Phase Transition Behavior in Combinatorics and Computation

Thesis submitted in partial fulfillment
of the requirements for the degree of
"Doctor of Philosophy"

by

Eran Omri

Submitted to the Senate of Ben-Gurion University
of the Negev

April 2009

Beer-Sheva

PHASE TRANSITION BEHAVIOR IN COMBINATORICS AND COMPUTATION

Thesis submitted in partial fulfillment
of the requirements for the degree of
"DOCTOR OF PHILOSOPHY"

by

Eran Omri

Submitted to the Senate of Ben-Gurion University
of the Negev

Approved by the advisor Prof. Amos Beimel________________
Approved by the advisor Prof. Menachem Kojman________________
Approved by the Dean of the Kreitman School of Advanced Graduate Studies________________

April 2009

Beer-Sheva

This work was carried out under the supervision of

Prof. Amos Beimel
and
Prof. Menachem Kojman

in the Department of Computer Science

Faculty of Natural Sciences

Acknowledgments

It is a pleasure to thank my advisors, Menachem Kojman and Amos Beimel, who put their hearts into my education as a researcher and into the completion of my thesis. I consider myself lucky to have had them as my advisors. I am grateful to Menachem for being much more than an advisor. Throughout the years, I have always felt his door was open to me for professional discussion, advice, and conversation about life. He has supported me in every possible way in the most fatherly fashion. Amos persistently worked on improving my skills as a researcher, always in a most patient, kind, and caring manner. His advice, his constructive criticism, and even his home were constantly made available to me (even while on sabbatical abroad). I thank Amos for all this and for being an advisor and a true friend (never confusing the two).

I am deeply grateful to Kobbi Nissim for his support and for being very generous in sharing his abundant knowledge and profound insights with me. I sincerely thank Andreas Weiermann for his kindness and for constantly inviting me to take part in his many research projects. I am grateful to Boaz Ben-Moshe for being my friend and for ever showering me with new research problems. I would like to thank Adam Smith, Yuval Ishai, and Cynthia Dwork for useful conversations.

I have enjoyed studying and conducting research with some great people. For that I thank Amos Beimel, Boaz Ben-Moshe, Paz Carmi, Michael Elkin, Menachem Kojman, Gyesik Lee, Kobbi Nissim, Andreas Weiermann, Noam Livne, Gil Segev, Ilan Orlov, Naomi Kirshner, and Enav Weinreb.

I deeply thank the people of the department of computer science at the Ben-Gurion University, including administrative staff, lab staff, students, and faculty members. Special thanks to Mazal Gagulashvili whose kindness and professional capabilities seem endless. I am also grateful to Shlomi Dolev, Klara Kedem, Eyal Shimony, and Michal Ziv-Ukelson for being very friendly and helpful. I thank Mike Codish for many talks about life and about coffee. I thank Matya Katz for sharing his ideas with me.

I am utterly grateful to my family and to my close friends for surrounding me with so much love and encouragement. I thank my parents-in-law Ligina and Yosef for their kindness and for hosting and feeding me on the many nights I had to stay in Be'er-Sheva.

I dedicate this thesis to four people who taught me the art of love. My dear parents, Tami and Omri, who have been there every step of the way, always eager to help and protect. To Anna, my wife, the love of my life, and my best friend – always most patient and supportive, allowing me to follow my dreams. Finally, to the little wonder, Eliya, who came into this world seven months ago and straight into my heart.

To my father and my mother,
to Anna, and to Eliya.

Contents

List of Figures

Abstract

In this thesis we investigate phase transition behaviors in several systems in combinatorics and in private computation. The term phase transition originated in physics, where it describes the transformation of a thermodynamic system from one phase to another. A phase transition of a system is usually characterized by a sharp threshold point at which a change in some underlying parameter of the system causes an abrupt change in one or more global properties of the system. It is now common in many research areas to use the term phase transition to describe similar phenomena of systems in those areas. Moreover, searching for such threshold behaviors in known systems have often led to new understanding of these systems. Examples are found in probability theory, statistics, computer science, combinatorics, logic, economics, and political science.

In this work we follow this paradigm investigating the behavior of fast growing combinatorial functions, and the limitations of some models of private data analysis. Generally speaking, we follow a few basic steps in our investigations of these systems. When interested in understanding the behavior of some property (e.g., the provability of Ramsey type propositions in some powerful axiomatic system $\mathcal{T}$), we consider two extreme cases (i.e., one in which the property holds and the other where it does not), and parameterize the system by emphasizing on the difference between the two extremes (e.g., if these Ramsey type propositions are provable when stated for a constant number of colors, but unprovable with linearly many colors, then we will parameterize on the number of colors). Finally, we consider the existence of a sharp threshold point and try to compute it if one exists.

The results of this thesis belong to three main areas that are described below.

Ramsey Theory

Ramsey Theory is a branch of combinatorics and graph theory that deals with the idea that within a sufficiently large system, however disordered, there must be some order. This theory is British mathematician Frank P. Ramsey's most celebrated contribution to mathematics. In this thesis we consider Ramsey functions (usually referred to as Ramsey numbers) for graph-colorings.

The most basic query of this area is known as the Party Puzzle: How many people, arbitrarily chosen, would suffice to ensure that amongst them there are either k of whom every two know each other or k of whom no two know each other? The basic Ramsey Theorem asserts that this question is legitimate, as for any natural number k, there exists a natural number N, so that any N given people contain k such that either all know each other or no two of them know each other.

Define the *complete graph* $K_N = (V, E)$ to be an undirected graph over N vertices, $V = \{0, \ldots, N-1\}$, containing all possible edges (i.e., the edges of K_N are the set $E = \{(i,j) : \quad 0 \leq i < j < N\}$). Ramsey's theorem for pair colorings asserts that for every two natural numbers k and c, there exists an N such that, for any coloring C of the edges of the complete graph K_N that uses no more than c colors, there is a sub-graph H of K_N that is a copy of the complete graph K_k (i.e., the complete graph over k vertices), such that all edges of H are assigned the same color by C (we then say that H is C-homogeneous). Furthermore, it is known that $N = c^{ck}$ suffices. We, however, are usually interested in the least N that suffices, which is called the standard Ramsey number for k, c and denoted by $R(k, c)$. By way of example, in the case of the party puzzle c is fixed to be 2 (i.e., know or don't-know).

In this thesis we consider the behavior of the Ramsey numbers of two variants of Ramsey's theorem. Both were originally proposed as natural Gödel sentences, that is, as examples of combinatorial propositions that are true but unprovable in some powerful axiomatic theory. These propositions are known as the regressive Ramsey theorem and the Paris-Harrington Ramsey theorem. We consider parameterized versions of these two theorems for pair colorings and prove that they have sharp phase transition thresholds.

g-regressive Ramsey numbers. Given a function $g : \mathbb{N} \to \mathbb{N}$, a coloring of the edges of the complete graph $C : E \to \mathbb{N}$ is g-regressive if for all edges $(i,j) \in E$ it holds that $C((i,j)) \leq i$ (where $i < j$). A subgraph H of K_N is called min-homogeneous for a coloring $C : E \to \mathbb{N}$ if C restricted to $E(H)$ (the edges of H) only depends on the minimum vertex in an edge, that is, for every two edges $(i, j_1), (i, j_2) \in E(H)$ it holds that $C((i, j_1)) = C((i, j_2))$.

For a given function g, the g-regressive Ramsey theorem for pairs states that for every natural number k there exists a natural number N, such that for every g-regressive coloring C of the edges of K_N, there exists a copy H of K_k inside K_n that is min-homogeneous for C. Let $R_g^{\text{reg}}(k)$ be the least natural number that satisfies the g-regressive Ramsey theorem for pairs for k.

Kanamori and McAloon introduced the regressive Ramsey theorem, which is exactly the g-regressive Ramsey theorem for $g = \text{Id}$, and showed that $R_{\text{Id}}^{\text{reg}}(\cdot)$ has an Ackermannian growth rate (i.e., grows faster than any primitive recursive function). On the other hand, for constant g, it is easy to see that $R_g^{\text{reg}}(\cdot)$

grows slower than the standard Ramsey number for pairs, and hence is primitive recursive.

Hence, we have for the two extremes that $R^{\mathrm{reg}}_{\mathrm{Id}}(\cdot)$ is Ackermannian, while if g is constant then $R^{\mathrm{reg}}_{g}(\cdot)$ is primitive recursive. We look for a threshold function for this transition from primitive recursive growth rate to Ackermannian growth rate. We consider the combinatorial behavior of these Ramsey numbers and show that a sharp threshold exists and lies above all functions $n^{1/f^{-1}(n)}$ obtained from a primitive recursive f and below $n^{1/\operatorname{Ack}^{-1}(n)}$. Worded differently, for a nondecreasing and unbounded g to have primitive recursive g-regressive Ramsey numbers it is necessary and sufficient that g is eventually dominated by $n^{1/t}$ for all $t > 0$ and that the rate at which g gets below $n^{1/t}$ is not too slow: if g gets below $n^{1/t}$ only after an Ackermannianly long time $M(t)$, then the g-regressive Ramsey numbers are still Ackermannian.

We also identify the threshold below which g-regressive colorings have usual Ramsey numbers, that is, admit homogeneous, rather than just min-homogeneous sets, and give a lower bound of $A_{53}(2^{2^{274}})$ on the Id-regressive Ramsey number of $k = 82$, where A_{53} is the 53-rd approximation of Ackermann's function.

g-large Ramsey numbers. Given a function $g : \mathbb{N} \to \mathbb{N}$, the g-large Ramsey theorem for pairs states that for every two natural numbers k, c there exists a natural number N, such that for every coloring C of the edges of K_N with c colors, there exists a C-homogeneous copy H of $K_{k'}$ (for $k' \geq k$) inside K_n, and the minimal vertex i in H satisfies $g(i) < k'$, i.e., $H = (V', E')$ such that for all $e_1, e_2 \in E'$ it holds that $C(e_1) = C(e_2)$ and $|V'| \geq \max\{k, g(\min(V'))\}$.

Let $R^*_g(k, c)$ be the least natural number that satisfies the g-large Ramsey theorem for pairs for k, c. The Paris-Harrington Ramsey theorem for pairs is exactly the g-large Ramsey theorem for pairs with $g = \mathrm{Id}$. Erdős and Mills have shown that $R^*_{\mathrm{Id}}(\cdot,\cdot)$ has an Ackermannian growth rate. The g-large Ramsey theorem for a constant function g is practically the standard Ramsey theorem.

Hence, in the two extremes we have that if g is constant then $R^*_g(\cdot,\cdot)$ is primitive recursive, and if $g = \mathrm{Id}$ then $R^*_g(\cdot,\cdot)$ is Ackermannian. We look for a threshold function for this transition from a primitive recursive growth rate to an Ackermannian growth rate. We consider the combinatorial behavior of these functions and show that a sharp threshold exists and lies above all functions $\log(n)/f^{-1}(n)$ obtained from an increasing primitive recursive f and below the function $\log(n)/\operatorname{Ack}^{-1}(n)$. Worded differently, for a nondecreasing and unbounded g to have primitive recursive g-large Ramsey numbers it is necessary and sufficient that g is eventually dominated by $\log(n)/t$ for all $t > 0$ and that the rate at which g gets below $\log(n)/t$ is not too slow, namely, is primitive recursive in t: if g gets below $\log(n)/t$ only after an Ackermannianly long time $M(t)$, then the g-large Ramsey numbers are still Ackermannian.

Iteration hierarchies

We consider functions defined via diagonalization from an iteration hierarchy (of Grzegorczyk type). Assume that we are given two functions $g, h : \mathbb{R} \cap [0, \infty) \to \mathbb{R} \cap [0, \infty)$. For $r \in \mathbb{R}$, let $\lfloor r \rfloor$ denote the largest integer not exceeding r. We then consider the following diagonalization: For $x \in \mathbb{N}$, let

$$\begin{aligned} B(g,h)_0(x) &\triangleq g(x), \text{and let} \\ B(g,h)_{k+1}(x) &\triangleq B(g,h)_k^{\lfloor h(x) \rfloor}(x) \quad \text{i.e. } \lfloor h(x) \rfloor \text{ many iterations,} \\ B(g,h)_\omega(x) &\triangleq B(g,h)_{\lfloor x \rfloor}(x). \end{aligned}$$

For example, the Ackermann function is defined as $\mathrm{Ack}(n) = B(g,h)_\omega(n)$, where $g(x) = x + 1$ and $h = \mathrm{Id}$, and that $\mathrm{A}_i(n) = B(g,h)_i(n)$. The question arises: for which pairs of functions g, h is the resulting diagonalization primitive recursive? When does it become Ackermannian?

Specifically, fixing $g(x) = x + 1$ (as in the Ackermann hierarchy), we investigate the threshold function h at which the resulting hierarchy stops being primitive recursive and becomes Ackermannian. We show that a sharp threshold for h exists and that it is intrinsically related to g-regressive Ramsey numbers.

We then consider a class of start functions g (starting with $g(x) = x + \varepsilon$ for $0 < \varepsilon \leq 1$ and constantly increasing them). We show appropriate classes of iteration modulus h for which the resulting classes of hierarchies are slow-growing and later we show appropriate classes of iteration modulus h for which the resulting classes of hierarchies are fast-growing. These two classes are very similar, yet result in extremely different growth rates.

Distributed differential privacy

In private data analysis our goal is to compute some function applied to a set of individual data with an overall privacy requirement of protecting the private information of individuals. The challenge is therefore imposed by the inherent tension between the need to compute some valuable estimation of the function (utility) on the one hand, and the need to protect the information of individuals (privacy) on the other. An exemplifying scenario for the notion of private data analysis is of a hospital database containing medical records of patients. The hospital, which would like to enable valuable medical research based on the information stored in the database, must ensure, for ethical as well as legal reasons, that the privacy of its patients is protected, i.e., no information regarding the medical condition of any specific patient can be “traced back” to that individual. This example illustrates the tension between supplying utility and preserving privacy.

We consider a criterion, suggested by Dwork et al., for analyses that preserve privacy of individuals, called *differential privacy*. Much consideration was given to constructing differentially private analyses in a setup where all the records of the database are held by some trusted entity. We consider a scenario where each record of the database is held by an individual party and the centralized trusted entity is implemented by a distributed protocol executed by the parties. The task of securely computing a given function in the distributed setting is known to be achievable using secure function evaluation (SFE) protocols. Thus, a natural paradigm for constructing distributed protocols for differentially private analyses is: first choose a differentially private analysis (i.e., choose *what* to compute) while abstracting away implementation issues, then construct an SFE protocol for this analysis (i.e., choose *how* to compute).

This paradigm for constructing protocols is both simple and modular; however, since SFE and differential privacy are significantly different requirements, it may result in non-optimal protocols. We initiate an examination of whether there are advantages to a paradigm where both a differentially private analysis and the distributed protocol for implementing it are constructed simultaneously. This examination highlights the relation between private data analysis and SFE. In particular, we examine the case of binary sum queries (where each of n parties holds a sensitive bit). We consider the communication complexity of distributed protocols computing an approximation for the binary sum, and show a sharp phase transition threshold of $\approx \sqrt{n}$ on the magnitude of error we allow in an analysis.

Keywords. Phase transition, Paris-Harrington theorem, Kanamori-McAloon theorem, Ackermannian functions, rapidly growing Ramsey numbers, Fast growing hierarchy, Threshold, Differential privacy, Secure Function Evaluation, Sum Queries.

Chapter 1

Introduction

1.1 Phase Transition

This thesis is an investigation of phase transition behaviors in combinatorics and in distributed private data analysis. The notion of phase transition originated in physics, where it is used to describe the transformation of a thermodynamic system from one phase to another. In physics, a phase transition of a system is characterized by a small change in an underlying parameter of the system, which causes an abrupt change in one or more physical properties of the system. A typical example is that of the transition of liquid water into vapor at boiling point. It is now common to use the terminology of phase transitions to describe systems exhibiting threshold behavior in other fields, such as probability theory, statistics, computer science, combinatorics, logic, economics, and political science. Many times, investigating the phase transition behavior of events in these fields by parameterizing the system, proving it exhibits some threshold behavior, and then finding the location of a sharp threshold point, yields profound understanding of some of the natural questions regarding these systems. This thesis follows the above paradigm in the fields of Ramsey theory, function hierarchies, and distributed private computation.

The investigation of phase transition behavior in problems in mathematics and computer science and applications of techniques from statistical physics to a probabilistic variant of a known mathematical model, have flourished in the last few decades. One classic example is found in the area of graph theory. In the late 1950s and early 1960s, Erdős and Rényi [36, 37] introduced the model of random graphs. A random graph $G(n, p)$ over n vertices is sampled by selecting edges independently, each with probability $p = p(n)$. Erdős and Rényi have already shown phase transition behaviors for properties of random graphs. For instance, they considered the size of the largest component of $G(n, c/n)$ for a constant $c > 0$, and showed that it has order $O(\log n)$ w.h.p.

if $c < 1$, and order $\theta(n)$ w.h.p. if $c > 1$.

In continuation to the work of Erdős and Rényi, much research was conducted investigating phase transition phenomena in random graph theory (see, e.g., [12, 13, 48, 11]). Friedgut and Kalai showed that every monotone graph property has a sharp threshold [41]. Interestingly for the context of this thesis, thresholds for Ramsey type properties of random graphs were considered (e.g., [42, 21]). Another topic where much research was done looking for sharp thresholds, which attracted the attention of computer scientists as well as physicists, is the area of computational complexity, and specifically the k-SAT problem (see, e.g., [40, 2]). We note that here, again, the idea is to consider the probability that an event occurs, i.e., a random CNF formula is satisfiable. That is, for a formula over boolean variables $x_1, \ldots, x_n$, obtained by uniformly choosing M clauses (each of size k) of the possible $2^n\binom{n}{k}$ and letting the formula be the disjunction of the chosen clauses, we ask whether there is an assignment of values to $x_1, \ldots, x_n$ that satisfies the formula. Sharp thresholds are known to exist for each $k \geq 3$ between satisfiability and unsatisfiability of a random formula, as a function of the ratio between M and n [40]. The question of how sharp these thresholds are is an open one. The reader is referred to [67] for more on the phase transition behavior of the above problems and many other probabilistic variants of mathematical and computation complexity problems, such as MAX-CUT, percolation theory, and hardness of approximation.

Phase transition phenomena are not confined to probabilistic properties of a system. By way of example, for the 3-SAT problem, we view the result of [61] (or rather a corollary of it) as a sharp phase transition of the MAX-3-SAT approximation problem at $7/8$; that is, their result asserts that 3-SAT cannot be efficiently approximated within a factor of $7/8 + o(1)$, unless P = NP, while approximating MAX-3-SAT by a factor of $7/8$ is known [50]. Much study of such phase transition phenomena was conducted in recent years in the area of logic and combinatorics. Less explicit phase transition results are at the heart of the recent line of rigorous investigation of private data analysis. We next mention a few of these works.

1.1.1 Phase Transition Phenomena in Logic and Combinatorics

The last few years have seen an unexpected series of results that bring together independence results in logic, analytic combinatorics, and Ramsey Theory. These results can be described intuitively as phase transitions from provability to unprovability of an assertion by varying a threshold parameter [4, 18, 57, 79, 81, 82, 75, 85, 84]. Another face of this phenomenon is the transition from slow-growing to fast-growing computable functions [80, 78, 83].

Ramsey theory. Ramsey theory is a branch of combinatorics that deals with the idea that within a sufficiently large system, however disordered, there must be some order. The evolution of this field was sparked by the paper "On a problem of formal logic" [71] by the British mathematician Frank P. Ramsey. In this paper Ramsey was addressing a special case of the decision problem for the first-order predicate calculus, proposed by David Hilbert, and proved his most famous "Ramsey theorem" as a means for proving that special case. It is now known that, not only did Ramsey not need his combinatorial argument for his proof, but that the general case of the decision problem cannot be solved. Interestingly, even though he put some effort into trying to solve the problem, Ramsey has criticized Hilbert saying that he had attempted to reduce mathematics to "...a meaningless game with marks on paper".

In Chapter 2 we state both the finite and the infinite versions of Ramsey's theorem and present a proof for the finite case. Let us describe a simple case of a Ramsey type problem called "the party puzzle". Given a natural number k, we consider the least natural number N, such that in any party of N people, there will always be either k of which all know each other, or k of which none know each other. We say that N is the Ramsey number of k. The Ramsey number of 3 is 6, the Ramsey number of 4 is 18, and the Ramsey number of 5 is only known to be within the interval $[43, 49]$.

Gödel's incompleteness. In 1931, Kurt Gödel [44], in a brilliant paper on formally undecidable propositions, proved that for any consistent, effectively generated formal theory that proves certain basic arithmetic truths, there is an arithmetic statement such that neither it nor its negation is provable in the theory. Gödel's work implied that the proposal for the foundation of classical mathematics known as Hilbert's Program cannot be carried out. However, as the propositions designed by Gödel to prove his incompleteness results were of a very unique form and specifically were self-referent, one might have questioned the relevance of these results to so called natural propositions. For instance, the question whether finite combinatorial theorems that are independent of powerful axiomatic systems such as first-order PA (Peano Arithmetic) can be discovered, was still valid at the time Gödel's work was published.

Paris-Harrington Gödel Sentence. J. Paris made some important advance in the late 1970s. Building on joint work with L. Kirby, he used model-theoretic techniques to investigate arithmetic incompleteness and proved theorems of finite combinatorics that were unprovable in PA [65]. Later, in [66], J. Paris and L. Harrington went on to present a proof that a straightforward variant of the familiar finite Ramsey theorem is independent of PA. We next state the main theorem of their work.

Theorem (Paris Harrington [66]): A nonempty $H = \{x_1, \ldots, x_\ell\} \subseteq \mathbb{N}$ (where $x_1 < x_2 < \ldots < x_\ell$) is *relatively large* if $|H| \geq x_1$ (i.e., $|H| \geq \min(H)$).

1. PH $\equiv$ For any natural numbers $k, c, d > 0$ there exists an $N \in \mathbb{N}$ such that for any coloring C of all d-tuples over $\{0, \ldots, N-1\}$ with colors $\{0, \ldots, c-1\}$, there is a relatively large $H \subseteq \{0, \ldots, N-1\}$ that is homogeneous for C and of cardinality at least k.

2. There is no proof from PA of PH.

The *Paris Harrington Ramsey number* of k and c, and d, denoted $R^*(k, c, d)$, is the least N that satisfies the requirement in the Paris Harrington theorem for k, c, and d. By way of example, if $N = R^*(k, 3, 2)$, then for every coloring of pairs over $\{0, \ldots, N-1\}$ with 3 colors (say, $red, green$, and $blue$) there exists a set $H = \{x_1, \ldots, x_\ell\} \subseteq \{0, \ldots, N-1\}$ (where $x_1 < x_2 < \ldots, < x_\ell$), such that $\ell > \max\{k, x_1\}$ and C restricted to H is constant (i.e., for some $c \in \{red, green, blue\}$ it holds that $f(x_i, x_j) = c$ for all $x_i, x_j \in H$).

We will mainly be interested in the case of pair colorings (i.e., $d = 2$) and we denote $R^*(k, c) = R^*(k, c, 2)$. Erdős and Mills [33] showed that $R^*(\cdot, \cdot)$ grows eventually faster than any primitive recursive function. On the other hand, for a fixed number of colors the resulting Ramsey function is primitive recursive. The result of [33] can also be phrased as an unprovability result by considering the fragment of Peano arithmetic $I\Sigma_1$ (induction restricted to Σ_1 arithmetical formulas). Recall that the provably total functions of $I\Sigma_1$ are exactly the primitive recursive functions. Thus, the proposition PH, restricted to colorings of pairs, is unprovable from $I\Sigma_1$.

By Paris Harrington [66], $R^*(k, c, d)$ is also not primitive recursive. Moreover, it grows much faster than standard examples of non-primitive recursive functions such as the Ackermann function, since it cannot be proven to be totally defined from the axioms of PA, while PA easily proves that the Ackermann function is well defined.

By omitting the relatively-largeness requirement from the PH proposition we get the standard Ramsey theorem, which is provable in PA. Similarly, by omitting the relatively-largeness requirement from the restriction of PH to pair colorings we get the standard Ramsey theorem for pairs, which is provable in $I\Sigma_1$. These observations call for some parameterization of the largeness requirement. Given a function $g : \mathbb{N} \to \mathbb{N}$, we say that H is g-large if $|H| \geq g(\min(H))$. Note that for our purpose it makes sense to consider g such that $g(n) < n$, since a relatively large H is exactly Id-large.

Now, replacing the relatively-largeness requirement in PH with g-largeness for different functions g, we ask: For which functions g do we obtain a proposition that is provable in PA? For a fixed tuple size d and a given g, is the resulting proposition provable in PA? is it provable in $I\Sigma_d$? is it provable in $I\Sigma_1$?

Some of these questions are considered in [81, 57] and sharp thresholds are presented for g in the considered cases. A survey of these results and more thresholds between provability and unprovability results as well as an illustration of the similarities in the techniques for achieving unprovability results is given by Bovykin [14]. Pudlák [70] notes some interesting connections of these unprovability results to the theory of computational complexity.

In Chapter 4 we consider the case of $d = 2$ and we define the appropriate g-large Ramsey number. We look for the threshold g between provability and unprovability somewhere between the two known extremes, i.e., one where g is constant and the other where $g = \mathrm{Id}$. When g is constant the resulting proposition is provable in $I\Sigma_1$ (i.e., the g-large Ramsey number is primitive recursive). If $g = \mathrm{Id}$ then the resulting proposition is unprovable in $I\Sigma_1$ (i.e., the g-large Ramsey number is not primitive recursive).

We stress that in Chapter 4 we only consider the combinatorial behavior of these Ramsey functions and show that a sharp threshold exists and lies above all functions $\log(n)/f^{-1}(n)$ obtained from an increasing primitive recursive f and below the function $\log(n)/\mathrm{Ack}^{-1}(n)$. Worded differently, for a nondecreasing and unbounded g to have primitive recursive g-large Ramsey numbers, it is necessary and sufficient that g is eventually dominated by $\log(n)/t$ for all $t > 0$ and that the rate at which g gets below $\log(n)/t$ is not too slow, namely, is primitive recursive in t: if g gets below $\log(n)/t$ only after an Ackermannianly long time $M(t)$, then the g-large Ramsey numbers are still Ackermannian.

Regressive Ramsey Gödel Sentences. A few years after the publication of the Paris Harrington work, A. Kanamori and K. McAloon [49], trying to avoid the notion of relative largeness and to give a simpler proof of independence of PA than the one presented in [66], introduced the regressive Ramsey theorem. This was yet another Ramsey type result, which can be stated in first order logic, and is independent of PA. Before presenting the main results of Kanamori and McAloon [49], we introduce the notion of *regressive* colorings. A coloring C of d-tuples over some set $X \subseteq \mathbb{N}$ with natural numbers (as colors) is regressive if for every d-tuple s of elements in X we have $C(s) \leq \min(s)$.

By way of example, let $d = 2$ and define a coloring of pairs of natural numbers $C(x, y) = x$ (assuming $x < y$). Note that C is regressive. On the other hand, there is no C-homogeneous set $H \subseteq \mathbb{N}$ such that $|H| > 2$.

Theorem (Kanamori and McAloon [49]): A set $H \subseteq X$ is min-homogeneous for a coloring C if for all d-tuples s, t over H, the equality $\min(s) = \min(t)$ implies $C(s) = C(t)$.

1. KM $\equiv$ For every $k > 0$ and $d > 0$ there exists an N such that for every regressive coloring of d-tuples from $\{0, \ldots, N-1\}$ there exists a min-homogeneous subset of size k.

2. KM cannot be proved from the axioms of PA (although it can be *phrased* in the language of PA).

3. Let $R^{\mathrm{reg}}(k)$ be the least N which satisfies KM for $d = 2$. The function $R^{\mathrm{reg}}(\cdot)$ eventually dominates every primitive recursive function.

4. Let $R'^{\mathrm{reg}}(k)$ be the least N which satisfies KM for $k = 2d$. The function R'^{reg} is not provably total in PA and eventually dominates every provably recursive function of PA.

Similarly to the case for PH, we can parameterize these statements. Here, g will be a parameter controlling the number of colors, requiring a coloring to be g-regressive instead of regressive. Informally, for a given function g we consider colorings C such that for every s it holds that $C(s) \leq g(\min(s))$. For constant g we obtain a weaker version of the standard Ramsey theorem (thus, provable in PA). Again, we will mainly be concerned with the case of $d = 2$ (where for constant g we obtain a weaker version of the standard Ramsey theorem for pairs, which is provable in $I\Sigma_1$).

Thus, questions very similar to the ones related to PH arise: for which g does the KM proposition for g-regressive colorings stop being provable from PA; for a given $d > 1$, for which g does the KM proposition for d-tuples and g-regressive colorings stop being provable from $I\Sigma_1$? (and from what fragments of PA is it still provable then?); are there sharp thresholds defining these transitions from provability to unprovability?

The case of a fixed d (the size of tuples we color) was considered by [57] and this result was later improved in [18]. In Chapter 3 we address the case of $d = 2$ and look for a threshold function for this transition from provability to unprovability, that is, from primitive recursiveness of the g-regressive Ramsey number (formally defined in Chapter 3) to its becoming Ackermannian. We consider the combinatorial behavior of these Ramsey numbers and show that a sharp threshold exists and lies above all functions $n^{1/f^{-1}(n)}$ obtained from a primitive recursive f and below $n^{1/\operatorname{Ack}^{-1}(n)}$. Worded differently, for a non-decreasing and unbounded g to have primitive recursive g-regressive Ramsey numbers it is necessary and sufficient that g is eventually dominated by $n^{1/t}$ for all $t > 0$ and that the rate at which g gets below $n^{1/t}$ is not too slow: if g gets below $n^{1/t}$ only after an Ackermannianly long time M_t, then the g-regressive Ramsey numbers are still Ackermannian.

We also identify the threshold below which g-regressive colorings have usual Ramsey numbers, that is, admit homogeneous, rather than just min-homogeneous sets, and give a lower bound of $A_{53}(2^{2^{274}})$ on the Id-regressive Ramsey number of $k = 82$, where A_{53} is the 53-rd approximation of Ackermann's function.

We remark that in both Chapter 3 and Chapter 4 we never explicitly mention the concept of provability and the discussion is purely combinatorial.

Phase transition in function hierarchies. As mentioned above, the phase transition between provability and unprovability can be defined in terms of growth rate of functions (which is the way we present our discussion in Chapter 3 and in Chapter 4). Another natural combinatorial question is to investigate the growth rate of functions obtained via diagonalization from an iteration hierarchy of Grzegorczyk type (see e.g., [52, 69]).

Assume that we are given two functions $g, h : \mathbb{R} \cap [0, \infty) \to \mathbb{R} \cap [0, \infty)$. For $r \in \mathbb{R}$, let $\lfloor r \rfloor$ denote the largest integer not exceeding r. We then consider the following diagonalization: For $x \in \mathbb{N}$, let

$$\begin{aligned} B(g,h)_0(x) &\triangleq g(x), \text{and let} \\ B(g,h)_{k+1}(x) &\triangleq B(g,h)_k^{\lfloor h(x) \rfloor}(x) \quad \text{i.e. } \lfloor h(x) \rfloor \text{ many iterations,} \\ B(g,h)_\omega(x) &\triangleq B(g,h)_{\lfloor x \rfloor}(x). \end{aligned}$$

For example, the Ackermann function is defined as $\mathrm{Ack}(n) = B(g,h)_\omega(n)$, where $g(x) = x + 1$ and $h = \mathrm{Id}$, and $\mathrm{A}_i(n) = B(g,h)_i(n)$. The question arises: for which pairs of functions g, h is the resulting diagonalization primitive recursive? When does it become Ackermannian?

In Chapter 5 we study for a given start function g iteration hierarchies with a sub-linear modulus h of iteration. In terms of g and h we classify the phase transition for the resulting diagonal function from being primitive recursive to being Ackermannian.

Specifically, fixing $g(x) = x + 1$ (as in the Ackermann hierarchy), we investigate the threshold function h at which the resulting hierarchy stops being primitive recursive and becomes Ackermannian. We show that a sharp threshold for h exists and that it is intrinsically related to g-regressive Ramsey numbers.

We then consider a class of start functions g (starting with $g(x) = x + \varepsilon$ for $0 < \varepsilon \leq 1$ and constantly increasing them). We show appropriate classes of iteration modulus h for which the resulting classes of hierarchies are slow-growing and later we show appropriate classes of iteration modulus h for which the resulting classes of hierarchies are fast-growing. These two classes are very similar, but yet result in extremely different growth rates.

1.2 Phase Transition Phenomena in Private Data Analysis

In private data analysis our goal is to compute some functionality $\hat{f}$ applied to a set of data gathered form individuals with an overall privacy requirement to protect the private information of individuals. The challenge is therefore imposed by the inherent tension between the need to compute some valuable estimation of $\hat{f}$ (utility) on the one hand, and the need to protect the information of individuals (privacy) on the other. A classic example is a hospital database containing medical records of patients. This example illustrates well the tension between utility and preserving privacy. The hospital, which would like to enable valuable medical research based on the information stored in the database, must ensure, for ethical as well as legal reasons, that the privacy of its patients is protected, i.e., no information regarding the medical condition of any specific patient can be "traced back" to that individual.

