an *arrow*, we mean a directed edge, i.e. an edge of a digraph.

4. The definitions in 5.3.2 differ slightly (though not essentially) from definitions in some other sources. For example, a walk is often defined as a sequence of edges only, while we include the nodes, mainly to avoid ambiguities caused by the fact that the end nodes of an edge are a priori unordered.

Definition 5.3.2. Let G be a graph; let H be a digraph.

1. A *walk* in G is a sequence of alternating nodes and edges, containing at least one node, such that each edge in the sequence has as its two neighbours precisely the endpoints of the edge, in any order. (So "the two neighbours" are equal iff the edge is a loop.) The sequence may be finite or infinite in either direction, but the first and last entries, if they exist, must be nodes.

   If the sequence is finite, say it starts with node a and ends with node b, then we say that it is a walk from a to b. The *length* of the walk is the number of edges in the walk, counted with multiplicity.

   A walk of length 0 is called a *trivial* walk; it consists of a single node.

2. A *trail* in G is a walk in which all edges are distinct.

3. A *path* in G is a walk in which all nodes are distinct.

4. A *cycle* in G is a finite walk with at least one edge, in which all nodes are distinct, except that the first and the last node coincide.

5. A *directed walk* in H is a walk in the underlying graph, such that all arrows 'point in the same direction'; concretely, when we write the walk as $(\ldots, x_{k-1}, e_{k-1}, x_k, e_k, x_{k+1}, \ldots)$, where the x_i are nodes and the e_i are arrows and the indices i are in some interval of integers, the walk is a directed walk iff each arrow e_k in the sequence is directed from x_k to x_{k+1}.

6. A *directed trail*, a *directed path*, a *directed cycle* are defined completely analogous as a directed walk, just replacing the word 'walk' in the definition by 'trail', 'path', 'cycle' respectively.

Fact 5.3.3. *For every node, there is the trivial path from the node to itself.*

Each path is a trail, and each directed path is a directed trail. (But not vice versa.)

Lemma 5.3.4. *Let G be a [directed] graph, and x and y nodes of G.*

1. *If there is a [directed] walk from x to y, then there is a [directed] path from x to y.*
2. *If there is a [directed] walk from x to x containing a node different from x, then there is a [directed] cycle from x to x containing a node different from x.*

Proof.

1. If the walk is a path, we are done. Otherwise, choose a node that is repeated in the walk, delete everything between the first and the last repetition of that node. Repeat this procedure for the resulting walk, until the resulting walk is a path.

If we started with a directed walk, then the result is a directed path.

2. Let y be the first node in the walk different from x, so the walk starts with a number of loops, then an edge e joining x and y, and then a walk from y to x. By part 1, there is a path p from y to x. The concatenation (x, e, p) is a cycle from x to x, containing the node $y \neq x$.

If we started with a directed walk, then the result is a directed cycle. □

Definition 5.3.5. Let H be a digraph, and x, y, z nodes of H.

On the set of nodes of H, we define the *successor* binary relation by calling z a successor of x if there is a directed path from x to z, and denote this by $z < x$ (or equivalently $x > z$).[1]

We call z a direct successor of x if there is an arrow from z to x.

Lemma 5.3.6. *The successor relation on the node set of a digraph H is reflexive and transitive; it is antisymmetric if and only if every directed cycle in H consists of loops.*

Proof. Reflexive: there is a trivial path from each node to itself.

Transitive: If $x > y$ and $y > z$, then by concatenation there is a directed walk from x to z, hence by Fact 5.3.4.1 there is a directed path from x to z, so $x > z$.

Antisymmetric: Suppose '$>$' is antisymmetric. For any directed cycle and any two nodes x and y in that cycle, we clearly have $x > y$ and $y > x$, hence $x = y$; hence, any directed cycle contains only one distinct node, and therefore consists of loops. Conversely, suppose '$>$' is not antisymmetric, so there are two distinct nodes x and y such that $x > y$ and $y > x$. By concatenation, there is a directed walk from x to x containing the node $y \neq x$. By Fact 5.3.4.2, there is a directed cycle from x to x containing a node different from x, so that cycle does not entirely consist of loops. □

Definition 5.3.7. Let S be a set.

1. A *function digraph on S* is a digraph with node set S and the property that each node has a unique direct successor.

[1] Our choice of '$<$' vs. '$>$' is such that it resembles an arrow in the right direction.

2. Let $\varphi : S \to S$ be a function. The *flow graph of* (S, φ) is the function digraph G_φ with node set S, and with an arrow from s to t iff $\varphi(s) = t$.

3. Let G be a function digraph on S. The *discrete dynamical system of* G is (S, φ_G), where for each $s \in S$, $\varphi_G(s)$ is the unique direct successor of s in G. 𝄽

***Remark* 5.3.8.** Clearly, the two operations from the previous definition that translate between the two kinds of structures – discrete dynamical systems and function digraphs – are inverse to each other, and properties of the one structure are reflected by the other structure. Thus, studying discrete dynamical systems on S is essentially equivalent to studying function digraphs on S.

Example 5.3.9. Let p be a prime number, and consider the polynomial function

$$\varphi : \mathbb{F}_p \to \mathbb{F}_p : \quad x \mapsto x^2 + 1.$$

The flow graphs of $(\mathbb{F}_p, \varphi)$ for $p = 17, 29, 31$ are depicted in Figure 5.1 on page 101. Note that the underlying graphs for $p = 17, 29$ are connected, while that for $p = 31$ has three connected components. ♪

Definition 5.3.10 (Converging arborescence)**.** Let G be a directed tree, let x be a node of G. We say that G an *arborescence converging to* x, and that x is *the convergence node of* G, if $y > x$ for all nodes y of G. 𝄽

***Remark* 5.3.11.** A converging arborescence is completely determined by its convergence node and its underlying graph: the directions of the arrows follow, because in a tree, for each pair of nodes there is a unique path from the one node to the other node.

Proposition 5.3.12 (The structure of flow graphs)**.** *Let (S, φ) be a discrete dynamical system. Let G be the flow graph of (S, φ), and let G' be a connected component of the underlying graph of G. We view G' as a digraph, inheriting the directions from G.*

1. Every two nodes x, y of G' have a common successor.

2. G' has at most one directed cycle (up to choice of starting point of a cycle).

3. (†) If G' has no directed cycle, then it is an infinite directed tree, but not an arborescence. Its node set is partially ordered by the successor relation. Each finite set of nodes has a unique first common successor.

4. (†) If G' has a directed cycle, then G' is the concatenation of that cycle with, for each node in the cycle, an arborescence converging to that node. In other words, G' could be viewed as the disjoint union of finitely many converging arborescences, together with a cyclic order on the set of their convergence nodes. 𝅗𝅥

Proof. For convenience, we represent walks in G by the sequence of directions (depicted by $\to$ and $\leftarrow$) of the edges in the walk.

1. Because G' is connected, there is a path p between x to y. The path can't contain '$\leftarrow, \to$' as a subsequence, because each node in a flow graph has only one direct successor. Hence, either p is a directed path, or p consists of a subsequence of only '$\rightarrow$' arrows followed by a subsequence of only '$\leftarrow$' arrows. In the former case, y is a common successor of x and y if p is directed from x to y, and vice versa. In the latter case, the node between the two subsequences is a common successor of x and y.

2. A path between two nodes residing in distinct directed cycles would necessarily start with '$\leftarrow$' and end with '$\rightarrow$', because the outgoing arrows from those nodes would result in a repeated node in the path. Hence, the path would contain a subsequence '$\leftarrow, \rightarrow$', contradicting that each node has a unique direct successor.

3. If the underlying graph of G' is not a tree, it contains a cycle. The directions of the arrows in the cycle must all be the same, for otherwise there would be a path from a node in the cycle to itself that contains a subsequence '$\leftarrow, \rightarrow$'. So G' contains a directed cycle.

Thus, suppose G' has no directed cycle, then G' is a directed tree. The nodes in G' are not periodic points, so their forward orbits are infinite, so G' is infinite. It is not an arborescence, because there cannot be an outgoing arrow from a convergence node. The node set is partially ordered by succession, by Fact 5.3.6.

4. For every node x in G', there exists a directed path from that node to any node y in the cycle. Namely, it is clear if x is in the cycle, so suppose x is not in the cycle. There is a path from y to x by connectedness, and it must start with '$\leftarrow$' because the outgoing arrow from y would lead to repetition in the path, hence all the directions in the path must be '$\leftarrow$', because subsequences '$\leftarrow, \rightarrow$' cannot be. So it is a directed path from x to y.

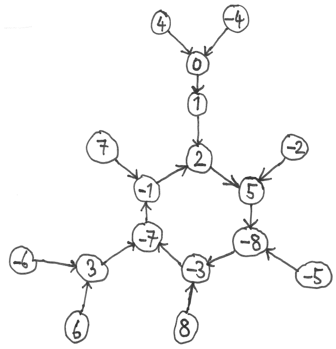
In other words, every node in G' is a preperiodic point, with as its eventual periodic orbit the set of the nodes in the cycle. For each node y in the cycle, we consider the full directed subgraph G'_y of G' containing those nodes x such that y is the first periodic element in the flow of x. Clearly, the node set of G' is the disjoint union of the node sets of the G'_y. Each G'_y is a tree, because otherwise it would contain a directed cycle (by the same argument as in part 3), which would contradict part 2. The only arrows in G' between the subgraphs G'_y are the arrows in the cycle, because if x is not a periodic point and y is the first periodic point in the flow of x (i.e. x is in G'_y), then y is also the first periodic point in the flow of the direct successor x' of x (i.e. x' is also in G'_y). Therefore, each graph G'_y is an arborescence converging to y, and G' is the concatenation of the graphs G'_y and the cycle. □

5.4 Invariants of flow graphs

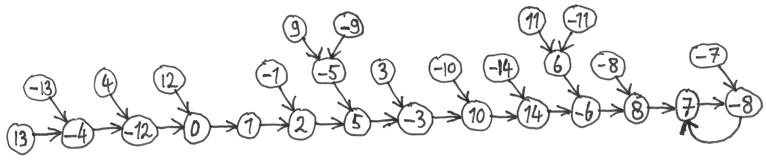
Definition 5.4.1. Let G be a flow graph, and let $k \in \mathbb{N}_1$.

1. γ_k denotes the number of cycles in G of length k.

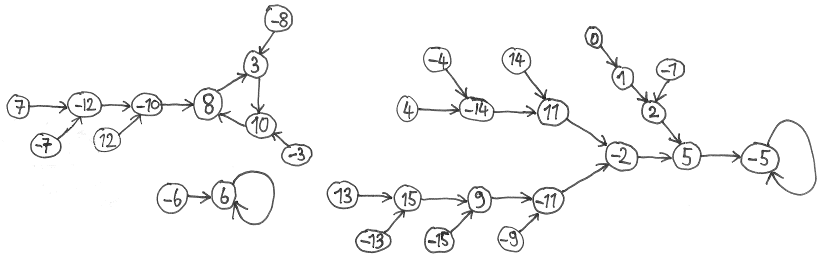
2. $\gamma := \sum_{k \in \mathbb{N}_1} \gamma_k$ denotes the number of cycles in G; equivalently, the number of connected components of the underlying graph.



(a) $p = 17$



(b) $p = 29$



(c) $p = 31$

Figure 5.1: Flow graphs of the polynomial map $x \mapsto x^2 + 1$ on some finite fields $\mathbb{F}_p$

3. $\overline{\gamma} := \sum_{k \in \mathbb{N}_1} k\gamma_k = |\mathrm{Per}(G)|$ denotes the number of periodic points in G. 𝄽

Example 5.4.2. Consider Figure 5.1, where the flow graphs of the function $x \mapsto x^2 + 1$ on the finite fields $\mathbb{F}_p$ for $p = 17, 29, 31$ are depicted.

1. For $p = 17$ we have $\gamma_6 = 1$, the other γ_k are zero; hence, $\gamma = 1$ and $\overline{\gamma} = 6$.

2. For $p = 29$ we have $\gamma_1 = 1$, the other γ_k are zero; hence, $\gamma = 1$ and $\overline{\gamma} = 1$.

3. For $p = 31$ we have $\gamma_1 = 2$ and $\gamma_3 = 1$, the other γ_k are zero; hence, $\gamma = 3$ and $\overline{\gamma} = 5$. ♪

Example 5.4.3. Many other invariants can be defined, but we have no space to study them here. For example:

- the size of the largest component;
- the size of the smallest component;
- the geometric mean of the sizes of the components (the arithmetic mean is trivial);
- the geometric mean of the lengths of the cycles (the arithmetic mean is already $\overline{\gamma}/n$);
- the size of the largest arborescence attached to a cycle;
- the average number of branches at a branch point;
- the average number of ancestors;
- the average number of descendants.

We do investigate some of these invariants in the final (experimental) section in the next chapter. ♪

Chapter 6

Random discrete dynamical systems

6.1 Introduction and statement of the theorem

Throughout this section, let $n \in \mathbb{N}$, let S be a set of cardinality n, and let $\varphi : S \to S$ be a transformation of S.

There are n^n different transformations φ of S. Equivalently, there are n^n flow graphs on S, although the number of isomorphy classes is smaller. In this chapter, we consider questions that can be phrased like this: Given some property of flow graphs that is invariant under isomorphism, which proportion of the n^n flow graphs on S have that property? In other words, imposing the uniform probability distribution on the set of flow graphs on S, with what probability does a random flow graph have the property?

Example 6.1.1. The property that the flow graph of φ consists of cycles only, i.e. that φ is a permutation, can be written in the notation from Definition 5.4.1 as "$\overline{\gamma} = n$". There are $n!$ permutations of S, so we have

$$\mathbf{P}(\overline{\gamma} = n) = n!/n^n. \tag{6.1}$$

By Stirling's formula, we have $n!/n^n = \sqrt{2\pi n}\, e^{-n}(1 + o(n))$ as $n \to \infty$. So the proportion of permutations decreases exponentially with n. ♪

***Remark* 6.1.2.** Formula (6.1) is a special case of part 1a of Theorem 6.1.5 below. The theorem, which is the main result of this chapter, contains exact and asymptotic formulas about the probability distributions of the invariants $\overline{\gamma}, \gamma, \gamma_k$ from Definition 5.4.1.

Theorem 6.1.5 is based on [Bol01, Theorem 14.33], which is a nice overview of combinatorial and asymptotical probabilistic formulas about flow graphs. Theorem 14.33 in [Bol01] contains some more results than our Theorem 6.1.5; in particular, it treats some additional properties of flow graphs.

Most proofs in [Bol01, Thm. 14.33] leave details to the reader; some proofs are left as exercises to the reader. In this chapter, we present detailed proofs of the (corrected)[1]

[1] One correction; see the statement of Theorem 6.1.5.2a

results, with error terms in the asymptotic formulas that are more accurate than in [Bol01]. The proofs of the exact formulas are detailed versions of those in [Bol01, Thm. 14.33]; the proofs of the asymptotic formulas are our own, although the overall ideas that led us to the proofs are contained in Chapter 5 and 14 of [Bol01].

Definition 6.1.3. Let $n, k \in \mathbb{N}_0$. We write $n^{\underline{k}} := \prod_{i=0}^{k-1}(n-i)$.

***Remark* 6.1.4.** In many sources, $n^{\underline{k}}$ is denoted by $(n)_k$, and is called *the falling factorial.* The notation $n^{\underline{k}}$ was invented by Donald Knuth [Knu97, p.50]. We chose for this notation to clarify the relation between n^k and $n^{\underline{k}}$ in formulas.

Theorem 6.1.5. ((*)[2] Exact and asymptotic formulas for random flow graphs)**.** *Let $n \in \mathbb{N}$, let S be a set of cardinality n, and let X be a uniformly random element of the set of transformation of S; that is, $\mathbf{P}(X = \varphi) = n^{-n}$ for all transformations φ of S.*

Let the random variables $\overline{\gamma}, \gamma, \gamma_k$ be the properties of the transformation φ as described in Definition 5.4.1. The probability distributions of $\overline{\gamma}, \gamma, \gamma_k$ have the following properties.

1a. The probability that X has precisely k periodic points (for any $k \geq 0$) is

$$\mathbf{P}(\overline{\gamma} = k) = \frac{k}{n}\frac{n^{\underline{k}}}{n^k} \tag{6.2}$$

$$= \frac{k}{n}\left(1 + O\left(\frac{k^2}{n}\right)\right) \quad \text{if } k = O(n^{1/2}). \tag{6.3}$$

1b. The expected number of periodic points is

$$\mathbf{E}(\overline{\gamma}) = \sum_{k=1}^{n} \frac{k^2}{n}\frac{n^{\underline{k}}}{n^k} \tag{6.4}$$

$$= \left(\frac{\pi}{2}\right)^{1/2} n^{1/2} + O(n^{\varepsilon}), \quad \text{for all } \varepsilon > 0. \tag{6.5}$$

2a. The probability that the flow graph of X is connected, is

$$\mathbf{P}(\gamma = 1) = \frac{1}{n}\sum_{k=1}^{n} \frac{n^{\underline{k}}}{n^k} \tag{6.6}$$

$$= \left(\frac{\pi}{2}\right)^{1/2} n^{-1/2} + O(n^{-1+\varepsilon}), \quad \text{for all } \varepsilon > 0. \tag{6.7}$$

2b. The expected number of connected components is[3]

$$\mathbf{E}(\gamma) = \sum_{k=1}^{n} \frac{1}{k}\frac{n^{\underline{k}}}{n^k} \tag{6.8}$$

$$= \frac{\log(n)}{2} + \frac{\log(2) + \Gamma}{2} + O(n^{-1/3}), \tag{6.9}$$

[2] (*) applies only to the error terms in (6.5), (6.7) and (6.9).