To model private computation we consider the setting of a *statistical database* containing n records $\mathbf{x} = (x_1, \ldots, x_n)$, where each x_i is the information of an individual, taken from some domain D. The interface between users and the database is defined by queries. Users can present a query q and get as reply an approximation of $q(\mathbf{x})$ (i.e., the result of the algorithm of the database applied to $q, \mathbf{x}$) that preserves individual privacy. A vast body of work on private data analysis has been published by researchers from different areas; however, we base our research on a line of rigorous investigation of what is safe to compute, which was initiated by the seminal paper of I. Dinur and K. Nissim [23].

In their work, Dinur and Nissim investigate the trade off between utility and privacy, and show a threshold of the amount of noise (perturbation) that must be added to queries in order to prevent strong violation of privacy. For the lowerbound, they consider an n-bit statistical database and show that any database algorithm for answering subset-sum queries that almost always answers with $o(\sqrt{n})$ perturbation, is strongly non-private (with respect to polynomial time adversaries). This impossibility result is not obtained under a specific definition of privacy, but rather by defining a database algorithm that enables an adversary (computationally bounded) to reconstruct all but a small fraction of the entries in the database.

In contrast to this impossibility result, Dinur, and Nissim show that once we allow added perturbation of magnitude $\widetilde{O}(\sqrt{n})$, it is possible to present a database that is private against polynomial adversaries in the strongest possible sense. That is, if the database is queried by a polynomial-time machine, then with extremely high probability it does not reveal any information about the data.

One way Dinur and Nissim suggest to understand their impossibility result is by viewing the set of query-response pairs as an encoding of the database, while the goal of the adversary is to efficiently decode this encoding even with the presence of some noise. Specifically, the lowerbound of [23] is obtained by considering the following error correction scheme. Given an n-bit string $\mathbf{x}$, the encoder computes $\mathbf{y} = A(\mathbf{x})$, where A is an $m \times n$ binary matrix (of which each entry is selected uniformly and independently). Then the encoded message is corrupted by adding some noise vector $\mathbf{e}$ of n entries, each with absolute value at most $\sqrt{n}$, and obtaining $\mathbf{y}' = \mathbf{x} + \mathbf{e}$. Finally, an adversary, knowing both A and $\mathbf{y}'$ is asked to approximate $\mathbf{x}$; that is, find $\mathbf{x}'$ which agrees with $\mathbf{x}$ on all but some ε fraction of the entries. The way the adversary finds such approximation is by solving a linear program.

To clarify the translation from the above scheme to the result of non-privacy in the statistical database model, we let each row of A be a subset-sum query and the appropriate entry in $\mathbf{e}$ be the perturbed response to that query. The adversary can sample the m rows of A and send them, one by one, to the database and obtain $\mathbf{e}$. Finally, by solving the linear program the adversary reconstructs most entries in the database. We note that for the result of [23], it suffices to take $m = n(\log n)^2$.

The work of Dinur and Nissim [23] initiated a line of works conducting rigorous investigation of the notion of private data analysis. The definition we work with – *differential privacy* – seems to entail some natural but essential properties, and proved to be robust. This definition has evolved in a sequence of works [23, 38, 30, 9, 27, 24, 25]. Informally, a computation is differentially private if any change in a single individual input may only induce a small change in the distribution of its outcomes (see Chapter 6 for the formal definition).

A few works have considered the limitations on what can be privately computed, much in the spirit of the lowerbound of [23]. The impossibility result of [23] has left open the question of whether a database can still preserve the privacy of individual records when adding much noise to a small fraction of the answers, while letting the remaining answers be fairly accurate. Dwork et al. [28] explored this direction by considering an error correcting scheme somewhat similar to that appearing in [23], with the difference that they allowed the entries of the matrix A to be normal random variables (rather than only allowing binary entries). They show that this attack, on any database mechanism answering at least 0.761 of the queries with an α-small additive error (perturbation), yields with very high probability a reconstruction of all but an $O(\alpha^2)$ fraction of the entries in the database. Taking $\alpha = o(\sqrt{n})$ we obtain similar parameters to those of [23] (only with many more unboundedly perturbed responses).

It is interesting to note that Dwork et al. [28] showed that their result is tight with respect to this attack. This stems from their more general result on the above error correcting scheme, where they show 0.761 to be a sharp threshold on the fraction of codeword coordinates that can sustain large corruption. Different attacks are considered by Dwork and Yekhanin in [31]. They use Fourier transforms for an attack that sharpens the result of [23] by reducing the number of required queries to n, the running time of the reconstruction algorithm, and by eliminating randomness (and hence the probability of failure). Dwork and Yekhanin [31] also present attacks that use polynomial interpolations to reconstruct a database that may answer up to $\frac{1}{2} - \varepsilon$ of the queries, adding an unbounded amount of noise (but, must answer all others using very little additive noise).

Once a definition of privacy has been formalized, the important and intriguing question of *what* set of analyses can be privately computed arises. The impossibility results described above were complemented by a long line of studies. Much work was done proving possibility results and constructing private data analysis mechanisms (see e.g. [30, 9, 27, 63, 5, 59, 10]) and some interesting and sometimes surprising relations were explored between private data analysis and other research areas, such as learning [9, 51] and mechanism design [59, 43]. The question of what can be privately computed in a distributed setup is also beginning to get some attention (see [25]). The next section briefly describes the distributed model.

1.2.1 Distributed Private Data Analysis

The distributed communication model has been considered in classic cryptography since the early 80s. For example, in *secure function evaluation* (SFE), we consider n parties $p_1, \ldots, p_n$, each holding a private input x_i, trying to distributively compute a function f applied to their inputs, with the requirement that an adversary may not gain any information other than the value of $f(x_1, \ldots, x_n)$. This requirement is insufficient for individual privacy if, for instance, f applied to private inputs $x_1, \ldots, x_n$ reveals "too much" information about one or more of its private inputs, then any implementation of f would do so too.

However, since any feasible functionality can also be securely computed in the distributed setup, it is natural to use the following paradigm for constructing distributed protocols for differentially private analyses: first choose an analysis (i.e., choose *what* to compute) while abstracting away implementation issues (e.g., by assuming that the computation is performed by a trusted server holding the data), then construct a secure protocol for this analysis (i.e., choose *how* to compute).

For example, this approach is the idea behind the SFE protocols presented in [25] for efficient generation of the noise. The protocols are secure in the presence of malicious, computationally bounded parties. Such noise generation can be used as part of a distributed computation of many private analyses.

On the other hand, the privacy requirement of SFE is sometimes too strict for our needs, e.g., consider the binary sum over n entries, where half of the entries are 1 and the other half are 0. An SFE protocol does not allow us to tell the difference between this database and its exact complement (obtained by flipping the bit of every entry). In differential privacy, we are allowed to tell the difference with very high probability.

Thus, the natural paradigm for constructing differentially private protocols may result in non-optimal protocols. In Chapter 7 we initiate an examination of whether indeed there are advantages in giving up the nice modularity property of the natural paradigm. Alternatively, we examine a paradigm wherein both a differentially private analysis and the distributed protocol (not necessarily SFE) for implementing it, are constructed simultaneously.

In particular, we consider the case of binary sum queries $f = \sum_{i=1}^{n} x_i$ (where each party holds a sensitive bit x_i) in the honest-but-curious setting, where coalitions of at most t players are allowed. Intuitively, in this setting, the parties follow the prescribed protocol, however, after the execution of the protocol terminates, coalitions of at most t players may work together and try to infer information about other players.

We show a phase transition behavior depending on the magnitude of error we allow in our analysis. Specifically, if we can settle for an approximation with additive error $\approx \sqrt{n}$, then the (non-SFE) *randomized response* protocol, suggested by Warner in 1965 [77], achieves exactly that, using a total of n messages. On the other hand, any secure protocol for a symmetric approximation using less than $nt/4$ messages yields an additive error that is linear in n.

In contrast, we show a lowerbound asserting that the randomized response protocol is optimal for low-communication protocols and thus, if we require an additive error of magnitude less than $\sqrt{n}$, we might as well follow the natural paradigm. Our lowerbound also yields a separation between the local and the global models, as well as a separation between the computational and the information theoretic settings.

1.3 Organization

This thesis consists of discussions both in the area of combinatorics and in the area of private data analysis. These two worlds coincide and share many conceptual notions. However, notations and definitions, as well as related works needed in each discussion, differ somewhat. The structure of the thesis as described next, is designed to allow the reader to easily recall relevant background.

In Chapter 2 we give necessary background in Ramsey theory, and the notations and definitions used in Chapter 3, Chapter 4, and Chapter 5. In Chapter 3 we study the phase transition behavior of g-regressive Ramsey numbers (these results are taken almost verbatim from [55]). In Chapter 4 we study the phase transition behavior of g-large Ramsey numbers (these results are taken almost verbatim from [55]). In Chapter 5 we study the phase transition behavior of function hierarchies (these results are taken almost verbatim from [64]). In Chapter 6 we give the notations and definitions used in Chapter 7. In Chapter 7 we study the phase transition behavior of distributed private data analysis protocols (these results are taken almost verbatim from [6]). Finally, in Chapter 8 we conclude our results and suggest some future work.

Chapter 2

Some Background in Ramsey Theory

This chapter contains the notations and basic definitions used in Chapter 3, Chapter 4, and Chapter 5, and it surveys some of the basic results in Ramsey theory offering a broader view for the discussion presented in these chapters.

The reader may want to skim through this chapter, since in most cases we follow the standard notations and definitions used in the literature (e.g., in [47]). Moreover, all propositions given in this chapter are known, and the proofs given for some of them are not essential for the understanding of our result. These proofs are presented here to give the reader some additional intuition for the propositions and the types of arguments used in proofs of Ramsey type propositions. Some definitions and notations that are more specific to this work are the ones concerning g-large and g-regressive Ramsey numbers and they are discussed in Section 2.3.3 and Section 2.3.5, respectively.

Let $\mathbb{N}$ denote the set of all natural numbers including 0 and let ω denote its cardinality. A number $d \in \mathbb{N}$ is identified with the set $\{n \in \mathbb{N} : n < d\}$, which may also be denoted by $[d]$. The set of all *d-element subsets* of a set X is denoted by $[X]^d$. For a function $C : [X]^d \to \mathbb{N}$ we write $C(x_1, \ldots, x_d)$ for $C(\{x_1, \ldots, x_d\})$ under the assumption that $x_1 < \cdots < x_d$. For an unbounded and nondecreasing function $g : \mathbb{N} \to \mathbb{N}$, define the inverse function $g^{-1} : \mathbb{N} \to \mathbb{N}$ by

$$g^{-1}(m) := \begin{cases} \ell & \text{if } \ell := \min\{i : g(i) \geq m\} > 0\,, \\ 1 & \text{otherwise (if } g(0) \geq m)\,. \end{cases}$$

2.1 Primitive Recursive Functions

Primitive recursive functions are functions from tuples of natural numbers to natural numbers. The set of primitive recursive functions is a proper subset of the recursive (or computable) functions.

Definition 2.1.1 (Primitive recursive functions). *A function that takes n arguments (an n-tuple) is called n-ary. The basic primitive recursive functions are given by the following three axioms:*

1. *The constant function 0 is primitive recursive.*
2. *The successor function S, which takes one argument and returns the succeeding number as given by the Peano postulates, is primitive recursive.*
3. *The projection functions $C_i^n(x_1, ..., x_n) = x_i$, which take n arguments and return one of them, are primitive recursive.*

More complex primitive recursive functions can be obtained by applying the operators given by the following pair of axioms:

1. Composition*: Given f, a k-ary primitive recursive function, and k l-ary primitive recursive functions $g_0, ..., g_{k-1}$, the composition of f with $g_0, ..., g_{k-1}$, i.e., the function $h(x_0, ..., x_{l-1}) = f(g_0(x_0, ..., x_{l-1}), ..., g_{k-1}(x_0, ..., x_{l-1}))$ is primitive recursive.*
2. Primitive recursion*: Given a k-ary primitive recursive function f and a $(k+2)$-ary primitive recursive function g, the $(k+1)$-ary function defined as the primitive recursion of f and g, i.e., the function h where $h(0, x_0, ..., x_{k-1}) = f(x_0, ..., x_{k-1})$ and $h(S(n), x_0, ..., x_{k-1}) = g(h(n, x_0, ..., x_{k-1}), n, x_0, ..., x_{k-1})$, is primitive recursive.*

A function is primitive recursive if it is one of the basic functions above, or can be obtained from the basic functions by applying the operations a finite number of times.

In other words, the class of primitive recursive functions is the smallest class of functions from $\mathbb{N}^d$ to $\mathbb{N}$ for all $d \geq 1$ that contains the constant functions, the projections, and the successor function and is closed under composition and recursion. This class is also closed under *bounded search*, frequently referred to as *bounded μ-operator*. See, e.g., [68, 17] for more details about the class of primitive recursive functions.

2.2 Ackermannian Functions

Given two functions $f, g : \mathbb{N} \to \mathbb{N}$, we say that g *eventually dominates* or *grows eventually faster than* f if there is some N so that for all $i \geq N$ it holds that $f(i) \leq g(i)$. In that case we also say that f is *eventually dominated* by g. We call f *nondecreasing* if for any $i < j$ we have $f(i) \leq f(j)$. A function

$h : \mathbb{N} \to \mathbb{N}$ is *unbounded* if for every $N \in \mathbb{N}$ there exists an i such that $h(i) > N$. For any function $f : \mathbb{N} \to \mathbb{N}$, the function $f^{(n)}$ is defined by $f^{(0)}(x) = x$ and $f^{(n+1)}(x) = f(f^{(n)}(x))$.

Definition 2.2.1 (The Ackermann function)**.** *The Ackermann function is defined as* $\mathrm{Ack}(n) = A_n(n)$ *for all* $n > 0$ *(and, say,* $\mathrm{Ack}(0) = 0$*) where each* A_n *is the standard* n-th approximation *of the Ackermann function, defined by:*

$$A_1(n) = n + 1$$
$$A_{i+1}(n) = A_i^{(n)}(n).$$

Let us record that $\mathrm{Ack}(1) = 2$, $\mathrm{Ack}(2) = 4$, $\mathrm{Ack}(3) = 24$, and $2^{2^{2^{70}}} < \mathrm{Ack}(4) < 2^{2^{2^{71}}}$. We remark that although Ack is not primitive recursive, its inverse Ack^{-1} is primitive recursive.

It is well known (see e.g., [17]) that each approximation A_n is *primitive recursive* and that every primitive recursive function is eventually dominated by some A_n. Thus, the Ackermann function eventually dominates every primitive recursive function.

Definition 2.2.2 (Ackermannian functions)**.** *A function* $g : \mathbb{N} \to \mathbb{N}$ *is said to be* Ackermannian *if it grows eventually faster than every primitive recursive function.*

There is no *smallest* Ackermannian function: if f is Ackermannian, then so is $i \mapsto f(i)/2$ or $i \mapsto f(i)^{1/2}$, etc. It is also important to note that there are functions $f : \mathbb{N} \to \mathbb{N}$ which are neither Ackermannian nor eventually dominated by any primitive recursive function.

Lemma 2.2.3. *If the composition* $f \circ g$ *of two nondecreasing functions is Ackermannian and one of* f *and* g *is primitive recursive, then the other is Ackermannian.*

Proof. If f is primitive recursive, then g should be Ackermannian. Assume now g is primitive recursive. Note that g is not bounded. And, given a primitive recursive function p, the function $h(n) := p(g(n+1))$ is primitive recursive too, so there is some N such that $f(g(n)) \geq h(n) = p(g(n+1))$ for all $n \geq N$. Since we can assume w.l.o.g. that p is nondecreasing, it holds for all $i \geq g(N)$ that $f(i) \geq f(g(n)) \geq p(g(n+1)) \geq p(i)$, where $g(n) \leq i \leq g(n+1)$ for some $n \geq N$. Hence f is Ackermannian. □

2.3 Ramsey Theory

Much of the discussion of Chapters 3, 4, and 5 deals with the behavior of Ramsey type propositions, i.e., propositions regarding the cardinality of various partition relations. In this section we describe some of the known results in Ramsey theory and their variants, obtained by parameterizing the partition relation with respect to functions $g : \mathbb{N} \to \mathbb{N}$.

Let us recall the standard Ramsey theorem.

Definition 2.3.1. *For a k-ary function f over a set A and for a given $B \subseteq A$ we use $f \restriction [B]^k$ to denote the restriction of f to $[B]^k$, i.e., $\{(x, f(x)) : x \in [B]^k\}$.*

Given a coloring $C : [X]^d \to \mathbb{N}$, a set $H \subseteq X$ is homogeneous *for C (or C-homogeneous) if it holds $C \restriction [H]^d$ is constant. The symbol*

$$X \to (k)^d_c$$

means: for every coloring $C : [X]^d \to c$ there exists $H \subseteq X$ such that $|H| \geq k$ and H is homogeneous for C. In case $d = 2$, we just write

$$X \to (k)_c.$$

Ramsey [71] established the Infinite Ramsey theorem:

$$\forall d, c > 0 \qquad \mathbb{N} \to (\mathbb{N})^d_c$$

as well as the finite Ramsey theorem:

$$\forall d, c, k > 0 \quad \exists m \in \mathbb{N} \qquad m \to (k)^d_c.$$

Below we present a proof for the finite version of the Ramsey theorem for pairs (i.e., the case for $d = 2$). The case for general d can be thereafter obtained by induction on d. The proof of the infinite Ramsey theorem follows the same technique and is in fact simpler than the proof below. Before going on to describe the proof, let us present a generalization of the notion of homogeneity.

Definition 2.3.2. *For a coloring $C : [X]^d \to \mathbb{N}$, a set $H \subseteq X$ is* min-homogeneous *for C, if $C(x, x_2, \ldots, x_d) = C(x, y_2, \ldots, y_d)$ for all $x, x_2, \ldots, x_d, y_2, \ldots, y_d \in H$ (i.e., for every $x \in [X]^d$ the color $C(x)$ is completely determined by $\min x$).*

The symbol

$$X \overset{\min}{\to} (k)^d_c$$

means: for every coloring $C : [X]^d \to c$ there exists $H \subseteq X$ such that $|H| \geq k$ and H is min-homogeneous for C. In case $d = 2$, we just write

$$X \overset{\min}{\to} (k)_c.$$

The *standard Ramsey number* denoted $R(k, c)$ is the least N such that $N \to (k)_c$. Let $R^{\min}(k, c)$ denote the least N so that $N \overset{\min}{\to} (k)_c$. Note that $R^{\min}(k, 1) = k$ and $R^{\min}(2, c) = 2$. We now recall the standard proof of the finite Ramsey theorem and show that it yields for $c, k \geq 2$:

$$(1) \quad c^k \overset{\min}{\to} (k)_c \quad \text{and} \quad (2) \quad c^{k \cdot c} \to (k)_c$$

That is, $R(k, c) \leq c^{k \cdot c}$ and $R^{\min}(k, c) \leq c^k$ for any $c, k \geq 2$.

Proof. (1) Given $c, k > 1$ and a coloring $C : [c^k]^2 \to c$, we construct a set $B \in [c^k]^k$ by forming two sequences, $\{b_i\}_{i=0}^k$ and $\{B_i\}_{i=0}^k$, recursively. We set $b_0 = 0$ and $B_0 = \{0, \ldots, c^k - 1\}$. Now, assume we have b_i and B_i for $i \geq 0$ such that $|B_i| \geq c^{k-i}$, we denote $S_j = \{a \in B_i : a \neq b_i \text{ and } C(b_i, a) = j\}$. By the pigeonhole principle and because $k > 1$, there exists some $0 \leq j \leq c$ so that $|S_j| \geq \frac{|B_i|}{c} \geq c^{k-i-1}$. We set B_{i+1} to be a largest S_j and we set b_{i+1} to be $\min(B_{i+1})$. Obviously, we can repeat the induction step at least k times. Thus, we end up having a sequence $B = \{b_0, b_1, ...b_k\}$, which is a min-homogeneous set for f of size $k + 1$.

To verify (2), we only need to notice that by the above proof of (1), given $c, k > 1$ and a coloring $C : [c^{kc}]^2 \to c$, we are guaranteed to have a set $B = \{b_1, b_2, \ldots, b_{kc}\}$, which is min-homogeneous for C. Let $\{c_1, c_2, \ldots, c_{kc}\}$ be the set of colors such that $c_i = C(b_i, b_j)$ for all $j > i$. Again, by the pigeonhole principle, there exists a set $I \subseteq \{1, 2, \ldots, kc\}$ of size k, such that for any $i, j \in I$, $c_i = c_j$. Hence, the set $A = \{b_i \in B : i \in I\}$ is homogeneous for C and is of size k. □

It it is also known for $R(k) = R(k, 2)$, that $2^{k/2} < R(k) < 2^{2k}$. The left inequality is shown by a simple probabilistic argument (see [3]), the right inequality stems from the proof above.

2.3.1 The Canonical Ramsey Theorem

In 1950, Erdős and Rado [34] proved a generalization of Ramsey's theorem with no restrictions on the number of colors. For a coloring $C : [X]^d \to \mathbb{N}$ where $X \subseteq \mathbb{N}$, we say that $H \subseteq X$ is *canonical* for C if there is a set of indices, $I \subseteq d$, so that for all $s, t \in [H]^d$ we have $C(s) = C(t) \leftrightarrow s \restriction I = t \restriction I$, where if $p = \{p_1, \ldots, p_d\}$, then $p \restriction I = \{(i, p_i) : i \in I\}$. We write $I = I(H)$ when I makes H canonical. By way of example, if $I(H) = \emptyset$, then H is homogeneous for C in the usual sense, and if $I(H) = d$ then C is injective on $[H]^d$.

Let us give a few examples of canonical sets:

- Let C be a coloring with domain $[\mathbb{N}]^3$, defined by $C(x_0, x_1, x_2) = x_0 + x_1 + x_2$. One can observe that if $H \subseteq \mathbb{N}$ is infinite and canonical for C, then $I(H) = 0, 1, 2$. One such set would be $\{2^k : k \in \mathbb{N}\}$. Now, for any $x_0 < x_1 < x_2, y_0 < y_1 < y_2 \in H$ so that there exists an index $i \in \{0, 1, 2\}$ such that $x_i \neq y_i$, let i be that maximum such index and assume, without loss of generality, that $x_i < y_i$. It holds that $x_0 + \cdots + x_i < y_i$ and thus, $C(x_0, x_1, x_2) < C(y_0, y_1, y_2)$.

- Let C be a function with domain $[\mathbb{N}]^3$, defined by $C(x_0, x_1, x_2) = x_0 + x_1 + x_2 \mod 10$. Here we may choose $H \subseteq \mathbb{N}$ to be $\{10^k : k \in \mathbb{N}\}$ and have $I(H) = \emptyset$, as for any $x_0 < x_1 < x_2 \in H$ it holds that $C(x_0, x_1, x_2) = 0$.

Using the infinite Ramsey theorem, Erdős and Rado proved:

Theorem 2.3.3 ([34]). *For all natural $d > 0$ and for every coloring $C : [\mathbb{N}]^d \to \mathbb{N}$, there exists an infinite $H \subseteq \mathbb{N}$ which is canonical for C.*

Theorem 2.3.3 is proved by induction on d and the infinite Ramsey theorem, which, in turn, might also be seen as a corollary of Theorem 2.3.3, since the '$\leftrightarrow$' requirement in the definition of canonical asserts that if the number of colors is finite, then $I = \emptyset$ is the only possible case.

Using compactness one can establish the finite version of the canonical Ramsey theorem.

Theorem 2.3.4. *For all natural $k, d > 0$ there exists $N \in \mathbb{N}$ such that for every coloring $C : [N]^d \to \mathbb{N}$ there exists $H \in [N]^k$ which is canonical for C.*

Before proving Theorem 2.3.4 we first state the canonical Ramsey theorem for pairs (i.e., the case where $d = 2$), and present a proof for this case to illustrate the basic idea used for the proof of the infinite case (i.e., Theorem 2.3.3).

Theorem 2.3.5. *For every coloring $C : [\mathbb{N}]^2 \to \mathbb{N}$ there exists an infinite $H \subseteq \mathbb{N}$ which is canonical for C.*

Proof. Let $\{\alpha_0, \alpha_1, \ldots, \alpha_5\}$ be the set of all unordered pairs over $\{0, 1, 2, 3\}$ and let $\{\beta_0, \beta_1, \ldots, \beta_{14}\}$ be the set of all unordered pairs over $\{\alpha_0, \alpha_1, \ldots, \alpha_5\}$. Now define $g : [\mathbb{N}]^4 \to 2^{15}$ as

$$g(x_0, x_1, x_2, x_3) = (y_{\beta_0}, y_{\beta_1}, \ldots, y_{\beta_{14}}),$$

where for $\beta_i = ((i_1, i_2), (j_1, j_2))$, we denote

$$y_{\beta_i} = \begin{cases} 1 & \text{if } C(x_{i_1}, x_{i_2}) = C(x_{j_1}, x_{j_2}); \\ 0 & \text{otherwise.} \end{cases}$$

In other words, g describes an equivalence relation induced by C on $[\mathbb{N}]^4$, which obviously has a finite number of classes. Thus, by the infinite Ramsey theorem, there exists an infinite $A \subseteq \mathbb{N}$ that is homogeneous for g. We now show that $f \restriction A$ is canonical.

We first examine two main cases:

1. There exist $a_1 < a_2 < a_3$ in A such that $C(a_1, a_2) = C(a_1, a_3)$. If this is the case, then by the definition of g, for all $x_1 < x_2 < x_3$ in A it holds that $C(x_1, x_2) = C(x_1, x_3)$. This is true because $g(a_1, a_2, a_3, z_1) = g(x_1, x_2, x_3, z_2)$, for every $z_1 > a_3$, $z_2 > x_3$.

2. There exist $a_1 < a_2 < a_3$ in A such that $C(a_1, a_3) = C(a_2, a_3)$. If this is the case, then by the definition of g, for all $x_1 < x_2 < x_3$ in A it holds that $C(x_1, x_3) = C(x_2, x_3)$.

Now, let us consider the possible sub-cases for A:

- If both Case 1 and Case 2 hold, then C is constant on A, that is, $C \restriction A$ is canonical with $I(A) = \emptyset$. This is true, as for any $x_1 < x_2 < x_3$ in A it holds that $C(x_1, x_3) = C(x_2, x_3) = C(x_1, x_2)$. Thus, for any $x_1 < x_2, x_1 \leq x_3 < x_4$ in A, if, either $x_1 = x_3$ or $x_2 = x_4$, then by Case 1 and Case 2 it holds that $C(x_1, x_2) = C(x_3, x_4)$. Otherwise $C(x_1, x_4) = C(x_2, x_4) = C(x_1, x_2)$ (regardless of the ordering of x_2 and x_4), and $C(x_1, x_4) = C(x_3, x_4) = C(x_1, x_3)$ and so, again we get that $C(x_1, x_2) = C(x_3, x_4)$.

- If Case 1 holds and Case 2 fails, then $C \restriction A$ is canonical with $I(A) = \{0\}$. To show this we only need verify that there do not exist $x_1 < x_2$, $x_3 < x_4$ in A such that $x_1 < x_3$ but $C(x_1, x_2) = C(x_3, x_4)$. Assume to the contrary that such x_1, x_2, x_3, x_4 do exist. Because g is constant on all 4-tuples, we can assume x_1 is not the smallest element in A. Let $x_0 < X_1$ be yet another element of A. Because $g(x_0, x_2, x_3, x_4) = g(x_1, x_2, x_3, x_4)$ (if $x_2 = x_3$ or $x_2 = x_4$ then look at $g(x_0, x_2, x_4, z) = g(x_1, x_2, x_4, z)$ for some $z \in A$ such that $z > x_4$) we may conclude that $C(x_0, x_2) = C(x_3, x_4) = C(x_1, x_2)$ contrary to failure of Case 2.

- If Case 1 does not hold, but Case 2 does, then $C \restriction A$ is canonical with $I(A) = \{1\}$. This is shown in a very similar manner to the former case.

- If neither Case 1 nor Case 2 hold, then $C \restriction A$ is canonical with $I(A) = \{0, 1\}$. Let us show that by assuming to the contrary that there exist $x_1, x_2, x_3, x_4 \in A$ such that $x_1 < x_3$ and $x_2 \neq x_4$ but $C(x_1, x_2) = C(x_3, x_4)$. Again, because g is constant on all 4-tuples, we may assume that x_1 is not the smallest element in A. Let $x_0 < x_1$ be an element of A. If $x_2 \neq x_3$, we get $g(x_0, x_2, x_3, x_4) = g(x_1, x_2, x_3, x_4)$, otherwise choose $z \in A$ so that $z > x_4$ and now we have $g(x_0, x_2, x_4, z) = g(x_1, x_2, x_4, z)$, and in either case, we may conclude that $C(x_0, x_2) = C(x_3, x_4) = C(x_1, x_2)$ contrary to failure of Case 2.

□

We now present a proof of the finite canonical Ramsey theorem (Theorem 2.3.4). We establish it from the infinite canonical Ramsey theorem by means of compactness. In order to do that, we first present the notion of equivalent colorings. Two colorings C_1, C_2 of $[X]^d$ are equivalent if for all $s_1, s_2 \in [X]^d$, it holds that $C_1(s_1) = C_1(s_2) \iff C_2(s_1) = C_2(s_2)$, that is, they induce the same partition of $[X]^d$. Obviously, if C_1, C_2 are equivalent and there exists no set $H \in [X]^k$ canonical for C_1, then there exists no set $H \in [X]^k$ canonical for C_2.

Proof of Theorem 2.3.4. Assume to the contrary that the finite canonical Ramsey theorem does not hold. Then there exist $k, d > 0$ such that for any $m \in \mathbb{N}$ there is a counterexample $C_m : [m]^d \to \mathbb{N}$ which is "bad" in the sense that there is no $H \in [m]^k$ canonical for C_m. Observe that for any such bad C_m there exists an equivalent c'_m such that for every $s \in [m]^d$, it holds that $c'_m(s) \leq \binom{\max(s)}{d}$. This is true since for any $l \in \mathbb{N}$ there may be at most $\binom{l}{d}$ colors used to color all elements of $[l]^d$. Let T be the set of all bad colorings

$$\begin{aligned} T = \{C : [m]^d \to \mathbb{N} \quad &: m \in \mathbb{N} \text{ and } C \text{ is such that } \forall s \in [m]^d,\, C(s) \leq \tbinom{\max(s)}{d} \\ &\text{and there is no } H \in [m]^k \text{ canonical for } C\}. \end{aligned}$$

Define a binary relation $\prec$ on colorings, by $C_1 \prec C_2$ iff $C_1 \subseteq C_2$ and there is no C_3 such that $C_1 \subseteq C_3 \subseteq C_2$. We claim that $\langle T, \prec \rangle$ infinite tree with a finite number of nodes at any level i. First, it is clear that T is infinite, as there exist infinitely many counterexamples. Now, observe that T is connected, since the empty coloring is a counterexample that is contained by any coloring. Furthermore, if $N_1 \leq N_2 \leq N$ and $C_1 : [N_1]^d \to \mathbb{N}$, $C_2 : [m_2]^d \to \mathbb{N}$ and $C : [N]^d \to \mathbb{N}$ such that $C_1 \subseteq C$ and $C_2 \subseteq C$, then necessarily $C_1 \subseteq C_2$ and therefore it holds that $\langle T, \prec \rangle$ is a tree.

Now, every bad coloring $C_N \in T$ with domain $[N]^d$ for $N > d$, when restricted to $[N']^d$ for $N' < N$, is a counterexample as well and of course it holds that $\forall s \in [N']^d$, it holds that $C_N(s) \leq \binom{\max(s)}{d}$ and therefore the restricted coloring is also in T. Therefore it holds that at any level i of $\langle T, \subseteq \rangle$, there are only colorings with domain $[i+d]^d$. Now, for every $N \geq d$, there is a finite number of colorings $f : [N]^d \to \mathbb{N}$ such that for every $s \in [N]^d$, it holds that $f(s) \leq \binom{\max(s)}{d}$ in general and thus a finite number of bad colorings $f' : [N]^d \to \mathbb{N}$ such that for every $s \in [N]^d$, it holds that $f'(s) \leq \binom{\max(s)}{d}$. Hence, there is a finite number of elements at any level of $\langle T, \subseteq \rangle$.

By König's Lemma [53], there is an infinite path $P = \{C_i\}_{i=0}^{\infty}$ in $\langle T, \prec \rangle$. Define a coloring $C : [\mathbb{N}]^d \to \mathbb{N}$ by, $C = \bigcup_{i \in \mathbb{N}} C_i$. We now observe that C is a legal coloring of d-tuples over $\mathbb{N}$ because of the relation $\subseteq$ and that there exists no $H \subseteq \mathbb{N}$ of size k that is canonical for C. Otherwise, if there were such H, then $C \restriction [\max(H)]^d$, which is in P, would not be a bad coloring and therefore would not be in T. Obviously, there does not exist infinite $H \subseteq \mathbb{N}$ which is canonical for C, contrary to the infinite canonical Ramsey theorem. □

Let $\mathrm{er}(k)$ be the least natural number m so that for every coloring $C : [m]^2 \to \mathbb{N}$, there exists $x \in [m]^k$ such that $f \restriction X$ is canonical. Leffman and Rödl [58] proved that there exist $c_1, c_2 > 0$ such that for every positive integer k, $2^{c_1 k^2} \leq \mathrm{er}(k) \leq 2^{2^{c_2^{k^3}}}$.