[3] In [Bol01, Thm. 14.33.iii], the asymptotic formula is incorrectly stated as $\mathbf{E}(\gamma) \sim \log(n)$, the corrected formula is $\mathbf{E}(\gamma) \sim \frac{1}{2}\log(n)$. The corrected formula, and a proof of it using the complex analysis of certain generating functions, are contained in [FO90, Theorem 2].

where $\Gamma \approx 0.58$ *is the Euler-Mascheroni constant.*[4]

3a. *For any fixed* $k \geq 1$,

$$\gamma_k \xrightarrow{d} \mathrm{Pois}(1/k),$$

that is, γ_k *tends in distribution to the Poisson distribution with mean* $1/k$, *as* $n \to \infty$. *Concretely, this means that for any integers* $k \geq 1$ *and* $\ell \geq 0$,

$$\lim_{n\to\infty} \mathbf{P}(\gamma_k = \ell) = \frac{e^{-\frac{1}{k}}}{k^\ell \ell!}.$$

3b. *For any* $k \geq 1$, *the expected number of cycles of length* k *is*

$$\mathbf{E}(\gamma_k) = \frac{1}{k}\frac{n^{\underline{k}}}{n^k} \tag{6.10}$$

$$= \frac{1}{k}\left(1 + O\left(\frac{k^2}{n}\right)\right) \quad \text{if } k = O(n^{1/2}). \tag{6.11}$$

6.2 Approximation of $n^{\underline{k}}/n^k$

The aim of this section is to state and prove Lemma 6.2.1, which we will need in our proof of the main theorem of this chapter, namely to derive the asymptotic formulas from the exact formulas. The lemma compares the quantity $n^{\underline{k}}/n^k$ – which is part of each of the exact formulas in the theorem – with the quantity $e^{-k^2/(2n)}$. The basis of the proof is the following variant of Stirling's formula:

$$m! = \sqrt{2\pi} m^{m+\frac{1}{2}} e^{-m}(1 + R(m)) \qquad \text{where } R(m) \text{ is of order } O(1/m), \tag{6.12}$$

which follows from the power series expansion of $\Gamma(m)$; there are explicit expressions for $R(m)$ up to a term of order $O(1/m^a)$, for any positive integer a, but we don't need such accuracy.

Lemma 6.2.1. (($\dagger$)[5] Approximation of $n^{\underline{k}}/n^k$)**.** *Let* n *and* k *be positive integers.*

1. $n^{\underline{k}}/n^k \leq e^{-k(k-1)/2n}$.

2. *Suppose* $k < n$. *There exists a real number* $\xi \in [0, \frac{k}{n-k}]$ *such that*

$$\frac{n^{\underline{k}}/n^k}{e^{-k^2/2n}} = \left(1 + \frac{k}{n-k}\right)^{1/2} \exp\left(-\frac{k^3}{2n(n-k)} + \frac{k^3}{3(n-k)^2(1+\xi)^3}\right)\frac{1+R(n)}{1+R(n-k)}, \tag{6.13}$$

where R *is as in Stirling's formula* (6.12).

[4]Usually the Euler-Mascheroni constant is denoted by γ, but we use the letter γ already.

[5]The results in this lemma are based on somewhat weaker statements that are stated without proof in [Bol01, Corollary 5.19]; the stronger formulation and the proof and are our own.

3. *We interpret k and n as variables, and suppose that $k < n$ and that $k = O(n^{2/3})$, in the sense that there exists $M \in \mathbb{R}$ such that $k < Mn^{2/3}$ for all k, n. Then*

$$\frac{n^{\underline{k}}/n^k}{e^{-k^2/2n}} = 1 + O\left(k^3/n^2\right).$$

4. *Suppose that $k < n$ and that $k = O(n^{1/2})$. Then*

$$n^{\underline{k}}/n^k = 1 + O\left(k^2/n\right).$$

Proof.
1. For any fixed n, this follows by induction on k. For convenience, we write $f(k) := n^{\underline{k}}/n^k$ and $g(k) := e^{-k(k-1)/2n}$, so we show that $f(k) \le g(k)$. Induction basis: $f(1) = 1 = g(1)$. Induction step: Suppose $f(k) \le g(k)$ for some $k \ge 1$. It suffices to show that

$$\frac{f(k+1)}{f(k)} \le \frac{g(k+1)}{g(k)}. \tag{6.14}$$

We have

$$\frac{f(k+1)}{f(k)} = 1 - \frac{k}{n} \qquad \text{and} \qquad \frac{g(k+1)}{g(k)} = e^{-k/n},$$

hence (6.14) follows by the fact that $1 + x \le e^x$ for any real x.

2. Writing $n^{\underline{k}}/n^k$ as $\dfrac{n!}{(n-k)!n^k}$, using Stirling's formula (6.12) for both $n!$ and $(n-k)!$, and simplifying the result, we get

$$\frac{n^{\underline{k}}/n^k}{e^{-k^2/2n}} = \left(\frac{n}{n-k}\right)^{n-k+\frac{1}{2}} e^{-k+(k^2/2n)} \frac{1+R(n)}{1+R(n-k)}.$$

Since $\frac{n}{n-k} = 1 + \frac{k}{n-k}$, to arrive at (6.13) it remains to show that

$$\left(1 + \frac{k}{n-k}\right)^{n-k} \exp\left(-k + \frac{k^2}{2n}\right) = \exp\left(-\frac{k^3}{2n(n-k)} + \frac{k^3}{3(n-k)^2(1+\xi)^3}\right) \tag{6.15}$$

for some $\xi \in [0, \frac{k}{n-k}]$. By taking the logarithm and simplifying, (6.15) is equivalent to

$$(n-k)\log(1 + \frac{k}{n-k}) = k - \frac{k^2}{2(n-k)} + \frac{k^3}{3(n-k)^2(1+\xi)^3} \qquad \text{for some } \xi \in [0, \tfrac{k}{n-k}].$$

That is a true statement, by the 2nd order Taylor approximation of $\log(1+x)$ at $x = 0$.

3. Because $k < n$ and $k = o(n)$ as $n \to \infty$, it follows that $\frac{1}{n-k}$ is $O(n^{-1})$. Hence, by (6.13),

$$\frac{n^{\underline{k}}/n^k}{e^{-k^2/2n}} = \left(1 + O\left(\frac{k}{n}\right)\right)^{\frac{1}{2}} \exp\left(O\left(\frac{k^3}{n^2}\right)\right) \frac{1+O\left(n^{-1}\right)}{1+R(n-k)}. \tag{6.16}$$

The function $R(m)$ is bounded away from -1, because $R(m) > -1$ for all m and $R(m) \to 0$ as $m \to \infty$, by (6.12). So $(1 + R(n-k))^{-1}$ is $O(1)$. Moreover, it is easily seen, by combining first order approximation and basic global properties, that $(1+x)^{\frac{1}{2}} = 1+O(x)$ when $x \geq 0$, and that $\exp(x) = 1 + O(x)$ when $x = O(1)$. Since $\frac{k}{n} \geq 0$ and $\frac{k^3}{n^2} = O(1)$, we can thus simplify (6.16) as

$$\frac{n^{\underline{k}}/n^k}{e^{-k^2/2n}} = \left(1 + O\left(\frac{k}{n}\right)\right)\left(1 + O\left(\frac{k^3}{n^2}\right)\right)\left(1 + O\left(\frac{1}{n}\right)\right).$$

After expanding the brackets, we can simplify the result a lot, because both $O(\frac{k}{n})$ and $O(\frac{1}{n})$ are also $O(1)$ and $O(\frac{k^3}{n^2})$. It can thus be simplified to the desired expression

$$\frac{n^{\underline{k}}/n^k}{e^{-k^2/2n}} = 1 + O\left(\frac{k^3}{n^2}\right).$$

4. We have

$$\frac{n^{\underline{k}}}{n^k} = \left(1 + O\left(\frac{k^3}{n^2}\right)\right)e^{-k^2/2n} = \left(1 + O\left(\frac{k^3}{n^2}\right)\right)\left(1 + O\left(\frac{k^2}{n}\right)\right) = 1 + O\left(\frac{k^2}{n}\right);$$

the first equality by part 3, the middle equality by first order approximation (using that $\frac{k^2}{n}$ is bounded), the last equality by expanding brackets and simplifying (using that $O(\frac{k^3}{n^2})$ is also $O(\frac{k^2}{n})$ and $O(1)$). □

***Remark* 6.2.2.** The accuracy of the approximation formula (6.13) is enough for our purposes, but it is clear from its proof that we could produce more accurate versions of the formula by using a higher order Taylor approximations of $\log(1+x)$. Concretely, if we upgrade from the 2nd to the 3th order Taylor approximation of $\log(1+x)$, then the argument of the exponential function in (6.13) gets replaced by

$$\frac{(3k-n)k^3}{6n(n-k)^2} - \frac{k^4}{4(n-k)^3(1+\xi)^4}, \tag{6.17}$$

where $\xi \in [0, \frac{k}{n-k}]$. If we upgrade from 3th to 4th order approximation, (6.17) gets replaced by

$$\frac{(3k-n)k^3}{6n(n-k)^2} - \frac{k^4}{4(n-k)^3} + \frac{k^5}{5(n-k)^4(1+\xi)^5}, \tag{6.18}$$

where $\xi \in [0, \frac{k}{n-k}]$. Etcetera. Thus, by using the mth order Taylor approximation of $\log(x+1)$, for any $m \geq 2$, we get an explicit expression $R_m(n,k)$ (not depending on an unknown number ξ) with the property that

$$\frac{n^{\underline{k}}/n^k}{e^{-k^2/2n}} \sim R_m(n,k) \quad \text{as } n \to \infty \text{ and } k = o\left(n^{m/(m+1)}\right).$$

For instance, by considering (6.13), (6.17) and (6.18) respectively, we can take

$$R_2(n,k) = 1, \quad R_3(n,k) = \exp\left(\frac{(3k-n)k^3}{6n(n-k)^2}\right), \quad R_4(n,k) = \exp\left(\frac{(3k-n)k^3}{6n(n-k)^2} - \frac{k^4}{4(n-k)^3}\right).$$

6.3 Proof of the theorem

Proof of Theorem 6.1.5. We prove the items in sequential order, except that the proof of 3b is right after the proof of 2a, because we use 3b in the proof of in 2b.

1a. (6.3) follows directly from (6.2) and part 4 of Lemma 6.2.1. We prove (6.2).

It is certain that X has at least 1 and at most n periodic points. Hence, (6.2) is true for $k = 0$ and for $k > n$, because it amounts to $\mathbf{P}(\overline{\gamma} = k) = 0$ in those cases. Further, (6.2) is clearly true for $k = n$, because it amounts to $\mathbf{P}(\overline{\gamma} = n) = n!/n^n$ in that case, which we already noted in (6.1).

Suppose that $1 \leq k < n$. Let $\mathcal{S}_k$ be the set of functions $S \to S$ with precisely k periodic points. Below, we will show that

$$|\mathcal{S}_k| = \binom{n}{k} k! \; kn^{n-k-1}.$$

From that, it follows that

$$\mathbf{P}(\overline{\gamma} = k) = \mathbf{P}(X \in \mathcal{S}_k) = \frac{|\mathcal{S}_k|}{n^n} = \frac{n!}{(n-k)!} \; kn^{n-k-1}n^{-n} = \frac{k}{n}\frac{n^{\underline{k}}}{n^k},$$

as was to be proved.

It remains to show that $|\mathcal{S}_k| = \binom{n}{k} k! \; kn^{n-k-1}$. We will show that the formula follows because $\mathcal{S}_k$ is in bijection with the cartesian product of three sets $\mathcal{P}_k, \mathcal{C}_k, \mathcal{F}_k$, with $|\mathcal{P}_k| = \binom{n}{k}$ and $\mathcal{C}_k = k!$ and $\mathcal{F}_k = kn^{n-k-1}$, namely:

- $\mathcal{P}_k$ is the set of subsets of S of size k;
- $\mathcal{C}_k$ is the set of permutations on $\{1, 2, \ldots, k\}$;
- $\mathcal{F}_k$ is the set $\{$ forests F on $\{1, 2, \ldots, n\}$: F consists of k trees $T_1, \ldots, T_k$, and each T_i contains precisely one of the points $1, 2, \ldots, k$ $\}$.

It is clear that $|\mathcal{P}_k| = \binom{n}{k}$ and $|\mathcal{C}_k| = k!$. For the result that $|\mathcal{F}_k| = kn^{n-k-1}$, we refer to [Bol01, Lemma 5.17], where two proofs of the result are presented and three other proofs of the result are cited. The first presented proof is elementary, although somewhat technical. It proceeds by induction on k, where the inductive basis ($k = 1$) is Cayley's formula for the number of trees on $\{1, 2, \ldots, n\}$, and the inductive step is based on Abel's generalisation of the binomial formula.

It remains to show that there is a bijection

$$\beta : \mathcal{P}_k \times \mathcal{C}_k \times \mathcal{F}_k \to \mathcal{S}_k.$$

We describe the bijection as a procedure to construct a function $s \in \mathcal{S}_k$ from a given triple $(p, c, f) \in \mathcal{P}_k \times \mathcal{C}_k \times \mathcal{F}_k$; more accurately, we construct the flow graph of s. The procedure is as follows, given an arbitrary but fixed total order[6] on S.

[6]The reason for this, and for the subsequent labeling, is to make the procedure well-defined, up to the choice of the total order on S.

1. Label the points of the subset $p \subset S$ as $1, 2, \ldots, k$, in the unique way that respects the order on S, and label the remaining points of S as $k+1, \ldots, n$, also respecting the order on S.

2. Let p and c together determine the cycles of s: the permutation c determines the cyclic structure on the subset $p = \{1, 2, \ldots, k\} \subset S$.

3. The forest f determines the remaining part of the flow graph of s, because by Proposition 5.3.12.4, the remaining part consists of k converging arborescences, converging to $1, \ldots, k$ respectively. Although the forest f consists of undirected trees, the directions of the arrows in s are uniquely determined by roots of the converging arborescences.

By Proposition 5.3.12.4, any flow graph can be represented by this procedure, so β is surjective. Clearly, distinct (p, c, f) correspond to distinct flow graphs, so β is injective.

1b. The equality (6.4) follows directly from the result of 1a (using that $\mathbf{P}(\overline{\gamma} = k) = 0$ for $k > n$). The harder part is to proof (6.5). We carry it out in detail, because it is prototypical for the proofs in part 2a and 2b.

Fix a number $\varepsilon > 0$, and fix any two numbers a, b such that

$$b \leq \frac{1}{2} + \frac{\varepsilon}{3} \quad \text{and} \quad \frac{1}{2} < a < b < \frac{2}{3}. \tag{6.19}$$

Because of the term $O(n^\varepsilon)$, it is enough to show (6.5) for all n above some bound (which may depend on ε), so we may assume that n is large enough such that

$$n^a + 1 \leq n^b.$$

As a first step, we show that

$$\sum_{k=1}^{n} \frac{k^2}{n} \frac{n^{\underline{k}}}{n^k} = o(n^{-1}) + \sum_{k=1}^{\lceil n^a \rceil} \frac{k^2}{n} \frac{n^{\underline{k}}}{n^k}. \tag{6.20}$$

Namely, let k be any integer with $\lceil n^a \rceil < k \leq n$. Let $c := \log_n(k-1)$, so $n^c = k - 1$. Because $n^a \leq k - 1 < n$, we have $a \leq c < 1$.

We make the following estimation, where the '$<$' follows from part 1 of Lemma 6.2.1:

$$\frac{k^2}{n} \frac{n^{\underline{k}}}{n^k} < \frac{n^2}{n} \exp\left(\frac{-(k-1)^2}{2n}\right) = n \exp\left(-\frac{n^{2c-1}}{2}\right) \leq n \exp\left(-\frac{n^{2a-1}}{2}\right);$$

It follows that

$$\sum_{k=\lceil n^a \rceil + 1}^{n} \frac{k^2}{n} \frac{n^{\underline{k}}}{n^k} < n^2 \exp\left(-\frac{n^{2a-1}}{2}\right),$$

because the summation has less than n terms. Because $2a - 1 > 0$, it is easily seen, by taking logarithms, that $n^2 \exp(-\frac{1}{2} n^{2a-1})$ is $o(n^{-1})$.[7] Hence (6.20) follows.

[7]Clearly, $o(n^{-1})$ is not the best estimate, but it is accurate enough for our purposes.

As a second step, we show that

$$\sum_{k=1}^{\lceil n^a \rceil} \frac{k^2}{n} \frac{n^{\underline{k}}}{n^k} = \Big(1 + O(n^{3b-2})\Big) \sum_{k=1}^{\lceil n^a \rceil} \frac{k^2}{n} e^{-k^2/2n}. \tag{6.21}$$

Namely, by part 3 of Lemma 6.2.1 we have

$$n^{\underline{k}}/n^k = \Big(1 + O\left(k^3/n^2\right)\Big) e^{-k^2/2n},$$

for all $k \leq \lceil n^a \rceil$, because $\lceil n^a \rceil \leq n^b < n^{2/3}$. Because $k^3/n^2 \leq n^{3b-2}$, it follows that

$$n^{\underline{k}}/n^k = \Big(1 + O\big(n^{3b-2}\big)\Big)\, e^{-k^2/2n},$$

where the bound which is implicit in $O(n^{3b-2})$ does not depend on k.[8] Therefore, we can place $(1 + O(n^{3b-2}))$ outside the summation, and (6.21) follows.

As a third step, we show that

$$\sum_{k=1}^{\lceil n^a \rceil} \frac{k^2}{n} e^{-k^2/2n} = O(1) + \int_0^{\lceil n^a \rceil} \frac{x^2}{n} e^{-x^2/2n} dx. \tag{6.22}$$

Namely, the summation in (6.22) is clearly a Riemann sum of the integral. It is easily seen that the integrand has a maximum at $x = \sqrt{2n}$, where it attains $2/e$ (note that this value does not depend on n), and that it is monotonic increasing from 0 to $2/e$ on the interval $[0, \sqrt{2n}]$, and monotonic decreasing from $2/e$ to 0 (asymptotically) on $[\sqrt{2n}, \infty]$. It is easily seen, especially pictorially, that the area that lies below the graph of the Riemann approximation[9] but above the graph of the integrand, is at most $2/e$, and that the same holds for the area above the graph of the Riemann approximation and below the graph of the integrand. Hence, the difference between the Riemann sum and the integral is at most $\pm 2/e$, which is $O(1)$. This proves (6.22).

As a final step, we show that

$$\int_0^{\lceil n^a \rceil} \frac{x^2}{n} e^{-x^2/2n} dx = \sqrt{n\pi/2} + o(1). \tag{6.23}$$

Writing $f(n) := \lceil n^a \rceil / \sqrt{2n}$ for convenience, we calculate (see below for explanation)

$$\begin{aligned}
\int_0^{\lceil n^a \rceil} \frac{x^2}{n} e^{-x^2/2n} dx &= \sqrt{2n} \int_0^{f(n)} y \cdot 2y e^{-y^2}\, dy \\
&= \sqrt{2n} \left(\Big[-y e^{-y^2} \Big]_0^{f(n)} + \int_0^{f(n)} e^{-y^2} dy \right) \\
&= \sqrt{2n} \Big(o(n^{-1}) + \sqrt{\pi}/2 \Big) = \sqrt{n\pi/2} + o(1).
\end{aligned}$$

[8] i.e., there exists a bound B, independent of k and n, and a function $b(k,n)$, such that $n^{\underline{k}}/n^k = (1 + b(k,n) n^{3b-2})) e^{-k^2/2n}$, and $|b(k,n)| < B$ for all k, n.

[9] i.e., the step function associated to the Riemann sum

In the third equality, we use the well-known identity $\int_0^\infty e^{-y^2}dy = \sqrt{\pi}/2$. Further, we use that the remaining integral from $f(n)$ to ∞ is $o(n^{-1})$, and that $-f(n)e^{-f(n)^2}$ is $o(n^{-1})$.[10] Essential for that is the fact that $f(n) \geq n^\delta$ for some $\delta > 0$, because $a > 1/2$. The remaining integral can be compared with $\int_{n^\delta}^\infty e^{-y}dy$ to see that it is $o(n^{-1})$, while for $-f(n)e^{-f(n)^2}$ we can see this by taking logarithms. This concludes the proof of (6.23).

Now we combine all the parts that we extracted: from (6.20)–(6.23), we conclude that

$$\sum_{k=1}^{n} \frac{k^2}{n} \frac{n^{\underline{k}}}{n^k} = o\big(n^{-1}\big) + \Big(1 + O\big(n^{3b-2}\big)\Big)\Big(o(1) + \sqrt{n\pi/2}\Big).$$

Expanding the brackets and simplifying the expression, using that $3b - 2 < 0$ so that $O(n^{3b-2})$ is also $o(1)$, yields

$$\sum_{k=1}^{n} \frac{k^2}{n} \frac{n^{\underline{k}}}{n^k} = \sqrt{n\pi/2} + O\big(n^{\frac{1}{2}+3b-2}\big) + o(1). \tag{6.24}$$

Because $b \leq \frac{1}{2} + \frac{\varepsilon}{3}$, we have that $\frac{1}{2} + 3b - 2 \leq \varepsilon$, hence (6.5) follows.

2a. Fortunately, we don't need to do that much work as for 1a and 1b, because the current proofs are similar to and build upon the previous proofs.

The proof of (6.6) builds on that of (6.2). There we used the set $\mathcal{S}_k$ of functions with the property that $\overline{\gamma} = k$. Let $\mathcal{S}'_k \subset \mathcal{S}_k$ be the subset of functions with the property that $\gamma = 1$. In the same way as that $\mathcal{S}_k$ is in bijection with $\mathcal{P}_k \times \mathcal{C}_k \times \mathcal{F}_k$, it follows that $\mathcal{S}'_k$ is in bijection with $\mathcal{P}_k \times \mathcal{C}'_k \times \mathcal{F}_k$, where $\mathcal{C}'_k$ consists of the *cyclic* permutations on $\{1, 2, \ldots, k\}$. Clearly, $|\mathcal{C}'_k| = (k-1)! = |\mathcal{C}_k|/k$, hence

$$\frac{|\mathcal{S}'_k|}{n^n} = \frac{|\mathcal{S}_k|}{n^n k} = \frac{1}{n} \frac{n^{\underline{k}}}{n^k} \tag{6.25}$$

where the last equality holds by the proof of 1a. Moreover, it is clear that

$$\mathbf{P}(\gamma = 1) = \mathbf{P}\Big(X \in \bigsqcup_{k=1}^{n} \mathcal{S}'_k\Big) = n^{-n} \sum_{k=1}^{n} |\mathcal{S}'_k|. \tag{6.26}$$

Combining (6.25) and (6.26) yields the desired result, (6.6).

The proof of (6.7) is analogous to that of (6.5). Namely, fix numbers ε, a, b satisfying the same restrictions as there, (6.19). By exactly copying the proofs of (6.20) and (6.21), but replacing k^2 with 1, we get the analogous formulas

$$\sum_{k=1}^{n} \frac{1}{n} \frac{n^{\underline{k}}}{n^k} = o\big(n^{-1}\big) + \sum_{k=1}^{\lceil n^a \rceil} \frac{1}{n} \frac{n^{\underline{k}}}{n^k} \tag{6.27}$$

[10] In both cases, $o(n^{-1})$ is not the best estimate, but accurate enough for the current purpose.

and

$$\sum_{k=1}^{\lceil n^a \rceil} \frac{1}{n} \frac{n^{\underline{k}}}{n^k} = \left(1 + O\left(n^{3b-2}\right)\right) \sum_{k=1}^{\lceil n^a \rceil} \frac{1}{n} e^{-k^2/2n} \tag{6.28}$$

respectively. Further, in the same way as we derived (6.22), but using instead that the next integrand monotonically decreases from n^{-1} towards 0 on the interval $[0, \infty]$, it follows that

$$\sum_{k=1}^{\lceil n^a \rceil} \frac{1}{n} e^{-k^2/2n} = O\left(n^{-1}\right) + \int_0^{\lceil n^a \rceil} \frac{1}{n} e^{-x^2/2n} dx. \tag{6.29}$$

Finally, writing $f(n) := \lceil n^a \rceil / \sqrt{2n}$ as in the proof of (6.23), we have

$$\int_0^{\lceil n^a \rceil} \frac{1}{n} e^{-x^2/2n} dx = \sqrt{\frac{2}{n}} \int_0^{f(n)} e^{-y^2} dy = \sqrt{\frac{\pi}{2n}} + o\left(n^{-1}\right); \tag{6.30}$$

the first equality follows by substituting $y = x/\sqrt{2n}$, the second equality follows by the fact that $\int_0^{f(n)} e^{-y^2} dy = \sqrt{\pi}/2 + o(n^{-1})$ which we already noted at (6.23).

Combining (6.27)–(6.30), we get

$$\sum_{k=1}^{n} \frac{1}{n} \frac{n^{\underline{k}}}{n^k} = o\left(n^{-1}\right) + \left(1 + O\left(n^{3b-2}\right)\right) \left(o\left(n^{-1}\right) + \sqrt{\pi/2n}\right).$$

Simplifying this in the same way as in the proof of (6.5), finishes the proof of (6.7).

3b. The asymptotic formula (6.11) follows directly from (6.10) and part 4 of Lemma 6.2.1. We proof (6.10).

Let $\mathbf{C}_k$ be the set of cycles of length k in S; hence $\mathbf{C}_k$ is in bijection with $\mathcal{P}_k \times \mathcal{C}'_k$, with notation as in the proof of 2a, and

$$|\mathbf{C}_k| = \binom{n}{k}(k-1)!.$$

For any cycle $\mathbf{c} \in \mathbf{C}_k$, we define the random variable $\gamma_{\mathbf{c}}$ as the indicator function for the event that $\mathbf{c}$ occurs in the flow graph of X. Hence,

$$\gamma_k = \sum_{\mathbf{c} \in \mathbf{C}_k} \gamma_{\mathbf{c}}.$$

Moreover, we have $\mathbf{E}(\gamma_{\mathbf{c}}) = \mathbf{P}(\mathbf{c}$ occurs in the flow graph of $X) = n^{-k}$, because the restriction that $\mathbf{c}$ occurs in the flow graph of X determines k of the function values of X, and leaves the other $n-k$ values free. It follows that

$$\mathbf{E}(\gamma_k) = \sum_{\mathbf{c} \in \mathbf{C}_k} \mathbf{E}(\gamma_{\mathbf{c}}) = \sum_{\mathbf{c} \in \mathbf{C}_k} n^{-k} = \binom{n}{k}(k-1)!\, n^{-k} = \frac{1}{k} \frac{n^{\underline{k}}}{n^k},$$

as desired.

2b. The equality (6.8) follows directly from the result of 3a and from the fact that $\gamma = \sum_{k=1}^{n} \gamma_k$. As usual, the difficulty lies in the asymptotic formula. Fortunately, the proof is for the most part similar to those of the asymptotic formulas in part 1b and 2a.

We denote the Euler-Mascheroni constant by Γ, that is,

$$\Gamma = \sum_{k=1}^{\infty} \Gamma_k \approx 0.58, \qquad \text{where} \quad \Gamma_k := \frac{1}{k} - \int_k^{k+1} \frac{1}{x} dx. \tag{6.31}$$

The reason for writing it this way will become clear below.

As in the proofs of 1b and 2a, fix numbers a, b such that $\frac{1}{2} < a < b < \frac{2}{3}$; however, we don't need a number ε. The following formulas (6.32)–(6.35) follow in a similar way as (6.20)–(6.23) from the proof of 1b, and also as and as (6.27)–(6.30) from the proof of 2a. Especially the first two formulas are proven in virtually the same way as the analogous ones, therefore we omit those proofs. The proofs of the last two formulas are more specific to the current situation, we provide those proofs below.

$$\sum_{k=1}^{n} \frac{1}{k} \frac{n^{\underline{k}}}{n^k} = o\left(n^{-1}\right) + \sum_{k=1}^{\lceil n^a \rceil} \frac{1}{k} \frac{n^{\underline{k}}}{n^k} \tag{6.32}$$

$$\sum_{k=1}^{\lceil n^a \rceil} \frac{1}{k} \frac{n^{\underline{k}}}{n^k} = \Big(1 + O\big(n^{3b-2}\big)\Big) \sum_{k=1}^{\lceil n^a \rceil} \frac{1}{k} e^{-k^2/2n} \tag{6.33}$$

$$\sum_{k=1}^{\lceil n^a \rceil} \frac{1}{k} e^{-k^2/2n} = O\big(n^{-1/3}\big) + \Gamma + \int_1^{\lceil n^a \rceil + 1} \frac{1}{x} e^{-x^2/2n} dx \tag{6.34}$$

$$\int_1^{\lceil n^a \rceil + 1} \frac{1}{x} e^{-x^2/2n} dx = O\left(n^{-1}\right) + \frac{\log(n) + \log(2) - \Gamma}{2} \tag{6.35}$$

Combining (6.32)–(6.35) results in

$$\sum_{k=1}^{n} \frac{1}{k} \frac{n^{\underline{k}}}{n^k} = o\big(n^{-1}\big) + \Big(1 + O\big(n^{3b-2}\big)\Big) \Big(O\big(n^{-1/3}\big) + \Gamma + \tfrac{1}{2}\log(n) + \tfrac{1}{2}\log(2) - \tfrac{1}{2}\Gamma\Big).$$

By taking b such that $3b - 2 < -1/3$, that is, $b < 5/9$, we can simplify this into the desired form (6.9).

It remains to prove (6.34) and (6.35).

We write for convenience

$$g(x) := \frac{1}{x} e^{-x^2/2n} \qquad \text{and} \qquad \Gamma'_k := g(k) - \int_k^{k+1} g(x) dx.$$

To prove (6.34), we must show that $\sum_{k=1}^{\lceil n^a \rceil} g(k) - \int_1^{\lceil n^a \rceil + 1} g(x) dx = \Gamma + O\big(n^{-1/3}\big)$, in other words, that

$$\sum_{k=1}^{\lceil n^a \rceil} \Gamma'_k = \Gamma + O\big(n^{-1/3}\big). \tag{6.36}$$

As a first step, we note that for any $k_0 \geq 1$, we have $\sum_{k=k_0}^{\infty} |\Gamma'_k| \leq g(k_0) \leq \frac{1}{k_0}$. This follows in the same way as for the comparison between Riemann sum and integral in the proofs of part 1b and 2a, using that the integrand $g(x)$ is monotonically decreasing. Analogously, we have $\sum_{k=k_0}^{\infty} |\Gamma_k| \leq \frac{1}{k_0}$. It follows that the statement (6.36) is equivalent to

$$\sum_{k=1}^{\lfloor n^{1/3} \rfloor} \Gamma'_k = \sum_{k=1}^{\lfloor n^{1/3} \rfloor} \Gamma_k + O\big(n^{-1/3}\big). \tag{6.37}$$

Note that

$$\begin{aligned} \Gamma'_k &= e^{-k^2/2n} \left(\frac{1}{k} - \int_k^{k+1} \frac{e^{(k^2-x^2)/2n}}{x} dx \right) \\ &= e^{-k^2/2n} \left(\Gamma_k + \int_k^{k+1} \frac{1 - e^{(k^2-x^2)/2n}}{x} dx \right). \end{aligned} \tag{6.38}$$

The remaining step in proving (6.37) is to show that the integral in (6.38), and the factor $\exp\big(-k^2/2n\big)$ in front, are "small enough". Concretely, for any integer k with $1 \leq k \leq n^{1/3}$, it is clearly true that

$$0 \leq \int_k^{k+1} \frac{1 - e^{(k^2-x^2)/2n}}{x} dx \leq 1 - e^{(-2k-1)/2n} \leq 1 - e^{-\left(2n^{-2/3}\right)},$$

and that $\exp\big(-\frac{1}{2}n^{-1/3}\big) \leq \exp\big(-k^2/2n\big) \leq 1$. Hence, the integral in (6.38) is in between two functions of n that are both $O(n^{-2/3})$, and the factor $e^{-k^2/2n}$ is in between two functions of n that are both $1 + O(n^{-1/3})$. From this and from (6.38) it follows that

$$\sum_{k=1}^{\lfloor n^{1/3} \rfloor} \Gamma'_k = \Big(1 + O(n^{-1/3})\Big)\Big(O(n^{-1/3}) + \sum_{k=1}^{\lfloor n^{1/3} \rfloor} \Gamma_k\Big),$$

from which (6.37) follows. This completes the proof of (6.34).

To prove (6.35), we need a basic fact about the so-called exponential integral E_1, which for $x \in \mathbb{R}_{>0}$ is defined by

$$E_1(x) := \int_x^{\infty} \frac{e^{-t}}{t} dt,$$

namely the fact that $E_1(x) = -\log(x) - \Gamma + O(x)$ as $x \to 0$; see [AS64, 5.1.11].

For convenience, we define $f(n) := (\lceil n^a \rceil + 1)^2/2n$. By the substitution $y = x^2/2n$, we have

$$\int_1^{\lceil n^a \rceil + 1} \frac{e^{-x^2/2n}}{x} dx = \int_{\frac{1}{2n}}^{f(n)} \frac{e^{-y}}{2y} dy = \frac{E_1\left(\frac{1}{2n}\right)}{2} - \int_{f(n)}^{\infty} \frac{e^{-y}}{2y} dy.$$

The integral on the right is smaller than $\int_{f(n)}^{\infty} e^{-y} dy = e^{-f(n)}$, and because $f(n) \geq \frac{1}{2} n^{2a-1}$ and $2a - 1 > 0$, it is clear that $e^{-f(n)}$ is of order $o(n^{-1})$. Moreover,

$$\frac{E_1\left(\frac{1}{2n}\right)}{2} = \frac{\log(2n) - \Gamma}{2} + O\left(n^{-1}\right).$$

This completes the proof of (6.35).

3a. We provide a proof only for the case that $k = 1$; we refer to [Bol01, Thm. 14.33.ii] for a proof[11] in the general case.

For $k = 1$ this is easy to prove, because each of the n points of S is with probability $1/n$ a fixed point of the function X, and these events are independent from each other. Therefore, the probability that precisely m points are fixed points, is

$$\mathbf{P}(\gamma_1 = m) = \binom{n}{m} \left(\frac{1}{n}\right)^m \left(1 - \frac{1}{n}\right)^{n-m}.$$

For any fixed m, we have

$$\left(1 - \frac{1}{n}\right)^{n-m} \to e^{-1} \quad \text{and} \quad \binom{n}{m} \left(\frac{1}{n}\right)^m = \frac{n^{\underline{m}}}{n^m m!} \to \frac{1}{m!} \quad \text{as } n \to \infty,$$

hence $\mathbf{P}(\gamma_1 = m) \to e^{-1}/m!$, corresponding to the Poisson distribution with mean 1. □

6.4 Comparing random and polynomial transformations

This section is entirely experimental in nature: we only show some concrete examples to illustrate the theory of the previous sections, and to illustrate the extent to which polynomial mappings on finite fields "look like" random mappings. It is not our intention here to state any conjecture or result about this.

Example 6.4.1. Consider Figure 6.1. The pictures on top depict the flow graphs of two random maps, the pictures on the bottom depict the flow graphs of two polynomial maps, given by the same polynomial $x^2 + 1$, but on the projective line over different fields.[12] Note that the numbers $7^3 + 1 = 344$ and $17^2 + 1 = 290$ are not far apart, therefore the pictures look comparable.