2.3.2 Paris-Harrington Ramsey Theorem

The question raised in 1931, by Gödel's incompleteness theorem [44], regarding the possibility of discovering finite combinatorial theorems that are independent of powerful axiomatic systems such as first-order PA (Peano Arithmetic), was settled in the late 1970s by J. Paris [65]. Using model-theoretic techniques to investigate arithmetic incompleteness, he proved theorems of finite combinatorics to be unprovable in PA, basing his results on a joint work with L. Kirby [54]. Later J. Paris and L. Harrington went on to present a proof that a straightforward variant of the finite Ramsey theorem is independent of Peano Arithmetic.

2.3.3 g-large Ramsey Numbers

Paris and Harrington [66] introduced the notion of a relatively large set of natural numbers. Here, we present the main theorem of their work by considering a parameterized version of this notion. We use a requirement on the size of the homogeneous set, relative to some (parameter) function g.

Definition 2.3.6. *A nonempty $H \subseteq \mathbb{N}$ is g-*large *for a function $g : \mathbb{N} \to \mathbb{N}$ if $|H| \geq g(\min H)$. The symbol*

$$X \to_g^* (k)_c^d$$

means: for every coloring $C : [X]^d \to c$ there is a g-large C-homogeneous $H \subseteq X$ such that $|H| \geq k$. That is, the restriction of C to $[H]^d$ is a constant function. In case $d = 2$, we just write

$$X \to_g^* (k)_c.$$

Fact 2.3.7. *Suppose $g : \mathbb{N} \to \mathbb{N}$ is any function. Then for every k, c, and d there is some N such that $N \to_g^* (k)_c^d$.*

The proof follows from the infinite Ramsey theorem and compactness. See Paris and Harrington [66] for more details. The principle of compactness is illustrated in the proof of Fact 2.3.9 at the end of the chapter.

Recall that the notion of a relatively large set introduced by Paris and Harrington is exactly g-large for $g = \text{Id}$. Paris and Harrington [66] proved that the statement:

$$\text{PH} \equiv (\forall d \geq 1, c > 0, k > 0)(\exists N)\ N \to_{\text{Id}}^* (k)_c^d$$

is a Gödel sentence over Peano Arithmetic, in the sense of [44]. A different proof of Paris and Harrington's theorem was given by Ketonen and Solovay [52].

The *g-large Ramsey number* of k and c, denoted $R_g^*(k, c)$, is the least N so that $N \to_g^* (k)_c$. Erdős and Mills showed in their seminal paper [33] that R_{Id}^* is not primitive recursive. For a fixed number of colors the resulting Ramsey function is primitive recursive. Erdős and Mills further showed that the Ramsey function becomes double exponential if the number of colors is restricted to two.

When these Ramsey functions are considered as a hierarchy indexed by the number of colors then it is cofinal in the Grzegorczyk hierarchy of primitive recursive functions.

2.3.4 Regressive Ramsey Theorem

In 1985, A. Kanamori and K. McAloon [49] introduced another Ramsey type proposition and showed it to be independent of PA. Trying to avoid the Paris-Harrington notion of relatively large finite sets and to obtain a simpler proof for the independence of PA, they introduced a new partition relation.

2.3.5 g-regressive Ramsey Numbers

Kanamori and McAloon [49] suggested the notion of a regressive coloring. We present their main result by considering a parameterized version of this notion, already introduced by Kanamori and McAloon at the end of [49].

Definition 2.3.8. *Given a set $X \subseteq \mathbb{N}$, a coloring $C : [X]^d \to \mathbb{N}$ is g-*regressive *for a function $g : \mathbb{N} \to \mathbb{N}$ if $C(x_1, \ldots, x_d) \leq g(x_1)$ for all $\{x_1, \ldots, x_d\} \subseteq X$. The symbol*

$$X \overset{\min}{\to} (k)^d_g$$

means: for every g-regressive coloring $C : [X]^d \to \mathbb{N}$ there exists $H \subseteq X$ such that $|H| \geq k$ and H is min-homogeneous for C, that is, $C(x, x_2, \ldots, x_d) = C(x, y_2, \ldots, y_d)$ for all $x, x_2, \ldots, x_d, y_2, \ldots, y_d \in H$. In case $d = 2$, we just write

$$X \overset{\min}{\to} (k)_g.$$

Fact 2.3.9. *Let $g : \mathbb{N} \to \mathbb{N}$ be arbitrary. Then*

1. *for every g-regressive coloring $C : [\mathbb{N}]^d \to \mathbb{N}$ there is an infinite $H \subseteq \mathbb{N}$ such that H is min-homogeneous for C;*

2. *for any d and k there is some N so that for every g-regressive coloring $C : [N]^d \to \mathbb{N}$ there is a min-homogeneous $H \subseteq N$ of size at least k.*

The first item follows from the infinite canonical Ramsey theorem, since the only two (out of 2^d) canonical colorings of d-tuples to which a g-regressive coloring may be equivalent on an infinite set are the minimum coloring and the constant coloring — both of which make the set min-homogeneous. The second item follows from the first via compactness. At the end of this section we give more detailed proofs for both items.

Recall that the notion of regressiveness is exactly g-regressiveness for $g =$ Id. Kanamori and McAloon [49] proved that the following statement,

$$\mathrm{KM} \equiv (\forall d \geq 1, k > 0)(\exists N)\ N \overset{\min}{\to} (k)^d_{\mathrm{Id}}$$

is a Gödel sentence in Peano Arithmetic.

The *g-regressive Ramsey number* of k, denoted $R^{\mathrm{reg}}_g(k)$, is the least N so that $N \overset{\min}{\to} (k)_g$. Kanamori and McAloon [49] also proved, using model-theoretic techniques, that $R^{\mathrm{reg}}_{\mathrm{Id}}$ is not primitive recursive. Purely combinatorial proofs of this can be found in [69] and in [56].

Note that there may be functions g, such that for any g-regressive coloring $C : [X]^2 \to \mathbb{N}$ there exists C-homogeneous $H \subseteq X$ of cardinality k (e.g., for constant functions $g = c$, the resulting requirement is a weaker one than the requirement of the standard Ramsey theorem). Therefore, it makes sense to add the following this notation. We let the symbol

$$X \to (k)_g$$

mean: for every g-regressive coloring $C : [X]^2 \to \mathbb{N}$ there exists $H \subseteq X$ such that $|H| \geq k$ and H is C-homogeneous.

Proof of Fact 2.3.9. We now present detailed proofs for the generalizations (via parameterization) of both the infinite and finite regressive Ramsey theorems. We find these proofs of some interest, since the proof of the infinite case makes use of the infinite canonical Ramsey theorem, while the proof of the finite case uses the compactness principle, which is a tool we do not possess in PA.

Proof of Fact 2.3.9– 1. The case where $d = 1$ is trivial as $\mathbb{N}$ is min-homogeneous for any coloring $C : \mathbb{N} \to \mathbb{N}$. Assume $d > 1$. Let $C : [\mathbb{N}]^d \to \mathbb{N}$ be regressive and let H be an infinite canonical set for C with $I = I(H)$. We claim that either $I = \emptyset$ or $I = \{0\}$. Suppose to the contrary that I contains $i \neq 0$. Let h be the minimal element in H. There are arbitrarily many d-tuples from H with h as first element and that disagree on i and hence are mapped by C to different values. Thus, there must be a d-tuple containing h that is mapped to $m > g(h)$, contrary to g-regressiveness of f. Now, if $I = \emptyset$, then H is homogeneous for C and if $I = \{0\}$, then H is min-homogeneous for C. □

It is worth noting that the finitistic version of the canonical Ramsey theorem does not induce the finite regressive theorem in the same way that the infinite version of the canonical Ramsey theorem gives the infinite version of the regressive Ramsey theorem in the proof above. A proof, following the track of the infinite version, would fail trying to assert that $I(H)$ for the finite canonical set H necessarily does not contain $i \neq 0$ as $g(\min(H))$ can be sufficiently large to allow all d-tuples containing $\min(H)$ be assigned a different color. Indeed, there are different bounds for canonical and regressive Ramsey numbers. Already in the case of pair colorings, canonical Ramsey numbers are exponential while regressive Ramsey numbers are Ackermannian. We will prove the finite version by means of compactness.

Proof of Fact 2.3.9– 2. Assume to the contrary that there exist $g : \mathbb{N} \to \mathbb{N}$ and $k, d \in \mathbb{N}$ such that for any $N \in \mathbb{N}$ there is a counterexample $C_N : [N]^d \to \mathbb{N}$ which is "bad" in the sense that C_N is g-regressive, but there exists no $H \in [N]^k$ min-homogeneous for C_N. Let T be the set of all bad colorings, that is, T is the set of all colorings $C : [N]^d \to \mathbb{N}$ for some $N \in \mathbb{N}$ such that C is g-regressive and there is no $H \in [N]^k$ that is min-homogeneous for C.

Define a binary relation $\prec$ on colorings, by $C_1 \prec C_2$ iff $C_1 \subsetneq C_2$ and there is no C_3 such that $C_1 \subsetneq C_3 \subsetneq C_2$. We claim that $\langle T, \prec\rangle$ is an infinite tree with a finite number of nodes at any level i. First, it is clear that T is infinite, as there exist infinitely many counterexamples. Now, observe that T is connected, since the empty coloring is a counterexample that is contained by any coloring. Furthermore, if $N_1 \leq N_2 \leq N$ and $C_1 : [N_1]^d \to \mathbb{N}$, $C_2 : [N_2]^d \to \mathbb{N}$ and $C : [N]^d \to \mathbb{N}$ such that $C_1 \subseteq C$ and $C_2 \subseteq C$, then necessarily $C_1 \subseteq C_2$ and therefore it holds that $\langle T, \prec\rangle$ is a tree.

Now, every bad coloring C_N with domain $[N]^d$ for $N > d$, when restricted to $[N']^d$ for $N' < N$, is a counterexample as well and therefore is also in T; therefore, it holds that at any level i of $\langle T, \prec\rangle$, there are only colorings with domain $[i + d]^d$. For every $N \geq d$, there is a finite number of g-regressive colorings of $[N]^d$ and thus a finite number of bad colorings of $[N]^d$. Hence, there is a finite number of elements at any level of $\langle T, \prec\rangle$.

By König's Lemma [53], there is an infinite path $P = \{C_i\}_{i=0}^{\infty}$ in $\langle T, \prec\rangle$. Define a coloring $C : [\mathbb{N}]^d \to \mathbb{N}$ by, $C = \bigcup_{i\in\mathbb{N}} C_i$. Observe that C is indeed a g-regressive coloring of d-tuples over $\mathbb{N}$ because of the relation $\subseteq$ and observe that there exists no $H \subseteq \mathbb{N}$ of size k that is min-homogeneous for C. Otherwise, if there were such H, then $C \restriction [\max(H)]^d$, which is in P, would not be a bad coloring and therefore would not be in T in the first place. Obviously, there exists no infinite $H \subseteq \mathbb{N}$ which is min-homogeneous for C, contrary to Fact 2.3.9–1. □

Chapter 3

Phase Transition Threshold of g-regressive Ramsey Numbers

In this chapter we show that the threshold for Ackermannian g-regressive Ramsey numbers lies above all functions $n^{1/f^{-1}(n)}$ obtained from a primitive recursive f and below $n^{1/\operatorname{Ack}^{-1}(n)}$.

Worded differently, for a nondecreasing and unbounded g to have primitive recursive g-regressive Ramsey numbers it is necessary and sufficient that g is eventually dominated by $n^{1/t}$ for all $t > 0$ and that the rate at which g gets below $n^{1/t}$ is not too slow: if g gets below $n^{1/t}$ only after an Ackermannianly long time M_t, then the g-regressive Ramsey numbers are still Ackermannian.

We also identify the threshold below which g-regressive colorings have usual Ramsey numbers, that is, admit homogeneous, rather than just min-homogeneous sets, and give a lower bound of $A_{53}(2^{2^{274}})$ on the Id-regressive Ramsey number of $k = 82$, where A_{53} is the 53-rd approximation of Ackermann's function.

We begin by recalling the following notations and definitions from Chapter 2. Given a set $X \subseteq \mathbb{N}$, a coloring $C : [X]^d \to \mathbb{N}$ is g-regressive for a function $g : \mathbb{N} \to \mathbb{N}$ if $C(x_1, \ldots, x_d) \leq g(x_1)$ for all $\{x_1, \ldots, x_d\} \subseteq X$. The symbol

$$X \overset{\min}{\to} (k)_g$$

means: for every g-regressive coloring $C : [X]^2 \to \mathbb{N}$ there exists $H \subseteq X$ such that $|H| \geq k$ and H is min-homogeneous for C. The g-regressive Ramsey number of k, denoted $R_g^{\text{reg}}(k)$, is the least N so that $N \overset{\min}{\to} (k)_g$.

Kanamori and McAloon introduced the notion of a g-regressive coloring in [49] and proved that the g-regressive Ramsey number for $g = \text{Id}$ is Ackermannian. On the other hand, it follows from the proof of the standard Ramsey theorem that R_g^{reg} is primitive recursive for every constant function g.

We next compute the sharp thresholds on g at which g-regressive Ramsey numbers cease to be primitive recursive and become Ackermannian. The results in this chapter are taken almost verbatim from [55].

3.1 g-regressive Lower Threshold

We begin with the following lemma which stems from Lemma 26.4 in [32].

Lemma 3.1.1. *$R^{\min}(k,c) \leq 2 \cdot c^{k-2}$ for any c, $k \geq 2$.*

Note that Lemma 26.4 in [32] talks about *end-homogeneous* sets. However, if we confine ourselves to the 2-dimensional case then it is just about min-homogeneous sets. Concerning n-dimensional min-homogeneous sets see [57].

Theorem 3.1.2. *Given $B : \mathbb{N} \to \mathbb{N}^+$ let $g_B(i) = \left\lfloor\, i^{1/B^{-1}(i)} \right\rfloor$. Assume B is nondecreasing and unbounded. Then for every $k \geq 2$ such that $B(k^2) \geq 2$ it holds that $(B(k^2))^{k+1} \overset{\min}{\to} (k)_{g_B}$.*

Proof. Given $k \geq 2$ such that $B(k^2) \geq 2$ set

$$N = (B(k^2))^{k+1} \quad \text{and} \quad \ell = 2 \cdot (B(k^2))^k \leq N\,.$$

Now let $C\colon [N]^2 \to \mathbb{N}$ be a g_B-regressive function. Consider the function $D\colon [B(k^2),\ell]^2 \to \mathbb{N}$ defined from C by restriction. For any $y \in [B(k^2),\ell]$ we have

$$y^{\frac{1}{B^{-1}(y)}} \leq (B(k^2))^{\frac{k+1}{B^{-1}(B(k^2))}} = (B(k^2))^{(k+1)\cdot k^{-2}}$$

which implies that $\mathrm{Im}(D) \subseteq (B(k^2))^{(k+1)\cdot k^{-2}} + 1$. On the other hand,

$$2 \cdot ((B(k^2))^{(k+1)\cdot k^{-2}} + 1)^{k-2} < ((B(k^2))^{(k+1)\cdot k^{-2}+1})^{k-1} < (B(k^2))^k.$$

By Lemma 3.1.1 there is some k-element set H which is min-homogeneous for D, and hence for C. □

Corollary 3.1.3. *Suppose $B : \mathbb{N} \to \mathbb{N}^+$ is unbounded, nondecreasing and $g(n) \leq g_B(n) = \left\lfloor\, n^{1/B^{-1}(n)} \right\rfloor$ for all n . If B is bounded by a primitive recursive function, then R^{reg}_g is bounded by a primitive recursive function. If, in addition, g itself is primitive recursive, then R^{reg}_g is primitive recursive.*

Proof. By the theorem above R^{reg}_g is eventually dominated by $(B(k^2))^{k+1}$ and thus is bounded by a primitive recursive function. If, in addition, g is primitive recursive, then the relation $N \overset{\min}{\to} (k)_g$ is a primitive recursive relation and the computation of R^{reg}_g requires only a bounded search for a solution for a primitive recursive relation and therefore R^{reg}_g is primitive recursive. □

3.2 g-regressive Upper Threshold

We provide now two different proofs for the upper threshold, by displaying two different "bad" colorings, each based on a different combinatorial proof of the fact the Id-regressive Ramsey numbers are Ackermannian [69, 56]. The first proof makes use of the idea from [69], and the second proof uses the idea of [56]. Both colorings are based on the idea of expanding the difference between two natural numbers by a "moving" base, depending on the position of the pair.

The first bad coloring we give codes "half" of the information that the second coloring codes: the color of $\{m, n\}$ according to the first coloring is the first different digit in the expansions of m and n, whereas according to the second it is the pair consisting of that digit and its position. The missing information in the first coloring is compensated by composing the regressive Ramsey function with the usual Ramsey function. The first proof is essentially asymptotic.

In the second proof we construct a single, simply computable $n^{1/\operatorname{Ack}^{-1}(n)}$-regressive, primitive recursive coloring of $[\mathbb{N}]^2$. It requires more detailed analysis of variants of approximations of Ackermann's function, but in return the result is less asymptotic and enables estimates of $R^{\text{reg}}_{\text{Id}}(k)$ for relatively small values of k. For instance, we show that $R^{\text{reg}}_{\text{Id}}(82)$ is larger than $A_{53}(2^{2^{274}})$.

3.2.1 g-regressive Upper Threshold – First Proof

We now begin working towards the first proof of the converse of Corollary 3.1.3: if f^{-1} is Ackermannian and $g(n) = n^{1/f(n)}$ then R^{reg}_g is Ackermannian. This proof generalizes the method developed in [69] and [56].

Definition 3.2.1. *For a given $t \in \mathbb{N} \setminus \{0\}$, we define a sequence of functions $(f_t)_i : \mathbb{N} \to \mathbb{N}$ as follows.*

$$(f_t)_1(n) = n + 1$$
$$(f_t)_{i+1}(n) = (f_t)_i^{(\lfloor n^{1/t} \rfloor)}(n)$$

Note that $(f_t)_i$ are strictly increasing. We also remark that $(f_1)_i = A_i$ and thus $(f_1)_k(k) = \operatorname{Ack}$. We would first like to show that the function $k \mapsto (f_t)_k(k)$ is Ackermannian for all $t > 0$. To do that, we show that although, for large t, the hierarchy $(f_t)_i$ grows more slowly than the Ackermann hierarchy $(f_1)_i$ (because functions are iterated only $n^{1/t}$ times instead of n times), one can compensate for this slowness by increasing the subscript i. The following computations show how much of an increase of i suffices for this purpose.

Claim 3.2.2. *For every $t, k, n > 0$ it holds that $(f_t)_k(n) \geq n + (\lfloor n^{1/t} \rfloor)^{k-1}$.*

Proof. We show the claim by induction on k. If $k = 1$, it follows by definition that $(f_t)_k(n) = n + 1 = n + (\lfloor n^{1/t} \rfloor)^{k-1}$. Let $k \geq 1$. By definition $(f_t)_{k+1}(n) = (f_t)_k^{(\lfloor n^{1/t} \rfloor)}(n)$ and by applying the induction hypothesis $\lfloor n^{1/t} \rfloor$ times we get that the right hand side of the equation is larger than $n + ((\lfloor n^{1/t} \rfloor)(\lfloor n^{1/t} \rfloor)^{k-1})$ which is $n + (\lfloor n^{1/t} \rfloor)^k$. □

Claim 3.2.3. *For every $t, k > 0$ and $n > 2^{t+1}$ it holds that $(f_{t+1})_{2t+3}(n^2) > n^2 + 2n + 1$.*

Proof. By Claim 3.2.2 we have that $(f_{t+1})_{2t+3}(n^2) \geq n^2 + (\lfloor n^{\frac{2}{t+1}} \rfloor)^{2t+2}$. Now

$$\begin{aligned} n^2 + (\lfloor n^{\frac{2}{t+1}} \rfloor)^{2t+2} &\geq n^2 + (n^{\frac{2}{t+1}} - 1)^{2(t+1)} \\ &\geq n^2 + (n^{\frac{4}{t+1}} - 2n^{\frac{2}{t+1}} + 1)^{t+1} \\ &> n^2 + (n^{\frac{2}{t+1}}(n^{\frac{2}{t+1}} - 2))^{t+1} \\ &> 2n^2 \\ &> n^2 + 2n + 1 \end{aligned}$$

for any $t, k > 0$ and $n > 2^{t+1}$. □

Claim 3.2.4. *Let $t > 0$. For all $n > 2^{t+1}$, $i > 0$ it holds that*

$$(f_{t+1})_{i+2t+2}(n^2) > ((f_t)_i(n))^2 .$$

Proof. We prove the claim simultaneously for all n, by induction on i. For $i = 1$, by Claim 3.2.3,

$$(f_{t+1})_{i+2t+2}(n^2) = (f_{t+1})_{2t+3}(n^2) > n^2 + 2n + 1 = ((f_t)_1(n))^2 = ((f_t)_i(n))^2 .$$

We now assume that Claim 3.2.4 is true for i (for all $n > 2^{t+1}$) and prove it for $i + 1$. To do that we need the following claim:

Claim 3.2.5. *For any $j \in \mathbb{N}^+$ it holds that $(f_{t+1})_{i+2t+2}^{(j)}(n^2) > ((f_t)_i^{(j)}(n))^2$.*

Proof. We show Claim 3.2.5 by induction on j. For $j = 1$ the claim is exactly the induction hypothesis for i. For $j > 1$ we have

$$(f_{t+1})_{i+2t+2}^{(j+1)}(n^2) = (f_{t+1})_{i+2t+2}((f_{t+1})_{i+2t+2}^{(j)}(n^2)).$$

The latter term is larger than $(f_{t+1})_{i+2t+2}(((f_t)_i^{(j)}(n))^2)$ by monotonicity and the induction hypothesis for j. Now, if we denote $n' = (f_t)_i^{(j)}(n)$, we easily see, by the induction hypothesis for i, that $(f_{t+1})_{i+2t+2}((n')^2) > ((f_t)_i(n'))^2$ which is, in fact, $((f_t)_i^{(j+1)}(n))^2$. □

We still need to show the induction step for Claim 3.2.4. We have

$$(f_{t+1})_{i+1+2t+2}(n^2) = (f_{t+1})_{i+2t+2}^{(\lfloor n^{\frac{2}{t+1}} \rfloor)}(n^2) \geq (f_{t+1})_{i+2t+2}^{(\lfloor n^{1/t} \rfloor)}(n^2)\,.$$

By Claim 3.2.5, the latter term is larger than $((f_t)_i^{(\lfloor n^{1/t} \rfloor)}(n))^2 = ((f_t)_{i+1}(n))^2$. □

Claim 3.2.6. *For all $t > 0$, $n > 4$ it holds that $(f_{t+1})_{i+t^2+3t}(n^{2^t}) > (A_i(n))^{2^t}$.*

Proof. Observe that $(A_i(n))^{2^t}$ is actually $((f_1)_i(n))^{2^t}$. By applying Claim 3.2.4 to the latter term, we get $((f_1)_i(n))^{2^t} < ((f_2)_{i+2+2}(n^2))^{2^{t-1}}$, since the parameter t of Claim 3.2.4 is 1 here. If we apply it now to the right hand side term, the parameter t of the claim would be 2 and we would find that this term is smaller than $((f_3)_{i+2+2+4+2}(n^2))^{2^{t-1}}$. Generally, if we apply the claim j times we get that $((f_1)_i(n))^{2^t} < ((f_{j+1})_{i+j^2+3j}(n^{2^j}))^{2^{t-j}}$ since we may replace $\sum_{l=1}^{j} 2j$ with $j^2 + j$. Thus, if we let $j = t$, we get the desired inequality. Note that we are allowed to apply Claim 3.2.4 t times, only if, for all $1 \leq j \leq t$ it holds that $n^{2^{j-1}} > 2^{j+1}$, which holds for every $n > 4$. □

Claim 3.2.7. *For every $t > 0$ and $n > 3^t$ it holds that $(f_t)_{4t+1}(n) > n^2$.*

Proof. Applying Claim 3.2.2 with $k = 4t+1$ we have $(f_t)_{4t+1}(n) \geq n + (\lfloor n^{1/t} \rfloor)^{4t}$ and the latter term is larger than $((n^{1/t} - 1)^2)^{2t}$ which equals $((n^{2/t} - 2n^{1/t} + 1))^{2t} > (n^{1/t}(n^{1/t} - 2))^{2t}$. Now, since $n > 3^t$ we know that $n^{1/t} - 2 > 1$ and thus, the latter term is larger than $(n^{1/t})^{2t} = n^2$. □

Claim 3.2.8. *For every $t > 0$ and $n > \max\{3^t, t^t\}$ it holds that $(f_t)_{4t+2}(n) > n^{2^t}$.*

Proof. By definition $(f_t)_{4t+2}(n) = (f_t)_{4t+1}^{(\lfloor n^{1/t} \rfloor)}(n)$ which is not less than $(f_t)_{4t+1}^{(t)}(n)$ since $n > t^t$. Now, applying Claim 3.2.7 t times, we get $(f_t)_{4t+1}^{(t)}(n) > n^{2^t}$ since f_t is monotone. □

Claim 3.2.9. *For any $t > 0$ and $n > \max\{4, 3^{t+1}, (t+1)^{t+1}\}$ it holds for any $i > 0$ that $(f_{t+1})_{i+t^2+4t+5}(n) > A_i(n)$.*

Proof. Since $n > 2^{t+1}$, we have that

$$(f_{t+1})_{i+t^2+4t+5}(n) = (f_{t+1})_{i+t^2+4t+4}^{(\lfloor n^{1/(t+1)} \rfloor)}(n) > (f_{t+1})_{i+t^2+4t+4}^{(2)}(n).$$

The latter term is clearly larger than $(f_{t+1})_{i+t^2+3t}((f_{t+1})_{4t+6}(n))$ since $i, t > 0$. By Claim 3.2.8 we have $(f_{t+1})_{4t+6}(n) > n^{2^{t+1}}$ and thus, by Claim 3.2.6 we get

$$(f_{t+1})_{i+t^2+3t}((f_{t+1})_{4t+6}(n)) > (f_{t+1})_{i+t^2+3t}(n^{2^t}) > (A_i(n))^{2^t}$$

which is clearly larger than $A_i(n)$. □

We are now ready to establish that the growth rate of $k \mapsto (f_t)_k(k)$ is Ackermannian in terms of k. We have already shown that every primitive recursive function is eventually dominated by $(f_t)_i$ for some i. We now use this and the fact that $(f_t)_i$ are increasing to establish that the growth rate of $k \mapsto (f_t)_k(k)$ is similar to that of the Ackermann function.

Claim 3.2.10. *For all* $0 < t \in \mathbb{N}$ *the function* $k \mapsto (f_t)_k(k)$ *is Ackermannian.*

Proof. For $t = 1$ the functions $(f_t)_k = A_k$, the standard k-th approximations of Ackermann's functions, so every primitive recursive function is eventually dominated by $(f_t)_k(k)$ (see e.g. [17]).

For $t > 1$ It suffices to show that for every $i \in \mathbb{N}$, the function $(f_t)_k(k)$ eventually dominates $A_i(k)$. Namely, that there exists $m_i \in \mathbb{N}$ such that for every $m > m_i$ it holds that $(f_t)_m(m) > A_i(m)$. But, by Claim 3.2.9, if we set $m_i = \max(\{(t+1)^{t+1}, i + t^2 + 4t + 5\})$, we get exactly that since for any $m > m_i$ it holds that $(f_t)_m(m) > (f_{t+1})_{i+t^2+4t+5}(m) > A_i(m)$. □

We now turn to the converse of Corollary 3.1.3.

Definition 3.2.11. *Given* $t \in \mathbb{N}^+$ *set,*

$$g_t(n) \triangleq \left\lfloor n^{1/t} \right\rfloor .$$

Lemma 3.2.12. $R^{\mathrm{reg}}_{g_t}(R(n+3, c)) \geq (f_t)_{c+1}(n)$ *for any* c *and* n.

Proof. Let $k = R(n+3, c)$ and define a function $C_t \colon [R^{\mathrm{reg}}_{g_t}(k)]^2 \to \mathbb{N}$ as follows:

$$C_t(x, y) = \begin{cases} 0 & \text{if } (f_t)_{c+1}(x) \leq y, \\ \ell & \text{otherwise,} \end{cases}$$

where the number ℓ is defined by

$$(f_t)_p^{(\ell)}(x) \leq y < (f_t)_p^{(\ell+1)}(x)$$

where $0 < p = \max\{q \colon (f_t)_q(x) \leq y\} < c + 1$. Note that C_t is g_t-regressive since $(f_t)_p^{(\left\lfloor x^{1/t} \right\rfloor)}(x) = (f_t)_{p+1}(x)$. Let H be a k-element subset of $R^{\mathrm{reg}}_{g_t}(k)$ which is min-homogeneous for C_t. Define a c-coloring $D_t \colon [H]^2 \to c$ by

$$D_t(x, y) = \begin{cases} 0 & \text{if } (f_t)_{c+1}(x) \leq y, \\ p - 1 & \text{otherwise,} \end{cases}$$

where p is as above. Then there is an $(n+3)$-element set $Y \subseteq H$ homogeneous for D_t. Let $x < y < z$ be the last three elements of Y. Then $n \leq x$ and thus it suffices to show that $(f_t)_{c+1}(x) \leq y$ since $(f_t)_{c+1}$ is an increasing function.

Assume $(f_t)_{c+1}(x) > y$. Then $(f_t)_{c+1}(y) \geq (f_t)_{c+1}(x) > z$ by the min-homogeneity. Let $C_t(x, y) = C_t(x, z) = \ell$ and $D_t(x, y) = D_t(x, z) = D_t(y, z) = p - 1$. Then

$$(f_t)_p^{(\ell)}(x) \leq y < z < (f_t)_p^{(\ell+1)}(x).$$

This implies that $z < (f_t)_p^{(\ell+1)}(x) \leq (f_t)_p(y) \leq z$. Contradiction. □

Corollary 3.2.13. *$R^{\text{reg}}_{g_t}$ is Ackermannian for any $t \in \mathbb{N}^+$.*

Proof. It is obvious by Claim 3.2.10 since $R^{\text{reg}}_{g_t}$ is nondecreasing. □

Theorem 3.2.14. *Suppose $f : \mathbb{N} \to \mathbb{N}$ is nonzero, nondecreasing and unbounded, and $f(i) \leq \text{Ack}(i)$ for all i. Let $g(i) = \left\lfloor\ i^{1/f^{-1}(i)} \right\rfloor$. It holds for all i that*

$$R^{\text{reg}}_g(R(4 + 3^{i+1} + (i + 1)^{i+1} + 3, i + i^2 + 4i + 5)) > f(i + 1)\,.$$

Proof. Let $p(i) = 4 + 3^{i+1} + (i + 1)^{i+1}$ and $q(i) = i + i^2 + 4i + 5$. Assume to the contrary that for some i

$$N(i) = R^{\text{reg}}_g(R(p(i) + 3, q(i))) \leq f(i + 1)\,.$$

For all $\ell \leq N(i)$ we have $f^{-1}(\ell) \leq i + 1$, and hence $\ell^{1/(i+1)} \leq \ell^{1/(f^{-1}(\ell))}$. Then

$$\begin{aligned} R^{\text{reg}}_g(R(p(i) + 3, q(i))) &\geq R^{\text{reg}}_{g_{i+1}}(R(p(i) + 3, q(i))) \\ &\geq (f_{i+1})_{q(i)+1}(p(i)) \\ &> A_{i+1}(p(i)) \\ &\geq \text{Ack}(i + 1) \\ &\geq f(i + 1) \end{aligned}$$

by Lemma 3.2.12 and Claim 3.2.9. Contradiction! □

Theorem 3.2.15. *Suppose $B : \mathbb{N} \to \mathbb{N}$ is positive, unbounded and nondecreasing. Let $g_B(i) = \left\lfloor\ i^{1/B^{-1}(i)} \right\rfloor$. Then $R^{\text{reg}}_{g_B}(k)$ is Ackermannian iff B is Ackermannian.*

Proof. Suppose B is Ackermannian. By replacing B with $\min\{B, \text{Ack}\}$, we may assume that $B(i) \leq \text{Ack}(i)$ for all $i \in \mathbb{N}$. That $R^{\text{reg}}_{g_B}$ is Ackermannian follows from the previous theorem, since $r(i) = R(4+3^{i+1}+(i+1)^{i+1}+3, i+i^2+4i+5)$ is primitive recursive.