[11] The proof in [Bol01, Thm. 14.33.ii] refers to Theorem 1.20 of the same book, but the reference should be to Theorem 1.21.

[12] The only consequence of considering the projective line instead of the field itself, is that it adds a component of cardinality 1 to the flow graph: the "point at infinity" is a fixed point, and no other point maps to it.

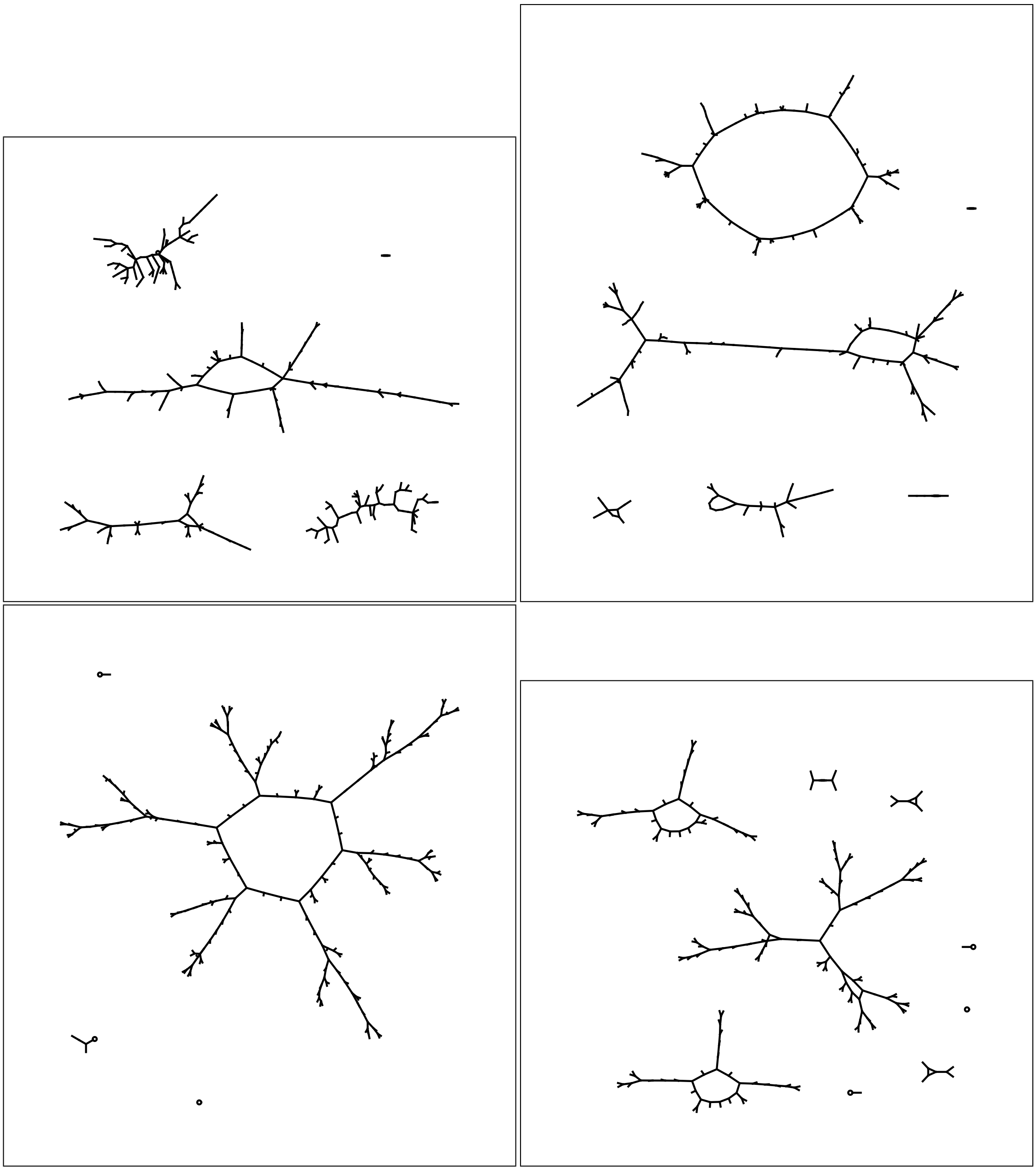


Figure 6.1: *Top:* Graph of two random maps on a set with $7^3 + 1$ elements. *Bottom:* Graphs of $x \mapsto x^2 + 1$ on $\mathbb{P}^1$ over a field with 7^3 and 17^2 elements (left to right).
These four pictures are reproduced with permission from [BCHvdM19]

The polynomial maps differ from random maps on a local scale: it is easily seen that every point has either no parents or two parents, except for the point 1 who has only 0 as a parent (because $-0 = 0$). For a random map, this seems to be a highly unlikely outcome, although we did not prove anything from which a concrete probability of the event would follow.

However, on a global scale, the polynomial maps look "random". We do not do any attempt to formalise or prove that. The only thing we do is to compute the invariants γ and γ_1, and compare them to the expected values from the theorem, just for the sake of illustration.

The number of connected components, γ: By Theorem 6.1.5.2b, we "expect" that $\gamma \approx 3.56$ in the case that $n = 7^3 + 1$, and that $\gamma \approx 3.47$ in the case that $n = 17^2 + 1$. But the error margins are likely to be high, because γ is a natural number and expected to be small, and because even $\mathbf{P}(\gamma = 1)$ is not so small, it is ≈ 0.074 in the case of $n = 17^2 + 1$. So the values of $\gamma = 5, 6, 4, 9$ don't seem to be extraordinary, although especially the last one is a bit large. (We always have the component at infinity, so perhaps our expectation should be "$E(\gamma) + 1$" for the polynomials.) Anyway, the random maps are not obviously different from the polynomial maps with respect to γ.

The number of 1*-cycles, i.e. fixed points, γ_1:* A fixed point is denoted in the pictures by a small circle (with or without an attached arborescence). We observe from the pictures the values $\gamma_1 = 1, 0, 3, 3$, from left to right and then top to bottom. We certainly have the fixed point at infinity for the polynomial maps, we prefer to leave it away, so we consider the map on the finite field only. With this correction, we have $\gamma_1 = 1, 0, 2, 2$. By Theorem 6.1.5.3a, the distribution of γ_1 is asymptotically Poisson with mean 1. Hence, the probabilities to observe $\gamma_1 = 0, 1, 2$ respectively, are $e^{-1}, e^{-1}, e^{-1}/2$ respectively, which are all quite large. So the observations of γ_1 don't differ from "what is to be expected" from a random map. ♪

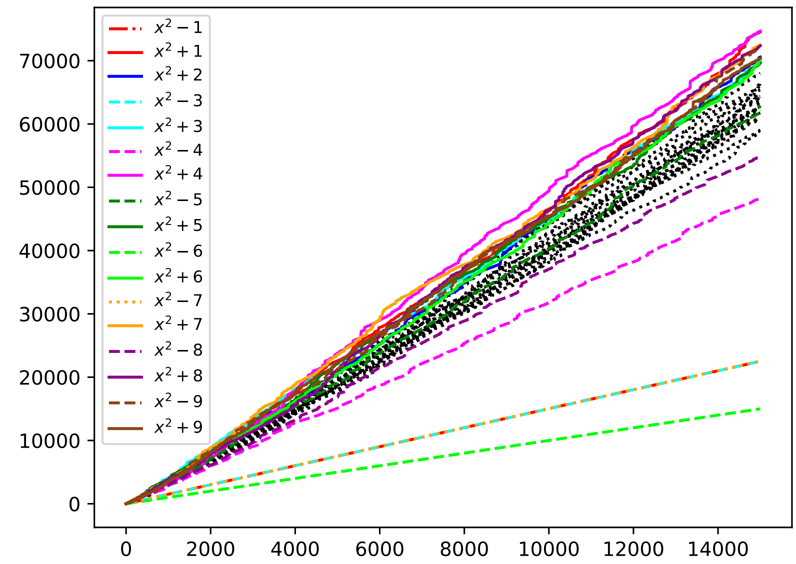
***Remark* 6.4.2.** There are however polynomials of which the flow graphs on finite fields are much more structured than random maps, for example $x^2 - 2$. The reason for that, (namely, $x \mapsto x^2 - 2$ is a "dynamically affine map"), is explained in [BCHvdM19, §1.1].

Example 6.4.3. To better visualise to what extend quadratic polynomial maps on finite fields $\mathbb{F}_p$ "look randomly", we fixed a polynomial $f_a(x) := x^2 + a$, and computed[13] invariants for the flow graph of the transformation f_a of $\mathbb{F}_p$, for the first 15000 prime numbers p. This gives hopefully more insight than in the previous example, where we considered $x^2 - 1$ on just two finite fields; when we do it for 15000 finite fields, we hope to see some more statistically significant behaviour. (Of course, if we observe behaviour, it does not follow that this is stable when p grows further.)

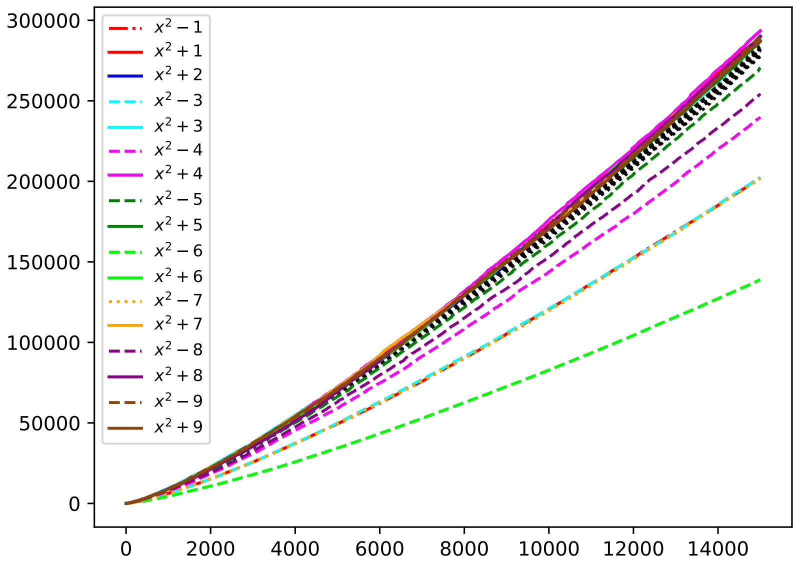
We did this for all 19 polynomials $f_a = x^2 + a$ with $a \in \mathbb{Z}$ and $-9 \leq a \leq 9$. However, we hide f_{-2} and f_0 from our plots, because they behave so differently that the contrasts between the remaining functions would be rendered invisible in the plots.

We did the same for 18 sequences of random functions: each sequence consists of 15000 random functions, independent from each other, on a set with p elements where p runs through the first 15000 primes. We did this so that we have an idea of the spread of the values for random functions, so that we can better judge whether a polynomial

[13] I wrote Python scripts for that purpose

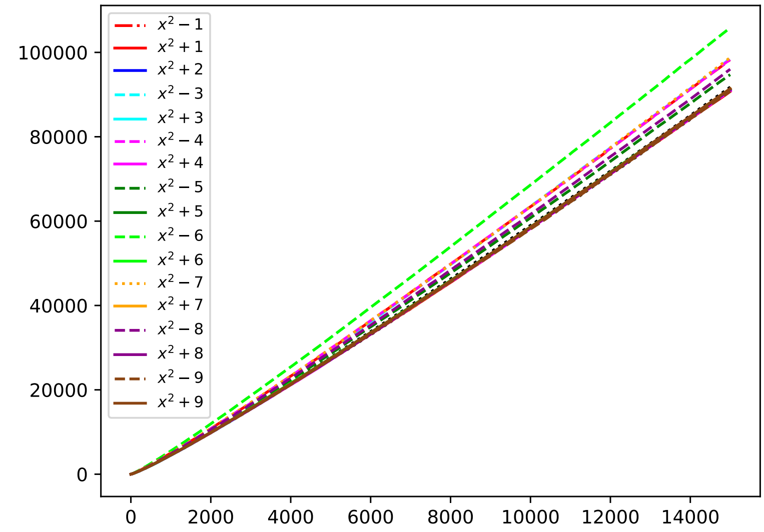


Cumulative plot for the length of the smallest cycle

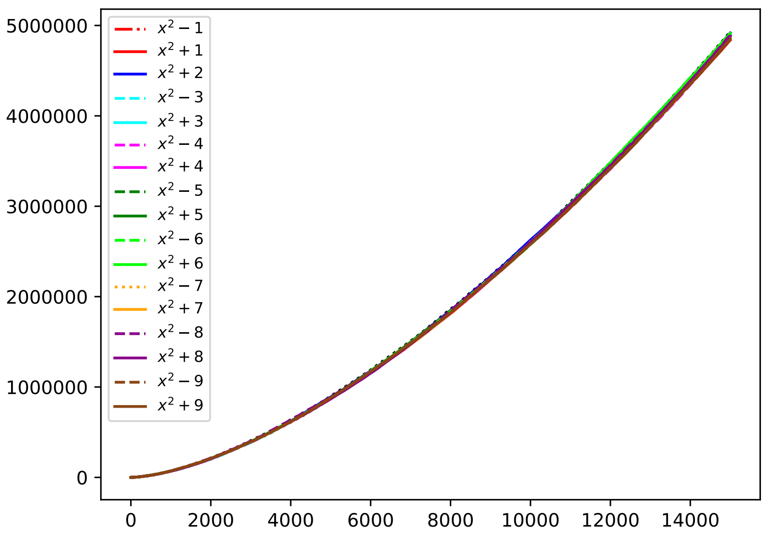


Cumulative plot for the geometric mean of the lengths of the cycles

Figure 6.2



Cumulative plot for γ, the number of components



Cumulative plot for $\overline{\gamma}$, the number of periodic points

Figure 6.3

"behaves non-randomly".

We did this computation for several invariants of flow graphs, and we made *cumu-*

lative plots of the invariants, with on the horizontal axis the number n such that p is the nth prime. For example, the plotted value at the number 1400 on the horizontal axis is the sum of the invariants for the flow graphs on $\mathbb{F}_p$, where p runs through the first 1400 prime numbers.

For a concrete example, the plot on top in Figure 6.2 shows a cumulative plot for the length of the smallest cycle. The fact that the plot for the polynomial $x^2 - 6$ follows the line $y = x$, means that the smallest cycle is always of length 1. This is easily verified: we have $3^2 - 6 = 3$, so 3 is a fixed point, for any p. The dotted black lines represent the 18 sequences of random functions. The fact that the bunch of random plots seem to follow the line $y = 60/14$, means that the average smallest cycle length seems to be roughly $60/14 \approx 4.3$, at least on the scale of the graph.

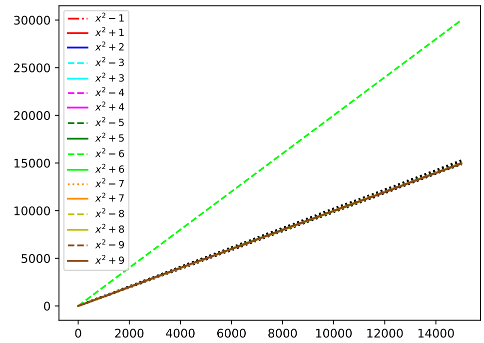
In Figure 6.2 (top), we see that most polynomials lie systematically a little bit above the bunch of the random plots; so their smallest cycle tend to be slightly larger than for random functions. This seems to be "compensated" by a few polynomials ($x^2 + a$ where $a = -6, -1, -3, -4$) whose plot is (much) below the bunch of random plots. The polynomial $x^2 + 5$ looks "random" in this respect.

Figure 6.2 (bottom) shows the geometric mean of the lengths of the cycles. Similar observations hold as for the plot on top.

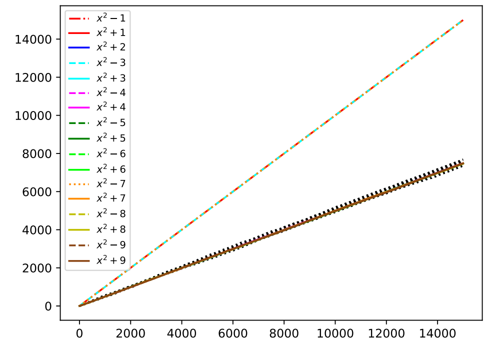
Figure 6.3 (top) plots γ, the number of components. The random plots (dotted black lines) are almost on a the same line on the scale of the plot; some polynomials, like $x^2 + 9$, are also almost on the line, so they "are like random" in this respect. Some other polynomials, like $x^2 - 6$ and $x^2 - 4$, are systematically above the plots, that is, have systematically slightly more components on average.

Figure 6.3 (bottom) plots $\overline{\gamma}$, the number of periodic points. The random and polynomial plots are approximately on the same curve, so to a good approximation, the number of periodic points seem to behave randomly for polynomials.