Suppose now that B is not Ackermannian, and fix an increasing primitive recursive function f so that for infinitely many $i \in \mathbb{N}$ it holds that $B(i) < f(i)$. On the other hand, it holds by Theorem 3.1.2 that $R^{\text{reg}}_{g_B}(k) \leq (B(k^2))^{k+1}$ for any $k \geq 2$ such that $B(k^2) \geq 2$. Hence it holds that $R^{\text{reg}}_{g_B}(i) \leq (f(i^2))^{i+1}$ for infinitely many $i \in \mathbb{N}$. This means that, for infinitely many $i \in \mathbb{N}$, $R^{\text{reg}}_{g_B}(i)$ is bounded by $f'(i)$ for some primitive recursive $f' : \mathbb{N} \to \mathbb{N}$. □

3.2.2 g-regressive Upper Threshold – Second Proof

We now begin the second proof by presenting a general method for constructing a "bad" g-regressive coloring which is a generalization of the method from [56]. In other words, given a function g and a natural number k, we present a g-regressive coloring C_g of pairs over a segment of size depending on g and k such that there is no min-homogeneous set for C_g of size $k+1$ within that segment. We then further show that if $g(n) = n^{1/r}$ for $r > 0$, then the size of the segment we may color is Ackermannian in terms of k. We then use this general coloring method to construct a single $n^{1/\operatorname{Ack}^{-1}(n)}$-regressive "bad" coloring of $[\mathbb{N}]^2$.

Let $g : \mathbb{N} \to \mathbb{N}$ a nondecreasing function such that for every $k \in \mathbb{N}$ there exists some $t \in \mathbb{N}$ such that $k \leq \frac{\sqrt{g(t)}}{2}$. Let $\mu_g : \mathbb{N} \to \mathbb{N}$ be a function which satisfies for all $k \in \mathbb{N}$ that $k \leq \frac{\sqrt{g(\mu_g(k))}}{2}$.

Definition 3.2.16. *We define a sequence of functions $(f_g)_i : \mathbb{N} \to \mathbb{N}$ as follows.*

$$(f_g)_1(n) = n + 1$$
$$(f_g)_{i+1}(n) = (f_g)_i^{(\lfloor \frac{\sqrt{g(n)}}{2} \rfloor)}(n)$$

Given $k > 2$, we define a sequence of semi-metrics $\langle (d_g)_i : i \in \mathbb{N} \rangle$ on $\{n : n \geq \mu_g(k)\}$ by setting, for $m, n \geq \mu_g(k)$,

$$(d_g)_i(m,n) = |\{\ell \in \mathbb{N} : m < (f_g)_i^{(\ell)}(\mu_g(k)) \leq n\}|$$

For $n > m \geq \mu_g(k)$ let $I_g(m,n)$ be the greatest i for which $(d_g)_i(m,n)$ is positive, and $D_g(m,n) = (d_g)_{I(m,n)}(m,n)$.

Let us fix the following (standard) pairing function Pr on $\mathbb{N}^2$:

$$\operatorname{Pr}(m,n) = \binom{m+n+1}{2} + n$$

$\operatorname{Pr} : \mathbb{N}^2 \to \mathbb{N}$ is bijective and monotone in each variable. Observe that if $m, n \leq \ell$ then $\operatorname{Pr}(m,n) < 4\ell^2$ for all $\ell > 2$.

Definition 3.2.17. *Given a natural number $k > 2$ and an unbounded nondecreasing function $g : \mathbb{N} \to \mathbb{N}$, we define a pair coloring C_g on $[\{n : n \geq \mu_g(k)\}]^2$ as follows:*

$$C_g(m,n) = \operatorname{Pr}(I_g(m,n), D_g(m,n))$$

Claim 3.2.18. $D_g(m,n) \leq \frac{\sqrt{g(m)}}{2}$ *for all $n > m \geq \mu_g(k)$.*

Proof. Let $i = I_g(m,n)$. Since $(d_g)_{i+1}(m,n) = 0$, there exist t and ℓ such that

$$t = (f_g)_{i+1}^{(\ell)}(\mu_g(k)) \leq m < n < (f_g)_{i+1}^{(\ell+1)}(\mu_g(k)) = (f_g)_{i+1}(t).$$

But $(f_g)_{i+1}(t) = (f_g)_i^{(\lfloor \frac{\sqrt{g(t)}}{2} \rfloor)}(t)$ and thus $\frac{\sqrt{g(t)}}{2} \geq (d_g)_i(t, (f_g)_{i+1}(t)) \geq D_g(m,n)$. □

Claim 3.2.19. *C_g is g-regressive on the interval $[\mu_g(k), (f_g)_k(\mu_g(k)))$.*

Proof. Clearly, $(d_g)_k(m,n) = 0$ for $\mu_g(k) \leq m < n < (f_g)_k(\mu_g(k))$ and therefore $I_g(m,n) < k \leq \frac{\sqrt{g(m)}}{2}$. From Claim 3.2.18 we know $D_g(m,n) \leq \frac{\sqrt{g(m)}}{2}$. Thus, $C_g(\{m,n\}) \leq \Pr(\lfloor \frac{\sqrt{g(m)}}{2} \rfloor, \lfloor \frac{\sqrt{g(m)}}{2} \rfloor)$, which is $< g(m)$ since $\frac{\sqrt{g(m)}}{2} > 2$. □

Claim 3.2.20. *For every $i \in \mathbb{N}$, every sequence $x_0 < x_1 < \cdots < x_i$ that satisfies $(d_g)_i(x_0, x_i) = 0$ is not min-homogeneous for C_g.*

Proof. The claim is proved by induction on i. If $i = 1$ then there are no $x_0 < x_1$ with $(d_g)_1(x_0, x_1) = 0$ at all. Let $i > 1$ and suppose to the contrary that $x_0 < x_1 < \cdots < x_i$ form a min-homogeneous sequence with respect to C_g and that $(d_g)_i(x_0, x_i) = 0$. Necessarily, $I_g(x_0, x_i) = j < i$. By min-homogeneity, $I(x_0, x_1) = j$ as well, and $(d_g)_j(x_0, x_i) = (d_g)_j(x_0, x_1)$. Hence, $\{x_1, x_2, \ldots x_i\}$ is min-homogeneous with $(d_g)_j(x_1, x_i) = 0$, contrary to the induction hypothesis. □

Corollary 3.2.21. *There exists no $H \subseteq [\mu_g(k), (f_g)_k(\mu_g(k)))$ of size $k+1$ that is min-homogeneous for C_g.*

Corollary 3.2.22. *Assume that the function $k \mapsto (f_g)_k(k)$ is Ackermannian. If there exists a function μ_g that is bounded by some primitive recursive function and satisfies for all k that $k \leq \mu_g(k)$ and that $k \leq \frac{\sqrt{g(\mu_g(k))}}{2}$, then R_g^{reg} is also Ackermannian.*

Proof. First consider the function $C'_g : [(f_g)_k(\mu_g(k))]^2 \to \mathbb{N}$ defined by

$$C'_g(m,n) = \begin{cases} 0 & \text{if } m < \mu_g(k), \\ C_g(m,n) & \text{otherwise.} \end{cases}$$

Note that C'_g is g-regressive and has, by Corollary 3.2.21, no min-homogeneous set of size $\mu_g(k) + k + 1$. Hence, we have $R_g^{\text{reg}}(\mu_g(k) + k + 1) > (f_g)_k(\mu_g(k))$.

On the other hand, the function $k \mapsto (f_g)_k(\mu_g(k))$ is obviously Ackermannian. Therefore, R_g^{reg} is also Ackermannian because $\mu_g(k)$ is bounded by some primitive recursive function (See the proof of Lemma 2.2.3). □

Lemma 3.2.23. *Given a real number $r > 0$ let $g(n) = \lfloor n^{1/r} \rfloor$. Then the function R_g^{reg} is Ackermannian.*

Proof. Given a real number $r > 0$ let $t = \lceil r \rceil$. We first observe that the function $k \mapsto \frac{k^{1/2t}}{2}$ grows eventually faster than the function $k \mapsto k^{1/4t}$ and therefore, by Claim 3.2.10, $k \mapsto (f_{g_t})_k(k)$ is Ackermannian. Set $\mu_{g_t}(k) = 4^t k^{2t}$. By Corollary 3.2.22, $R^{\mathrm{reg}}_{g_t}$ is Ackermannian. Therefore R^{reg}_{g} is Ackermannian too. □

We conclude with a single primitive recursive procedure for coloring all of $[\mathbb{N}]^2$ whose Ramsey function is Ackermannian.

Theorem 3.2.24. *Suppose* $g(n) = \left\lfloor n^{1/\operatorname{Ack}^{-1}(n)} \right\rfloor$ *for* $n > 0$ *and* $g(0) = 0$. *There exists a g-regressive, primitive recursive coloring* $C : [\mathbb{N}]^2 \to \mathbb{N}$ *such that for every primitive recursive function* $f : \mathbb{N} \to \mathbb{N}$ *there exists* $N_f \in \mathbb{N}$ *such that for all* $m > N_f$ *and* $H \subseteq m$ *which is min-homogeneous for* C *it holds that* $f(|H|) < m$.

Proof. We define a g-regressive coloring C by dividing $\mathbb{N}^+$ into disjoint intervals of the form $(\mu_{t-1}, \mu_t]$, defining a g-regressive coloring C_t for all pairs over each such interval. For each t, we specify an upper bound k_t on the sizes of C_t-min-homogeneous subsets of $(\mu_{t-1}, \mu_t]$. For the first interval we fix an ad-hoc coloring and for all other intervals we use the definition of C_g as described above. Finally, we integrate all colorings to a single coloring of all pairs over $\mathbb{N}$, by simply setting $C(m,n) = 0$ for m, n from different intervals and $C(0,n) = 0$ for all $n \in \mathbb{N}^+$. For notational convenience we start with $\mu_2 = 0$. We set $\mu_3 = 2^{61}$ and $\mu_t = \operatorname{Ack}(t)$ for $t \geq 4$.

On $(\mu_2 = 0, \mu_3]$ fix C_3 as follows. Since $g(n) \geq 1$ for all $n > 0$ we may color pairs from $(0, 2^{61}]$ g-regressively by 2 colors. Using a simple probabilistic argument it may be shown that for any $k \geq 4$, there exists a 2-coloring of $\left[2^{k/2}\right]^2$ with no min-homogeneous set of size k. We set $k_3 = 122$ and let C_3 be a restriction of such a coloring to $(0, 2^{61}]$.

Now we need to define C_t for all $t > 3$. Let $k_4 = 98$ and $k_t = 16t^2 + 9t + 2$ for all $t > 4$. We color pairs over the interval $[\mu_{t-1}, (f_{g_t})_{k_t}(\mu_{t-1}))$ by C_g as defined above (Definition 3.2.17), using as parameters, $g = g_t$, as defined in Definition 3.2.11, and $k = k_t$. For formality, we fix the function $\mu_{g_t}(k) = \mu_{t-1}$ iff t is the least number such that $3 < t$ and $k \leq k_t$. For our needs, however, it suffices to observe that for all $t > 3$ it holds that $k_t \leq \frac{\sqrt{g_t(\mu_{g_t}(k_t))}}{2}$, which can easily be verified. We set C_t, for $t > 3$, to be the restriction of C_{g_t} to $(\mu_{t-1}, \mu_t]$ (See Claim 3.2.25 to observe that it is a restriction).

The following claim shows that the union of all intervals, indeed covers all $\mathbb{N}$.

Claim 3.2.25. $\operatorname{Ack}(t) < (f_{g_t})_{k_t}(\mu_{t-1})$ *for all* $t > 3$.

Proof. We first prove Claim 3.2.25 for $t = 4$. Note that $k_4 = 98$. Observe that $\frac{61}{8} - 1 > \frac{61}{10}$ and hence for all $n \geq 2^{61}$ and for every $i \in \mathbb{N}$ it holds that $(f_{g_4})_i(n) \geq (f_{10})_i(n)$. By Claim 3.2.2 we know that $(f_{10})_{97}(2^{61}) > (\lfloor (2^{61})^{1/10} \rfloor)^{96} > 2^{576}$.

Using the same argument again we also know that $(f_{10})_{97}(2^{576}) > 2^{5472}$. Thus, $(f_{g_4})_{k_4}(\mu_3) = (f_{g_4})_{98}(2^{61})$ is much larger than

$$(f_{g_4})_{97}^{(3)}(2^{61}) > (f_{g_4})_{97}((f_{10})_{97}^{(2)}(2^{61})) > (f_{g_4})_{97}(2^{5472})$$

Now, since $\frac{5472}{8} - 1 > \frac{5472}{9}$ it holds for all $m \geq 2^{5472}$ and for every $i \in \mathbb{N}$ that $(f_{g_4})_i(m) \geq (f_9)_i(m)$. Hence we have $(f_{g_4})_{97}(2^{5472}) \geq (f_9)_{97}(2^{5472})$ and by Claim 3.2.6, we have that $(f_9)_{97}(2^{5472}) > (f_9)_{9+8^2+24}(5^{2^8}) > (A_9(5))^{2^8}$ and thus, obviously larger than $A_4(4)$. Now, let $t > 4$. Observe that $\mu_{t-1} > A_4(t-1)$ and therefore larger than 2^{4t}. Since for all $n \geq 2^{4t}$ and for every $i \in \mathbb{N}$ it holds that $(f_{g_t})_i(n) \geq (f_{4t})_i(n)$, we have that $(f_{g_t})_{k_t}(\mu_{t-1}) \geq (f_{4t})_{k_t}(\mu_{t-1})$. It also holds that $\mu_{t-1} > (4t)^{4t}$. Hence, by Claim 3.2.9 $(f_{4t})_{k_t}(\mu_{t-1}) > A_{k_t-16t^2-8t-2}(\mu_{t-1}) = A_t(Ack(t-1))$ which is obviously larger than $A_t(t)$. □

Finally, we define C as follows.

$$C(m,n) = \begin{cases} C_t(m,n) & \text{if } 0 < m,n \in (\mu_{t-1}, \mu_t], \\ 0 & \text{otherwise.} \end{cases}$$

Claim 3.2.26. *The coloring C is g-regressive.*

Proof. Let $m, n \in \mathbb{N}$ be such that $m < n$. If $C(m,n) = 0$ then we have $C(m,n) \leq g(m)$. Otherwise, m and n are in the same interval. If $m, n \in (\mu_2, \mu_3]$ then $C(m,n) \leq 1 \leq \left\lfloor\, m^{1/Ack^{-1}(m)} \right\rfloor$ by definition of C_3. If $m, n \in (\mu_{t-1}, \mu_t]$ for some $t > 3$, then we have $Ack^{-1}(m) = t$. We also know C_t is g_t regressive on that interval and thus $C(m,n) = C_t(m,n) \leq \lfloor\, m^{1/t} \rfloor = \left\lfloor\, m^{1/Ack^{-1}(m)} \right\rfloor$. □

Claim 3.2.27. *The coloring C is primitive recursive.*

Proof. It is primitive recursive to compute for an input n the last value of Ack below n. Thus, given input m, n one can determine whether there is some $t \geq 3$ so that $m, n \in (\mu_{t-1}, \mu_t]$. The computation of C on each $(\mu_{t-1}, \mu_t]$ is uniform and primitive recursive. So altogether, C is primitive recursive. □

Claim 3.2.28. *For any given $N \in \mathbb{N}$ with $\mathrm{Ack}^{-1}(N) < j$ for some $j > 3$, there is no C-min-homogeneous $H \subseteq [N]$ of size $(k_j)^2 + 123$.*

Proof. Clearly, for all $t > 3$ it holds that $k_t < k_{t+1}$ and that $k_t > t$. Thus, since at any interval $(\mu_{t-1}, \mu_t]$ for $3 < t \leq j$, the largest min-homogeneous subset may be of size k_t and hence, no more than k_j. Therefore, in the union of all those intervals there is no min-homogeneous subset larger than $k_j(j-3) < (k_j)^2$. Now, in the first interval there can be no min-homogeneous of size 122. Thus, as we allow 0 to be an element of any min-homogeneous subset, so there is no min-homogeneous $H \subseteq [N]$ of size $(k_j)^2 + 123$ in the union of all intervals before $\mathrm{Ack}(j)$, of which $[N]$ is a subset. □

To conclude the proof of Theorem 3.2.24, fix $f_1(i) = i^2 + 123$ and $f_2(i) = 16i^2 + 9i + 2$. Now, given some primitive recursive function f, let f' be some increasing primitive recursive function which bounds f. Note that the composition $h = f' \circ (f_1 \circ f_2)$ is also primitive recursive. Let $t_0 > 4$ be the least natural number such that for all $t \geq t_0$ it holds that $\mathrm{Ack}(t-1) > h(t)$. Let $N_f = \mathrm{Ack}(t_0)$. Given $m > N_f$ such that $m \in (\mu_{t-1}, \mu_t]$ and $H \subseteq m$ which is min-homogeneous for C, by Claim 3.2.28 we know that $|H| < k_t^2 + 123 = f_1(f_2(t))$. By monotonicity of f', we have $f'(|H|) < f'(f_1(f_2(t)))) = h(t)$. Since $N_f < m$ and by monotonicity of Ack, we have $t \geq t_0$ and thus $h(t) < \mathrm{Ack}(t-1) < m$. Now, $f(i) \leq f'(i)$ for all $i \in \mathbb{N}$ and therefore $f(|H|) \leq f'(|H|) < m$

This completes the proof of Theorem 3.2.24. □

3.2.3 The Id-regressive Ramsey Number of 82 is Larger than $A_{53}(2^{2^{274}})$

We provide now a (huge) lower estimate on an Id-regressive Ramsey number for a reasonably small $k = 82$. The point to stress is that the bad colorings we had above work not only asymptotically but may be used to estimate small values. For more on small regressive Ramsey numbers see Blanchard [8].

Claim 3.2.29. *For $g = Id$ it holds that $R_g^{\mathrm{reg}}(82) > A_{53}(2^{2^{274}})$.*

Proof. Let $\mu = 2^{14}$ and $k = 64$. By Claims 3.2.19 and 3.2.20 we know that there is a g-regressive coloring C_{Id} on the interval $[\mu, (f_{\mathrm{Id}})_k(\mu))$ which yields no $H \subseteq [\mu, (f_{\mathrm{Id}})_k(\mu))$ of size $k+1$ which is min-homogeneous for C_{Id}. Let us now examine the magnitude of $(f_{\mathrm{Id}})_k(\mu)$. By definition

$$(f_{\mathrm{Id}})_k(\mu) = (f_{g_1})_{64}(2^{14}) = (f_{g_1})_{63}^{(64)}(2^{14}).$$

Since for all $x > 2^6$ it holds that $\frac{x^{1/2}}{2} > x^{1/3}$ and by monotonicity, we may look at $(f_3)_{63}^{(64)}(2^{14})$ which, by Claim 3.2.2, is larger than $(f_3)_{63}^{(63)}((\lfloor 2^{14/3} \rfloor)^{62}) > (f_3)_{63}^{(63)}(2^{285})$. By applying the same argument again we get $(f_3)_{63}^{(63)}(2^{285}) > (f_3)_{63}^{(62)}(2^{5889})$. We go on applying Claim 3.2.2 in the straightforward manner until we establish that the latter term is larger than $(f_3)_{63}^{(59)}(2^{51981110})$ and then we start using 60 instead of 62 at the exponent which enables us to lose the rounding operation. Thus, we know $(f_3)_{63}^{(59)}(2^{51981110}) > (f_3)_{63}^{(1)}(2^{51981110*20^{58}}) > (f_3)_{63}(2^{2^{276}})$. By applying Claim 3.2.6 to the latter term we get $(f_3)_{63}(2^{2^{276}}) = (f_3)_{53+2^2+6}((2^{2^{274}})^{2^2}) > (A_{53}(2^{2^{274}}))^{2^2}$ which is obviously larger than $A_{53}(2^{2^{274}})$.

On $[0, 13)$ there is an Id-regressive coloring with no min-homogeneous set with more than 4 elements (see [8]). On $[13, 2^{14})$ let $C(m, n)$ be the largest position of a different digit in the base 2 expansions of m and n. This coloring is Id-regressive, since $C(m, n) \leq 13$ for all such m, n and admits no min-homogeneous set of size 14. Coloring m, n from different intervals by 0 produces then a coloring on the interval $[0, A_{53}(2^{2^{274}}))$ with no min-homogeneous set of size larger than $4 + 13 + 64 = 81$. □

3.3 Phase Transition From Homogeneous Ramsey Numbers to Min-Homogeneous Ramsey Numbers

We now look at the threshold g at which one can guarantee the usual Ramsey theorem for g-regressive colorings, that is, have homogeneous rather than just min-homogeneous sets.

Theorem 3.3.1. *Suppose $f : \mathbb{N} \to \mathbb{N}^+$ is nondecreasing and unbounded, and let $g(x) = \left\lfloor \frac{\log(x)}{f(x)\log(\log(x))} \right\rfloor$ for $x \geq 4$ and $g(x) = 0$ for $x < 4$. Then for all k there exists some N so that $N \to (k)_g$.*

Proof. Given $k \geq 4$, find N_1 such that $f(N_1) > k$. Observe that for all $N_1 \leq m_1 \leq m_2$, it holds that $g(m_1) \leq \left\lfloor \frac{\log(m_2)}{k\log(\log(m_2))} \right\rfloor$. This is because the function $\frac{z}{\log z}$ is not decreasing for $z \geq 2$. Let $N = \max\{2N_1, 2^{2^k}\}$. Clearly, $\left\lfloor \frac{\log(N)}{k\log(\log(N))} \right\rfloor \geq 1$. We claim that any g-regressive function defined on $[N]^2$ admits a k sized homogeneous set.

Let $C : [N]^2 \to \mathbb{N}$ be g-regressive and $C' : [N_1, N]^2 \to c$ be its restriction, where $c = \left\lfloor \frac{\log(N)}{k\log(\log(N))} \right\rfloor + 1$. Note that we have, since $k \leq \log(\log(N))$,

$$\begin{aligned} c^{c \cdot k} &\leq \left(\frac{2\log(N)}{k\log(\log(N))} \right)^{k\left(\frac{\log(N)}{k\log(\log(N))}+1\right)} \\ &\leq N \cdot \frac{(\log(N))^k}{N^{\log(\log(\log(N)))/\log(\log(N))}} \cdot \frac{1}{\log(\log(N))} \\ &< \frac{N}{\log(\log(N))} < \frac{N}{2} \leq N - N_1 \end{aligned}$$

By the standard Ramsey Theorem, there is a k sized C'-homogeneous set $H \subseteq [N_1, N]$. Hence C admits a k sized homogeneous set.

It should be noted that this is of interest when f grows slowly (e.g. $f(m) = \log^*(m)$). □

Theorem 3.3.2. *Suppose $j \in \mathbb{N}$ and $g(i) = \frac{\log(i)}{j}$. Then there exists some k such that for all N it holds that $N \nrightarrow (k)_g$.*

Proof. Given $j \geq 2$ we set $s = 2^j$ and $k = 2s+1$ and construct a g-regressive coloring $C : \mathbb{N}^2 \to \mathbb{N}$ where there exists no $H \subseteq \mathbb{N}$ of size $\geq k$ that is homogeneous for C. For any $n \in \mathbb{N}$, let $r_s(n) = (n_0, \dots, n_{\ell-1})$, where $\ell = \lfloor\ \log_s(n) \rfloor + 1$ and $n_i < s$, be the representation of n in s basis, i.e. $n = n_0 \cdot s^{\ell-1} + \cdots + n_{\ell-1} \cdot s^0$.

For any $m, n \in \mathbb{N}$ such that $m < n$ and $\ell = \lfloor \log_s(m) \rfloor + 1 = \lfloor \log_s(n) \rfloor + 1$, let $f(m,n) = \min\{i < \ell : m_i < n_i\}$, where $r_s(m) = (m_0, \dots, m_{\ell-1})$ and $r_s(n) = (n_0, \dots, n_{\ell-1})$. We define C as

$$C(m,n) = \begin{cases} \lfloor \log_s(m) \rfloor & \text{if } \lfloor \log_s(m) \rfloor \neq \lfloor \log_s(n) \rfloor; \\ f(m,n) & \text{if } \lfloor \log_s(m) \rfloor = \lfloor \log_s(n) \rfloor. \end{cases}$$

Note that C is g-regressive since for all $m, n \in \mathbb{N}$ it holds that $C(m,n) \leq \log_s(m) = \frac{\log(m)}{j}$.

Observation 3.3.3. *Let $Y = \{y_1, y_2, ..., y_{s+1}\}$ where $y_1 < y_2 < ... < y_{s+1}$, be a homogeneous set for C. Then $\lfloor \log_s(y_1) \rfloor < \lfloor \log_s(y_{s+1}) \rfloor$.*

To show Observation 3.3.3, let Y be a homogeneous set for C and suppose to the contrary that $\lfloor \log_s(y_1) \rfloor = \lfloor \log_s(y_{s+1}) \rfloor$. From the definition of C we get that f is constant on Y. Thus elements of Y, pairwise differ in the ith value in their s basis representation for some index i, which is impossible since there are only s possible values for any index. Contradiction.

Now let $H = \{x_1, x_2, ..., x_{2s+1}\}$, where $x_1 < x_2 < ... < x_{2s+1}$, and suppose to the contrary that H is homogeneous for C. By Observation 3.3.3 we get that $\lfloor \log_s(x_1) \rfloor < \lfloor \log_s(x_{s+1}) \rfloor < \lfloor \log_s(x_{2s+1}) \rfloor$ and therefore $C(x_1, x_{s+1}) < C(x_{s+1}, x_{2s+1})$ contrary to homogeneity. □

Chapter 4

Phase Transition Threshold of g-large Ramsey Numbers

In this chapter we show that the threshold for Ackermannian g-large Ramsey numbers lies above all functions $\log(n)/f^{-1}(n)$ obtained from an increasing primitive recursive f and below the function $\log(n)/\operatorname{Ack}^{-1}(n)$.

Worded differently, for a nondecreasing and unbounded g to have primitive recursive g-large Ramsey numbers it is necessary and sufficient that g is eventually dominated by $\log(n)/t$ for all $t > 0$ and that the rate at which g gets below $\log(n)/t$ is not too slow, namely, is primitive recursive in t: if g gets below $\log(n)/t$ only after an Ackermannianly long time $M(t)$, then the g-large Ramsey numbers are still Ackermannian.

Here, in this thesis, $\log$ denotes the logarithm to base 2.

We begin by recalling some notations and definitions from Chapter 2. A nonempty $H \subseteq \mathbb{N}$ is g-large for a function $g : \mathbb{N} \to \mathbb{N}$ if $|H| \geq g(\min H)$. The symbol

$$X \to^*_g (k)_c$$

means: for every coloring $C : [X]^2 \to c$ there is a g-large C-homogeneous $H \subseteq X$ such that $|H| \geq k$. The g-large Ramsey number of k and c, denoted $R^*_g(k, c)$, is the least N so that $N \to^*_g (k)_c$.

The notion of g-largeness is a generalization of the notion of a relatively large set introduced by Paris and Harrington in [66]. Erdős and Mills [33] proved that R^*_{Id} is Ackermannian. On the other hand, for a constant function $g = t$, the g-large Ramsey number is just the standard Ramsey number (that is, $R^*_g(k, c) = R^{(}\max\{k, t\}, c))$ and thus primitive recursive.

We next compute the sharp thresholds on g at which g-large Ramsey numbers cease to be primitive recursive and become Ackermannian.

In this chapter we shall work with a new hierarchy of functions F_m. It is similar to that of A_m, only it starts with a faster growing function than the successor function:

$$F_m(i) \triangleq \begin{cases} 2^i & \text{if } m = 0, \\ F_{m-1}^{(i \dot{-} 1)}(i) & \text{otherwise.} \end{cases}$$

Here $i \dot{-} 1 = i - 1$ if $i > 0$ and 0 otherwise. This is merely done for technical convenience and helps us handle the logarithm much better. For any $m \in \mathbb{N}$, F_m is an increasing primitive recursive function. The function $F \colon \mathbb{N} \to \mathbb{N}$, defined by $F(i) \triangleq F_i(i)$, is Ackermannian. In fact, F and Ack have almost the same growth rate. The results in this chapter are taken almost verbatim from [55].

4.1 g-large Lower Threshold

We employ classical bounds by Erdős and Rado for the lower bound and a result by Abbott [1] for the upper bound which relies on the probabilistic method of Erdős. The following lemma follows e.g. from Theorem 1 in [35].

Lemma 4.1.1. *$R(k,c) \leq c^{c \cdot k - 1} = 2^{(c \cdot k - 1) \cdot \log(c)}$ for any c, $k \geq 2$.*

For $m \in \mathbb{N}$ and a function $B : \mathbb{N} \to \mathbb{N}$ set

$$f_m(i) = \left\lfloor \frac{\log(i)}{F_m^{-1}(i)} \right\rfloor \quad \text{and} \quad f_B(i) = \left\lfloor \frac{\log(i)}{B^{-1}(i)} \right\rfloor .$$

Lemma 4.1.2. *Let $B : \mathbb{N} \to \mathbb{N}$ be a nondecreasing and unbounded positive function. Then for every c, $k \geq 2$ it holds that $R(t,c) + B(c \cdot \lceil \log(c) \rceil) \to^*_{f_B} (k)_c$, where $t = \max\{k, B(c \cdot \lceil \log(c) \rceil)\}$.*

Proof. Given c, $k \geq 2$ let $N = R(t,c)$. By Lemma 4.1.1

$$N + B(c \cdot \lceil \log(c) \rceil) \leq 2^{(c \cdot t - 1) \cdot \log(c)} + B(c \cdot \lceil \log(c) \rceil) \leq 2^{c \cdot \log(c) \cdot t} .$$

Now, let $C \colon [N + B(c \cdot \lceil \log(c) \rceil)]^2 \to c$ be given. Since $N + B(c \cdot \lceil \log(c) \rceil) - B(c \cdot \lceil \log(c) \rceil) = N$, there is an $H \subseteq [B(c \cdot \lceil \log(c) \rceil), N + B(c \cdot \lceil \log(c) \rceil))$ homogeneous for C, such that $|H| \geq t$. Therefore, we have

$$\frac{\log(\min(H))}{B^{-1}(\min(H))} \leq \frac{c \cdot \log(c) \cdot t}{B^{-1}(B(c \cdot \lceil \log(c) \rceil))} \leq t \leq |H| .$$

Thus H is f_B-large. □

Theorem 4.1.3. *For every fixed m the function $R^*_{f_m}$ is primitive recursive.*

Proof. By Lemma 4.1.2, $R^*_{f_m}$ is bounded by a primitive recursive function and thus is itself primitive recursive, as the class of primitive recursive functions is closed under the bounded μ-operator. □

4.2 g-large Upper Threshold

We now turn to establish a complement for Theorem 4.1.3.

Lemma 4.2.1. *There are positive integers c_0 and M such that for all $c \geq 2$ and $k \geq M$ it holds that $R(k,c) \geq 2^{\frac{1}{c_0}\cdot c\cdot k}$.*

Proof. See Abbott [1]. □

Lemma 4.2.2. *Let $M \geq 2$ and c_0 be the constants from Lemma 4.2.1. Let $d \geq 4$ be arbitrary, but fixed. Put $\varepsilon = \frac{1}{d}$ and $K = 2 \cdot d \cdot M + 1$. Then*

$$R^*_{\hat{f}_\varepsilon}(k, c_0 \cdot M \cdot d \cdot 2) > 2^{2^{k \cdot d}}$$

for all $k \geq K$, where $\hat{f}_\varepsilon(i) = \varepsilon \cdot \log(i)$.

Proof. Pick $k \geq K$. Let $n_0 = 0$, $n_1 = R(k, c_0) - 1$, and for $1 \leq i < k-1$

$$n_{i+1} = n_i + R(\lfloor \varepsilon \cdot \log(n_i) \rfloor, c_0 \cdot M \cdot d \cdot 2 - 1) - 1.$$

Finally put $n = n_{k-1}$. We claim:

$$n \not\to^*_{\hat{f}_\varepsilon} (k)_{c_0 \cdot M \cdot d \cdot 2}$$

Choose $C_0 : [n_0, n_1)^2 \to c_0$ such that every C_0-homogeneous $H \subseteq [n_0, n_1)$ satisfies $|H| < k$. For $1 \leq i < k-1$ choose

$$C_i : [n_i, n_{i+1})^2 \to c_0 \cdot M \cdot d \cdot 2 - 1$$

such that if H is C_i-homogeneous then $|H| < \lfloor \varepsilon \cdot \log(n_i) \rfloor$.

Define $C \colon [n]^2 \to c_0 \cdot M \cdot d \cdot 2$ as follows:

$$C(u,v) = \begin{cases} C_i(u,v) + 1 & \text{if } n_i \leq u < v < n_{i+1}, \\ 0 & \text{otherwise.} \end{cases}$$

Let H be C-homogeneous. If the color of H is 0 then $|(H \cap [n_i, n_{i+1}))| \leq 1$, hence $|H| \leq k-1 < k$. If the color of H under C is greater than 0 then $H \subseteq [n_j, n_{j+1}]$ for some j and H is homogeneous for C_j. If $j = 0$ then $|H| < k$ by choice of C_0. If $j > 0$ then

$$|H| < \lfloor \varepsilon \cdot \log(n_j) \rfloor \leq \lfloor \varepsilon \cdot \log(\min(H)) \rfloor \ .$$

This implies that $n < R^*_{\hat{f}_\varepsilon}(k, c_0 \cdot M \cdot d \cdot 2)$.