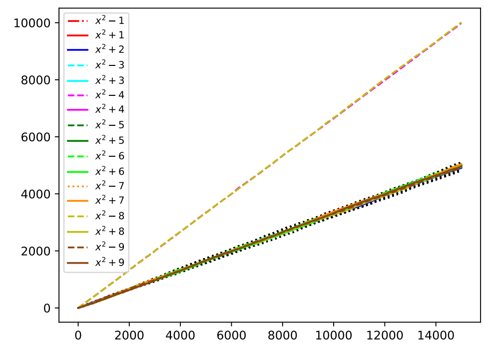
Finally, Figure 6.4 plots γ_k, for $k = 1, 2, 3, 4, 5, 6$: the number of cycles of length k. Something peculiar is visible: Most polynomials follow the (apparent) line of the random plots, and behave thus rather "randomly" in this respect. (The line is asymptotically given by $y = x/k$, because $1/k$ is the mean of the Poisson random variable in Theorem 6.1.5.3a.) But some of the polynomials that we look at (except for $k = 5$) follow approximately a line of twice that slope. In the case of $k = 1$, the exceptional polynomial is $x^2 - 6$, which is understandable, as we already noted that it has one 1-cycle "extra", for each p. For $k = 2$, the exceptional polynomials are $x^2 - 7, x^2 - 1, x^2 - 3$; for $k = 3$, it are $x^2 - 8$ and $x^2 - 4$; for $k = 4$ it is $x^2 - 5$; for $k = 6$, it is $x^2 - 4$. Note that all the exceptions (also in previous figures) are f_a with negative a. Perhaps it is true that the exceptional polynomials at k always have a k-cycle "extra", as in the case with $a = -6, k = 1$? I did not try to check it. ♪



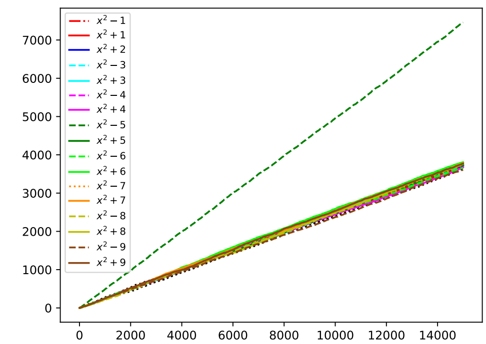
Cumulative plot of γ_1



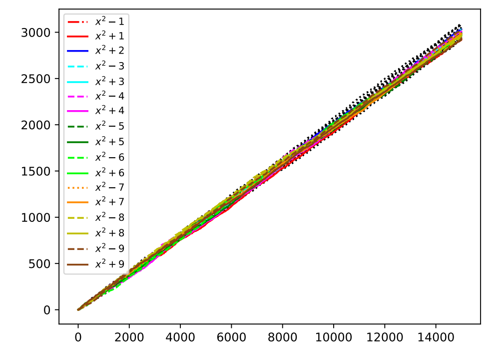
Cumulative plot of γ_2



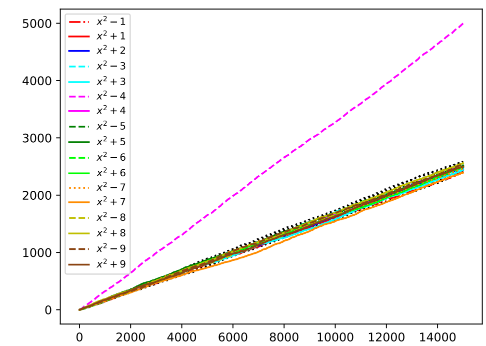
Cumulative plot of γ_3



Cumulative plot of γ_4



Cumulative plot of γ_5



Cumulative plot of γ_6

Figure 6.4

Chapter 7

The $3x + 1$ dynamical system and generalisations

7.1 The discrete dynamical systems $(\mathbb{N}, \mathcal{C}_{a,b,c})$

The subject of study in this chapter are the discrete dynamical systems $(\mathbb{N}, \mathcal{C}_{a,b,c})$, where the transformations $\mathcal{C}_{a,b,c}$ are as in the next definition.

Definition 7.1.1. Let $a, b, c \in \mathbb{N}$.

1. We define the map $\delta_c : \mathbb{N} \to \mathbb{N}$ by letting $\delta_c(x) := x/c^k$, where k is the largest number in $\mathbb{N}_0$ such that c^k divides x.

2. We define the map $\mathcal{C}_{a,b,c} : \mathbb{N} \to \mathbb{N}$ by letting $\mathcal{C}_{a,b,c}(x) := \delta_c(ax + b)$.

Example 7.1.2. Consider the discrete dynamical system $(\mathbb{N}, \mathcal{C}_{3,1,2})$. The transformation $\mathcal{C}_{3,1,2}$ maps $x \in \mathbb{N}$ to the largest odd integer that divides $3x + 1$. Thus, to compute for example the flow[1] of $x := 29$ under $\mathcal{C}_{3,1,2}$, we do the next calculation, where each arrow means "Divide by 2 if possible in $\mathbb{N}$, else multiply by 3 and add 1":

$$\begin{aligned} &29 \to 88 \to 44 \to 22 \to 11 \to 34 \to 17 \to 52 \to 26 \to 13 \\ &\quad \to 40 \to 20 \to 10 \to 5 \to 16 \to 8 \to 4 \to 2 \to 1 \to 4 \to 2 \to 1 \to \cdots \end{aligned}$$

The odd numbers in this sequence constitute the flow of 29; that is, the flow equals $(29, 11, 17, 13, 5, 1, 1, 1, \ldots)$. (This flow is embedded in Figure 7.1.)

Thus, 29 is preperiodic with eventual cycle $\langle 1 \rangle$.

7.2 The Collatz conjecture

Before we study the dynamics of the maps $\mathcal{C}_{a,b,c}$ for general $(a, b, c) \in \mathbb{N}^3$, we consider the particular case $(a, b, c) = (3, 1, 2)$ that we introduced in Example 7.1.2. This is the

[1] "flow" is defined in Definition 5.1.1

most famous case, because it is the subject of the *Collatz conjecture*, also known as the $3x + 1$ conjecture or problem. We postpone historical notes to §7.3.2, after we state the conjecture and illustrate it with an example.

***Remark* 7.2.1.** Clearly, the function $\mathcal{C}_{3,1,2}$ maps into the odd natural numbers. So every flow of $(\mathbb{N}, \mathcal{C})$ consists of odd numbers, except possibly for the starting point. Because we are merely interested in the "long-term behaviour" of the system, we could just as well consider the system $(\mathbb{N}_{\text{odd}}, \mathcal{C})$, where $\mathbb{N}_{\text{odd}} := 2\mathbb{N} - 1$.

Definition 7.2.2 (Collatz transformation and flow graph)**.**

1. We write $\mathcal{C} := \mathcal{C}_{3,1,2}$.
2. We call the discrete dynamical system $(\mathbb{N}_{\text{odd}}, \mathcal{C})$ the *Collatz system*, and its flow graph the *Collatz flow graph* (see Figure 7.1).

***Conjecture* 7.2.3** (Collatz conjecture, 3 equivalent formulations)**.**

1. Every point of the Collatz system is preperiodic with eventual cycle $\langle 1 \rangle$.
2. The only cycle of the Collatz system is $\langle 1 \rangle$, and there is no $x \in \mathbb{N}_{\text{odd}}$ such that $\lim_{n\to\infty} \mathcal{C}^{\circ n}(x) = \infty$.
3. The underlying graph of the Collatz flow graph is connected.

Proof that the statements 1–3 in Conjecture 7.2.3 are equivalent.
"1 $\Longrightarrow$ 2": Clear. *"1 $\Longleftarrow$ 2":* Suppose statement 2. For all $x \in \mathbb{N}_{\text{odd}}$, there exists $M \in \mathbb{N}$ such that there are infinitely many $k \geq 0$ with $\mathcal{C}^{\circ k}(x) \leq M$. Because $\mathbb{N}_{\leq M}$ is finite, it follows that the flow of x under $\mathcal{C}$ is not injective. By Fact 5.2.2.2, it follows that x is preperiodic.

"2 $\Longleftrightarrow$ 3": Let G' be a connected component of the Collatz flow graph, viewed as a digraph by inheriting the directions from the flow graph. By Proposition 5.3.12, G' is either an infinite directed tree, or it consists of a cycle and arborescences converging to nodes in the cycle. Hence, the existence of a connected component other than the component of the cycle $\langle 1 \rangle$, is equivalent to the existence of either a cycle other than $\langle 1 \rangle$, or an infinite directed tree. Clearly, a point $x \in \mathbb{N}_{\text{odd}}$ is in an infinite directed tree iff $\lim_{n\to\infty} \mathcal{C}^{\circ n}(x) = \infty$. □

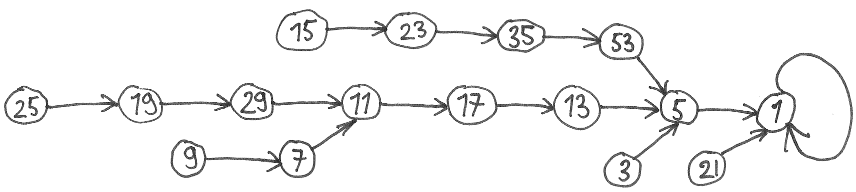


Figure 7.1: Part of the Collatz flow graph

Example 7.2.4. For $m \in \mathbb{N}_{\text{odd}}$, let's write $\mathcal{C}_m$ for the smallest directed subgraph of the Collatz flow graph that contains all the odd numbers up to m, i.e. all the nodes in the set $(\mathbb{N}_{\text{odd}})_{\leq m}$. Figure 7.1 depicts the digraph $\mathcal{C}_{25}$.

For $m \geq 3$, we can construct $\mathcal{C}_m$ from $\mathcal{C}_{m-2}$ as follows: Let x be the first element in the flow of m that is a node of $\mathcal{C}_{m-2}$; then $\mathcal{C}_m$ is obtained from $\mathcal{C}_{m-2}$ by "attaching to the node x the branch $m \to \mathcal{C}(m) \to \mathcal{C}^{\circ 2}(m) \to \cdots \to x$".

Thus, by letting m grow, we get an increasing sequence $\mathcal{C}_1, \mathcal{C}_3, \mathcal{C}_5, \ldots$ of directed subgraphs of the Collatz flow graph. Clearly, the Collatz conjecture is equivalent to the statement that $\mathcal{C}_m$ is connected, for all $m \in \mathbb{N}_{\text{odd}}$.

The digraph $\mathcal{C}_{25}$ (Figure 7.1) and the numbers occurring in it are rather small, but that changes relatively much at the next odd number: $\mathcal{C}_{27}$ is constructed from $\mathcal{C}_{25}$ by attaching to the node 23 the branch

$$\begin{aligned}
&27 \to 41 \to 31 \to 47 \to 71 \to 107 \to 161 \to 121 \to 91 \to 137 \to 103 \to 155 \to 233 \to 175 \\
&\quad \to 263 \to 395 \to 593 \to 445 \to 167 \to 251 \to 377 \to 283 \to 425 \to 319 \to 479 \to 719 \\
&\quad \to 1079 \to 1619 \to 2429 \to 911 \to 1367 \to 2051 \to 3077 \to 577 \to 433 \to 325 \to 61 \to 23,
\end{aligned}$$

which contains 37 arrows and in which 3077 is the largest occurring number.[2]

***Remark* 7.2.5.** We stated the Collatz conjecture in terms of the discrete dynamical system $(\mathbb{N}_{\text{odd}}, \mathcal{C})$, but usually (see [Lag12]) the conjecture is stated in terms of either $(\mathbb{N}, C)$ or $(\mathbb{N}, T)$, where

$$C(x) = \begin{cases} 3x+1 & \text{if } x \equiv 1 \pmod 2 \\ x/2 & \text{if } x \equiv 0 \pmod 2, \end{cases} \qquad T(x) = \begin{cases} (3x+1)/2 & \text{if } x \equiv 1 \pmod 2 \\ x/2 & \text{if } x \equiv 0 \pmod 2. \end{cases}$$

We refer to the maps $\mathcal{C} : \mathbb{N}_{\text{odd}} \to \mathbb{N}_{\text{odd}}$ and $T : \mathbb{N} \to \mathbb{N}$ and $C : \mathbb{N} \to \mathbb{N}$ as *Collatz transformations*.

Clearly, the flow graph of $(\mathbb{N}, T)$ can be obtained from the flow graph of $(\mathbb{N}, C)$ by "omitting the child (an even number) of each odd number"; and the flow graph of $(\mathbb{N}_{\text{odd}}, \mathcal{C})$ can be obtained from the flow graph of $(\mathbb{N}, C)$ and also from the flow graph of $(\mathbb{N}, T)$ by "omitting all even numbers".

Thus, the three statements that the flow graphs of $(\mathbb{N}, C)$ and $(\mathbb{N}, T)$ and $(\mathbb{N}_{\text{odd}}, \mathcal{C})$ respectively are connected, are all equivalent to the Collatz conjecture. The (conjecturally unique) cycle is $\langle 1, 4, 2 \rangle$ in the case of C, and $\langle 1, 2 \rangle$ in the case of T, and $\langle 1 \rangle$ in the case of $\mathcal{C}$.

7.3 Notes about the Collatz conjecture

7.3.1 Progress related to the Collatz conjecture

Currently[3], the Collatz conjecture is still an unsolved problem [Roo20].

[2] The flow of 27 is currently printed as an example on
`https://en.wikipedia.org/wiki/Collatz_conjecture`.

[3] August 17, 2020

There has been progress in several directions. The most straightforward direction is by increasing the bound M such that it is known that all numbers $n \in \mathbb{N}_{\text{odd}}$ with $n \leq M$ are preperiodic under $\mathcal{C}$ with eventual cycle $\langle 1 \rangle$; equivalently, in the notation of Example 7.2.4, that the directed subgraph $\mathcal{C}_M$ of the Collatz flow graph is connected.

According to Eric Roosendaal's webpage [Roo20], it has been verified in 2020 by an open source computer program by David Barina that we can take $M \geq 2^{68} \approx 3 \cdot 10^{20}$. The largest bound M claimed in the published scientific literature, as far as I found, is approximately $5.76 \cdot 10^{18}$ [Lag10, §6.1, "world record" W1].

Apart from the straightforward direction of increasing the bound M, much work has been done with a more theoretical flavour, although frequently assisted by "brute force computation" to arrive at concrete numeric results. To mention one example of such a result, "world record" W2 in [Lag10, §6.1]: *If the Collatz flow graph contains a cycle other than* $\langle 1 \rangle$*, then the length of such a cycle exceeds* 6.58 *billion.* Some other examples of results are "world records" W3–W5 in [Lag10, §6.1].

The most significant result about the Collatz conjecture in a very long time was recently proved in 2019 by Terence Tao.[4] The main result in his paper, [Tao19, Theorem 2], is as follows.

Let $\mathrm{Col}_{\mathrm{Min}}(N) := \inf_{n \in \mathbf{N}} \mathrm{Col}^n(N)$ be the minimal element of the Collatz orbit of N. (In our notation: $\mathrm{Col}^n(N) = \mathcal{C}^{\circ n}(N)$, so $\mathrm{Col}_{\mathrm{Min}}(N)$ is the smallest element in the flow of the point N.)

Theorem 7.3.1 ("Almost all orbits of the Collatz map attain almost bounded values")**.** *Let* $f : \mathbb{N} \to \mathbb{R}$ *be any function with* $\lim_{N \to \infty} f(N) = +\infty$*. Then we have* $\mathrm{Col}_{\mathrm{Min}}(N) < f(N)$ *for almost all* N *(in the sense of logarithmic density).* ⌟

Thus for instance, for all $n \in \mathbb{N}$ we have $\mathrm{Col}_{\mathrm{Min}}(N) < \log^{\circ n}(N)$ for almost all N (in the sense of logarithmic density).

Because I learned about this important result short before the deadline of this thesis, I don't have time and space left to do more than mentioning the result and the reference to the proof.

There has been written a lot about the Collatz conjecture, and about generalisations and related problems. Jeffrey Lagarias, himself one of the foremost contributors to research on the problem, published some clarifying overview papers. The papers [Lag11] and [Lag12] are richly annotated bibliographies of work done on the $3x + 1$ problem and related problems; the first one covers 197 papers from the years 1963–1999, the second one 134 papers from 2000–2009. Lagarias further wrote two introductory papers on the subject, [Lag10] and [Lag85], the one recent and the other not so recent; the older one contains several interesting results in the direction of the conjecture, including proofs, and both contain many references and historical and philosophical notes.

[4] Quoting from [Har19]: "It's a great advance in our knowledge of what's happening on this problem," said [Jeffrey] Lagarias. "It's certainly the best result in a very long time."

7.3.2 Historical and philosophical notes

As [Lag10] and [Lag85] already provide excellent historical and philosophical notes, I keep it relatively brief. Lothar Collatz, as a student in Hamburg in the early 1930's, got interested in flow graphs of arithmetic functions similar to the function T from Remark 7.2.5. He never published any related problems nor research on them, but it has been stated (by himself and others) that he circulated the $3x+1$ problem, along with similar iteration problems, at a conference in the early 1950's. Among others, Helmut Hasse, Shizuo Kakutani and Stanislaw Ulam got interested in the problem and spread it further.

The problem became known to be notoriously intractable. Lagarias [Lag85, §1] quotes private communication with Kakutani: 'For about a month everybody at Yale worked on it, with no result. A similar phenomenon happened when I mentioned it at the University of Chicago. A joke was made that this problem was part of a conspiracy to slow down mathematical research in the U.S.' Paul Erdős, also in private communication with Lagarias, said about the $3x + 1$ problem: 'Mathematics is not yet ready for such problems', and called it 'Hopeless. Absolutely hopeless.' [Lag10]

Lagarias gave the following reason that the problem is so intractable. 'We face this dilemma: On the one hand, to the extent that the problem has structure, we can analyze it—yet it is precisely this structure that seems to prevent us from proving that it behaves "randomly". On the other hand, to the extent that the problem is structureless and "random", we have nothing to analyze and consequently cannot rigorously prove anything.' [Lag85, §4]

A related reason for the intractability is that the $3x + 1$ problem seems similar to certain other recursion problems that turn out to be undecidable problems. For example, it has been shown that the $3x + 1$ problem can be formulated as a reachability problem (Theorem 7.4.7 below) for a certain "Post Tag System", named after Emil Post who started around 1920 to study such systems [Lag10, §7]. We say more about this in §7.4.