Now we use induction on $1 \leq i < k$ to prove that $n_i \geq 2^{2^{i \cdot d \cdot M}}$. For $i = 1$ we have, by Lemma 4.2.1,

$$n_1 \geq 2^{\frac{1}{c_0} \cdot k \cdot c_0} - 1 \geq 2^{2 \cdot d \cdot M}$$

since $k \geq K = 2 \cdot d \cdot M + 1$. The induction hypothesis yields for $i < k - 1$

$$\lfloor \varepsilon \cdot \log(n_i) \rfloor \geq \lfloor \varepsilon \cdot 2^i \cdot d \cdot M \rfloor = 2^i \cdot M \,.$$

Thus by Lemma 4.2.1 we have for the induction step

$$n_{i+1} \geq R(\lfloor \varepsilon \cdot \log(n_i) \rfloor, c_0 \cdot M \cdot d \cdot 2 - 1) \geq 2^{\frac{1}{c_0} \cdot (c_0 \cdot M \cdot d \cdot 2 - 1) \cdot 2^i \cdot M} \geq 2^{2^{i+1} \cdot d \cdot M}$$

and hence it holds that $R^*_{\hat{f}_\varepsilon}(k, c_0 \cdot M \cdot d \cdot 2) > n = n_{k-1} \geq 2^{2^{k-1} \cdot d \cdot M} \geq 2^{2^k \cdot d}$ since $M \geq 2$. □

In the proof above, the growth rate of R is high enough to compensate for the logarithms and ensures that n_i grows faster than the double exponential function. We can furthermore show that $R^*_{\hat{f}_\varepsilon}$ is Ackermannian:

Lemma 4.2.3. *With the notation of Lemma 4.2.2 we have for all $m \in \mathbb{N}$:*

$$R^*_{\hat{f}_\varepsilon}(k, c_0 \cdot d \cdot M \cdot 2 + m) > 2^{d \cdot F_m(k)}$$

Proof. We prove the claim simultaneously for all $k \geq K$, by induction on m. If $m = 0$, it is simply Lemma 4.2.2, since $F_0(k) = 2^k$. Now assume that the claim is true for $m \geq 0$.

Put $n_0 = 0$ and $n_1 = R^*_{\hat{f}_\varepsilon}(k, c_0 \cdot d \cdot M \cdot 2 + m) - 1$. By recursion on $i > 0$ define

$$n_{i+1} = R^*_{\hat{f}_\varepsilon}(\lfloor \varepsilon \cdot \log(n_i) \rfloor, c_0 \cdot d \cdot M \cdot 2 + m) - 1.$$

Finally put $n = n_{k-1}$. We claim that

$$[0, n) \not\to^*_{\hat{f}_\varepsilon} (k)_{c_0 \cdot d \cdot M \cdot 2 + m + 1}.$$

Choose $C_0 : [0, n_1)^2 \to c_0 \cdot d \cdot M \cdot 2 + m$ such that every C_0-homogeneous H satisfies $|H| < \max\{k, \hat{f}_\varepsilon(\min H)\}$. And for each $1 \leq i < k - 1$ choose $C_i : [0, n_{i+1})^2 \to c_0 \cdot d \cdot M \cdot 2 + m$ such that every C_i-homogeneous $H \subseteq [n_0, n_{i+1})$ satisfies $|H| < \max\{\lfloor \varepsilon \cdot \log(n_i) \rfloor, \hat{f}_\varepsilon(\min H)\}$.

Define $C \colon [0, n)^2 \to c_0 \cdot M \cdot d \cdot 2 + m + 1$ as follows:

$$C(u, v) = \begin{cases} C_i(u, v) + 1 & \text{if } n_i \leq u < v < n_{i+1}, \\ 0 & \text{otherwise.} \end{cases}$$

Let H be C-homogeneous. If the color of H is 0 then $|H \cap [n_i, n_{i+1})| \leq 1$ for every $i < k - 1$, hence we have $|H| \leq k - 1 < k$. If the color of H is greater than 0 then $H \subseteq [n_j, n_{j+1})$ for some j and H is homogeneous for C_j. If $j = 0$ then $|H| < \max\{k, \hat{f}_\varepsilon(\min H)\}$. If $j > 0$ then

$$|H| < \max\{\varepsilon \cdot \log(n_j), \hat{f}_\varepsilon(\min H)\} \leq \hat{f}_\varepsilon(\min H) \,.$$

By induction on $1 \leq i < k$ we show $n_i \geq 2^{d \cdot F_m^{(i)}(k)}$. First note that we have

$$n_1 = R^*_{\hat{f}_\varepsilon}(k, c_0 \cdot d \cdot M \cdot 2 + m) - 1 \geq 2^{d \cdot F_m(k)}$$

by the main induction hypothesis. Now the induction hypothesis yields for $i \geq 1$ that $\varepsilon \cdot \log(n_i) \geq F_m^{(i)}(k) \geq k \geq K$. Hence again the main induction hypothesis yields

$$n_{i+1} = R^*_{\hat{f}_\varepsilon}(\lfloor \varepsilon \cdot \log(n_i) \rfloor, c_0 \cdot d \cdot M \cdot 2 + m) - 1 \geq 2^{d \cdot F_m^{(i+1)}(k)}$$

Therefore $R^*_{\hat{f}_\varepsilon}(k, c_0 \cdot d \cdot M \cdot 2 + m + 1) > n = n_{k-1} \geq 2^{d \cdot F_m^{(k-1)}(k)} = 2^{d \cdot F_{m+1}(k)}$. □

Theorem 4.2.4. *Suppose $B : \mathbb{N} \to \mathbb{N}$ is nonzero, nondecreasing and unbounded, and $B(i) \leq F(i)$ for all i. Let c_0 and M be as in Lemma 4.2.1. Then*

$$N(d) = R^*_{f_B}(2 \cdot d \cdot M + 1, c_0 \cdot d \cdot M \cdot 2 + d) > B(d)$$

for all $d \geq 4$.

Proof. Assume to the contrary that it is not so for some $d \geq 4$. Then for any $i \leq N(d)$ we have

$$\frac{\log(i)}{B^{-1}(i)} \geq \frac{1}{d} \cdot \log(i)$$

since $B^{-1}(i) \leq d$. Set $K_d = 2 \cdot d \cdot M$, $\varepsilon = \frac{1}{d}$, and denote $\hat{f}_\varepsilon(i) = \varepsilon \cdot \log(i)$. Clearly, every f_B-large set for a given coloring $C : [N(d)]^2 \to c_0 \cdot K_d + d$ is also a $\hat{f}_\varepsilon$-large set for C. Thus, we have

$$\begin{aligned} R^*_{f_B}(K_d + 1, c_0 \cdot K_d + d) &\geq R^*_{\hat{f}_\varepsilon}(K_d + 1, c_0 \cdot K_d + d) \\ &\geq F_d(K_d) \\ &> F(d) \\ &\geq B(d) \end{aligned}$$

by Lemma 4.2.3. Contradiction! □

Theorem 4.2.5. *Suppose $B : \mathbb{N} \to \mathbb{N}$ is positive, unbounded and nondecreasing. Then the function $R^*_{f_B}$ is Ackermannian iff B is Ackermannian.*

Proof. Suppose B is Ackermannian. By replacing B with $\min\{B, F\}$, we assume that $B(i) \leq F(i)$ for all $i \in \mathbb{N}$. This is done with no loss of generality, since clearly, if $B'(i) \leq B(i)$ for all $i \in \mathbb{N}$ and $R^*_{f_{B'}}$ is Ackermannian, then $R^*_{f_B}$ is Ackermannian too.

By the previous theorem, $R^*_{f_B}$ composed with the primitive recursive functions $r_1(i) = 2 \cdot i \cdot M + 1$ and $r_1(i) = c_0 \cdot i \cdot M \cdot 2 + i$ is Ackermannian. Therefore, $R^*_{f_B}$ itself is Ackermannian.

Conversely, suppose that B is not Ackermannian, and fix an increasing primitive recursive function f so that for infinitely many $i \in \mathbb{N}$ it holds that $B(i) < f(i)$. For each such i, let $c_i = \max\{c : c \cdot \log(c) \leq i\}$ and let $k_i = B(i)$. By Lemma 4.1.2, it holds that $R(k_i, c_i) + B(c_i \cdot \lceil \log(c_i) \rceil) \to^*_{f_B} (k_i)_{c_i}$. Since $f(i) \geq B(c_i \cdot \lceil \log(c_i) \rceil)$, it holds that $R^*_{f_B}(k_i, c_i) \leq R(k_i, c_i) + f(i)$. This is true for infinitely many c_i and infinitely many k_i. Thus, $R^*_{f_B}$ is not Ackermannian. □

Chapter 5

Phase Transition Threshold of Function Hierarchies

In this chapter we investigate phase transition phenomena that are related to natural subclasses of the class of recursive functions. In particular we take a closer look at the Grzegorczyk hierarchy from the phase transition perspective. For this purpose, let us assume that we are given two functions $g, h : \mathbb{R} \cap [0, \infty) \to \mathbb{R} \cap [0, \infty)$. For $r \in \mathbb{R}$, let $\lfloor r \rfloor$ denote the largest integer not exceeding r.

Define for $x \in \mathbb{N}$

$$\begin{aligned} B(g,h)_0(x) &\triangleq g(x), \\ B(g,h)_{k+1}(x) &\triangleq B(g,h)_k^{\lfloor h(x) \rfloor}(x) \quad \text{i.e., } \lfloor h(x) \rfloor \text{ many iterations,} \\ B(g,h)_\omega(x) &\triangleq B(g,h)_{\lfloor x \rfloor}(x). \end{aligned}$$

We allow here for real number values in the range of $B(g,h)_k$ to avoid messy rounding to integers at every step of the calculation. This would be necessary if we would deal with number-theoretic functions only. We recall that the Ackermann function is defined as $\mathrm{Ack}(n) = B(g,h)_\omega(n)$ where $g(x) = x+1$ and $h = \mathrm{Id}$, and that $\mathrm{A}_i(n) = B(g,h)_i(n)$ is called the ith approximation of the Ackermann function. It is well known (see e.g., [17]) that each approximation A_i is *primitive recursive* and that every primitive recursive function is eventually dominated by some A_i. Thus the Ackermann function eventually dominates every primitive recursive function.

To avoid trivialities we assume that for some $\varepsilon > 0$ we have $g(x) \geq x + \varepsilon$ for all but finitely many x [an iteration of the identity map would, in our context, of course be senseless] and we assume that h is weakly increasing and unbounded. Now, fixing g, one may ask for which h the function $B(g,h)_\omega$ becomes Ackermannian. Similarly, fixing h, one may ask for which g the function $B(g,h)_\omega$ becomes Ackermannian. So in contrast to the situations previously considered, the phase transition depends on two order parameters and we will indicate that the phase transition has a surprisingly rich structure.

In the rest of the chapter we use the following notation to define the function hierarchies we consider.

Notation. Let $|x|$ be the non-negative part of the logarithm function with respect to base two, that is $|x| \triangleq \max\{\log_2(x), 0\}$. Let $|x|_{l+1} \triangleq ||x||_l$ where $|x|_0 \triangleq x$. Then $|\cdot|_l$ is the lth iterate of $|\cdot|$, hence we have $|2_l(x)|_l = x$ and $|x|_2 = ||x|| = |(|x|)|$.

The results in this chapter are taken almost verbatim from [64].

5.1 Iteration Hierarchies for $g(x) = x + 1$

In this section we fix $g(x) = x + 1$ and present a rather sharp threshold on the behavior of such function hierarchies. This particular case is resolved using results presented in Chapter 3; thus besides being interesting in its own right, this phase transition investigation reveals a somewhat surprising intrinsic relation between regressive Ramsey functions and parameterized iteration hierarchies. We note that the results needed for this section appear in Section 3.1 and in Section 3.2.2. The reader is referred to the appropriate results whenever they are used.

Using the notation of Chapter 3, we denote $B(g, x^{1/t})$, where $t \in \mathbb{N}$ is a constant, by (f_t). Namely, $(f_t)_i^j(x) = B(g, h)_i^j(x)$ for all i, j and x, where $g(x) = x + 1$ and $h(x) = x^{1/t}$.

We recall that Claim 3.2.9 asserts that for every $t > 0$ and $n > \max\{4, 3^t, t^t\}$ it holds that

$$(f_t)_{i+t^2+2t+2}(n) > A_i(n).$$

Claim 5.1.1. *For every $i \in \mathbb{N}$ and for every $n \in \mathbb{N}$ such that:*

1. $n > i + (||n||)^2 + 2||n|| + 2$ *and*

2. $\mathrm{Ack}(||n||) > \mathrm{A}_i(n)$

it holds for $h_{\mathrm{Ack}}(n) \triangleq n^{\frac{1}{\mathrm{Ack}^{-1}(n)}}$ *that*

$$B(g, h_{\mathrm{Ack}})_{i+(||n||)^2+2||n||+2}(n) > \mathrm{A}_i(n).$$

Proof. To show that, we examine two cases. First, if it holds that $B(g, h_{\mathrm{Ack}})_{i+(||n||)^2+2||n||+2}(n) \geq \mathrm{Ack}(||n||)$, then we are done by demand 2. Otherwise, we may fix $t = ||n||$ and we have that for all $y \in \{0, \ldots, \mathrm{Ack}(t) - 1\}$ it holds that $y^{\frac{1}{t}} < y^{\frac{1}{\mathrm{Ack}^{-1}(y)}}$. Note that $B(g, h_{\mathrm{Ack}})_{i+t^2+2t+2}$ is non-decreasing since h_{Ack} is non-decreasing. Thus, $B(g, h_{\mathrm{Ack}})_{i+t^2+2t+2}(n) \geq (f_t)_{i+t^2+2t+2}(n)$ and by Claim 3.2.9, the latter term is larger than $A_i(n)$. □

We remark that the choice of $t = ||n||$ is arbitrary and any α^{-1}, such that α is a monotone increasing primitive recursive function and $\alpha(x) > x^x$ for large enough x, would do the job.

Theorem 5.1.2. *Let $g(x) = x + 1$ and $h_\alpha(x) = x^{\frac{1}{B(g,id)_\alpha^{-1}(x)}}$. Then $B(g, h_\alpha)_\omega$ is Ackermannian iff $\alpha = \omega$.*

Proof. The 'if' direction is in fact the claim that if $h_\alpha(x) = x^{\frac{1}{\mathrm{Ack}^{-1}(x)}}$ then $B(g, h_\alpha)_\omega$ eventually grows faster than any primitive recursive function. It would suffice to show that for every $i \in \mathbb{N}$, there exists x_0 such that for all $x > x_0$, it holds that $B(g, h_\alpha)_\omega(x) > \mathrm{A}_i(x)$. Now, this is a direct corollary of Claim 5.1.1, since it is clear that for every such i there exists some $x_0 \in \mathbb{N}$ such that for all $x > x_0$ it holds that $x \geq i + (||x||)^2 + 2||x|| + 2$ and that $\mathrm{Ack}(||x||) > \mathrm{A}_i(x)$ and thus $B(g, h_\alpha)_x(x) \geq B(g, h_\alpha)_{i+(||x||)^2+2||x||+2}(x)$, which by Claim 5.1.1 is larger than $\mathrm{A}_i(x)$. In other words, for every primitive recursive function f, $B(g, h_\alpha)_x(x)$ eventually dominates f.

The 'only if' direction is the claim that if $\alpha = i$ for some $i \in \mathbb{N}$, and therefore $h_\alpha(x) = x^{\frac{1}{\mathrm{A}_i^{-1}(x)}}$, then $B(g, h_\alpha)_\omega(x)$ is not Ackermannian in terms of x. Note this implies the same for any h_α of the form $h_\alpha(x) = x^{\frac{1}{\beta^{-1}(x)}}$ where β is a non-decreasing unbounded primitive recursive function. To show this direction, for $\alpha = i > 3$ and $h_\alpha(x) = x^{\frac{1}{\mathrm{A}_i^{-1}(x)}}$, fix $h_\beta(x) = 4(h_\alpha(x))^2 = x^{\frac{1}{\beta^{-1}(x)}}$ where $\beta^{-1}(x) = \frac{|x| \, \mathrm{A}_i^{-1}(x)}{2|x|+2 \, \mathrm{A}_i^{-1}(x)}$. We again refer to Chapter 3. Corollary 3.1.3 states that the h_β-regressive Ramsey number $R_{h_\beta}^{\mathrm{reg}}(k)$ is primitive recursive in k since A_i is primitive recursive. On the other hand, Corollary 3.2.22 asserts that if $B(g, h_\alpha)_\omega(k)$ is Ackermannian in k, using the function $\mu_{h_\beta}(k) = k^k$, we may obtain an Ackermannian lower bound also for $R_{h_\beta}^{\mathrm{reg}}(k)$, but this would be a contradiction. For the case of $\alpha \leq 3$, observe that $h_\alpha \leq h_{\alpha+1}$ and thus $B(g, h_\alpha)_\omega(k) \leq B(g, h_{\alpha+1})_\omega(k)$. □

5.2 Slow Growing Iteration Hierarchies

For the rest of this section let $F_0(x) \triangleq 2^x$ and $F_{k+1}(x) \triangleq F_k^x(x)$. Then F_k is primitive recursive (in each k). Further let $F(x) \triangleq F_x(x)$. Then F is a slight variant of the Ackermann function, hence Ackermannian and of course not primitive recursive. In addition let $2_l(x) \triangleq F_0^l(x)$.

For the rest of the chapter fix $\varepsilon > 0$, let $g_0(x) = x + \varepsilon$ and define recursively $g_{k+1}(x) = 2^{g_k(|x|)}$. Then

$$g_l(x) = 2_l(|x|_l + \varepsilon).$$

These scaling functions grow faster and faster when l becomes larger but no g_l is of exponential growth. The following result classifies slow-growing iteration hierarchies for a rather large class of order parameters.

Theorem 5.2.1. *Let $1 \geq \varepsilon > 0$ and let d be a natural number.*

Define $h[d,l](x) \triangleq |x|_l^{\frac{1}{F_d^{-1}(|x|_l)}}$ and

$$B[d,l]_k(x) \triangleq B(g_l, h[d,l])_k(x).$$

Let $C = \max\left\{2_l(F_d(2^{k+2}))\right\}$. Then for all $x \geq C$ and all $i \leq |x|_l^{\frac{1}{F_d^{-1}(|x|_l)}}$ we have

$$B[d,l]_k^i(x) \leq 2_l(|x|_l + |x|_l^{\frac{2^{k+1}}{F_d^{-1}(|x|_l)}} \cdot i).$$

Hence the diagonal function $B[d,l]$ is primitive recursive.

Proof. Since g_l and hence $B[d,l]_k$ are monotone in ε we may assume that $\varepsilon = 1$. We prove the claim by main induction on k. If $k = 0$ then $B[d,l]_0^i(x) = g_l^i(x)$. We prove the claim by subsidiary induction on i. Assume first that $i = 1$. We prove the claim by another subsidiary induction on l. Assume $l = 0$. Then for $x \geq C$:

$$\begin{array}{rcl} B[d,0]_0^1(x) & = & g_0(x) \\ & = & x+1 \\ & \leq & 2_0(|x|_0 + |x|_0^{\frac{2^1}{F_d^{-1}(|x|_0)}}). \end{array}$$

Assume now $l > 0$. Then the induction hypothesis for $l-1$ yields for $x \geq C$:

$$\begin{array}{rcl} B[d,l]_0^1(x) & = & g_l(x) \\ & = & 2^{g_{l-1}(|x|)} \\ & \leq & 2^{2_{l-1}(||x||_{l-1}+||x||_{l-1}^{\frac{2}{F_d^{-1}(||x||_{l-1})}})} \\ & = & 2_l(|x|_l + |x|_l^{\frac{2}{F_d^{-1}(|x|_l)}}). \end{array}$$

Now consider the case $1 \leq i < |x|_l^{\frac{1}{F_d^{-1}(|x|_l)}}$. Then we obtain by the subsidiary induction hypothesis

$$\begin{array}{rcl} B[d,l]_0^{i+1}(x) & = & B[d,l]_0\big(B[d,l]_0^i(x)\big) \\ & \leq & B[d,l]_0(2_l\big(|x|_l + |x|_l^{\frac{2}{F_d^{-1}(|x|_l)}} \cdot i)\big) \\ & = & 2_l(|2_l(|x|_l + |x|_l^{\frac{2}{F_d^{-1}(|x|_l)}} \cdot i)|_l + 1) \\ & = & 2_l(|x|_l + |x|_l^{\frac{2}{F_d^{-1}(|x|_l)}} \cdot i + 1) \\ & \leq & 2_l\big(|x|_l + |x|_l^{\frac{2}{F_d^{-1}(|x|_l)}} \cdot (i+1)\big) \end{array}$$

since by assumption $x \geq C = 2_l(F_d(2^{k+2}))$.

Now assume that $k > 0$. We prove the claim by subsidiary induction on i. If $i = 1$ then the main induction hypothesis yields

$$\begin{array}{rcl} B[d,l]_k(x) & = & B[d,l]_{k-1}^{\left\lfloor |x|_l^{\frac{1}{F_d^{-1}(|x|_l)}} \right\rfloor}(x) \\ & \leq & 2_l\Big(|x|_l + |x|_l^{\frac{2^k}{F_d^{-1}(|x|_l)}} \cdot \left\lfloor |x|_l^{\frac{1}{F_d^{-1}(|x|_l)}} \right\rfloor\Big) \\ & \leq & 2_l\Big(|x|_l + |x|_l^{\frac{2^{k+1}}{F_d^{-1}(|x|_l)}}\Big). \end{array}$$

If $1 \leq i < |x|_l^{\frac{1}{F_d^{-1}(|x|_l)}}$ then we obtain by the subsidiary induction hypothesis

$$\begin{array}{rcl} B[d,l]_k^{i+1}(x) & = & B[d,l]_k(B[d,l]_k^i(x)) \\ & \leq & B[d,l]_k(2_l(|x|_l + |x|_l^{\frac{2^{k+1}}{F_d^{-1}(|x|_l)}} \cdot i)). \end{array}$$

Now set $y = 2_l(|x|_l + |x|_l^{\frac{2^{k+1}}{F_d^{-1}(|x|_l)}} \cdot i)$. Then we obtain from the main induction hypothesis and $i < |x|_l^{\frac{1}{F_d^{-1}(|x|_l)}}$ that

$$\begin{array}{rcl} B[d,l]_k^{i+1}(x) & \leq & B[d,l]_{k-1}^{\left\lfloor |y|_l^{\frac{1}{F_d^{-1}(|y|_l)}} \right\rfloor}(y) \\ & \leq & 2_l\Big(|y|_l + |y|_l^{\frac{2^k}{F_d^{-1}(|y|_l)}} \cdot \left\lfloor |y|_l^{\frac{1}{F_d^{-1}(|y|_l)}} \right\rfloor\Big) \\ & \leq & 2_l\Big(|x|_l + |x|_l^{\frac{2^{k+1}}{F_d^{-1}(|x|_l)}} \cdot i + |y|_l^{\frac{2^k+1}{F_d^{-1}(|y|_l)}}\Big). \end{array}$$

The claim would now follow from

$$|y|_l^{\frac{2^k+1}{F_d^{-1}(|y|_l)}} \leq |x|_l^{\frac{2^{k+1}}{F_d^{-1}(|x|_l)}}.$$

Since $F_d^{-1}(|x|_l + |x|_l^{\frac{2^{k+1}}{F_d^{-1}(|x|_l)}} \cdot i) \geq F_d^{-1}(|x|_l)$ and $i < |x|_l^{\frac{1}{F_d^{-1}(|x|_l)}}$ this would follow from

$$\Big(|x|_l + |x|_l^{\frac{2^{k+1}+1}{F_d^{-1}(|x|_l)}}\Big)^{\frac{2^k+1}{F_d^{-1}(|x|_l)}} \leq |x|_l^{\frac{2^{k+1}}{F_d^{-1}(|x|_l)}}$$

hence from

$$|x|_l + |x|_l^{\frac{2^{k+1}+1}{F_d^{-1}(|x|_l)}} \leq |x|_l^{\frac{2^{k+1}}{2^k+1}}.$$

This finally follows from the assumption that $x \geq C = 2_l(F_d(2^{k+2}))$. □

5.3 Fast Growing Iteration Hierarchies

In this section we show that slightly faster-growing functions in the role of the functions $h[d,l]$ from Theorem 5.2.1 yield Ackermannian growth of the induced iteration hierarchies. Let us recall the definition of the Ackermann hierarchy from Section 2.2. We put $A_0(x) \triangleq x+1$ and $A_{k+1}(x) \triangleq A_k^x(x)$. Thus, if we put $\mathrm{Ack}(x) \triangleq A_x(x)$, then Ack is the Ackermann function that eventually dominates every primitive recursive function. Further recall that our scale functions are defined as follows: $g_0(x) = x + \varepsilon$ and $g_{k+1}(x) = 2^{g_k(|x|)}$. Let us further assume from now on that $d > 0$.

Let us fix constants $C_{k,l}$ for $k > 0$ and $l \geq 0$ such that

$$\left\lfloor |x|_l^{\frac{1}{d}} \right\rfloor \cdot |x|_l^{\frac{k-1}{d}} \geq |x|_l^{\frac{k}{d}} \cdot \frac{1}{2}$$

for $x \geq C_{k,l}$. We may assume that the function $k \mapsto C_{k,l}$ is primitive recursive in k for any fixed l.

Theorem 5.3.1. *Assume $1 \geq \varepsilon > 0$ and let d be a natural number. Let*

$$C[d] = \max\left\{C_{3\cdot d,l}, 2_l\left(\left\lfloor \frac{2^{3\cdot d}}{\varepsilon}\right\rfloor + 1\right)\right\}.$$

Define

$$h[\![d,l]\!](x) = \sqrt[d]{|x|_l}$$

and

$$B[\![d,l]\!]_k(x) = B(g_l, h[\![d,l]\!])_k(x).$$

Then we have

$$B[\![d,l]\!]_{3\cdot d+i+1}(2_l(x^d)) \geq 2_l((A_i(x))^d)$$

for $x \geq C[d]$.

Proof. Recall that that $g_l(x) = 2_l(\varepsilon + |x|_l)$. By induction on i one verifies that $B[\![d,l]\!]_0^i(x) = g_l^i(x) = 2_l(\varepsilon \cdot i + |x|_l)$. Let $\varepsilon_k = \frac{\varepsilon}{2^k}$, we now claim that the following equation holds,

$$B[\![d,l]\!]_k^i(x) \geq 2_l(\varepsilon_k \cdot i \cdot |x|_l^{\frac{k}{d}} + |x|_l) \tag{5.1}$$

for $i, k \geq 1$ and $x \geq C_{k,l}$. We prove (Equation (5.1)) by main induction on k and subsidiary induction on i. Assume that $k = 1$. Then we obtain for $i = 1$ that

$$\begin{array}{rcl} B[\![d,l]\!]_1^1(l)(x) & = & B[\![d,l]\!]_0^{\left\lfloor |x|_l^{\frac{1}{d}} \right\rfloor}(x) \\ & \geq & 2_l(\varepsilon \cdot \left\lfloor |x|_l^{\frac{1}{d}} \right\rfloor + |x|_l) \\ & \geq & 2_l(\varepsilon_1 \cdot |x|_l^{\frac{1}{d}} + |x|_l) \end{array}$$

since $x \geq C_{1,l}$. The subsidiary induction hypothesis yields

$$\begin{array}{rcl} B[\![d,l]\!]_1^{i+1}(x) & = & B[\![d,l]\!]_1^1(B[\![d,l]\!]_1^i(x)) \\ & \geq & B[\![d,l]\!]_1^1(2_l(\varepsilon_1 \cdot i \cdot |x|_l^{\frac{1}{d}} + |x|_l)) \\ & \geq & 2_l(\varepsilon_1 \cdot (|2_l(\varepsilon_1 \cdot i \cdot |x|_l^{\frac{1}{d}} + |x|_l)|_l)^{\frac{1}{d}} + |2_l(\varepsilon_1 \cdot i \cdot |x|_l^{\frac{1}{d}} + |x|_l)|_l) \\ & \geq & 2_l(\varepsilon_1 \cdot |x|_l^{\frac{1}{d}} + \varepsilon_1 \cdot i \cdot |x|_l^{\frac{1}{d}} + |x|_l). \end{array}$$

Assuming Equation (5.1) for k we show it for $k+1$ by subsidiary induction on i as follows: First let $i = 1$. Then

$$\begin{array}{rcl} B[\![d,l]\!]_{k+1}(x) & = & B[\![d,l]\!]_k^{\left\lfloor |x|_l^{\frac{1}{d}} \right\rfloor}(x) \\ & \geq & 2_l(\varepsilon_k \cdot \left\lfloor |x|_l^{\frac{1}{d}} \right\rfloor \cdot |x|_l^{\frac{k}{d}} + |x|_l) \\ & \geq & 2_l(\varepsilon_{k+1} \cdot |x|_l^{\frac{k+1}{d}} + |x|_l) \end{array}$$

since $x \geq C_{k+1,l}$. For the induction step of the subsidiary induction we obtain

$$\begin{array}{rl} & B[\![d,l]\!]_{k+1}^{i+1}(x) = B[\![d,l]\!]_{k+1}(B[\![d,l]\!]_{k+1}^i(x)) \\ \geq & B[\![d,l]\!]_{k+1}(2_l(\varepsilon_{k+1} \cdot i \cdot |x|_l^{\frac{k+1}{d}} + |x|_l)) \\ \geq & 2_l(\varepsilon_{k+1} \cdot (|2_l(\varepsilon_{k+1} \cdot i \cdot |x|_l^{\frac{k+1}{d}} + |x|_l)|_l)^{\frac{k+1}{d}} + |2_l(\varepsilon_{k+1} \cdot i \cdot |x|_l^{\frac{k+1}{d}} + |x|_l)|_l) \\ \geq & 2_l(\varepsilon_{k+1} \cdot |x|_l^{\frac{k+1}{d}} + \varepsilon_{k+1} \cdot i \cdot |x|_l^{\frac{k+1}{d}} + |x|_l). \end{array}$$

Equation (5.1)) yields $B[\![d,l]\!]_{3 \cdot d}(x) \geq 2_l(|x|_l^2)$ for $x \geq C[d]$.

By induction on i this yields

$$B[\![d,l]\!]_{3 \cdot d}^i(x) \geq 2_l(|x|_l^{2^i}) \tag{5.2}$$

for $x \geq C[d]$.

We claim now that

$$B[\![d,l]\!]_{d \cdot 3+i+1}(2_l(x^d)) \geq 2_l((A_i(x))^d)$$

for $x \geq C[d]$. The proof is by induction on i. For $i = 0$ we find by (Equation (5.2))

$$\begin{array}{rcl} B[\![d,l]\!]_{3 \cdot d+1}(2_l(x^d)) & \geq & B[\![d,l]\!]_{3 \cdot d}^x(2_l(x^d)) \\ & \geq & 2_l((|2_l(x^d)|_l)^{2^x}) \\ & \geq & 2_l((A_0(x))^d). \end{array}$$

Assuming the claim for i we obtain it for $i+1$ as follows:

$$\begin{aligned} B[\![d,l]\!]_{3\cdot d+1+i}(2_l(x^d)) &\geq B[\![d,l]\!]^x_{3\cdot d+i}(2_l(x^d)) \\ &\geq 2_l((A_i^x(x))^d) \\ &= 2_l((A_{i+1}(x))^d). \end{aligned}$$

□

Theorem 5.3.2. *Assume* $1 \geq \varepsilon > 0$. *Let* $C[d] = \max\left\{C_{3\cdot d,l}, 2_l\left(\left\lfloor \frac{2^{3\cdot d}}{\varepsilon_k}\right\rfloor + 1\right)\right\}$.

Define $h[\![l]\!]^\star(x) = |x|_l^{\frac{1}{\mathrm{Ack}^{-1}(x)}}$. *Let*

$$B[\![l]\!]^\star_k(x) = B(g_l, h[\![l]\!]^\star)_k(x)$$

and

$$B[\![l]\!]^\star(x) = B[\![d,l]\!]^\star_{\lfloor x\rfloor}(x).$$

Then we have

$$B[\![l]\!]^\star(2_l((4\cdot d + C[d])^d)) > \mathrm{Ack}(d).$$

Hence $B[\![l]\!]^\star$ *is not primitive recursive.*

Proof. Assume for a contradiction that $\mathrm{Ack}(d) \geq B[\![l]\!]^\star(2_l((4\cdot d+C[d])^d))$. Then for any $i \leq B[\![l]\!]^\star_{4\cdot d+C[d]}(2_l((4\cdot d + C[d])^d))$ we have $\mathrm{Ack}^{-1}(i) \leq d$ hence $|i|_l^{\frac{1}{d}} \leq |i|_l^{\frac{1}{\mathrm{Ack}^{-1}(i)}}$ and therefore by Theorem 5.3.1

$$\begin{aligned} B[\![l]\!]^\star(2_l((4\cdot d + C[d])^d)) &\geq B[\![d,l]\!]^\star_{4\cdot d+C[d]}(2_l(4\cdot d + C[d])^d) \\ &\geq B[\![d,l]\!]_{4\cdot d+C[d]}(2_l(4\cdot d + C[d])^d) \\ &> 2_l(A_d(4\cdot d + C[d]))^d \\ &> \mathrm{Ack}(d). \end{aligned}$$

Contradiction! Hence $B[\![l]\!]^\star$ is not primitive recursive since $d \mapsto C[d]$ is primitive recursive. □

It seems plausible that Theorems 5.2.1, 5.3.1 and 5.3.2 hold for all start functions g_l where $x + \varepsilon \leq g_0(x) \leq x + x^c$ for some fixed $c < 1$ and the same functions $h(d)_l$ and $h(l)^\star$. So we expect that our phase transition results will be structurally stable under small perturbations of the starting function g.