Nevertheless, Lagarias stresses (in both introductory papers) that 'No problem is so intractable that something interesting cannot be said about it. Study of the $3x + 1$ problem has uncovered a number of interesting phenomena' [Lag85].

Remarkable about the history of the $3x+1$ problem is that, despite of intense labour that started in the 1950's, the problem led an "underground existence", and there was no published mathematical literature about it until the 1970's. This may partly have been, according to [Lag10], because of the dominance of the Bourbaki-style in that time, with its emphasis on theory with rich internal structure. In addition, 'The results that could be proved appeared pathetically weak, so that it could seem damaging to one's professional reputation to publish them. In some mathematical circles it might have seemed in bad taste even to show interest in such a problem, which appears déclassé.'

Is the $3x + 1$ problem an important problem? Lagarias answers this as follows: 'Perhaps not for its individual sake, where it merely stands as a challenge. It seems to be a prototypical example of an extremely simple to state, extremely hard to solve, problem. A middle of the road viewpoint is that this problem is representative of a large class of problems, concerning the behavior under iteration of maps that are expanding on part of their domain and contracting on another part of their domain. This general class of problems is of definite importance, and is currently of great interest as an area of mathematical (and physical) research'. [Lag10]

7.4 A Post tag system encompassing the Collatz system

Definition 7.4.1. A *Post tag system* is a discrete dynamical system $(\mathcal{W}, \varphi)$, where $\mathcal{W}$ is the set of finite words over some finite alphabet $\mathcal{A}$, and where the transformation $\varphi := \varphi_{\nu,\lambda}$ is as defined below and depends on an arbitrary but fixed number $\nu \in \mathbb{N}$ and an arbitrary but fixed function $\lambda : \mathcal{A} \to \mathcal{W}$.

The transformation φ of $\mathcal{W}$ is defined as follows. The empty word is by definition a fixed point of φ. Let $w \in \mathcal{W}$ be a non-empty word, let $a \in \mathcal{A}$ be the first letter of w. We obtain $\varphi(w)$ by attaching the word $\lambda(a)$ at the end of w, followed by deleting the first ν letters from the start of the resulting word—unless there are less than ν letters to delete, in which case the empty word is returned.

Lastly, ν is called the *deletion number* and λ is called the *production function* of the Post tag system.

***Remark* 7.4.2.** Clearly, the only relevant aspect about the alphabet $\mathcal{A}$ is its cardinality $\mu := |\mathcal{A}|$.

Definition 7.4.3. For any pair $(\mu, \nu) \in \mathbb{N}^2$, we denote by $\mathcal{T}(\mu, \nu)$ the class of Post tag systems for which the alphabet has cardinality μ and the deletion number is ν.

***Remark* 7.4.4.** We wrote 'class' rather than 'set' because the alphabet could be *any* set of cardinality μ.

Notation 7.4.5. We adopt the following conventions.

1. Let w be a word, let $n \in \mathbb{N}$. We write w^n for the word that is the n-fold concatenation of w. Thus, $(do)^2 = dodo$.

2. The empty word is, in accordance with multiplicative notation, written as 1. For any word w, we let $w^0 := 1$.

Example 7.4.6. Consider the Post tag system $(\mathcal{W}, \varphi)$ in $\mathcal{T}(3, 2)$ that is determined by the alphabet $\mathcal{A} := \{\alpha, \beta, \gamma\}$ and by the production function

$$\lambda : \mathcal{A} \to \mathcal{W} : \qquad \alpha \mapsto \beta\gamma, \quad \beta \mapsto \alpha, \quad \gamma \mapsto \alpha^3.$$

The transformation $\varphi := \varphi_{2,\lambda}$ of $\mathcal{W}$ is as in Definition 7.4.1.

It turns out that if a word $w \in \mathcal{W}$ is a power of α, say $w = \alpha^n$ with $n \geq 0$, then the flow of w under φ contains another power of α. Specifically, if $n \geq 1$, then the flow contains $\alpha^{T(n)}$, where $T : \mathbb{N} \to \mathbb{N}$ is the Collatz transformation as in Remark 7.2.5. This is the content of Theorem 7.4.7, which was discovered by Liesbeth de Mol [Mol08].

We illustrate the above statement by a few concrete examples. For $w = \alpha^2$, we have

$$\begin{aligned} w &= \alpha\ \alpha \\ \varphi(w) &= \quad \beta\ \gamma \\ \varphi^{\circ 2}(w) &= \qquad \alpha \\ \varphi^{\circ 3}(w) &= \qquad\quad \gamma \\ \varphi^{\circ 4}(w) &= \qquad\qquad \alpha\ \alpha. \end{aligned}$$

Thus, $\varphi^{\circ 2}(\alpha^2) = \alpha = \alpha^{T(2)}$, and $\varphi^{\circ 2}(\alpha) = \alpha^2 = \alpha^{T(1)}$. Further,

$$\begin{aligned} w &= \alpha\ \alpha\ \alpha\ \alpha \\ \varphi(w) &= \quad \alpha\ \alpha\ \beta\ \gamma \\ \varphi^{\circ 2}(w) &= \qquad \beta\ \gamma\ \beta\ \gamma \\ \varphi^{\circ 3}(w) &= \qquad\quad \beta\ \gamma\ \alpha \\ \varphi^{\circ 4}(w) &= \qquad\qquad \alpha\ \alpha. \end{aligned}$$

Thus, $\varphi^{\circ 4}(\alpha^4) = \alpha^2 = \alpha^{T(4)}$. Lastly,

$$\begin{aligned} w &= \alpha\ \alpha\ \alpha\ \alpha\ \alpha \\ \varphi(w) &= \quad \alpha\ \alpha\ \alpha\ \beta\ \gamma \\ \varphi^{\circ 2}(w) &= \qquad \alpha\ \beta\ \gamma\ \beta\ \gamma \\ \varphi^{\circ 3}(w) &= \qquad\quad \gamma\ \beta\ \gamma\ \beta\ \gamma \\ \varphi^{\circ 4}(w) &= \qquad\qquad \gamma\ \beta\ \gamma\ \alpha\ \alpha\ \alpha \\ \varphi^{\circ 5}(w) &= \qquad\qquad\quad \gamma\ \alpha\ \alpha\ \alpha\ \alpha\ \alpha\ \alpha \\ \varphi^{\circ 6}(w) &= \qquad\qquad\qquad \alpha\ \alpha\ \alpha\ \alpha\ \alpha\ \alpha\ \alpha\ \alpha. \end{aligned}$$

Thus, $\varphi^{\circ 6}(\alpha^5) = \alpha^8 = \alpha^{T(5)}$. ♪

Theorem 7.4.7 ([Mol08]). *Let $(\mathcal{W}, \varphi)$ be the Post tag system that we considered in Example 7.4.6. Let $n \in \mathbb{N}$. Then*

$$\varphi^{\circ m}(\alpha^n) = \alpha^{T(n)} \qquad \textit{where } m := 2\lceil n/2 \rceil,$$

and where T is the Collatz transformation as in Remark 7.2.5. 𝅗𝅥

Our statement of this theorem is based on [Mol08, Theorem 2.1]. The proof of Theorem 7.4.7 is basically left as an exercise to the reader in [Mol08].

Proof of Theorem 7.4.7. The proof is probably most easily understood by viewing it as a generalisation of the calculations from Example 7.4.6.

Let $m \in \mathbb{N}_0$ be the unique number such that either $n = 2m$ or $n = 2m + 1$.

After m iterations of φ, the word α^n is transformed into $(\beta\gamma)^m$ in case $n = 2m$, and into $\alpha(\beta\gamma)^m$ in case $n = 2m + 1$.

The word $(\beta\gamma)^m$ is after m iterations of φ transformed into α^m. Thus, if $n = 2m$, it follows that α^n is transformed after $2m$ iterations of φ into α^m, which equals $\alpha^{T(n)}$.

The word $\alpha(\beta\gamma)^m$ is transformed by φ into $(\gamma\beta)^m\gamma$, which after m iterations is transformed into $\gamma\alpha^{3m}$, which is transformed into α^{3m+2}. Thus, if $n = 2m+1$, it follows that α^n is transformed after $2m + 2$ iterations of φ into α^{3m+2}, which equals $\alpha^{T(n)}$. □

***Remark* 7.4.8.** By Theorem 7.4.7, and by the embedding $\mathbb{N} \to \mathcal{W} : n \mapsto \alpha^n$, we can informally say that the Post tag system $(\mathcal{W}, \varphi)$ "encompasses" the Collatz system. Clearly, the Collatz conjecture is equivalent to the statement that for all $n \in \mathbb{N}$, the orbit of the word α^n contains the word α.

***Remark* 7.4.9** (Complexity of Post tag systems)**.** Citing from [Lag10, §7]: 'Starting in the 1920's, Emil Post uncovered great complexity in studying (...) "Post Tag Systems".' Lagarias illustrates this with the 'halting problems' and 'reachability problems' for a Post tag system $(\mathcal{W}, \varphi)$.

The *reachability problem* for a pair $(w, w') \in \mathcal{W}^2$ is the question:
Does the orbit of w under φ contain w'?

The *halting problem* for a word $w \in \mathcal{W}$ is the reachability problem for the pair $(w, 1)$; in other words, it is the question whether w eventually reaches the fixed point 1.

Thus, by Remark 7.4.8, the Collatz conjecture is equivalent to the conjunction of the reachability problems for (α^n, α) for all $n \geq 1$, for the Post tag system from Theorem 7.4.7. Note that an affirmative answer to the reachability problem for (α^n, α), implies that the answer to the halting problem of α^n is 'no', because it then follows that the eventual cycle of α^n is the cycle that contains $\{\alpha, \alpha^2\}$, and not the cycle $\langle 1 \rangle$.

Thus, the apparent complexity of the Collatz conjecture is illustrated by the hardness of halting problems for Post tag systems in general. The hardness of the halting problems is elaborated on in [Lag10, §7]. The strongest mentioned result in this direction is Theorem 7.4.10 below, which was proved independently (according to [Lag10, §7]) by Hao Wang [Wan63], John Cook and Marvin Minsky [CM64], and S.J. Maslov [Mas64].

Theorem 7.4.10. *There is no recursive decision procedure for the halting problems of the Post tag systems in the class $\mathcal{T}(2, \mu)_{\mu \in \mathbb{N}}$.*

That is, there is no recursive decision procedure that decides for all $\mu \in \mathbb{N}$, for all Post tag systems $(\mathcal{W}, \varphi)$ in $\mathcal{T}(2, \mu)$, and for all $w \in \mathcal{W}$, whether or w is under φ preperiodic with eventual cycle $\langle 1 \rangle$.

For example, the proof in [CM64] proceeds by representing any two-symbol Turing machine as some Post tag system in the class $\mathcal{T}(2, \mu)_{\mu \in \mathbb{N}}$. The theorem follows from the existence of universal two-symbol Turing machines, which was proved by Claude Shannon [Sha56, pp.164–165], and from the undecidability of the halting problem for universal Turing machines, which was proved by Alan Turing in his classical paper [Tur37]. □

***Remark* 7.4.11.** After remarking that 'we now have a deeper appreciation of exactly how simple a problem can be and still simulate a universal computer', Lagarias [Lag10, §7] argues that 'there are, however, reasons to suspect that the 3x + 1 function is not complicated enough to be universal, i.e. to allow the encoding of a universal computer in its input space.'

7.5 Relations between elements in a flow in $(\mathbb{N}, \mathcal{C}_{a,b,c})$

Throughout this section, let $(a, b, c) \in \mathbb{N}^3$.

In Definition 7.1.1 we defined the dynamical system $(\mathbb{N}, \mathcal{C}_{a,b,c})$, but up to this point, we only studied it in the case that $(a,b,c) = (3,1,2)$. In the remainder of this chapter, we turn our attention to general triples (a,b,c).

***Remark* 7.5.1.** Let $x_0 \in \mathbb{N}$, let $(x_0, x_1, x_2, \ldots)$ be the flow of x_0 under $\mathcal{C}_{a,b,c}$. By Definition 7.1.1, there exists for all $m \in \mathbb{N}$ a number $k_m \in \mathbb{N}_0$ such that $x_m = (ax_{m-1} + b)/c^{k_m}$. Hence, we can make the following definition.

Definition 7.5.2. Let $x \in \mathbb{N}$. The *signature* $\mathrm{sig}_{(a,b,c)}(x)$ of the point x of the dynamical system $(\mathbb{N}, \mathcal{C}_{a,b,c})$ is the sequence in $\mathbb{N}_0$ defined as follows. Let $(x_0, x_1, x_2, \ldots)$ be the flow of x under $\mathcal{C}_{a,b,c}$, where $x_0 := x$. Then

$$\mathrm{sig}_{(a,b,c)}(x) = (k_1, k_2, k_3, \ldots) \in \mathbb{N}^{\mathbb{N}}, \qquad \text{where} \quad c^{k_m} = \frac{ax_{m-1} + b}{x_m}.$$

Further, for $n \in \mathbb{N}$, we write for

$$\mathrm{sig}_{(a,b,c),n}(x) := (k_1, k_2, \ldots, k_n) \in \mathbb{N}^n.$$

Example 7.5.3 (Signatures in Collatz flows)**.**

1. By Example 7.1.2, we have $\mathrm{sig}_{(3,1,2)}(29) = (3,1,2,3,4,2,2,2,2,2,2,\ldots)$.
2. By a similar calculation, using Example 7.2.4, we have
$$\begin{aligned}\mathrm{sig}_{(3,1,2)}(27) = (&1,2,1,1,1,1,2,2,1,2,1,1,2,1,1,1,2,3,1,1,2,1,2,\\ &1,1,1,1,1,3,1,1,1,4,2,2,4,3,1,1,5,4,2,2,2,2,2,2,\ldots)\end{aligned}$$
3. The Collatz conjecture is equivalent to the statement that $\mathrm{sig}_{(3,1,2)}(x)$ ends with the infinite tail $(2,2,2,2,2,\ldots)$, for all $x \in \mathbb{N}$.

The following theorem relates flows in $(\mathbb{N}, \mathcal{C}_{a,b,c})$ to certain symmetric means: to the arithmetic mean, and to the translation mean that we defined and studied in §2.7.

Theorem 7.5.4 ((*) Relation between flows under $\mathcal{C}_{a,b,c}$ and means)**.** *Let $n \in \mathbb{N}$. Let $x_0 \in \mathbb{N}$, let $(x_0, x_1, x_2, \ldots)$ be the flow of x_0 under $\mathcal{C}_{a,b,c}$. Let $\mathbf{x}_n := (x_0, \ldots, x_{n-1}) \in \mathbb{N}^n$, and let $\mathbf{k}_n := (k_1, \ldots, k_n) := \mathrm{sig}_{(a,b,c),n}(x) \in \mathbb{N}^n$. Let*

$$r := \frac{a}{b}, \qquad R_r(\mathbf{x}_n) := \frac{1}{\mathrm{TM}_{-r,n}(\mathbf{x}_n)} + r.$$

1. *("Ratios in the flow":) We have*
$$\left(\frac{x_n}{x_0}\right)^{1/n} = \frac{b\,R_r(\mathbf{x}_n)}{c^{\mathrm{AM}_n(\mathbf{k}_n)}}.$$
2. *("Logarithmic version of part 1":) We have*
$$\log(x_n) = \log(x_0) + n\Big(\log(b\,R_r(\mathbf{x}_n)) - \mathrm{AM}_n(\mathbf{k}_n)\log(c)\Big).$$

3. *("Average signature of cycle":) The condition that $x_n = x_0$, i.e. that $\langle x_0, x_1, \ldots, x_{n-1}\rangle$ is a cycle of $(\mathbb{N}, \mathcal{C}_{a,b,c})$, is equivalent to*

$$\mathrm{AM}_n(\mathbf{k}_n) = \frac{\log(b\,R_r(\mathbf{x}_n))}{\log(c)}. \tag{7.1}$$

4. *("Closeness of $R_r(\mathbf{x}_n)$ and r":) We have*

$$0 < \frac{1}{\mathrm{GM}_n(\mathbf{x}_n)} \le R_r(\mathbf{x}_n) - r \le \frac{1}{\mathrm{HM}_n(\mathbf{x}_n)} \le \frac{1}{\mathrm{Min}(\mathbf{x}_n)}, \tag{7.2}$$

and $$(\lim_{n\to\infty} x_n = \infty) \implies \lim_{n\to\infty} R_r(\mathbf{x}_n) = r. \tag{7.3}$$

5. *("$\log(x_n)$ asymptotically linear":) Suppose that $\lim_{n\to\infty} \mathrm{AM}_n(\mathbf{k}_n) =: k$ exists and is > 0, and that x_0 is not preperiodic. Then*

$$\lim_{n\to\infty} \log(x_n) - \log(x_0) - n\log(a/c^k) = 0.$$ ⅃

Proof. **1.** We have $c^{k_i} = (ax_{i-1}+b)/x_i$, for all $1 \le i \le n$. Hence,

$$c^{\sum \mathbf{k}_n} = \prod_{i=1}^{n} \frac{ax_{i-1}+b}{x_i} = \frac{x_0}{x_n}\prod_{i=1}^{n}\frac{ax_{i-1}+b}{x_{i-1}} = \frac{x_0}{x_n}\prod_{i=0}^{n-1} a + bx_i^{-1} = \frac{x_0}{x_n} b^n \prod(\mathbf{x}_n^{-1}+r). \tag{7.4}$$

By definition (Defn. 2.7.1), we have $\prod(\mathbf{x}_n^{-1}+r) = \big(\mathrm{TM}_{r,n}(\mathbf{x}_n^{-1})+r\big)^n$. By duality (Theorem 2.7.4.1), we have $\mathrm{TM}_{r,n}(\mathbf{x}_n^{-1}) = 1/\mathrm{TM}_{-r,n}(\mathbf{x}_n)$. Hence, $\prod(\mathbf{x}_n^{-1}+r) = \big(R_r(\mathbf{x}_n)\big)^n$. Thus, the desired result follows by taking nth roots in (7.4), and noting that $\frac{1}{n}\sum \mathbf{k}_n = \mathrm{AM}_n(\mathbf{k}_n)$.

2. Follows directly form part 1 by taking logarithms.

3. By part 2, the formula (7.1) is equivalent to $\log(x_n) = \log(x_0)$, which is equivalent to $x_n = x_0$.

4. (7.2) follows directly from the inequalities $\mathrm{Min}_n \le \mathrm{HM}_n \le \mathrm{TM}_{-r,n} \le \mathrm{GM}_n$ (Corollary 2.7.6), using that $-r \le 0$. If $\lim_{n\to\infty} x_n = \infty$, then by Corollary 1.3.5 we have $\lim_{n\to\infty} 1/\mathrm{HM}_n(\mathbf{x}_n) = 0$; thus, (7.2) follows from (7.3).

5. Because x_0 is not preperiodic, it follows that $\lim_{n\to\infty} x_n = \infty$, in the same way as in the proof of "1 $\Longleftarrow$ 2" in 7.2.3. Hence, the result follows by combining part 2 and (7.3). □

Example 7.5.5 (Illustrations of Theorem 7.5.4)**.** Let $(a,b,c) = (3,1,2)$.

1. We have $b\,R_r(\mathbf{x}_n) = \big(\mathrm{TM}_{-\frac{1}{3},n}(\mathbf{x}_n)\big)^{-1} + 3$.

2. Part 3 of Theorem 7.5.4, applied to $\mathbf{x}_1 = (1)$ and $\mathbf{k}_1 = (2)$, states that $2 = \log(1+3)/\log(2)$, which is correct.

3. We noted in Example 7.2.4 that when $x_0 = 27$, then $x_{37} = 23$. Let $n = 37$ and $\mathbf{x}_n = (x_0, \ldots, x_{36})$. It is clear from Example 7.2.4 that the numbers occurring in $\mathbf{x}_n$ are relatively large; hence, $\mathrm{TM}_{-\frac{1}{3},n}(\mathbf{x}_n)$ must be relatively close to 0, so $b\,R_r(\mathbf{x}_n) \approx 3$. Moreover, we clearly have $(x_n/x_0)^{1/n} \approx 1$. Hence, by Theorem 7.5.4.1, we must have that $\mathrm{AM}_n(\mathbf{k}_n) \approx \log(3)/\log(2)$. We can verify this by calculating $\mathrm{AM}_n(\mathbf{k}_n)$ from Example 7.5.3.2:

$$\mathrm{AM}_{37}(1,2,1,1,1,1,2,2,1,2,1,1,2,1,1,1,2,3,1,1,2,1,2,1,1,1,1,1,3,1,1,1,4,2,2,4,3)$$
$$= 1.5945\ldots \approx \log(3)/\log(2) = 1.5849\ldots. \quad ♪$$

7.5.1 Cycles of $(\mathbb{N}, \mathcal{C}_{a,b,c})$

Example 7.5.6. Let $(a, b, c) = (5, 7, 2)$. So we have $r = 5/7$.

It is easily verified that $\langle 119, 301, 189\rangle$ is a cycle of $(\mathbb{N}, \mathcal{C}_{5,7,2})$.[5] Writing $\mathbf{x}_3 = (119, 301, 189)$, we clearly have correspondingly $\mathbf{k}_3 = (1, 3, 3)$. By Theorem 7.5.4.3, it follows that

$$\frac{7}{3} = \frac{\log\left(5 + \left(\mathrm{TM}_{-\frac{5}{7}}(119, 301, 189)\right)^{-1}\right)}{\log 2}. \tag{7.5}$$

Explicitly, we have

$$\mathrm{TM}_{-\frac{5}{7}}(119, 301, 189)\big)^{-1} = \mathrm{TM}_{\frac{5}{7}}(119^{-1}, 301^{-1}, 189^{-1})$$
$$= \Big((119^{-1} + 5/7)(301^{-1} + 5/7)(189^{-1} + 5/7)\Big)^{1/3} - 5/7 \;=\; 0.0056\ldots$$

Also without this computation, we know that $\mathrm{TM}_{-\frac{5}{7}}(119, 301, 189)\big)^{-1} \le 119^{-1}$ (because $\mathrm{TM}_{-\frac{5}{7}}$ is a mean) and hence is "small" compared to 5; hence it follows from (7.5) that $7/3 \approx \log(5)/\log(2)$. We verify that:

$$\frac{7}{3} = 2.3333\ldots, \qquad \frac{\log 5}{\log 2} = 2.3219\ldots. \quad ♪$$

The observations in Example 7.5.6 naturally lead to the following.

Corollary 7.5.7. (†).[6] *Let $B \in \mathbb{R}_{>0}$. Let $\langle x_0, \ldots, x_{n-1}\rangle$ be a cycle of $(\mathbb{N}, \mathcal{C}_{a,b,c})$ such that $\mathrm{Min}(\mathbf{x}_n) \ge B$, where $\mathbf{x}_n := (x_0, \ldots, x_{n-1})$. Let $\mathbf{k}_n := \mathrm{sig}_{(a,b,c),n}(x_0)$. Then we have*

$$0 < \frac{\sum \mathbf{k}_n}{n} - \frac{\log(a)}{\log(c)} \le \frac{\log(a + \frac{b}{B})}{\log c} - \frac{\log a}{\log c} < \frac{b}{Ba\log c}, \qquad \textit{and} \quad \frac{\sum \mathbf{k}_n}{n} \in \tfrac{1}{n}\mathbb{Z}. \quad ♩$$

Proof. The first two inequalities follow from Theorem 7.5.4.3, the last inequality by first-order approximation of log. □

[5] I found this cycle, and many other cycles of dynamical systems $(\mathbb{N}, \mathcal{C}_{a,b,c})$, by automating the cycle search with Python.

[6] Only in this generality I did not find the result in the published literature; for the special case $c = 2$, the result is well-established, see Remark 7.5.8.

***Remark* 7.5.8.** Suppose a is not a rational power of c; so $\log(a)/\log(c)$ is irrational.

Corollary 7.5.7 can be applied as follows. For some (large) B, compute the flows of all numbers up to B; in particular, we find this way all cycles that contain a number below B. So if there exists any cycle $\langle x_0, \ldots, x_{n-1}\rangle$ that has not been found yet, then this implies by Corollary 7.5.7 the existence of a number in $\frac{1}{n}\mathbb{Z}$ that is particularly close to $\log(a)/\log(c)$. It is known, in a precise sense, that the continued fraction expansion of $\log(a)/\log(c)$ provides the best rational approximations of the irrational number; see for instance [Eli93, §3]. Thus, there follows a concrete lower bound for the length n of the cycle.

Shalom Eliahou performed such analysis (with more details than we sketched) for $(a, b, c) = (3, 1, 2)$, and concluded the following:

Theorem 7.5.9. *[Eli93, Thm. 1.1] If there is a cycle other than $\langle 1, 2\rangle$ of the Collatz transformation T as in Remark 7.2.5, then the length of such a cycle is*

$$301994a + 17087915b + 85137581c$$

for some $a, b, c \in \mathbb{N}_0$ with $b > 0$ and $ac = 0$. ⅃

Actually, the numbers 301994, 17087915, 85137581 follow by taking $B = 2^{40}$; since the Collatz conjecture has nowadays been verified up to 2^{68} (see §7.3.1), the three numbers can be replaced by larger numbers, as Eliahou remarks in [Eli93, §4].

John Simons and Benne de Weger performed an even more detailed analysis for $(a, b, c) = (3, 1, 2)$, which results in the detailed Theorem 3 in [SdW05].

Simons generalises that theorem to $(a, b, c) = (a, b, 2)$, with arbitrary a, b, in [Sim07, §4], and includes concrete numerical bounds for several (a, b).

A very detailed analysis of cycles for $(a, b, c) = (3, b, 2)$, with arbitrary b, can be found in the papers [BM00] and [BM06] by Edward Belaga and Maurice Mignotte.

7.6 Expected large-scale behaviour of $(\mathbb{N}, \mathcal{C}_{a,b,c})$

Throughout this section, let $(a, b, c) \in \mathbb{N}^3$.

This section is merely experimental; we do some heuristic "reasoning", and illustrate the "outcomes" of the heuristics with some self-produced pictures. But the heuristic arguments are naive and should not be attributed any truth. However, we think that the thoughts behind the "heuristic reasoning" are valuable for enhancing the (intuitive) understanding of the behaviour of the systems $(\mathbb{N}, \mathcal{C}_{a,b,c})$. Moreover, the pictures seem to agree with the "outcomes of the arguments".

Heuristic considerations 7.6.1. Let $x := x_0 \in \mathbb{N}$ be a *non-periodic* point of the dynamical system $(\mathbb{N}, \mathcal{C}_{a,b,c})$. We investigate how we expect that the growth or decline of the flow $(x_0, x_1, x_2, \ldots)$ behaves on a large scale.

The most important ingredient of the heuristic argument is Theorem 7.5.4.5, which gives an asymptotically linear formula for $\log(x_n)$, in the case that the "average exponent" $\mathrm{AM}_n(\mathbf{k}_n)$ converges as $n \to \infty$, where $\mathbf{k}_n = \mathrm{sig}_{(a,b,c),n}(x)$. We argue that it is "plausible" that the limit $\lim_{n\to\infty} \mathrm{AM}_n(\mathbf{k}_n)$ exists.

First, we reduce somewhat the triples (a, b, c) that we consider. The cases $c = 1$ and $a = 1$ are trivially analysed; therefore, we suppose that $a, c \geq 2$.

If a and b have a common divisor $d > 1$, then it is not hard to analyse the system with (a, b, c) in terms of the system with $(a/d, b/d, c)$. So we suppose that $\gcd(a, b) = 1$.

Further, if $\gcd(a, c) =: d > 1$, then we have $ax + b \equiv b \not\equiv 0 \pmod{d}$ for all $x \in \mathbb{N}$, so it follows that we never divide by c when we perform the transformation $\mathcal{C}_{a,b,c}$, in other words, $\mathcal{C}_{a,b,c}$ equals the trivially analysed map $x \mapsto ax + b$. Therefore, we suppose $\gcd(a, c) = 1$.

It turns out to be convenient to write the numbers x_n in their "c-adic expansion", so $x_n = a_0 + a_1 c + a_2 x^2 + a_3 x^3 + \ldots$, for some $0 \leq a_i \leq c - 1$, depending on n. (The sum terminates because $x_n \in \mathbb{N}$, but we can extend it by an infinite tail of $a_i = 0$.) This is convenient because $\mathcal{C}_{a,b,c}$ acts on x_n by performing first the affine map $x_n \mapsto ax_n + b$, and then *shifts* the sequence of resulting digits $(a_0, a_1, \ldots)$ to the left by erasing the (optional) tail of k_{n+1} zeros $(a_0, \ldots)$ at the start. Thus, the map $\mathcal{C}_{a,b,c}$ is very easily represented by a combination of an affine map and a shifting map.

We do the following "simplifying supposition": *After sufficiently many iterations, the digits $a_1, a_2, a_3, \ldots$ can be considered to be uniformly random variables on $\{0, \ldots, c-1\}$, independent from each other and from n. That is, $\mathbf{P}(a_i = \ell) = 1/c$ for all $\ell \in \{0, \ldots, c-1\}$, independently for $i = 1, 2, 3, \ldots$.* The reason that we only consider a_i for $i \geq 1$ to be random, and not a_0, is that the value of a_0 can have influenced the behaviour of the flow "in the past", i.e. for smaller n than the current n; thus, we can get "stuck in a cycle", as we show below. However, the other a_i can be considered "genuinely random", because the map $\mathcal{C}_{a,b,c}$ did never "use" the values of those digits in determining the next point in the flow, so we can do "as if we never saw a_i with $i \geq 1$ before".

Thus, it follows that *if after applying the affine part $x_n \mapsto ax_n + b$ of the map $\mathcal{C}_{a,b,c}$ the first digit of $ax_n + b$ is zero (i.e. in the case that $k_n \geq 1$), then the length of the tail of zeros $a_0, a_1, a_2, \ldots$ is geometrically distributed,* with "succes probability" equal to $(c-1)/c$ (where "success" means encountering the first non-zero digit). Thus, if $k_n \geq 1$, then the expected value of k_n is $c/(c-1)$.

We distinguish two cases.

Case 1: The permutation $x \mapsto ax + b$ of $\mathbb{Z}/c\mathbb{Z}$ is not a cycle. Then it is easily seen that after each iteration of $\mathcal{C}_{a,b,c}$ for which $k_n \geq 1$, we arrive with positive probability at a number x' that ends with a digit in $\mathbb{Z}/c\mathbb{Z}$ that is not in the same cycle as the number 0. Such numbers x' are mapped to $ax' + b$ by $\mathcal{C}_{a,b,c}$; this keeps being so indefinitely when we iterating $\mathcal{C}_{a,b,c}$, the first digit is "trapped forever in the cycle without zero". So the flow of x ends with the "trivial" recursion $x_{n+1} = ax_n + b$.

As the possibility to become trapped in a cycle that does not contain zero, is positive at each iteration, x is "almost surely" eventually trapped.

Case 2: The permutation $x \mapsto ax + b$ of $\mathbb{Z}/c\mathbb{Z}$ is a cycle.

It is not hard to see that when we consider any "newly arrived" digit a_0 (that we "did not see before, because we just shifted the digit sequence") to be uniformly random, then it takes on average $2/c$ iterations before it happens again that the first digit is zero. Hence, we have (as always only heuristically, not as a truth) that $\mathbf{P}(k_n \geq 1) = 2/c$;

hence, by the conditional expectation $\mathbf{E}(k_n : k_n \geq 1) = c/(c-1)$ that we derived before, it follows that $a_0\ k \geq 1$, and then the average k is

$$\mathbf{E}(k_n) = \frac{2}{c} \cdot \frac{c}{c-1} = \frac{2}{c-1}$$

(for all sufficiently large k). It follows by the law of large numbers that

$$\lim_{n\to\infty} \mathrm{AM}_n(\mathbf{k}_n) = \frac{2}{c-1}.$$

It follows by Theorem 7.5.4.5 that

$$\lim_{n\to\infty} \log(x_n) - \log(x_0) - n\log\left(\frac{a}{c^{(2/(c-1))}}\right) = 0. \tag{7.6}$$

Note that the assumption that x is not cyclic implies that $\lim_{n\to\infty} x_n = \infty$, and hence that

$$\frac{a}{c^{(2/(c-1))}} > 1, \qquad \text{that is,} \quad c^{2/(c-1)} < a. \tag{7.7}$$

Note that $c^{2/(c-1)}$ is a decreasing function of c, and that only the values $(a, c) = (2, 3), (3, 2), (5, 2)$ don't satisfy (7.7). satisfy $c^{2/(c-1)} > a$. we also have the restriction that $x \mapsto ax + b$ should be a cyclic permutation modulo c, and it is easily seen that from the three tuples above, this holds only for $(a, c) = (3, 2)$. In other words, only for maps $\mathcal{C}_{3,b,2}(x)$ we arrive at a "contradiction", so the assumption that x is not in a cycle seems to be false. Thus, for those maps, we expect that all points are cyclic. Moreover, for extremely large starting points x, we still expect that the heuristic argument that led to (7.6) is valid, up to some point where x_n becomes too "small". So we expect for very large x that the flow of x follows a straight line with downward slope $\log\left(\frac{a}{c^{(2/(c-1))}}\right)$. We illustrated this for six concrete (random) large starting points in Figure 7.3, for $b = 1007$. The graph of $\log(x_n)$ is depicted, as well as the expected line given by (7.7).

For tuples (a, b, c) with $(a, b) \neq (3, 2)$, on the other hand, we expect that the above argument makes it plausible that the asymptotic equation (7.6) is "increasingly likely" when the starting point x gets larger, because it seems 'increasingly unlikely" that the point is preperiodic. Figure 7.2 shows $\log(x_n)$ for flow graphs of some (random) small starting points, and also the expected asymptotic line with upward slope $\log\left(\frac{a}{c^{(2/(c-1))}}\right)$. From top to bottom in the figure, the scale of the axes is multiplied by 10; note that the of graphs of $\log(x_n)$ in the bottom figures are really relatively close to the expected line.

Combining the heuristic arguments with Corollary 7.5.7, it seems unlikely that there are arbitrary large cycles. Thus, we expect that there are finitely many cycles, and that the density of starting points below x that are preperiodic, converges to zero as $x \to \infty$.

This expected behaviour about the distribution of preperiodic points is illustrated in Figure 7.4; each coloured graph represents a cycle (all cycles that contain numbers smaller than 10^6 are depicted), and the graph for each cycle is a cumulative plot of the number of points that eventually arrive in that cycle. The upper two graphs are examples of the case where we expect convergence of all points to cycles, because $(a, b) = (3, 2)$;

the distribution over the different cycles looks at first glance asymptotically linear. The graphs suggest that in both examples, there is a cycle in which roughly half of the points eventually land, but also cycles that are much “rarer” as an eventual cycle (to some of them, less than 1% of the points seem to converge). In the four graphs below them, we have $(a, b) \neq (3, 2)$, so we are in the situation that we expect the asymptotic density of preperiodic points to be zero.