For the record, let us consider the situation when one starts with an exponential or double exponential function. This leads rather quickly to Ackermannian growth

Theorem 5.3.3. 1. *Let* $g(x) = 2^x$ *and* $h(x) = |x|_k$. *Then* $B(g,h)_\omega$ *is Ackermannian.*

2. *Let* $g(x) = 2^{2^x}$ *and* $h(x) = \min\{l : |x|_l \leq 1\}$. *Then* $B(g,h)_\omega$ *is Ackermannian.*

Proof. 1. By induction on k one easily shows $B(g,h)_k(2_k(x)) \geq 2_k(A_k(x))$.
2. By induction on k one easily shows $B(g,h)_k(2_k(x)) \geq 2_{A_k(x)}(A_k(x))$.

□

Chapter 6

Some Background in Private Data Analysis

Research in the area of private data analysis is mainly concerned with the collision of two interests that emerges when dealing with large sets of sensitive individual data. On the one hand, it may be highly beneficial in many ways to analyze these large data sets; on the other hand it is usually sensible and even essential to require that the privacy of the individual be preserved. Real life examples of such scenarios are abundant and large collections of individual data records are collected and maintained by statistical agencies such as the U.S. Census Bureau, health care organizations, financial organizations, search engines. There are two main questions posed by this scenario. The first is *what* analyses are both useful and privacy preserving. The second question, is *how* to compute such analyses in a given setup. We continue a line of rigorous investigation of these questions, which started in the work of Dinur and Nissim [23].

In this chapter we give some notations and basic definitions from the area of private data analysis, and survey some of the basic results mostly concerned with answering the what question. In Chapter 7 we combine the discussions regarding these two questions. While Chapter 7 is almost completely self-contained, here we aim at presenting a slightly broader view of the abstract model of computation, of which the distributed and local models considered in Chapter 7 are possible realizations.

For a much broader view and an in depth consideration of the main questions and results in the area of private data analysis, we refer the reader to [62].

6.1 The Abstract Model – A Statistical Database

To model private computation we use the setting of a *statistical database*, containing records of sensitive information accompanied by some algorithmic mechanism. The users of the database can issue queries about the records stored in the database and, in turn, get the result of the database mechanism applied to the given query. We think of the result of the mechanism as being some approximation of the value of the issued query. For a more formal definition of the model, we first recall the definition of a randomized function (see Figure 6.1).

Definition 6.1.1. *Let D, D_R, and R be sets. An n-ary* randomized function *is a function $\hat{f} : D^n \times D_R \to R$, where D is the domain of $\hat{f}$ and D_R is the set of random inputs. For $\mathbf{x} = (x_1, \ldots, x_n) \in D^n$ we usually write $\hat{f}(\mathbf{x})$ with the underlying convention that $\hat{f}(x_1, \ldots, x_n) = \hat{f}(x_1, \ldots, x_n, r)$, where r is uniformly selected from D_R. Following this convention, we also usually omit D_R from the notation and write $\hat{f} : D^n \to R$.*

Definition 6.1.2. *[The statistical database model] Let $\mathcal{D}$ and R be two sets. An n-entry* database $\mathbf{x} = (x_1, \ldots, x_n)$ *is an element in $\mathcal{D}^n$. A* statistical database *is a pair $\langle \mathbf{x}, \mathcal{S} \rangle$, where $\mathbf{x}$ is a database and $\mathcal{S}$ is some (abstract) randomized algorithmic mechanism. A* query *is a function $q : \mathcal{D}^n \to R$. The interface between users and the database is defined by an interplay of queries and responses. Upon a query q issued by a user, the database mechanism responds with $\mathcal{S}(\mathbf{x}, q)$. We will usually omit $\mathcal{S}$ from the notation and simply refer to the database $\mathbf{x}$.*

By way of example, consider a setup where a query $q = (q_1, \ldots, q_n)$ is a sequence of predicates (i.e., $q_i(x_i) \in \{0, 1\}$) and the mechanism replies with the vector $(y_1, \ldots, y_n)$, where $y_i = q_i(x_i)$ w.p. α and $y_i = 1 - q_i(x_i)$ w.p. $1 - \alpha$. This mechanism is known as the randomized response mechanism.

6.1.1 Interactive vs. non-interactive interplay

The exchange of queries and responses between a user and the database may be limited to a single round, in which the user issues a single query answered by a single response of the database. We call this type of interplay *non-interactive*. We sometimes think of the non-interactive model as a setup in which the database is not presented with any specific query, but rather releases a *sanitization* of the information. Thereafter, users can compute queries freely on this sanitized version of $\mathbf{x}$ (usually from a certain class of queries). For example, this is the way the U.S. Census Bureau operates for most releases. However, see "U.S. factfinder" for an interactive service.

Alternatively, an *interactive* interplay consists of multiple rounds. In each round the user issues a new (adaptive) query to the the database and gets an

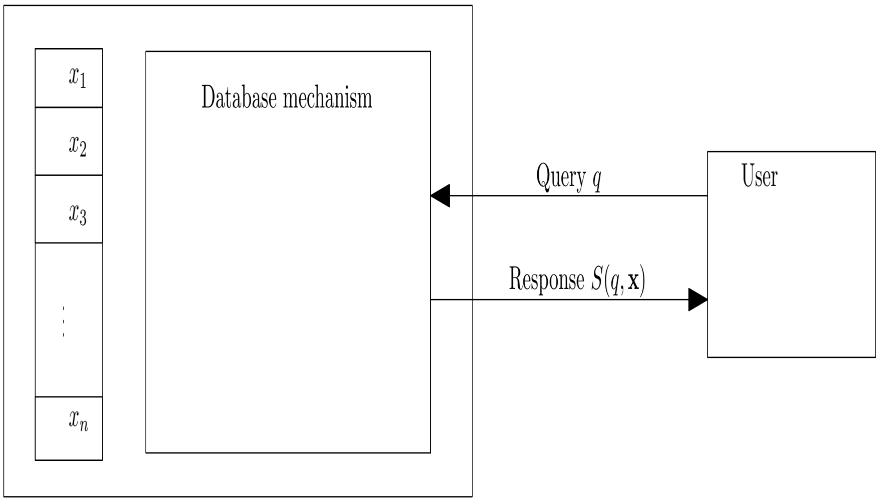

Figure 6.1: The statistical database model.

appropriate response. In this case we view the concatenation of all responses as the output of the database.

6.1.2 Communication models

The setting of a statistical database is sometimes an abstraction of the interface between the database and its users and it applies to different setups. Specifically, the database mechanism may be realized in different ways. We next list a few possible models of communication that imply different realizations of the database mechanism.

The centralized (global) model. In the centralized model the database mechanism is controlled by a trusted entity, which has access to all records of the database. Upon an issued query q, this entity can simply apply $\mathcal{S}$ to q and x (see Figure 6.2).

The local model. In the local model we have n parties $p_1, \ldots, p_n$, where party p_i holds an input x_i. Similarly to the centralized model, the interface of the user to the database mechanism is through a central entity C with the difference that this entity is no longer considered to be trusted (see Figure 6.3). Upon a query q, this C can issue a single query q_i to each party p_i and receive a sanitized response $\mathcal{S}_i(x_i, q_i)$ from party p_i. Thereafter, C applies some algorithm G to $(\mathcal{S}_1(x_1, q_1), \ldots, \mathcal{S}_n(x_n, q_n))$ and returns the output to the user as the response of the query q. This setup is called the *local non-interactive* model.

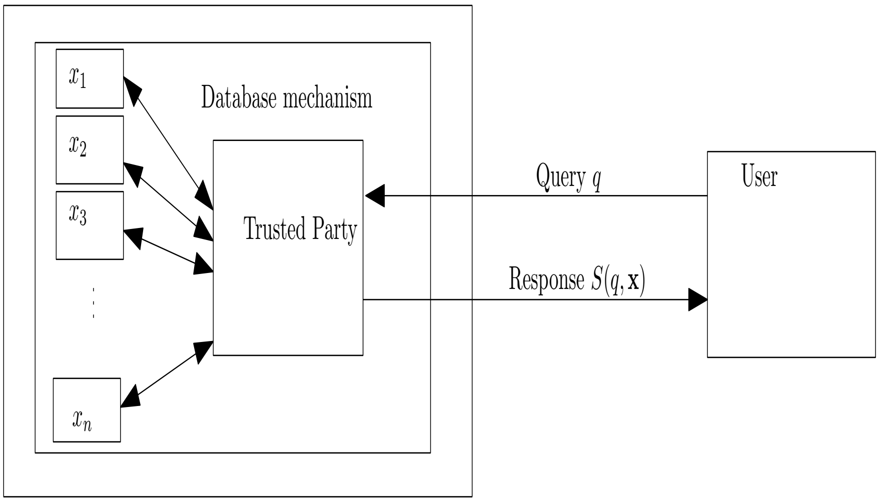

Figure 6.2: The global model.

Alternatively, upon a query q, there can be multiple rounds, where in each round C issues a query q_i to each party p_i and receives an appropriate response. The query C sends to p_i at each round depends on answers C got from all parties in previous rounds. Finally, at the end of this execution, C applies some algorithm G to the communication transcript it viewed and returns the output to the user as the response of the query q. This setup is called the *local interactive* model.

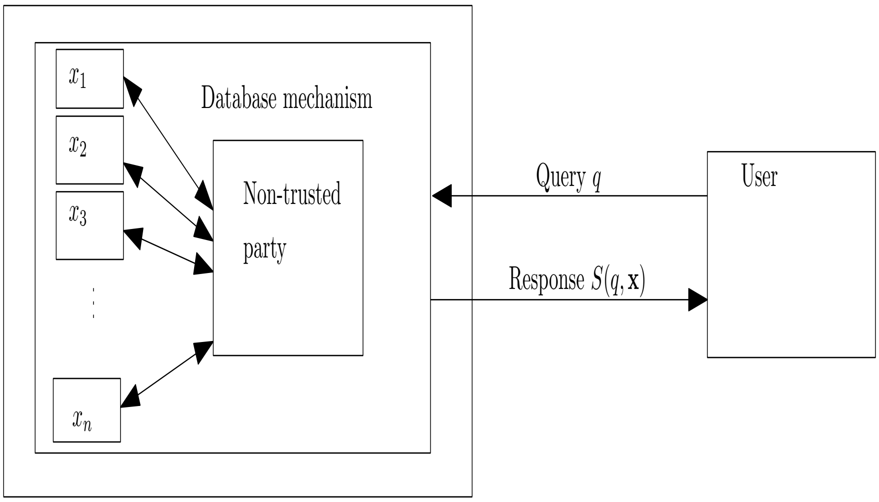

Figure 6.3: The local model.

The distributed model. In the distributed model there are n parties, each holding an input x_i. The parties engage in a protocol and interact via some underlying communication network. For a given query q of a user, we think of $\mathcal{S}(\mathbf{x}, q)$ as the functionality that should be computed by the protocol (see Figure 6.4).

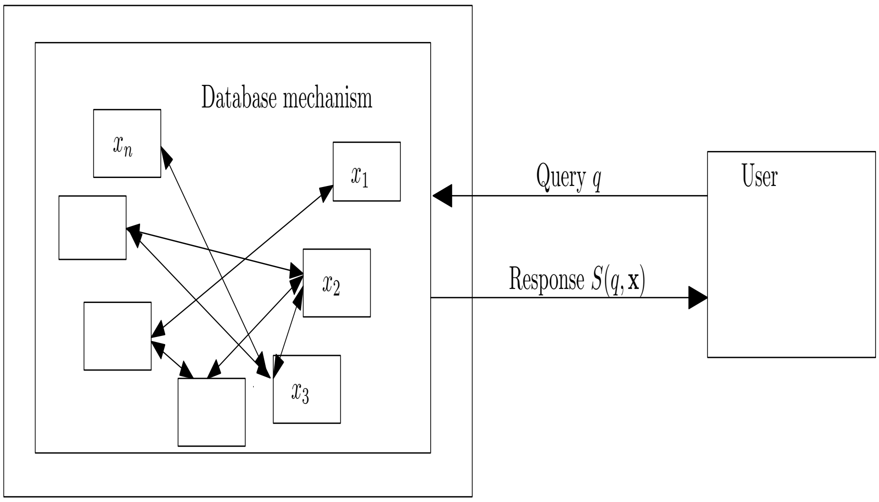

Figure 6.4: The distributed model.

Remark 6.1.3 (Interactive protocols). *In Chapter 7 we discuss interactive protocols. We stress that an interactive protocol is conceptually different from the interactive interplay discussed earlier. In interactive interplay a user poses multiple queries to the database and these queries need not be related. In contrast, a distributed protocol is merely a tool for implementing the role of the database mechanism in the interplay between the database and the user. Thus, the goal of a protocol (interactive or non-interactive) is to compute a single query q posed by the user to the mechanism (possibly, out of a sequence of interactive interplay).*

6.2 Differential Privacy

We use the definition of *differential privacy* suggested in [27] to capture the notion of individual privacy. The privacy is defined to be a property of the database mechanism (rather than, say, the output of the computation or the knowledge of the adversary). Informally, we require that a change of any single entry in the database may only slightly change the distribution of the responses of the database seen by the user (i.e., the view of a possible adversary). Define the *Hamming distance* between two databases $\mathbf{x}, \mathbf{x}'$ as

$$\mathrm{d}_H(\mathbf{x}, \mathbf{x}') = |\{i : x_i \neq x'_i\}|.$$

We say that two databases x, x' are a *neighboring pair* if they differ in exactly one entry, i.e., $\mathrm{d}_H(\mathrm{x}, \mathrm{x}') = 1$. For simplicity, in the definition below (and in the rest of this chapter), we think of a fixed query q and denote $\mathcal{S}(\cdot) = \mathcal{S}(\cdot, q)$, i.e., the output of the mechanism of the database is a (randomized) function of only the database x.

Definition 6.2.1 (ε-differential privacy [27]). *Let $\hat{f} : \mathcal{D}^n \to R$ be a randomized function (an analysis). We say that $\hat{f}$ is* ε-differentially private *if for all neighboring vectors x, x', and for all possible sets of outcomes $\mathcal{V} \subseteq R$ it holds that*

$$\Pr[\hat{f}(\mathrm{x}) \in \mathcal{V}] \le e^{\varepsilon} \cdot \Pr[\hat{f}(\mathrm{x}') \in \mathcal{V}]. \tag{6.1}$$

The probability is taken over the randomness of $\hat{f}$.

We say that a mechanism $\mathcal{S}$ is ε-differentially private if the randomized function it computes is ε-differentially private.

One way to understand the definition is as a mental game, in which we let an adversary pick i and pick all entries in the database except for x_i; we fix x_i and apply the mechanism to the database, and let the adversary try to distinguish which of the two values of x_i we chose. Let p be the probability that the adversary succeeds; we say that $p - \frac{1}{2}$ is the advantage of the adversary (over an adversary that simply guesses by tossing a fair coin). The definition above says that the advantage will be ε-small (where ε is the privacy parameter). This seems to be a very strict notion of privacy.

Surprisingly, some powerful techniques exist for constructing analyses that yield useful outcome, and yet preserve differential privacy. In the next section we present one basic (and simple) technique, which belongs to a class of techniques for constructing analyses via output perturbation. For more on this class and on other techniques, see, e.g., [30, 9, 27, 63, 5, 59, 10].

Before moving on to describing ways for constructing differentially private analyses, let us make a few remarks about Definition 6.2.1 and mention some nice properties it entails (for a deeper consideration of this definition the reader is referred to [62]).

A relaxed privacy definition. A natural relaxation of Definition 6.2.1 allows for events occuring with negligible probability, for which the definition does not hold (i.e., the ratio between probabilities of these events occurring with some neighboring inputs is not bounded by e^{ε}). The next two examples give some motivation to this relaxation. Let us first recall some properties of the Laplace distribution $\mathrm{Lap}(\lambda)$ with mean $\mu = 0$ and variance $2\lambda^2$. Denote by $h(\cdot)$ the probability density function of this distribution and by $H(\cdot)$ the cumulative distribution function. We have,

$$h(t) = \frac{1}{2\lambda} e^{-\frac{|t|}{\lambda}} \qquad \text{and} \qquad H(t) = 0.5\left[1 + \mathrm{sgn}(t)\left(1 - \exp(-|t|/\lambda)\right)\right].$$

The following holds for all t, t':

$$\frac{h(t)}{h(t')} = \frac{e^{\frac{-|t|}{\lambda}}}{e^{\frac{-|t'|}{\lambda}}} = e^{\frac{|t'|-|t|}{\lambda}} \leq e^{\frac{|t-t'|}{\lambda}} \ , \tag{6.2}$$

where the inequality follows from the triangle inequality.

Example 6.2.2. *Consider a mechanism that given $x \in \{0,1\}$ outputs $x + Y$ where Y is sampled according to* $\text{Lap}(1/\varepsilon)$*. For every $x, x' \in \{0,1\}$ it holds by Equation (6.2) that $\frac{h_x(v)}{h_{x'}(v)} \leq e^{\varepsilon}$. It is easy to see that this implies the requirement of Definition 6.2.1. Hence, this mechanism is ε-differentially private.*

We remark that while this mechanism yields almost no usefulness, it is shown in Section 6.3 how to generalize these ideas for constructing highly useful differentially private analyses when dealing with larger databases.

Example 6.2.3. *Consider a very similar mechanism to that of Example 6.2.2, which given $x \in \{0,1\}$ outputs $x + Y'$ where Y' is obtained by limiting the random variable Y, sampled as before (i.e., Y is sampled according to* $\text{Lap}(1/\varepsilon)$*), to be within the interval $[-k/\varepsilon, k/\varepsilon]$, for some large k (that is, if $Y > k/\varepsilon$ we set $Y' = k/\varepsilon$, similarly, if $Y < -k/\varepsilon$ we set $Y' = -k/\varepsilon$, and otherwise we set $Y' = Y$).*

Note that the resulting mechanism is no longer ε-differentially private since it holds that $\Pr[\mathcal{S}(0) > k/\varepsilon] = 0$ *while* $\Pr[\mathcal{S}(1) > k/\varepsilon] > 0$*, hence the ratio between these two probabilities is unbounded. However, note that the probability that $\mathcal{S}(1) > k/\varepsilon$ is exponentially small in k; hence, the overall probability that an adversary is able to distinguish between the two cases stays practically the same as in Example 6.2.2. It is therefore only natural to still call this mechanism private.*

Definition 6.2.4 ((ε, δ)-differential privacy [25])**.** *A mechanism $\mathcal{S}$ is said to be (ε, δ)-differentially private if for all neighboring pairs of databases $\mathbf{x}, \mathbf{x}' \in \mathcal{D}^n$, and for all subsets of possible answers $\mathcal{V}$:*

$$\Pr[\mathcal{S}(\mathbf{x}) \in \mathcal{V}] \leq \Pr[\mathcal{S}(\mathbf{x}') \in \mathcal{V}]e^{\varepsilon} + \delta \ . \tag{6.3}$$

The probability is taken over the coin tosses of the mechanism.

Privacy of sets. While differential privacy is intended to capture the notion of individual privacy and furthermore is defined with respect to a change in a single entry, it would be somewhat disappointing to find out that it allows a change in, say, two or three entries to cause a massive change in output distribution. Fortunately, as we next show, this is not the case, but rather privacy of sets may deteriorate only linearly in the size of the set (for small sets and for small enough ε). Obviously, for an analysis to be meaningful, the distribution on the outputs must change with a change of many of the entries in the database the, thus privacy of sets must deteriorate, at least for large enough sets.

Lemma 6.2.5. *Let $\mathcal{S}$ be an ε-differentially private mechanism and let $\mathbf{x}, \mathbf{x}'$ be two databases such that $\mathrm{d}_H(\mathbf{x}, \mathbf{x}') = c$. Then*

$$\frac{\Pr[\mathcal{S}(\mathbf{x}) \in \mathcal{V}]}{\Pr[\mathcal{S}(\mathbf{x}') \in \mathcal{V}]} \leq e^{\varepsilon c} .$$

Proof. We prove the lemma by induction on c. For $c = 1$ it is simply the ε-differential privacy of $\mathcal{S}$. Assume correctness for c and let $\mathbf{x}, \mathbf{x}'$ be two databases such that $\mathrm{d}_H(\mathbf{x}, \mathbf{x}') = c + 1$. There exists a database $\mathbf{x}''$ such that $\mathrm{d}_H(\mathbf{x}, \mathbf{x}'') = c$ and $\mathrm{d}_H(\mathbf{x}'', \mathbf{x}') = 1$. By Equation (6.1) and by the induction hypothesis, it follows that

$$\frac{\Pr[\mathcal{S}(\mathbf{x}) \in \mathcal{V}]}{\Pr[\mathcal{S}(\mathbf{x}') \in \mathcal{V}]} = \frac{\Pr[\mathcal{S}(\mathbf{x}) \in \mathcal{V}]}{\Pr[\mathcal{S}(\mathbf{x}'') \in \mathcal{V}]} \cdot \frac{\Pr[\mathcal{S}(\mathbf{x}'') \in \mathcal{V}]}{\Pr[\mathcal{S}(\mathbf{x}') \in \mathcal{V}]} \leq e^{\varepsilon c} e^{\varepsilon} = e^{\varepsilon(c+1)} .$$

□

Composition. Another useful property of differential privacy is that even in the presence of an adaptive adversary privacy stays meaningful after k rounds when $k\varepsilon$ is not too big (i.e., privacy degrades in a linear fashion, as long as k and ε are small enough). We recall it was defined that the output of the mechanism in an adaptive interplay is the sequence of answers it supplied; this does not fall into the scope of Definition 6.2.1, which assumes a single known query. The following theorem shows what can be guaranteed if at each round we activate a differentially private mechanism.

We next present a theorem from [25] asserting that even the relaxed (ε, δ)-differential privacy is quite robust in this sense.

Theorem 6.2.6 ([25])**.** *A mechanism that permits k adaptive interactions, each with an (ε, δ)-differentially private mechanism, is $(k\varepsilon, k\delta)$-differentially private.*

6.3 Private Data Analysis via Output perturbation and Global Sensitivity

Given a function (query) $q : \mathcal{D}^n \to \mathbb{R}$, it is natural to ask if there exists some randomized approximation $\hat{q}$ of q that is differentially private. Clearly, this depends on our definition of approximation, but staying on the intuitive level, the answer to this question is correlated with the *sensitivity* of q, namely, the magnitude of change in the output of q, caused by a change in one entry of the input. We next show that for queries that have low sensitivity (in a very strong sense), it is enough to mask the value of $q(\mathbf{x})$ by some carefully selected random variable.

Query sensitivity [27]. Given a query $q : \mathcal{D}^n \to \mathbb{R}$, the *local sensitivity* is a function of both q and a given database $\mathbf{x}$.

$$\mathrm{LS}_q(\mathbf{x}) = \max_{\{\mathbf{x}':\mathrm{d}_H(\mathbf{x},\mathbf{x}')=1\}} |q(\mathbf{x}) - q(\mathbf{x}')|.$$

The *global sensitivity* is a function of q taken to be the maximum local sensitivity over all databases $\mathbf{x}$, i.e.,

$$\mathrm{GS}_q = \max_{\mathbf{x}' \in \mathcal{D}^n} \left(\mathrm{LS}_q(\mathbf{x})\right).$$

The framework of output perturbation via *global sensitivity* was suggested in [27]. In this framework we consider queries of the form $q : \mathcal{D}^n \to \mathbb{R}$. The outcome is obtained by adding to $q(\mathbf{x})$ noise sampled from the Laplace distribution, calibrated to GS_q. Formally, $\hat{q}$ is defined as

$$\hat{q}(\mathbf{x}) = q(\mathbf{x}) + Y, \text{ where } Y \sim \mathrm{Lap}(\mathrm{GS}_q/\varepsilon). \tag{6.4}$$

This results in an ε-differentially private mechanism. To verify this, for a database $\mathbf{y}$, denote by $h_{\mathbf{y}}(\cdot)$ the probability density function of the distribution on the output of $\hat{q}(\mathbf{y})$. For every pair of neighboring databases $\mathbf{x}, \mathbf{x}'$ and for every possible outcome $v \in \mathbb{R}$, we have that,

$$\begin{aligned} \frac{h_{\mathbf{x}}(v)}{h_{\mathbf{x}'}(v)} &\leq \frac{h(v - q(\mathbf{x}))}{h(v - q(\mathbf{x}'))} \\ &\leq e^{\frac{\varepsilon|(v-q(\mathbf{x}))-(v-q(\mathbf{x}'))|}{\mathrm{GS}_q}} \qquad \text{(by Equation (6.2))} \\ &\leq e^{\varepsilon}\,. \end{aligned}$$

Example 6.3.1. *The* binary sum *function* $\mathrm{SUM} : \{0,1\}^n \to \mathbb{R}$ *is defined as*

$$\mathrm{SUM}(\mathbf{x}) = \sum_{i=1}^{n} x_i.$$

For every two neighboring $\mathbf{x}, \mathbf{x}' \in \{0,1\}^n$ *we have that* $|\mathrm{SUM}(\mathbf{x}) - \mathrm{SUM}(\mathbf{x}')| = 1$ *and hence* $\mathrm{GS}_{\mathrm{SUM}} = 1$. *Applying Equation (6.4), we get an ε-differentially private approximation,* $\hat{f}(\mathbf{x}) = \mathrm{SUM}(\mathbf{x}) + Y$, where $Y \sim \mathrm{Lap}(1/\varepsilon)$*; that is, we get a differentially private approximation of* SUM *with* $O(1)$ *additive error.*

Distributed protocols. In Chapter 7 we consider a few possible realizations of database mechanisms by distributed protocols. We recall that any analysis that can be computed in the presence of a trusted party can be distributively computed using a secure function evaluation protocol, such that the computation itself gives no information to any coalition. However, this translation may be costly and result in non-efficient protocols. We investigate the implied tradeoff between efficiency and accuracy.

Chapter 7

Phase Transition Threshold of Differentially Private Distributed Protocols

In this chapter we observe a phase transition behavior of distributed private protocols, where the communication complexity of such protocols changes abruptly with a minor change in the magnitude of noise we allow to be created by the system.

7.1 Distributed Private Data Analysis: Simultaneously Solving How and What

We consider the combination of two directions in the field of privacy concerning distributed private inputs – secure function evaluation [86, 46, 19, 7] and differential privacy [27, 24]. While in both the goal is to privately evaluate some function of individual inputs, the privacy requirements are significantly different.

Secure function evaluation (SFE) allows n parties $p_1, \ldots, p_n$, sharing a common interest in distributively computing a function $f(\cdot)$ of their inputs $\mathbf{x} = (x_1, \ldots, x_n)$, to compute $f(\mathbf{x})$ while making sure that no coalition of t or fewer curious parties learns more than the outcome of $f(\mathbf{x})$, i.e., for every such coalition, executing the SFE protocol is equivalent to communicating with a trusted party that is given the private inputs x and releases $f(\mathbf{x})$. SFE has been the subject of extensive cryptographic research (initiated in [86, 46, 19, 7]), and SFE protocols exist for any feasible function $f(\cdot)$ in a variety of general settings.

SFE is an important tool for achieving privacy of individual entries – no information about these entries is leaked beyond the outcome $f(\mathbf{x})$. However, this guarantee is insufficient in many applications, and care must be taken in choosing the function $f(\cdot)$ to be computed – any implementation, no matter how secure, of a function $f(\cdot)$ that leaks individual information would not preserve individual privacy.

A criterion for functions that preserve the privacy of individual entries, *differential privacy*, has evolved in a sequence of recent works [23, 38, 30, 9, 27, 24, 25]. It has been demonstrated that differentially private analyses exist for a variety of tasks including the approximation of numerical functions (by means of adding carefully chosen random noise that conceals any single individual's contribution) [27, 9, 63, 43], non-numerical analyses [59], datamining [9, 63], learning [9, 51], non-interactive sanitization [10, 29, 39], and statistical analysis [26, 74].

7.1.1 Constructing Protocols that Preserve Differential Privacy

Combining these two lines of research – SFE and differential privacy – we get a very natural paradigm for constructing protocols that preserve differential privacy, making use of the generality of SFE:

1. Decide on *what* to compute, e.g., a differentially private analysis $\hat{f}(\cdot)$ that approximates a desired analysis $f(\cdot)$. This can be done while abstracting out all implementation issues, assuming the computation is performed by a trusted party that only announces the outcome of the analysis.

2. Decide on *how* to compute, e.g., construct an SFE protocol for computing $\hat{f}(\mathbf{x})$ either by using one of the generic transformations of the feasibility results mentioned above, or by crafting an efficient protocol that utilizes the properties of $\hat{f}(\cdot)$.

This natural paradigm yields a conceptually simple recipe for constructing distributed analyses preserving differential privacy, and, furthermore, allows a valuable separation of our examinations of the *what* and *how* questions. However, comparing the privacy requirements of SFE protocols with differential privacy suggests that this combination may result in sub-optimal protocols. For example, differential privacy is only concerned with how the view of a coalition changes when one (or only few) of the inputs are changed, whereas SFE protocols are required to keep these views indistinguishable even when significant changes occur, if these changes do not affect the function's outcome. Hence, it is interesting to learn whether there are advantages to a paradigm where the analysis to be computed and the protocol for computing it are chosen simultaneously.

The main distributed model we consider is of honest-but-curious parties $p_1, \ldots, p_n$ that perform a computation over their private inputs $x_1, \ldots, x_n$, while maintaining differential privacy with respect to coalitions of size up to t (for formal definitions see Section 7.2 below). The model of honest-but-curious parties has been examined thoroughly in cryptography, and was shown to enable SFE in a variety of settings [86, 46, 7, 19]. While it is probably most natural to consider a setting where the players are computationally limited, we present our results in an information theoretic setting, as this setting allows us to prove lowerbounds on protocols, and hence demonstrate rigorously when constructing differentially private protocols is better than using the natural paradigm.

The second model we consider is the *local model*. This model is also referred to in the literature as *randomized response* and *input perturbation*. This model was originally introduced by Warner [77] to encourage survey responders to answer truthfully, and has been studied extensively since. Protocols executing in the local model have a very simple communication structure, where each party p_i can only communicate with a designated honest-but-curious party C, referred to as a *curator*. The communication can either be *non-interactive*, where each party sends a single message to the curator, which replies with the protocol's outcome, or *interactive*, where several rounds of communication may take place.

7.1.2 Our Results

We initiate an examination of the paradigm where an analysis and the protocol for computing it are chosen simultaneously. We begin with two examples that present the potential benefits of using this paradigm: it can lead to simpler protocols, and more importantly it can lead to more efficient protocols. For the latter we consider the Binary Sum function,

$$\mathrm{SUM}(x_1, \ldots, x_n) = \sum_{i=1}^{n} x_i \quad \text{for } x_i \in \{0, 1\}.$$

The major part of this work examines whether constructing non-SFE protocols for computing an approximation $\hat{f}(\cdot)$ to $\mathrm{SUM}(\cdot)$ yields an efficiency gain[1]. Ignoring the dependency on the privacy parameter, our first observation is that for approximations with additive error $\approx \sqrt{n}$ there is a gain – for a natural class of *symmetric* approximation functions (informally, functions where the outcome does not depend on the order of inputs), it is possible to construct differentially private protocols that are much more efficient than any SFE protocol for a function in this class. Moreover, these differentially private protocols are secure against coalitions of size up to $t = n - 1$, and need not rely on secure channels.

The picture changes when we consider additive error smaller than $\sqrt{n}$. This follows from a sequence of results:

[1] We only consider *oblivious protocols* where the communication pattern is independent of input and randomness (see Section 7.2).

1. We prove first that no such non-interactive protocols in the local model exist. Furthermore, no local protocols with $\ell \leq \sqrt{n}$ rounds and additive error $\sqrt{n}/\tilde{O}(\ell)$ exist.

2. We show that in particular, no local interactive protocol with $o(\sqrt{n/\log n})$ rounds exists for computing $\mathrm{SUM}(\cdot)$ within constant additive error (this is in contrast to the centralized setup where $\mathrm{SUM}(\cdot)$ can be computed within $O(1)$ additive error).

3. Finally, we prove that the bounds on local protocols imply that no distributed protocols exist that use $nt/4$ messages, and approximates $\mathrm{SUM}(\cdot)$ within additive error $\sqrt{n}/\tilde{O}(\ell)$ in ℓ rounds.

Considering the natural paradigm, i.e., computing a differentially-private approximation to $\mathrm{SUM}(\cdot)$ using SFE, we get a protocol for approximating $\mathrm{SUM}(\cdot)$ with $O(1)$ additive error, and sending $O(nt)$ messages. Thus, for protocols with error $o(\sqrt{n}/\varepsilon)$ and small number of rounds, there is no gain in using the paradigm of a simultaneous design of the function and its protocol.

Our results imply that differentially private protocols constructed under computational hardness assumptions, yielding a computational version of differential privacy (see Definition 7.2.2), are provably more efficient than protocols that do not make use of computational hardness. For instance, the phase transition we observe at $\theta(\sqrt{n}/\varepsilon)$ additive error *does not* hold in a computational setting. See Example 7.2.3 for details.

7.1.3 Techniques

We prove our lowerbound for the distributed model in a sequence of reductions. We begin with a simple reduction from any differentially private protocol for SUM to a gap version of the threshold function, denoted $\mathrm{GAP\text{-}TR}$. Henceforth, it is enough to prove our lowerbound for $\mathrm{GAP\text{-}TR}$.

In the heart of our lowerbound for $\mathrm{GAP\text{-}TR}$ is a transformation from efficient distributed protocols into local interactive protocols, showing that if there are distributed differentially-private protocols for $\mathrm{GAP\text{-}TR}(\cdot)$ in which half of the parties interact with less than $t+1$ parties, then there exist differentially-private protocols for $\mathrm{GAP\text{-}TR}(\cdot)$ in the local interactive model. This allows us to prove our impossibility results in the local model, which is considerably simpler to analyze.

In analyzing the local non-interactive model, we prove lowerbounds borrowing from analyses in [23, 30]. The main technical difference is that our analysis is a lowerbound and hence should hold for general protocols, whereas the work in [23, 30] was concerned with proving feasibility of privacy-preserving computations (i.e., upperbounds), and hence they analyze of very specific protocols.

To extend our lowerbounds from the local non-interactive to interactive protocols, we decompose an ℓ-round interactive protocol to ℓ one-round protocols, analyze the ℓ protocols, and use composition to obtain the lowerbound.

7.1.4 Related Work

Secure function evaluation and private data analysis were first tied together in the *Our Data, Ourselves (ODO)* protocols [25]. The constructions in [25] – distributed SFE protocols for generating shares of random noise used in private data analyses – follow the natural paradigm discussed above (however, they avoid utilizing generic SFE feasibility results to gain on efficiency). We note that a difference between the protocols in [25] and the discussion herein is that ODO protocols are secure against malicious parties, in a computational setup, whereas we deal with honest-but-curious parties, and mostly in an information theoretic setup. Following our work, computational differential privacy was considered in [60]; they present several definitions of computational differential privacy, study the relationships between these definitions, and construct efficient 2-party computational differentially private protocols for approximating the distance between two vectors. In this work, we supply a definition of computationally (t, ϵ)-differentially private protocols which is close to the definition of IND-CDP privacy in [60].

Lowerbounds on the local non-interactive model were previously presented implicitly in [27, 72, 51], and explicitly in [23, 28]. The two latter works are mainly concerned with what is called the global (or centralized) interactive setup, but have also implications to approximation to SUM in the local *non-interactive* model, namely, that it is impossible to approximate it within additive error $c\sqrt{n}$ (for some constant $c > 0$), a slightly weaker result compared to our lowerbound of $c\sqrt{n}/\varepsilon$ for ε-differentially private local non-interactive protocols. However, (to the best of our understanding) these implications of [23, 28] do not imply the lowerbounds we get for local interactive protocols and distributed protocols.

Chor and Kushilevitz [20] consider the problem of securely computing modular sum when the inputs are distributed. They show that this task can be done while sending roughly $n(t+1)/2$ messages. Furthermore, they prove that this number of messages is optimal for a family of protocols that they call oblivious. These are protocols where the communication pattern is fixed and does not depend on the inputs or random inputs. In our work we extend their lowerbound result and prove that with $n(t+1)/4$ messages no symmetric approximation for SUM with sub-linear additive error can be computed in an oblivious protocol.

7.1.5 Organization

The rest of the chapter is organized as follows: In Section 7.2 we define an extension of differential privacy definition to differentially private protocols, describe the local model of communication, and define the binary sum and gap threshold functions. In Section 7.3, we present two motivating examples for our new methodology of simultaneously solving how and what. In Section 7.4 we prove lowerbounds on the error of differentially private protocols for computing the binary sum and gap threshold functions in the local model, and in Section 7.5 we extend these lowerbounds to the distributed model. Finally, in Section 7.6 we prove that an SFE protocol for computing a symmetric approximation of the sum function with less than $nt/4$ messages has an error of $\Omega(n)$ (compared to a non-SFE protocol that approximates the sum function with $O(n)$ messages and an error of $\Omega(\sqrt{n})$).

7.2 Differentially Private Protocols

Our privacy definition for distributed protocols (Definition 7.2.1 below) can be viewed as a distributed variant of ε-differential privacy. Informally, a computation is differentially private if any change in a single individual input may only induce a small change in the distribution of its outcomes.

We consider a distributed setting, where n parties $p_1, \ldots, p_n$ hold private inputs $x_1, \ldots, x_n$, respectively, and engage in a protocol Π in order to compute (or approximate) a function $f(\cdot)$ of their joint inputs. Parties are *honest-but-curious*, which means they follow the prescribed randomized protocol; however, as the execution of the protocol terminates, parties can collide and try to infer information about inputs of parties outside the coalition. The protocol Π is executed in a synchronous environment with point-to-point secure (untappable) communication channels, and is required to preserve privacy with respect to coalitions of size up to t. Following [20], we only consider a *fixed-communication* protocol Π (also called an oblivious protocol) where every channel is either (i) active in every run of Π (i.e., at least one bit is sent over the channel), or (ii) never used[2]. Parties that are adjacent to at least $t+1$ active channels are called *popular*, other parties are called *lonely*.

The main definition we present is an extension of Definition 6.2.1 to a distributed setting. Informally, we require that differential privacy is preserved with respect to any coalition of size up to t.

[2] Our proofs also work in a relaxed setting where every channel is either (i) used in at least a constant fraction of the runs of Π (where the probability is taken over the coins of Π), or (ii) is never used.

Notation. A *vector* $\mathbf{x} = (x_1, \ldots, x_n)$ is an ordered sequence of n elements of some domain D. Vectors $\mathbf{x}, \mathbf{x}'$ are *neighboring* if they differ on exactly one entry, and are *T-neighboring* if they differ on a single entry whose index is *not* in $T \subset [n]$.

Definition 7.2.1 (Distributed differential privacy). *Let Π be a protocol between n (honest-but-curious) parties. For a set $T \subseteq [n]$ and fixed inputs $\mathbf{x} = (x_1, \ldots, x_n)$, let $\mathrm{View}_T(x_1, \ldots, x_n)$ be the random variable containing the inputs of the parties in T (i.e., $\{x_i\}_{i \in T}$), the random inputs of the parties in T, and the messages that the parties in T received during the execution of the protocol with private inputs $\mathbf{x} = (x_1, \ldots, x_n)$ (the randomness is taken over the random inputs of the parties).*

*We say that Π is (t, ε)-*differentially private *if for all $T \subset [n]$, where $|T| \leq t$, for all T-neighboring $\mathbf{x}, \mathbf{x}'$, and for all possible sets $\mathcal{V}_T$ of views of the parties in T:*

$$\Pr[\mathrm{View}_T(\mathbf{x}) \in \mathcal{V}_T] \leq e^{\varepsilon} \cdot \Pr[\mathrm{View}_T(\mathbf{x}') \in \mathcal{V}_T], \tag{7.1}$$

where the probability is taken over the random inputs of the parties in the protocol Π.

An equivalent requirement is that for all $T \subset [n]$, where $|T| \leq t$, for all T-neighboring $\mathbf{x}, \mathbf{x}'$, and for all distinguishers D (i.e., functions, not necessarily efficiently computable, from views to $\{0, 1\}$),

$$\Pr[D(\mathrm{View}_T(\mathbf{x})) = 1] \leq e^{\varepsilon} \cdot \Pr[D(\mathrm{View}_T(\mathbf{x}')) = 1].$$

This requirement can be relaxed to only consider distinguishers that are computationally bounded:

Definition 7.2.2 (Computational distributed differential privacy). *We say that Π is* computationally (t, ε)-differentially private *if for every probabilistic polynomial-time algorithm D, and for every polynomial $p(\cdot)$, there exists k_0 such that for all $k \geq k_0$, for all $T \subset [n]$, where $|T| \leq t$, and for all T-neighboring inputs $\mathbf{x}, \mathbf{x}' \in \left(\{0,1\}^k\right)^n$:*

$$\Pr[D(\mathrm{View}_T(\mathbf{x})) = 1] \leq e^{\varepsilon} \cdot \Pr[D(\mathrm{View}_T(\mathbf{x}')) = 1] + \frac{1}{p(n \cdot k)} ,$$

where the probabilities are taken over the random inputs of the parties in protocol Π and the randomness of D.

Example 7.2.3. *We next describe a computationally $(n/2, \varepsilon)$-differentially private protocol for computing* SUM *with $O(\log n/\epsilon)$ additive error, $O(n)$ messages, and constant number of rounds. This protocol uses a homomorphic encryption scheme with threshold decryption (that is, only the sets of all parties can decrypt messages). For example, if we use ElGamal encryption, the distributed key generation and decryption require one round in which each party sends one message. The protocol works in three phases:*

Key Generation. *The parties generate public and private keys for the homomorphic encryption scheme with threshold decryption.*

Encryption. *Each party p_i chooses a random noise_i (according to a distribution that will be defined later), computes $y_i = x_i + \mathrm{noise}_i$, encrypts y_i using the public encryption key and sends the encryption to p_1.*

Decryption. *Party p_1 computes z, an encryption of $y = \sum_{i=1^n} y_i$ (this is possible as the encryption scheme is homomorphic). p_1 sends z to each p_i, which in return sends a decryption message back to p_1. Finally, p_1 decrypts y from the decryption messages and sends y to all parties.*

One way to generate each party's noise is for each party to sample from the Normal distribution with mean zero and variance $6\log^2 n/(n\varepsilon^2)$. Since the sum of normal random variables is a normal random variable, $y = \sum_{i=1^n} x_i + \mathrm{noise}$ where noise is sampled from a normal distribution with mean zero and variance $6\log^2 n/\varepsilon^2$. Furthermore, even if a coalition of $n/2$ parties subtracts the noise that its parties added to y, the variance of the remaining noise is $3\log^2 n/\varepsilon^2$. Using the analysis of [25], the protocol is a computationally $(n/2,\varepsilon)$-differentially private protocol which with constant probability has error $O(\log n/\epsilon)$.

The above protocol is a computationally $(n/2,\varepsilon)$-differentially private protocol for computing SUM *with $O(\log n/\epsilon)$ additive error, $O(n)$ messages, and constant number of rounds. In contrast, we prove that $(n/2,\varepsilon)$-differentially information-theoretically private protocol for computing* SUM *with $o(\sqrt{n})$ additive error and constant number of rounds must send $\Omega(n^2)$ messages. Thus, our results shows that requiring only computational differentially-privacy does result in more efficient protocols.*

Using standard SFE feasibility results (in the computational setting), it is possible now to prove that the natural paradigm presented in Section 7.1.1 yields protocols that adhere to Definition 7.2.2. Consider an ϵ-differentially private data analysis $\hat{f}$ and a computationally bounded distinguisher D, trying distinguish between a computation of an SFE protocol computing $\hat{f}$ with neighboring inputs x and x′. Since, $\hat{f}$ preserves differential privacy the distributions on the outputs must be ε close, the same must hold for the random variables describing the adversary's view (up to some negligible function in the length of the (concatenated) inputs). We get:

Lemma 7.2.4 (Informal)**.** *Let $\hat{f}$ be ε-differentially private, and let Π be a t-secure protocol computing $\hat{f}$, then Π is computationally (t,ε)-differentially private.*

In the above lemma, the if the t-secure protocol Π computing $\hat{f}$ has perfect security, then Π is information-theoretically (t,ε)-differentially private.

Remark 7.2.5. *We will only consider protocols computing a (randomized) function $\hat{f}(\cdot)$ resulting in all parties computing the* same *outcome of $\hat{f}(\mathbf{x})$. This can be achieved, e.g., by having one party compute $\hat{f}(\mathbf{x})$ and send the outcome to all other parties.*

7.2.1 Basic Facts about Distributed Protocols

Throughout this chapter we use some basic facts that apply to all randomized distributed protocols. We start with a general notation.

Notation 7.2.6. *Fix an n-party randomized protocol Π, assume that each p_i holds an input x_i, and fix some communication transcript c. We define $\alpha_i^c(x_i)$ as the probability that in each round p_i with input x_i sends messages according to c provided that in previous rounds it sees messages according to c (that is get messages according to c and sends messages according to c). The probability is taken over the random string of party p_i.*

Let c be an ℓ-round transcript in which, without loss of generality, p_i sends a message at each round. We note that $\alpha_i^c(x_i) = \prod_{k=1}^{\ell} \beta_k$, where β_k is probability that p_i with input x_i sends in round k the message according to c provided that in previous rounds it sees messages according to c. This is verified simply by fixing the inputs and the randomness of all other parties to be consistent with c and considering the conditional probabilities.

The following lemma captures the intuition that the probability that a transcript c is exchanged in a given distributed protocol Π with input vector $\mathbf{x} = (x_1, \ldots, x_n)$ is the product of the probabilities that each party p_i chooses a random input r_i such that r_i and x_i are consistent with c.

Lemma 7.2.7. *Fix an n-party randomized protocol Π, assume that each p_i holds an input x_i, and fix some communication transcript c. Then, the probability that c is exchanged is $\prod_{i=1}^{n} \alpha_i^c(x_i)$.*

Proof. Since the random inputs of the parties are independently chosen, we have that $\alpha_i^c(x_i)$ for all is are mutually independent, hence the lemma follows. □

7.2.2 The Local Model

The local model (previously discussed in [27, 51]) is a simplified distributed communication model where the parties communicate via a designated party – a *curator* – denoted C (with no local input). We will consider two types of differentially private local protocols. In *non-interactive* local protocols each party p_i applies an ε-differentially private algorithm S_i on its private input x_i and randomness r_i, and sends $S_i(x_i, r_i)$ to C that then performs an arbitrary computation and publishes its result.

In *interactive* local protocols the protocol proceeds in *rounds*, where in each round j the curator sends to each party p_i a "query" message $q_{i,j}$ and party p_i responds with the jth "answer" $A_i(x_i, q_{i,1}, \ldots, q_{i,j}, r_i)$; the answer is a function of the party's input x_i, its random input r_i, and the first j queries. I.e., each round consists of two communication phases: first, the query messages are sent by the curator, then, each party sends the appropriate response message. We note that in the honest-but-curious setting we can assume, without loss of generality, that the curator is deterministic, as randomness for the curator may be provided by parties in their first message.

Definition 7.2.8 (Differential privacy in the local model)**.** *We say that a protocol* Π *in the local model is* ε-differentially private *if the curator's view preserves* ε*-differential privacy. Formally, for all neighboring* $\mathbf{x}, \mathbf{x}'$ *and for every possible set* $\mathcal{V}_C$ *of views of the curator:*

$$\Pr[\mathrm{View}_C(\mathbf{x}) \in \mathcal{V}_C] \quad \leq \quad e^{\varepsilon} \cdot \Pr[\mathrm{View}_C(\mathbf{x}') \in \mathcal{V}_C],$$

where $\mathrm{View}_C(\mathbf{x})$ *is the random variable containing the messages that* C *receives during the execution of the protocol with private inputs* $\mathbf{x} = (x_1, \ldots, x_n)$ *and the probability is taken over the random inputs of the parties.*

We note that $\mathrm{View}_C(\mathbf{x})$ is defined in accordance with Definition 7.2.1 (with some abuse of notation, as C is not a set). However, since C has no initial input and since C is assumed to be deterministic, $\mathrm{View}_C(\mathbf{x})$ only contains the messages that C receives during the execution of the protocol with inputs $\mathbf{x} = (x_1, \ldots, x_n)$.

The differential privacy requirement in the local model may be equivalently phrased as a requirement to preserve the privacy of each party independently of other parties. We next give a definition in this spirit by considering the probabilities that a party p_i replies in a certain way to a given sequence of queries with $x_i = 0$ and with $x_i = 1$. Any communication transcript c in an execution of the protocol defines a transcript c_i, where $c_i = q_{i,1}, a_{i,1}, \ldots, q_{i,\ell}, a_{i,\ell}$ is the restriction of c to the messages transferred between party p_i and the curator (recall that in the local model every party communicates solely with the curator). Thus, we can use $\alpha_i^{c_i}(x_i)$ (see Notation 7.2.6) to denote the probability that p_i with private input x_i replies by $a_{i,1}, \ldots, a_{i,\ell}$ provided the curator has sent queries $q_{i,1}, \ldots, q_{i,\ell}$. Using this notation, we formally present the alternative definition of privacy in the local model.

Definition 7.2.9 (Differential privacy in the local model – Individual privacy requirement)**.** *We say that a protocol* Π *in the local model is* ε-differentially private *if the curator's view preserves* ε*-differential privacy with respect to each party separately. Formally, for every* $i \in [n]$ *and for any possible communication transcript* $c_i = q_{i,1}, a_{i,1}, \ldots, q_{i,\ell}, a_{i,\ell}$ *between party* p_i *and the curator (i.e., there exist inputs* $x'_1, \ldots, x'_n$ *and random inputs* $r'_1, \ldots, r'_n$ *consistent with* c_i*), and for every* $x_i, y_i \in D$ *it holds that* $\alpha_i^{c_i}(x_i) \leq e^{\varepsilon} \cdot \alpha_i^{c_i}(y_i)$*, where the probabilities are taken over the random input of* p_i*.*

We next show the equivalence of the two privacy definitions of local protocols given above.

Claim 7.2.10. *The privacy requirements in Definition 7.2.8 and in Definition 7.2.9 are equivalent.*

Proof. For a communication transcript $c_i = q_{i,1}, a_{i,1}, \ldots, q_{i,j}, a_{i,j}$ between party p_i and the curator, denote by $\alpha_i^{c_i}(x_i)$ the probability that p_i is consistent with c_i with input x_i (namely, the probability that p_i with input x_i replies with messages $a_{i,1}, \ldots, a_{i,j}$ provided that query messages sent by the curator were $q_{i,1}, \ldots, q_{i,j}$).

First assume that Π is ε-differentially private according to Definition 7.2.8. Let c_i be a communication transcript exchanged between p_i and the curator. We are interested in the maximum ratio between the probabilities $\alpha_k^{c_k}(x_k)$ and $\alpha_k^{c_k}(x'_k)$ for any p_k, any pair of inputs x_k, x'_k, and any such transcript c_k. Let c be any full communication transcript of all parties, such that c_k is the part of c which is exchanged between the curator and p_k. Denote c_i, the part of c which is exchanged between the curator and each party p_i for $i \neq k$. Fix any set of inputs $(x_1, \ldots, x_{k-1}, x_{k+1}, \ldots, x_n)$ for all parties other than p_k, by Lemma 7.2.7 we have that

$$\frac{\alpha_k^{c_k}(x_k)}{\alpha_k^{c_k}(x'_k)} = \frac{\alpha_k^{c_k}(x_k) \cdot \prod_{i \neq k} \alpha_i^{c_i}(x_i)}{\alpha_k^{c_k}(x'_k) \cdot \prod_{i \neq k} \alpha_i^{c_i}(x_i)} = \frac{\alpha_k^{c}(x_k) \cdot \prod_{i \neq k} \alpha_i^{c}(x_i)}{\alpha_k^{c}(x'_k) \cdot \prod_{i \neq k} \alpha_i^{c}(x_i)} \leq e^{\varepsilon}.$$

The last inequality holds by Definition 7.2.8.

For the other direction, assume that Definition 7.2.9 holds, i.e., that for any party p_i and for any possible communication transcript $c_i = q_{i,1}, a_{i,1}, \ldots, q_{i,j}, a_{i,j}$ between party p_i and the curator, it holds that $\frac{\alpha_k^{c_k}(x_k)}{\alpha_k^{c_k}(x'_k)} \leq e^{\varepsilon}$. Let $\mathbf{x}, \mathbf{x}'$ be some pair of neighboring inputs such that $x_k \neq x'_k$, and $x_i = x'_i$ for all $i \neq k$. For any possible view v of the curator, denote by c the full communication transcript explaining v. Denote by c_i the part of c which is exchanged between the curator and each party p_i, specifically, c_k is the part of c which is exchanged between the curator and each party p_k. Hence, by Lemma 7.2.7,

$$\frac{\Pr[\mathrm{View}_C(\mathbf{x}) = v]}{\Pr[\mathrm{View}_C(\mathbf{x}') = v]} = \frac{\prod_{i=1}^{n} \alpha_i^{c}(x_i)}{\prod_{i=1}^{n} \alpha_i^{c}(x'_i)} = \frac{\alpha_k^{c_k}(x_k) \cdot \prod_{i \neq k} \alpha_i^{c_i}(x_i)}{\alpha_k^{c_k}(x'_k) \cdot \prod_{i \neq k} \alpha_i^{c_i}(x_i)} = \frac{\alpha_k^{c_k}(x_k)}{\alpha_k^{c_k}(x'_k)} \leq e^{\varepsilon}.$$

□

7.2.3 Approximation

We will construct protocols whose outcome approximates a function $f : D^n \to \mathbb{R}$ by a probabilistic function, according to the following definition:

Definition 7.2.11 (Approximation). *A randomized function $\hat{f} : D^n \to \mathbb{R}$ is an* additive *(γ, τ)-approximation for a (deterministic) function f if*

$$\Pr\left[|f(\mathbf{x}) - \hat{f}(\mathbf{x})| > \tau(n)\right] \leq \gamma(n)$$

for all $\mathbf{x} \in D^n$*. The probability is over the randomness of* $\hat{f}$*.*

For example, by the properties of the Laplace distribution, Equation (6.4) yields an additive $(e^{-k}, k \cdot \mathrm{GS}_f/\varepsilon)$-approximation to f, for every $k > 0$.

7.2.4 The Binary Sum and Gap Threshold Functions

We consider the binary sum function defined to be $\mathrm{SUM}_n(x_1, \ldots, x_n) = \sum_{i=1}^{n} x_i$ for $x_i \in \{0, 1\}$. When n is clear from the context, we omit the subscript n. We next define a gap version of the threshold function:

Definition 7.2.12 (Gap Threshold)**.** *We define* GAP-TR *by cases on* SUM*:*

If $\mathrm{SUM}_n(x_1, \ldots, x_n) \le \kappa$ *then* $\text{GAP-TR}_{\kappa,\tau}(x_1, \ldots, x_n) = 0.$

If $\mathrm{SUM}_n(x_1, \ldots, x_n) \ge \kappa + \tau$ *then* $\text{GAP-TR}_{\kappa,\tau}(x_1, \ldots, x_n) = 1.$

Note that there are no requirements on the output of $\text{GAP-TR}_{\kappa,\tau}$ when $\kappa < \mathrm{SUM}_n(x_1, \ldots, x_n) < \kappa + \tau$. Clearly, a $(\gamma, \tau/2)$-approximation $\hat{f}$ to SUM can be translated into an $(\gamma, 0)$-approximation $\hat{g}$ to $\text{GAP-TR}_{\kappa,\tau}$, by adding a simple calculation step, that is, given an approximation $y = \hat{f}(\mathbf{x})$ for $\mathrm{SUM}_n(\mathbf{x})$, set $\text{GAP-TR}_{\kappa,\tau}(\mathbf{x})$ to be 0 if $y \le \kappa + \tau/2$ and 1 otherwise. Thus, the following claim is straightforward in both the distributed and the local models.

Claim 7.2.13. *If there exists an* ℓ*-round,* (t, ε)*-differentially private (respectively,* ε*-differentially private in the local model) protocol that* $(\gamma, \tau/2)$*-approximates* SUM_n *sending* ρ *messages, then for every* κ *there exists an* ℓ*-round,* (t, ε)*-differentially private (respectively,* ε*-differentially private in the local model) protocol that correctly computes* $\text{GAP-TR}_{\kappa,\tau}$ *with probability at least* $1 - \gamma$*, sending at most* ρ *messages.*

Specifically, non-existence of (t, ε)-differentially private protocols for computing $\text{GAP-TR}_{0,\tau}$ correctly with $n(t+1)/4$ messages implies that there exists no (t, ε)-differentially private protocols for computing SUM_n with $n(t+1)/4$ messages and additive error magnitude $\tau/2$. The next claim asserts that the same non-existence also implies that, for any $0 \le \kappa \le n - \tau$, there exists no (t, ε)-differentially private protocol for computing $\text{GAP-TR}_{\kappa,\tau}$ correctly with $n(t+1)/8$ messages. Again, it applies to both the distributed and the local models.

Claim 7.2.14. *If for some* $0 \le \kappa \le n - \tau$ *there exists an* ℓ*-round,* (t, ε)*-differentially private (respectively,* ε*-differentially private in the local model)* n*-party protocol that correctly computes* $\text{GAP-TR}_{\kappa,\tau}$ *with probability at least* γ *sending at most* ρ *messages, then there exists an* ℓ*-round,* $(t/2, \varepsilon)$*-differentially private (respectively,* ε*-differentially private in the local model)* $n/2$*-party protocol that correctly computes* $\text{GAP-TR}_{0,\tau}$ *with probability at least* γ *sending at most* ρ *messages.*

Proof. First assume $\kappa \leq n/2$. Given an n-party protocol Π that correctly computes $\text{GAP-TR}_{\kappa,\tau}$, define an $n/2$-party protocol Π' for computing $\text{GAP-TR}_{0,\tau}$ by simulating parties $p_{\frac{n}{2}+1}, \ldots, p_n$ where $x_{\frac{n}{2}+1}, \ldots, x_{\frac{n}{2}+\kappa}$ are set with value 1 and $x_{\frac{n}{2}+\kappa+1}, \ldots, x_n$ are set to 0. In the local model, a designated party, say p_1, can simulate these $n/2$ parties. In the distributed model, however, we let each party p_i simulate party $p_{i+n/2}$ and, thus, the resulting protocol is only $(t/2, \varepsilon)$-differentially private.

To verify that the resulting protocol is indeed $(t/2, \varepsilon)$-differentially private, observe that any view v of a coalition T' of size $t' \leq t/2$ in some execution of the resulting protocol is exactly the view of the coalition T of size $2t' \leq t$, implied by T' (for $p_i \in T'$ we have $p_i, p_{i+n/2} \in T$), in the appropriate computation of the original protocol. Moreover, any T'-neighboring $\mathbf{x}, \mathbf{x}'$ define T-neighboring $\mathbf{xy}, \mathbf{x}'\mathbf{y}$ (where $\mathbf{y} = 1^{\kappa}0^{\frac{n}{2}-\kappa}$), such that $\Pr[\text{View}_T(\mathbf{xy}) = v] = \Pr[\text{View}_{T'}(\mathbf{x}) = v]$ and $\Pr[\text{View}_T(\mathbf{x}'\mathbf{y}) = v] = \Pr[\text{View}_{T'}(\mathbf{x}') = v]$. Thus, by the privacy of the original protocol, the resulting protocol is $(t/2, \varepsilon)$-differentially private.

Otherwise, if $\kappa > n/2$, we can transform the original protocol to one that correctly computes $\text{GAP-TR}_{n-\kappa-\tau,\tau}$, by flipping all input bits before engaging in the execution, running the original protocol, and finally flipping the result of the computation. □

7.3 Motivating Examples

We begin with two examples manifesting benefits of choosing an analysis together with a differentially private protocol for computing it. In the first example, this paradigm yields more efficient protocols than the natural paradigm; in the second example, it yields simpler protocols.

7.3.1 Binary Sum – $\sqrt{n}$ Additive Error

We begin with a simple protocol for approximating SUM_n within $O(\sqrt{n}/\varepsilon)$-additive approximation. This protocol is well known as *Randomized Response* [77] (randomized response protocols under some related definition of privacy were also recently studied in [38]). We describe the protocol in the (non-interactive) local model, and it can be easily translated to a two round (and $2n$ messages) (n, ε)-differentially private distributed protocol by letting some arbitrarily designated party (say p_1) play the role of C.

Let $\text{flip}_{\alpha}(x)$ be a randomized bit flipping operator returning x with probability $0.5 + \alpha$ and $1 - x$ otherwise, where $\alpha = \frac{\varepsilon}{4+2\varepsilon}$. The protocol proceeds as follows:

1. Each party p_i with private input $x_i \in \{0, 1\}$ sends $z_i = \text{flip}_{\alpha}(x_i)$ to C.
2. C locally computes and publishes $k = \sum_{i=1}^{n} z_i$.

3. Each party locally computes $\hat{f} = (k - (0.5 - \alpha)n)/2\alpha$.

A total of $O(n)$ messages and $O(n \log n)$ bits of communication are exchanged. To see that the protocol satisfies the privacy requirement of Definition 7.2.9, note that

$$\frac{\Pr[\text{flip}_\alpha(1) = 1]}{\Pr[\text{flip}_\alpha(0) = 1]} = \frac{0.5 + \alpha}{0.5 - \alpha} = 1 + \varepsilon \leq e^\varepsilon,$$

and similarly $\Pr[\text{flip}_\alpha(0) = 0]/\Pr[\text{flip}_\alpha(1) = 0] \leq e^\varepsilon$. To see that the protocol approximates the sum function, note that

$$\mathbb{E}[z_i] = \mathbb{E}[\text{flip}_\alpha(x_i)] = \begin{cases} 0.5 + \alpha & \text{if } x_i = 1 \\ 0.5 - \alpha & \text{if } x_i = 0. \end{cases}$$

Thus,

$$\mathbb{E}[k] = (0.5 + \alpha) \cdot \text{SUM}(\mathbf{x}) + (0.5 - \alpha) \cdot (n - \text{SUM}(\mathbf{x})) = 2\alpha \cdot \text{SUM}(\mathbf{x}) + (0.5 - \alpha)n,$$

and hence,

$$\mathbb{E}[\hat{f}] = \mathbb{E}\left[\frac{k - (0.5 - \alpha)n}{2\alpha}\right] = \text{SUM}(\mathbf{x}).$$

we apply the following Chernoff bound: Given n zero-one random variables $X_1, \ldots, X_n$ and $0 < t < 1$, $\Pr\left[\sum_{i=1}^n X_i \leq (1 - t)\mu\right] < \exp\left(-\frac{t^2\mu}{2}\right)$, where $\mu = \sum_{i=1}^n \mathbb{E}[X_i]$. Hence, we get that $\hat{f}$ is an additive $(O(1), O(\sqrt{n}/\varepsilon))$-approximation to $\text{SUM}(\cdot)$, that is, with constant probability, the error is $O(\sqrt{n}/\varepsilon)$.

Remark 7.3.1. *We next sketch an alternative ε-differentially private protocol that $(O(1), \sqrt{n}/\varepsilon)$-approximates SUM_n:*

1. *Each party p_i with private input $x_i \in \{0, 1\}$ samples $y_i \sim \text{Lap}(1/\varepsilon)$ and sends $z_i = x_i + y_i$ to C.*

2. *C locally computes $\hat{f} = \sum_{i=1}^n z_i$ and publishes the result.*

The privacy of the protocol follows from the arguments in Section 6.3.

Remark 7.3.2. *The above constructions result in* symmetric *approximations to* $\text{SUM}(\cdot)$ *(i.e., the output distribution depends solely on* $\text{SUM}(\cdot)$ *and not on the specific assignment). While these differentially private protocols use $O(n)$ messages, it can be shown that for such symmetric functions that no efficient SFE protocols for such functions exist (see Section 7.6 for more details).*

7.3.2 Distance from a Long Subsequence of 0's

Our second function measures how many bits in a sequence x of n bits should be set to zero to get an all-zero consecutive subsequence of length n^α. In other words, the function should return the minimum weight over all substrings of x of length n^α bits:

$$\mathrm{DIST}_\alpha(\mathbf{x}) = \min_i \left(\sum_{j=i}^{i+n^\alpha-1} x_j \right).$$

For $t \le n/2$ we present a (t, ε, δ)-differentially private protocol[3] approximating $\mathrm{DIST}_\alpha(\mathbf{x})$ with additive error $\tilde{O}(n^{\alpha/3}/\varepsilon)$.

In our protocol, we treat the n-bit string x (where x_i is held by party p_i) as a sequence of $n^{1-\alpha/3}$ disjoint intervals $I_1, \ldots, I_{n^{1-\alpha/3}}$, each $n^{\alpha/3}$ bit long. Let i_k be the index of the first bit in the interval I_k, and observe that $\min_{i_k}(\sum_{j=i_k}^{i_k+n^\alpha-1} x_j)$ is an $n^{\alpha/3}$ additive approximation of DIST_α. The protocol for computing an approximation $\hat{f}$ to DIST_α is sketched below.

1. Every party p_i generates independent random variables $Y_i^1, \ldots Y_i^{n^{1-\alpha/3}}$, each distributed according to the normal distribution $N(\mu = 0, \sigma^2 = 2R/n)$ where $R = \frac{2\log(\frac{2}{\delta})}{\varepsilon^2}$. The random variable Y_i^k is the noise contributed by p_i to the perturbed sum over the interval I_k.

   Party p_i then shares x_i and $Y_i^1, \ldots Y_i^{n^{1-\alpha/3}}$ between the parties $p_1, \ldots, p_{t+1}$ using an additive $(t+1)$-out-of-$(t+1)$ secret sharing scheme[4].

2. Every party p_i, where $1 \le i \le t+1$, sums, for every interval I_k of length $n^{\alpha/3}$, the shares of x_i's it got from the parties in the interval together with the shares of Y_i^k's it got from all parties, and sends this sum to p_1.