The graphs make the conclusions of the heuristic argument sound plausible. But the arguments provide in no way a proof, since the assumption that the digits are “random” is not really true. ♪

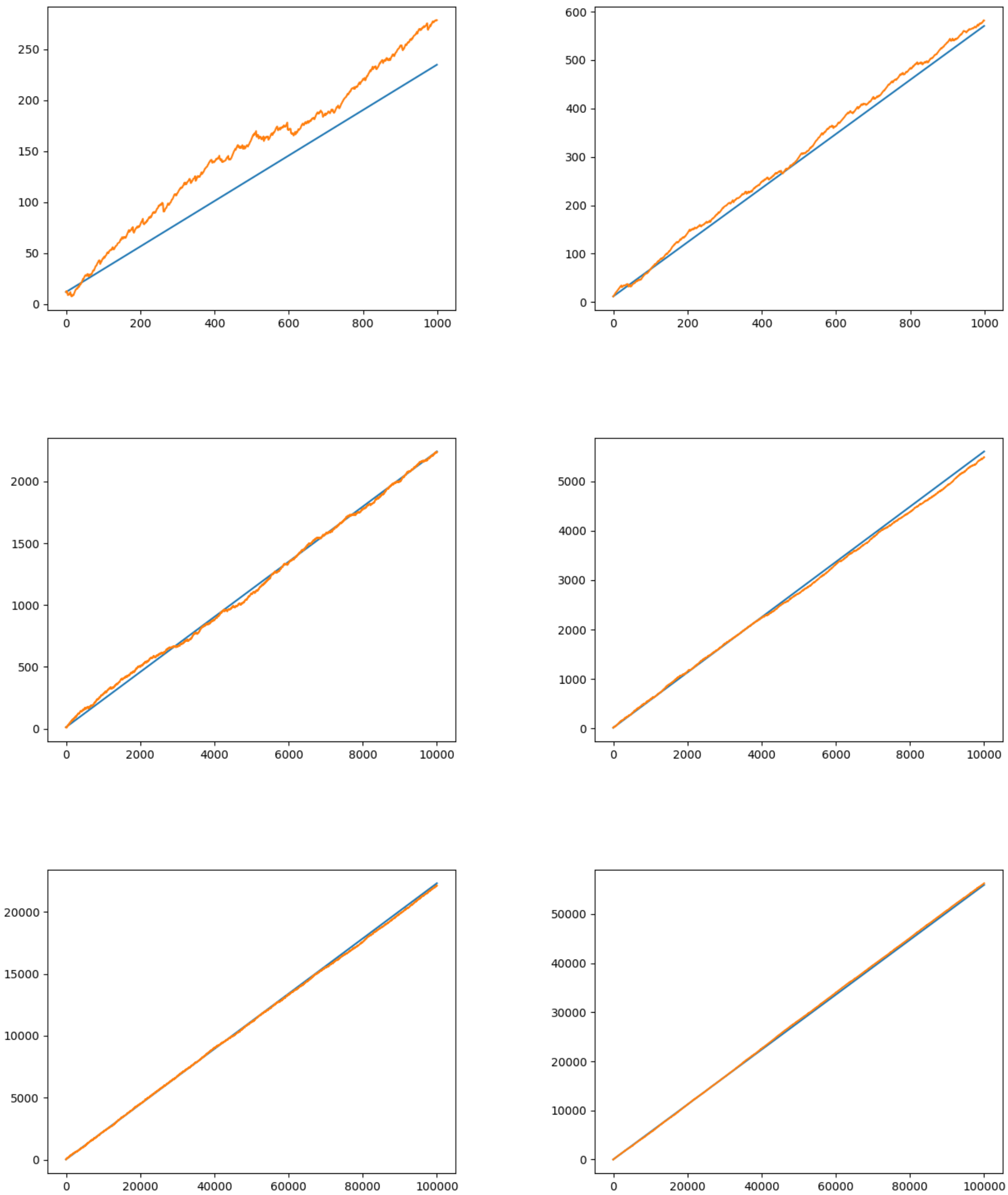


Figure 7.2: Ascending behaviour for $\mathcal{C}_{a,b,c}$ with $(a, b, c) = (5, 1, 2)$ (left three graphs) and with $(a, b, c) = (7, 1, 2)$ (right three graphs). From top to bottom, the scale of the axes increases with factors of approximately 10.

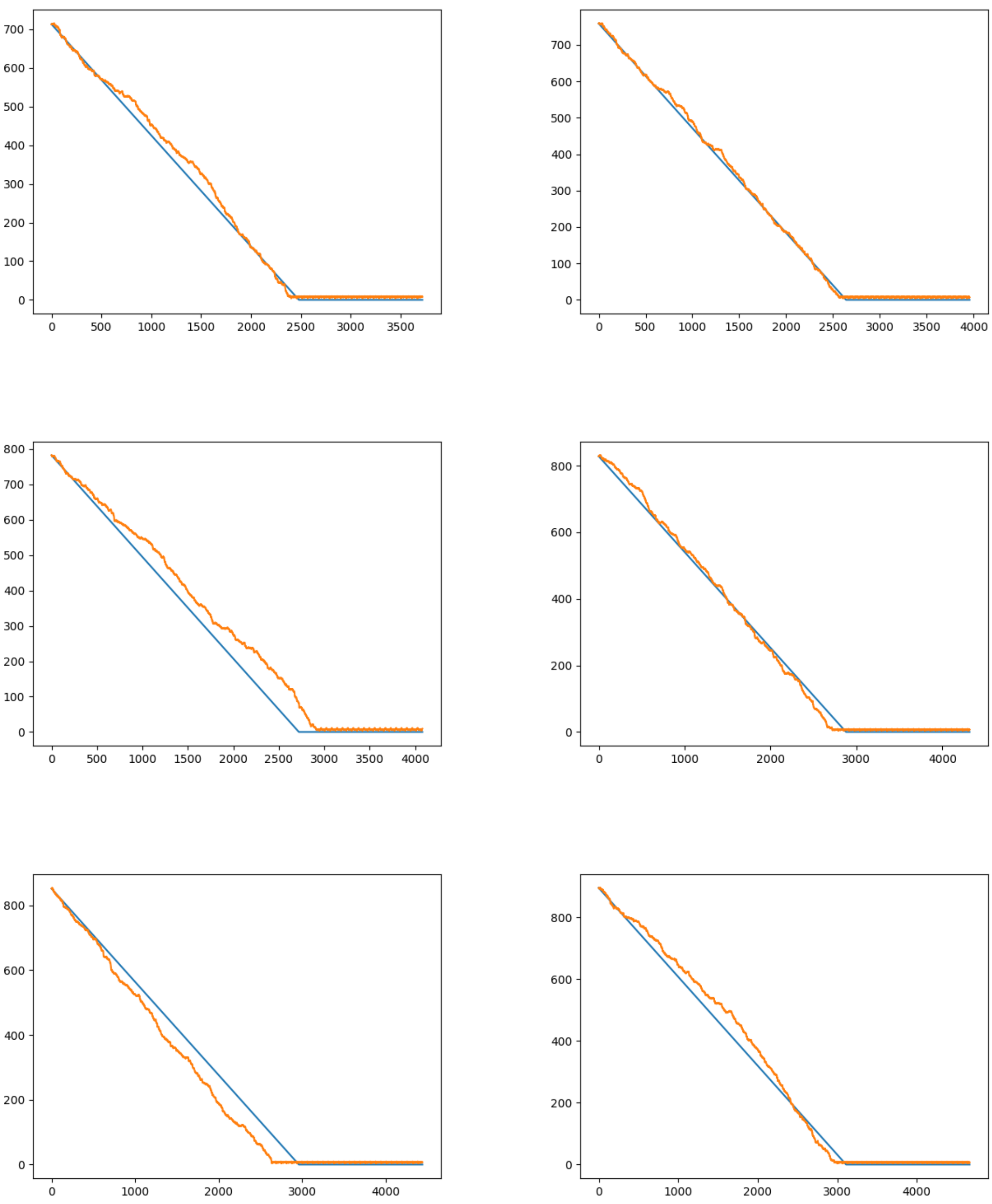
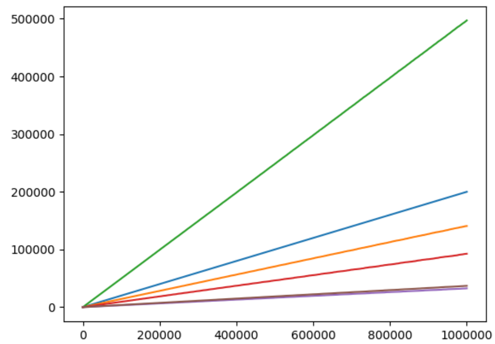
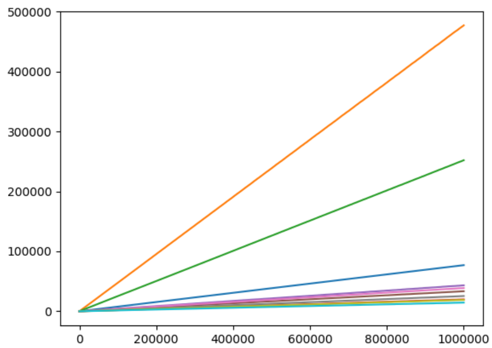


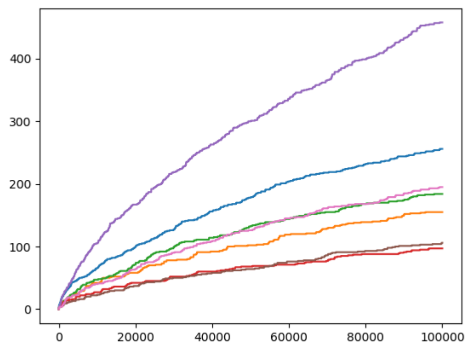
Figure 7.3: Descending behaviour for $\mathcal{C}_{3,1007,2}$, for (random) large starting points.



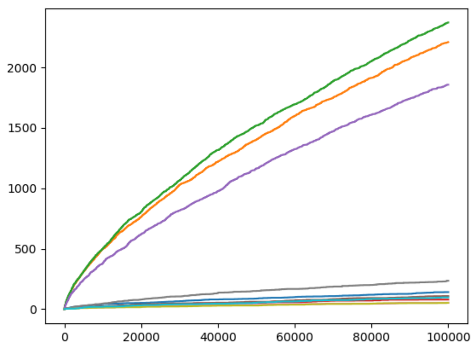
$\mathcal{C}_{3,5,2}$



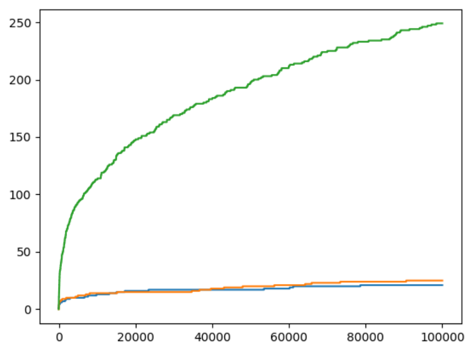
$\mathcal{C}_{3,13,2}$



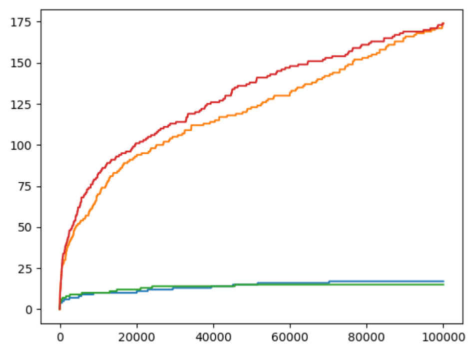
$\mathcal{C}_{5,3,2}$



$\mathcal{C}_{5,9,2}$



$\mathcal{C}_{7,5,2}$



$\mathcal{C}_{7,15,2}$

Figure 7.4: Cumulative plots of the distribution numbers over eventual cycles (in case the numbers is preperiodic). Each cycle is represented by a coloured graph.

Bibliography

[AD89] J. Aczél and J. Dhombres. *Functional Equations in Several Variables*, volume 31 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, 1989.

[AS64] Milton Abramowitz and Irene A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover Books on Mathematics. Dover Publications, 1964.

[BCHvdM19] Jakub Byszewski, Gunther Cornelissen, Marc Houben, and Lois van der Meijden. Dynamically affine maps in positive characteristic. *arXiv: Number Theory*, 2019.

[BM00] Edward G. Belaga and Maurice Mignotte. Cyclic structure of dynamical systems associated with $3x + d$ extensions of Collatz problem. *hal-00129656*, 2000.

[BM06] Edward G. Belaga and Maurice Mignotte. The Collatz problem and its generalizations: Experimental data. table 1. primitive cycles of $(3n + d)$-mappings. *hal-00129727*, 2006.

[Bol01] Béla Bollobás. *Random Graphs*. Number 73 in Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2nd edition, 2001.

[Bul03] P.S. Bullen. *Handbook of Means and Their Inequalities*, volume 560 of *Mathematics and Its Applications*. Springer, 2003.

[CFKT14] Peter Csiba, Ferdinánd Filip, Attila Komzsíkb, and János T. Tóth. On the existence of the generalized Gauss composition of means. *Annales Mathematicae et Informaticae*, 43:55–65, 2014.

[CM64] John Cook and Marvin Minsky. Universality of tag systems with $p = 2$. *Journal of the ACM*, 11(1):15–20, 1964.

[Cox84] David A. Cox. The arithmetic-geometric mean of Gauss. *L'Enseignement Mathématique*, 30, 1984.

[DMP05] Zoltán Daróczy, Gyula Maksa, and Zsolt Páles. Functional equations involving means and their Gauss composition. *Proceedings of the American Mathematical Society*, 134(2):521–530, 2005.

[Eli93] Shalom Eliahou. The $3x+1$ problem: new lower bounds on nontrivial cycle lengths. *Discrete Mathematics*, 118(1–3):45–56, 1993.

[Fla73] Harley Flanders. Differentiation under the integral sign. *The American Mathematical Monthly*, 80(6):615–627, 1973.

[FO90] Philippe Flajolet and Andrew M. Odlyzko. Random mapping statistics. *Advances in Cryptology – Eurocrypt 1989*, Lecture Notes in Computer Science, volume 434, 1990.

[Haj13] Mowaffaq Hajja. Some elementary aspects of means. *International Journal of Mathematics and Mathematical Sciences*, 2013, 2013.

[Har19] Kevin Hartnett. Mathematician proves huge result on 'dangerous' problem. `https://www.quantamagazine.org/mathematician-terence-tao-and-the-collatz-conjecture-20191211/`, december 2019. Accessed: 2020-08-21.

[HLP34] G.H. Hardy, J.E. Littlewood, and G. Pólya. *Inequalities*. Cambridge University Press, 1st edition, 1934.

[Knu97] Donald E. Knuth. *The Art of Computer Programming*, volume 1. Addison-Wesley Professional, 3th edition, 1997.

[Lag85] Jeffrey C. Lagarias. The $3x + 1$ problem and its generalizations. *The American Mathematical Monthly*, 92(1):3–23, 1985.

[Lag10] Jeffrey C. Lagarias. *'The $3x + 1$ problem: an overview', paper published in book 'The Ultimate Challenge: The $3x + 1$ Problem'*. American Mathematical Society, 2010.

[Lag11] Jeffrey C. Lagarias. The $3x + 1$ problem: An annotated bibliography (1963–1999) (sorted by author) (january 1, 2011 version). *Unpublished. Preprint: arXiv:math/0309224v13*, 2011.

[Lag12] Jeffrey C. Lagarias. The $3x + 1$ problem: An annotated bibliography, II (2000-2009). *Unpublished. Preprint: arXiv:math/0608208v6*, 2012.

[Liv03] Mario Livio. *The Golden Ratio: The Story of phi, the World's Most Astonishing Number*. Broadway Books, 2003.

[Mas64] S.Ju. Maslov. On E. L. Post's "tag problem" (Russian). *Trudy Matematicheskogo Instituta imeni V.A. Steklova*, 72:5–56, 1964.

[Mat99] Janusz Matkowski. Iterations of mean-type mappings and invariant means. *Annales Mathematicae Silesianae*, 13(13):211–226, 1999.

[Mol08] Liesbeth De Mol. Tag systems and Collatz-like functions. *Theoretical Computer Science*, 390:92–101, 2008.

[MP15] Janusz Matkowski and Zsolt Páles. Characterization of generalized quasi-arithmetic means. *Acta Scientiarum Mathematicarum*, 81(34), 2015.

[Roo20] Eric Roosendaal. On the $3x + 1$ problem. `www.ericr.nl/wondrous/`, 2020. Accessed: 2020-08-17.

[Ros10] Sheldon M. Ross. *Introduction to Probability Models*. Academic Press, 10th edition, 2010.

[Sal76] Eugene Salamin. Computation of π using arithmetic-geometric mean. *Mathematics of Computation*, 30(135):565–570, 1976.

[SdW05] John Simons and Benne de Weger. Theoretical and computational bounds for m-cycles of the $3n+1$ problem – version 1.44 (2010). *Acta Arithmetica*, 117(1):51–70, 2005.

[Sha56] Claude Shannon. A universal Turing machine with two internal states. *Automata Studies*, 34:157–165, 1956.

[Sim07] John Simons. On the (non)-existence of m-cycles for generalized Syracuse sequences. *Acta Arithmetica*, 131(3), 2007.

[SV11] József Sándor and Antoine Verroken. On a limit involving the product of prime numbers. *Notes on Number Theory and Discrete Mathematics*, 17:1–3, 2011.

[Tao19] Terence Tao. Almost all orbits of the Collatz map attain almost bounded values. *arXiv:1909.03562 [math.PR]*, 2019.

[Tur37] A.M. Turing. On computable numbers, with an application to the Entscheidungsproblem. *Proc. London Math. Soc.*, 42:230–265, 1937.

[Wan63] Hao Wang. Tag systems and lag systems. *Mathematische Annalen*, 152(1):65–74, 1963.