3. For every interval I_k, party p_1 computes the sum of the $t+1$ sums it got for the interval. By the additivity of the secret sharing scheme, this sum is equal to

$$S_k = \sum_{j=i_k}^{i_k+n^{\alpha/3}-1} x_j + \sum_{\ell=1}^{n} Y_\ell^k = \left(\sum_{j=i_k}^{i_k+n^{\alpha/3}-1} x_j \right) + Z_k,$$

   where $Z_k = \sum_{\ell=1}^{n} Y_\ell^k$ (notice that $Z_k \sim N(\mu = 0, \sigma^2 = 2R)$).

[3] (ε, δ)-differential privacy is a generalization, defined in [25], of ε-differential privacy where it is only required that $\Pr[\hat{f}(\mathbf{x}) \in \mathcal{V}] \le e^\varepsilon \cdot \Pr[\hat{f}(\mathbf{x}') \in \mathcal{V}] + \delta$.

[4] Shared secrets are taken from a large enough finite domain by rounding the numbers $\log n$ digits after the point. This yields no breach in privacy and adds a small magnitude of error.

4. p_1 computes $\min_k \sum_{j=k}^{k+n^{2\alpha/3}} S_k$ and sends this output to all parties.

We take Z_k to be with variance $2R$ (rather than R) to ensure that the sum of the noise generated by $n - t \geq n/2$ of the parties is enough to conceal the x_i's in interval I_k. Using the analysis of [25], this protocol is a (t, ε, δ)-differentially private protocol when $2t < n$. Furthermore, we next give a sketch proof that with high probability the additive error is $\tilde{O}(n^{\alpha/3}/\varepsilon)$. It suffices to consider the probability of error with respect to the $n^{\alpha/3}$ additive approximation of DIST_α, implied by $\min_{i_k}(\sum_{j=i_k}^{i_k+n^\alpha-1} x_j)$. Denote by β_{i_k}, the total noise added to the $n^{2\alpha/3}$ sums of the intervals, starting at I_k, that is, $\beta_{i_k} = \sum_{j=k}^{k+n^{2\alpha/3}} Z_k$. We consider the probability that for some k it holds that $\beta_{i_k} < -\gamma$, where $\gamma = dn^{\alpha/3}\sqrt{2R}$ is a bound on the approximation's error. Given k, this probability at most $\frac{\exp(-\frac{d^2}{2})}{d\sqrt{2\pi}}$ since the noise we add to this summation has standard deviation $\sigma = n^{\alpha/3}\sqrt{2R}$. Hence, the probability that at least for one summation we add noise of magnitude $-\gamma$ is at most $\frac{\exp(-\frac{d^2}{2})}{d\sqrt{2\pi}} n^{1-\frac{\alpha}{3}} = \frac{\exp\left(\frac{2\ln n(1-\frac{\alpha}{3})-d^2}{2}\right)}{d\sqrt{2\pi}}$.

To conclude, we showed a simple 3 round protocol for DIST_α. This protocol demonstrates two advantages of the paradigm of choosing what and how together. First, we choose an approximation of DIST_α (i.e., we compute the minimum of subsequences starting at a beginning of an interval). This approximation reduces the communication in the protocol. Second, we leak information beyond the output of the protocol, as p_1 learns the sums S_k's[5].

7.4 Lowerbounds on the Error of Binary Sum and GAP-TR in the Local Model

In this section we consider protocols in the local model computing the binary sum, and prove that in any such ℓ-round, ε-differentially private protocol the additive error cannot be less than $\sqrt{n}/\tilde{O}(\ell)$. To this end, we consider protocols in the local model for computing the GAP-TR function, as defined in Definition 7.2.12. Specifically, we prove that such a protocol in the local model can only compute $\mathrm{GAP\text{-}TR}_{0,\tau}$ for $\tau = \Omega(\sqrt{n}/\tilde{O}(\ell))$ and obtain the impossibility result for SUM using Claim 7.2.13.

Towards this goal, we show that there are two inputs for which the curator sees similar distributions on the messages, thus, has to return similar answers. However, one input contains $\Omega(\sqrt{n})$ ones and the other is the all-zero input, and the algorithm errs on at least one of the inputs. We will prove the existence of such input with $\Omega(\sqrt{n})$ ones, by considering a distribution $\mathcal{A}$ on inputs and later proving that such input taken from the distribution $\mathcal{A}$ exists.

[5] One can use the techniques of [22] to avoid leaking these sums while maintaining a constant number of rounds, however the resulting protocol is less efficient.

Notation 7.4.1. *Let* $\alpha \triangleq \frac{1}{\varepsilon\sqrt{dn}}$ *for* d *to be determined later (*d *is a function of* ℓ *– the number of rounds in the protocol). Define the distribution* $\mathcal{A}$ *on inputs from* $\{0,1\}^n$ *as follows: a vector* $\mathbf{x} = (x_1, \ldots, x_n)$ *is chosen, where* $x_i = 1$ *with probability* α *and* $x_i = 0$ *with probability* $(1 - \alpha)$ *(each input* x_i *is chosen independently).*

From here on, we use X to identify the random variable representing the input and X_i for its ith coordinate. When considering the random variable over $\mathcal{A}$, we use the notation $\Pr_{\mathcal{A}}[\cdot]$. For a set of views of the curator D, we use the notation $\Pr_{\mathcal{A}}[D]$ to denote the probability of the event that the view of the curator falls in D when X is generated according to the probability distribution $\mathcal{A}$.

Recall that our goal is to prove lowerbounds on τ for differentially-private protocols computing the function $\mathrm{GAP\text{-}TR}_{0,\tau}$ in the local model. Notice that we take $\kappa = 0$, namely, we want to prove that the curator cannot distinguish between the all-zero input and inputs of weight at least τ (for small values of τ). Towards this goal, in Section 7.4.1, we analyze properties of non-interactive differentially private protocols in the local model, and show that a curator, trying to distinguish between input chosen according to distribution $\mathcal{A}$ and the all zero input, fails with constant probability. In Section 7.4.2 we generalize this analysis to interactive protocols in the local model. In Section 7.4.3, we complete the proof of the lowerbound on the gap-threshold function in the local model.

7.4.1 Differentially Private Protocols in the Non-Interactive Local Model

We consider protocols in the non-interactive local model where each party holds an input $x_i \in \{0,1\}$ and independently applies an algorithm S_i (also called a sanitizer) before sending the sanitized result c_i to the curator. More formally, we want to prove that if each S_i is 2ε-differentially private for some $0 < \varepsilon \leq 1^6$, then the curator errs with constant probability when trying to distinguish between an input chosen according to distribution $\mathcal{A}$ and the input vector 0^n. We remark that here we consider protocols that are 2ε-differentially private as this is required in the analysis of Section 7.4.2, i.e., for proving lowerbounds for interactive protocols.

We denote for every possible view $\mathbf{c} = (c_1, \ldots, c_n)$ of the curator C.

$$r(\mathbf{c}) \triangleq \frac{\Pr_{\mathcal{A}}\left[\mathrm{View}_C(\mathbf{X}) = \mathbf{c}\right]}{\Pr\left[\mathrm{View}_C(\mathbf{0}) = \mathbf{c}\right]} \quad \text{and} \quad r_i(c_i) \triangleq \frac{\Pr_{\mathcal{A}}\left[S_i(X_i) = c_i\right]}{\Pr\left[S_i(0) = c_i\right]}. \tag{7.2}$$

[6] We can replace the condition $\varepsilon \leq 1$ by a condition that $\varepsilon \leq \varepsilon_0$ for any constant $\varepsilon_0 \geq 1$. This would change some of the constants in the calculations below.

Note that in a non-interactive protocol, $\mathrm{View}_C(\mathbf{x})$ is the random variable which is a concatenation of the messages C receives from $p_1, \ldots, p_n$ with private inputs $\mathbf{x} = (x_1, \ldots, x_n)$, i.e., $(S_1(x_1), \ldots, S_n(x_n))$. Thus, $r(\mathbf{c}) = \prod_{i=1}^{n} r_i(c_i)$. We next show that for views of the curator c that are likely with inputs selected according to $\mathcal{A}$, the ratio $r(c)$ is bounded by a constant. I.e., with constant probability (over the choice of c), we get a constant ratio between the probabilities of c being the view of the curator when the protocol is executed with $\mathbf{x}$, selected according to $\mathcal{A}$, and when the protocol is executed with $\mathbf{0}$.

Define a random variable $\mathbf{C} = (C_1, \ldots, C_n)$ where $C_i = S_i(X_i)$ and X_i is chosen according to the distribution $\mathcal{A}$. In Lemma 7.4.4 we bound $\mathrm{Pr}_{\mathcal{A}}[r(\mathbf{C}) > \delta]$ using the Hoeffding bound. Towards proving this bound, we define the random variables $V_i \triangleq \ln r_i(C_i)$. For every $\eta > 0$, we have that

$$\Pr_{\mathcal{A}}[r(\mathbf{C}) > \eta] = \Pr_{\mathcal{A}}\left[\prod_{i=1}^{n} r_i(C_i) > \eta\right] = \Pr_{\mathcal{A}}\left[\sum_{i=1}^{n} V_i > \ln \eta\right], \tag{7.3}$$

where the first equality holds since the X_is are chosen independently. To apply the Hoeffding bound, we need to show that each variable V_i is bounded, and to compute the expectation of V_i. Both tasks are achieved using the 2ε-differential privacy of the sanitizers, that is,

$$e^{-2\varepsilon} \leq \frac{\Pr[S_i(1) = c_i]}{\Pr[S_i(0) = c_i]} \leq e^{2\varepsilon}. \tag{7.4}$$

Lemma 7.4.2. *For every i and for any $0 < \varepsilon \leq 1$, with probability 1,*

1. $1 - 2\alpha\varepsilon \leq r_i(c_i) \leq 1 + 4\alpha\varepsilon$
2. $-4\alpha\varepsilon \leq V_i \leq 4\alpha\varepsilon$.

Recall that $\alpha = \frac{1}{\varepsilon\sqrt{dn}}$ (see Notation 7.4.1).

Proof. For every i and every value c_i,

$$\begin{aligned} r_i(c_i) &= \frac{\mathrm{Pr}_{\mathcal{A}}[S_i(X_i) = c_i]}{\Pr[S_i(0) = c_i]} \\ &= \frac{\alpha \Pr[S_i(1) = c_i] + (1 - \alpha)\Pr[S_i(0) = c_i]}{\Pr[S_i(0) = c_i]} \\ &= 1 + \alpha \frac{\Pr[S_i(1) = c_i] - \Pr[S_i(0) = c_i]}{\Pr[S_i(0) = c_i]}. \end{aligned}$$

Using $\Pr[S_i(1) = c_i] \leq e^{2\varepsilon}\Pr[S_i(0) = c_i]$ we get, on one hand, that

$$r_i(c_i) \leq 1 + \alpha\frac{\Pr[S_i(0) = c_i]e^{2\varepsilon} - \Pr[S_i(0) = c_i]}{\Pr[S_i(0) = c_i]} = 1 + \alpha(e^{2\varepsilon} - 1) \leq 1 + 4\alpha\varepsilon,$$

since $e^{2x} < 1+4x$ for every $0 < x \leq 1$. Thus, $V_i = \ln r_i(C_i) \leq \ln(1+4\alpha\varepsilon) \leq 4\alpha\varepsilon$, since $\ln(1+x) \leq x$ for every $0 \leq x \leq 1$ (recalling that $\alpha = \frac{1}{\varepsilon\sqrt{dn}}$). Using the fact that $e^{-2\varepsilon}\Pr[S_i(0) = c_i] \leq \Pr[S_i(1) = c_i]$ we get, on the other hand, that

$$r_i(c_i) \;\geq\; 1+\alpha\frac{\Pr[S_i(0)=c_i]e^{-2\varepsilon} - \Pr[S_i(0)=c_i]}{\Pr[S_i(0)=c_i]} \;=\; 1+\alpha(e^{-2\varepsilon}-1) \geq 1-2\alpha\varepsilon,$$

since $1-e^{-2x} < 2x$ for every $0 < x \leq 1$. Thus, $V_i = \ln r_i(C_i) \geq \ln(1-2\alpha\varepsilon) \geq -4\alpha\varepsilon$, since $\ln(1-x) \geq -2x$ for every $0 \leq x \leq 0.5$. □

Lemma 7.4.3. *For every i and for any $0 < \varepsilon \leq 1$, $\mathbb{E}[V_i] \leq 32\alpha^2\varepsilon^2$.*

Proof. In this proof we assume that the output of S_i is a countable set. Denote $B_b \triangleq \{c_i : r_i(c_i) = 1+b\}$ for every $-2\alpha\varepsilon \leq b \leq 4\alpha\varepsilon$ (by Lemma 7.4.2, these are the only values possible for b). Note that by the definition of r_i, for every $c_i \in B_b$

$$\Pr_{\mathcal{A}}[S_i(X_i) = c_i]/\Pr[S_i(0) = c_i] = 1+b.$$

Thus, $\Pr[S_i(0) \in B_b] = \frac{\Pr_{\mathcal{A}}[S_i(X_i)\in B_b]}{1+b} \leq (1-b+2b^2)\Pr_{\mathcal{A}}[S_i(X_i) \in B_b]$. Let $\beta = 2\alpha\varepsilon$. We next bound $\mathbb{E}[V_i]$.

$$\begin{aligned}
\mathbb{E}[V_i] &= \mathbb{E}_{\mathcal{A}}[\ln r(C_i)] = \sum_{-\beta\leq b\leq 2\beta} \Pr_{\mathcal{A}}[S_i(X_i) \in B_b]\ln(1+b) \\
&\leq \sum_{-\beta\leq b\leq 2\beta} \Pr_{\mathcal{A}}[S_i(X_i) \in B_b]b \qquad \text{(since } \ln(1+b) \leq b) \\
&= \sum_{-\beta\leq b\leq 2\beta} \Pr_{\mathcal{A}}[S_i(X_i) \in B_b] - \sum_{-\beta\leq b\leq 2\beta} \Pr_{\mathcal{A}}[S_i(X_i) \in B_b](1-b+2b^2) \\
&\quad + \sum_{-\beta\leq b\leq 2\beta} \Pr_{\mathcal{A}}[S_i(X_i) \in B_b](2b^2) \\
&\leq \sum_{-\beta\leq b\leq 2\beta} \Pr_{\mathcal{A}}[S_i(X_i) \in B_b] - \sum_{-\beta\leq b\leq 2\beta} \Pr[S_i(0) \in B_b] \\
&\quad + \sum_{-\beta\leq b\leq 2\beta} \Pr_{\mathcal{A}}[S_i(X_i) \in B_b](2b^2) \\
&\leq 1-1+8\beta^2 \sum_{-\beta\leq b\leq 2\beta} \Pr_{\mathcal{A}}[S_i(X_i) \in B_b] \;=\; 8\beta^2 = 32\alpha^2\varepsilon^2.
\end{aligned}$$

□

By Lemma 7.4.3, $\mathbb{E}[\sum_{i=1}^n V_i] = \sum_{i=1}^n \mathbb{E}[V_i] \leq 32\alpha^2\varepsilon^2 n = 32/d$. We next prove Lemma 7.4.4 which shows that $\sum_{i=1}^n V_i$ is concentrated around this value.

Lemma 7.4.4. $\Pr_{\mathcal{A}}[r(\mathbf{C}) > \exp(\nu/d)] < \exp(-(\nu-32)^2/32d)$ *for every $\nu > 32$.*

Proof. We apply the Hoeffding bound: Let $V_1, \ldots, V_n$ be independent random variables such that $V_i \in [a, b]$. Then, $\Pr\left[\sum_{i=1}^{n} V_i - \mu \geq t\right] \leq \exp\left(-\frac{2\,t^2}{n(b-a)^2}\right)$ for every $t > 0$ (where $\mu = \sum_{i=1}^{n} \mathbb{E}[V_i]$).

By (7.3), Lemma 7.4.2, Lemma 7.4.3, and by the notation $\alpha = \frac{1}{\varepsilon\sqrt{dn}}$:

$$\begin{aligned}
\Pr_{\mathcal{A}}[r(\mathbf{C}) > \exp(\nu/d)] &= \Pr_{\mathcal{A}}\left[\sum_{i=1}^{n} V_i > \frac{\nu}{d}\right] \\
&= \Pr_{\mathcal{A}}\left[\sum_{i=1}^{n} V_i - \sum_{i=1}^{n} \mathbb{E}V_i > \frac{\nu}{d} - \sum_{i=1}^{n} \mathbb{E}V_i\right] \\
&\leq \Pr_{\mathcal{A}}\left[\sum_{i=1}^{n} V_i - \sum_{i=1}^{n} \mathbb{E}V_i > \frac{\nu}{d} - n \cdot 32\alpha^2\varepsilon^2\right] \\
&\leq \exp\left(-\frac{2\,(\frac{\nu}{d} - n \cdot 32\alpha^2\varepsilon^2)^2}{64\,n\,\alpha^2\,\varepsilon^2}\right) \\
&= \exp\left(-(\nu - 32)^2/32d\right).
\end{aligned}$$

□

The following corollary is a rephrasing of Lemma 7.4.4 that follows from the definition of r in (7.2).

Corollary 7.4.5. *Let Π be a 2ε-private, non-interactive, local protocol, for $0 < \varepsilon \leq 1$. Assume we execute Π with input $\mathbf{x}$, sampled according to distribution $\mathcal{A}$ and denote the view of the curator in that execution by $\mathbf{c}$. Then, for every $\nu > 32$ with probability at least $1 - \exp(-(\nu - 32)^2/32d)$, over the choice of x and the random inputs of the parties,*

$$\Pr_{\mathcal{A}}[\mathrm{View}_C(\mathbf{Z}) = \mathbf{c}] \leq \exp(\nu/d) \Pr[\mathrm{View}_C(\mathbf{0}) = \mathbf{c}],$$

where in the left hand side the probability is taken over the choice of Z according to the distribution $\mathcal{A}$ and the randomness of the sanitizers and in the right hand side the probability is taken over the randomness of the sanitizers.

7.4.2 Differentially Private Protocols in the Interactive Local Model

In this section we generalize Corollary 7.4.5 to interactive local protocols where each party holds an input $x_i \in \{0, 1\}$ and communicates with the curator in rounds. Towards this goal, we decompose an ℓ-round ε-differentially private protocol into ℓ protocols in the non-interactive local model, and prove that each protocol is 2ε-differentially private. Thus, we can apply Corollary 7.4.5 to each protocol, and then apply a composition lemma. Intuitively, we show that an ℓ-round protocol Π in the local model can be viewed as a composition of ℓ protocols, for each of which the curator cannot always distinguish between the case that inputs are all zeros and the case that inputs are sampled according to distribution $\mathcal{A}$. We view the original protocol as a protocol between the curator and a single party, simulating all n parties. In this protocol the curator's goal is to determine if inputs are all zero or sampled according to $\mathcal{A}$. We then apply the following composition lemma to show that the curator's success probability does not increase by too much as ℓ grows.

A Composition Lemma

We consider interactive protocols, where a (deterministic) curator C sends adaptive queries to a single party p holding a private input $x \in \{0,1\}$ in a similar setup to that of the local model (only we make no requirement for ε-differential privacy), i.e., for $0 \leq i \leq \ell$, in the first phase of round i the curator sends p a message $q_i = C(i, \mathcal{V}_1, \ldots, \mathcal{V}_{i-1})$ computed over the transcript of messages $\mathcal{V}_1, \ldots, \mathcal{V}_{i-1}$ previously received from p. We denote by A_i the randomized algorithm that is defined by the message q_i; in the second phase of round i party p computes $\mathcal{V}_i = A_i(x)$ (using fresh random coins for each A_i) and sends $\mathcal{V}_i$ to C.

Definition 7.4.6. *We say that a possible outcome $\mathcal{V}$ is ε-good for algorithm A if $\Pr[A(1) = \mathcal{V}] \leq e^{\varepsilon} \Pr[A(0) = \mathcal{V}]$, where the probabilities are taken over the randomness of algorithm A. An algorithm A is (ε, δ)-good if $\Pr[A(1)$ is ε-good for $A] \geq 1-\delta$, where the probability is taken over the randomness of A.*

Let Π be a protocol, as defined above, in which for any transcript of messages $\mathcal{V}_1, \ldots, \mathcal{V}_{i-1}$, sent by p in previous rounds, C replies with a query q_i, defining an (ε, δ)-good algorithm A_i. Define a randomized algorithm $\hat{A}$ that simulates the interaction between p and C, i.e., given input $x \in \{0,1\}$ it outputs a transcript $(A_1, \mathcal{V}_1, A_2, \mathcal{V}_2, \ldots, A_\ell, \mathcal{V}_\ell)$ sampled according to $\Pi(x)$.

Lemma 7.4.7. *$\hat{A}$ is $(\ell\varepsilon, 1-(1-\delta)^\ell)$-good.*

Proof. With probability at least $(1-\delta)^\ell$, the result of $\hat{A}(1)$ is a transcript

$$\hat{\mathcal{V}} = (A_1, \mathcal{V}_1, A_2, \mathcal{V}_2, \ldots, A_\ell, \mathcal{V}_\ell)$$

such that $\mathcal{V}_i$ is ε-good for A_i for all $i \leq \ell$. It suffices, hence, to prove that when that happens the transcript $\hat{\mathcal{V}}$ is $\ell\varepsilon$-good for $\hat{A}$, and indeed: $\Pr[\hat{A}(1) = (A_1, \mathcal{V}_1, A_2, \mathcal{V}_2, \ldots, A_\ell, \mathcal{V}_\ell)] = \prod_{i=1}^{\ell} \Pr[A_i(1) = \mathcal{V}_i] \leq \prod_{i=1}^{\ell} e^{\varepsilon} \cdot \Pr[A_i(0) = \mathcal{V}_i] = e^{\ell\varepsilon} \cdot \Pr[\hat{A}(0) = (A_1, \mathcal{V}_1, A_2, \mathcal{V}_2, \ldots, A_\ell, \mathcal{V}_\ell)]$. □

Using the composition lemma

Lemma 7.4.8. *Let Π be an ℓ-round, local, ε-differentially private protocol. Suppose we execute Π with an input vector $\mathbf{x}$, sampled according to distribution $\mathcal{A}$, and set $\mathbf{c}$ to be the view of the curator C in that execution. Then, for every $\nu > 32$, with probability at least $1 - \ell \exp\left(-(\nu - 32)^2/32d\right)$, over the choice of x and the random inputs of the parties,*

$$\Pr_{\mathcal{A}}[\mathrm{View}_C(\mathbf{Z}) = \mathbf{c}] \leq \exp\left(\ell\nu/d\right) \Pr[\mathrm{View}_C(\mathbf{0}) = \mathbf{c}],$$

where in the left side the probability is taken over the choice of Z according to the distribution $\mathcal{A}$ and the randomness of the sanitizers and in the right side the probability is taken over the randomness of the sanitizers.

Proof. Fix an ℓ-round, ε-differentially private, local protocol $\mathcal{P}$. Recall that in the interactive local model, a protocol is composed of ℓ-rounds where in each round the curator sends a query to each party and the party sends an answer.

Our first goal is to make the parties stateless. Fix a party p_i. First, we assume that in interaction j the curator sends all queries and answers $q_1, a_1, \dots, a_{j-1}, q_j$ it sent and received from p_i in previous rounds[7]. Second, we assume that party p_i chooses a fresh random string in each round, that is, in round j, party p_i chooses with uniform distribution a random string that is consistent with the queries and answers it got in the previous rounds (since we assume that the parties are unbounded, such choice is possible). Party p_i uses this random string to answer the jth query. In other words, we can consider p_i as applying an algorithm A_j to compute the jth answer; this algorithm depends on the previous queries and answers and uses an independent random string r_j.

We next claim that A_j is 2ε-differentially private. That is, we claim that the probability that a_j is generated given the previous queries and answers is roughly the same when p_i holds the bit 0 and when p_i holds the bit 1. For a transcript c of the first j rounds between p_i and the curator C and for $x_i \in \{0, 1\}$, we denote by $R_c^{x_i}$ the set of all random strings r, such that p_i with private input x_i and random input r sends at each round messages according to c, provided it received all messages according to c in previous rounds. Recall that $\Pr[r_j \in R_c^{x_i}]$ is denoted $\alpha_i^c(x_i)$. Let $c_j = q_1, a_1, \dots, q_{j-1}, a_{j-1}, q_j, a_j$ be messages sent in the first j rounds and let $c'_j = q_1, a_1, \dots, q_{j-1}, a_{j-1}, q_j$ be the prefix of c_j without the jth round answer message a_j (that is $c_j = c'_j, a_j$). Note that since r_j must be consistent with the c'_j, it holds for every $x_i \in \{0, 1\}$ that $\Pr[A_j(x_1) = a_j] = \Pr[r_j \in R_{c_j}^{x_1} | r_j \in R_{c'_j}^{x_1}]$. We therefore need to show that

$$e^{-2\varepsilon} \leq \frac{\Pr[A_j(1) = a_j]}{\Pr[A_j(0) = a_j]} \quad = \quad \frac{\Pr[r_j \in R_{c_j}^1 | r_j \in R_{c'_j}^1]}{\Pr[r_j \in R_{c_j}^0 | r_j \in R_{c'_j}^0]} \leq e^{2\varepsilon}, \tag{7.5}$$

[7] To simplify notation, we omit the subscript i from the queries and answers.

To show that, we use the following two facts, which follow from Definition 7.2.9:

$$e^{-\varepsilon} \leq \frac{\alpha_i^{c_j}(1)}{\alpha_i^{c_j}(0)} = \frac{\Pr[r_j \in R^1_{c_j}]}{\Pr[r_j \in R^0_{c_j}]} \leq e^{\varepsilon}, \tag{7.6}$$

and

$$e^{-\varepsilon} \leq \frac{\alpha_i^{c'_j}(1)}{\alpha_i^{c'_j}(0)} = \frac{\Pr[r_j \in R^1_{c'_j}]}{\Pr[r_j \in R^0_{c'_j}]} \leq e^{\varepsilon}. \tag{7.7}$$

Hence, we have

$$\begin{aligned} r \triangleq \frac{\Pr[A_j(1) = a_j]}{\Pr[A_j(0) = a_j]} &= \frac{\Pr[r_j \in R^1_{c_j} \wedge r_j \in R^1_{c'_j}]}{\Pr[r_j \in R^1_{c'_j}]} \cdot \frac{\Pr[r_j \in R^0_{c'_j}]}{\Pr[r_j \in R^0_{c_j} \wedge r_j \in R^0_{c'_j}]} \\ &= \frac{\Pr[r_j \in R^1_{c_j}]}{\Pr[r_j \in R^1_{c'_j}]} \cdot \frac{\Pr[r_j \in R^0_{c'_j}]}{\Pr[r_j \in R^0_{c_j}]} = \frac{\alpha_i^{c_j}(1)}{\alpha_i^{c_j}(0)} \cdot \frac{\alpha_i^{c'_j}(0)}{\alpha_i^{c'_j}(1)}. \end{aligned}$$

By using the right inequality in (7.6) and the left inequality in (7.7), we get that $r \leq e^{2\varepsilon}$ and similarly, by using the left inequality in (7.6) and the right inequality in (7.7), we get that $r \geq e^{-2\varepsilon}$. Thus, the answers of the n parties in round j are 2ε-private, and we can apply Corollary 7.4.5 to the concatenation of the n answers.

We now use the above protocol to construct a protocol $\mathcal{P}_1$ between a single party, holding a one bit input x and a curator. Throughout the execution of the protocol the party simulates all n parties as specified by the original protocol $\mathcal{P}$ (i.e., sends messages to the curator with the same distribution as the n parties send them). If the bit of the party in $\mathcal{P}_1$ is 1 it chooses the n input bits of the n parties in $\mathcal{P}$ according to distribution $\mathcal{A}$. If the bit of the party in $\mathcal{P}_1$ is 0 it chooses the n input bits of the n parties in $\mathcal{P}$ to be the all-zero vector. By Corollary 7.4.5 we can apply the composition lemma – Lemma 7.4.7 – to the composition of the ℓ non-interactive, 2ε-differentially private protocols and the lemma follows. □

Corollary 7.4.9. *Let $0 < \varepsilon \leq 1$. For every $\nu > 32$ and for every set D of views in an ℓ-round, ε-differentially private, local protocol,*

$$\Pr[\mathrm{View}_C(\mathbf{0}) \in D] \geq \frac{\Pr_{\mathcal{A}}[\mathrm{View}_C(\mathbf{X}) \in D] - \ell \exp\left(-(\nu - 32)^2/32d\right)}{\exp\left(\ell\nu/d\right)}.$$

Proof. Let

$$D_1 = \left\{\mathbf{c} \in D : \Pr_{\mathcal{A}}[\mathrm{View}_C(\mathbf{X}) = \mathbf{c}] \leq \exp\left(\ell\nu/d\right)\Pr[\mathrm{View}_C(\mathbf{0}) = \mathbf{c}]\right\}$$

and

$$D_2 = \left\{\mathbf{c} \in D : \Pr_{\mathcal{A}}[\mathrm{View}_C(\mathbf{X}) = \mathbf{c}] > \exp\left(\ell\nu/d\right)\Pr[\mathrm{View}_C(\mathbf{0}) = \mathbf{c}]\right\}.$$

That is, $D_2 = D \setminus D_1$. By Lemma 7.4.8,

$$\Pr_{\mathcal{A}}[\mathrm{View}_C(\mathbf{X}) \in D_2] \leq \ell \exp\left(-(\nu - 32)^2/32d\right).$$

Furthermore, $\Pr[\mathrm{View}_C(\mathbf{0}) \in D_1] \geq \frac{\Pr_{\mathcal{A}}[\mathrm{View}_C(\mathbf{X}) \in D_1]}{\exp(\ell\nu/d)}$. Thus,

$$\begin{array}{rcl} \Pr[\mathrm{View}_C(\mathbf{0}) \in D] & \geq & \Pr[\mathrm{View}_C(\mathbf{0}) \in D_1] \\ & \geq & \dfrac{\Pr_{\mathcal{A}}[\mathrm{View}_C(\mathbf{X}) \in D_1]}{e^{\ell\nu/d}} \\ & = & \dfrac{\Pr_{\mathcal{A}}[\mathrm{View}_C(\mathbf{X}) \in D] - \Pr_{\mathcal{A}}[\mathrm{View}_C(\mathbf{X}) \in D_2]}{e^{\ell\nu/d}} \\ & \geq & \dfrac{\Pr_{\mathcal{A}}[\mathrm{View}_C(\mathbf{X}) \in D] - \ell e^{-(\nu-32)^2/32d}}{e^{\ell\nu/d}}. \end{array}$$

□

7.4.3 Completing the Lowerbound for $\mathrm{GAP\text{-}TR}_{0,\tau}$ and SUM in the Local Model

In this section we complete the proof that in any ℓ-round, ε-differentially private, local protocols for the gap-threshold function, namely, $\mathrm{GAP\text{-}TR}_{0,\tau}$, the curator errs with constant probability when $\tau \ll \sqrt{n}$ and ℓ is small. Recall that for proving this result, we construct the distribution $\mathcal{A}$ which chooses each bit in the input independently at random where it is one with probability α and zero with probability $1 - \alpha$. Lemma 7.4.10, which follows from a standard Chernoff bound argument, states that when generating a vector $(X_1, \ldots, X_n)$ according to $\mathcal{A}$, the sum $\sum_{i=1}^n X_i$ is concentrated around its expected value, which is αn.

Lemma 7.4.10. $\Pr_{\mathcal{A}}\left[\sum_{i=1}^n X_i \leq (1-\gamma)\alpha n\right] < \exp\left(-\frac{\gamma^2\sqrt{n}}{2\varepsilon\sqrt{d}}\right)$ *for every* $0 \leq \gamma < 1$.

Proof. We apply the following Chernoff bound: Given n zero-one random variables $X_1, \ldots, X_n$ and $0 < t < 1$, $\Pr\left[\sum_{i=1}^n X_i \leq (1-t)\mu\right] < \exp\left(-\frac{t^2\mu}{2}\right)$, where $\mu = \sum_{i=1}^n \mathbb{E}[X_i]$. In our case, $\mu = \frac{\sqrt{n}}{\varepsilon\sqrt{d}}$. Thus, $\Pr_{\mathcal{A}}\left[\sum_{i=1}^n X_i \leq (1-\gamma)\alpha n\right] < \exp\left(-\frac{\gamma^2\sqrt{n}}{2\varepsilon\sqrt{d}}\right)$. □

By Corollary 7.4.9 the distributions on the outputs when the input vector is taken from $\mathcal{A}$ and when it is the all zero vector, are not far apart. By Lemma 7.4.10, with high probability the number of ones in the inputs distributed according to $\mathcal{A}$ is fairly big. These facts are used in Theorem 7.4.11 to prove the lowerbound.

Theorem 7.4.11. *Let $0 < \varepsilon \leq 1$. There exist constants $c > 0$ and $p > 0$ such that in any ℓ-round, ε-differentially private, local protocol for computing $\mathrm{GAP\text{-}TR}_{0,\tau}$ for $\tau = c\frac{\sqrt{n}}{\varepsilon\ell\sqrt{\log \ell}}$ the curator errs with probability at least p.*

Proof. Fix any ℓ-round, ε-differentially private, local protocol for computing $\mathrm{GAP\text{-}TR}_{0,\tau}$. Recall that in the local model the curator is assumed to be deterministic. Thus, the curator, having seen its overall view of the execution of the protocol c, applies a deterministic algorithm G to c, where $G(c)$ is the output of the protocol (which supposed to answer $\mathrm{GAP\text{-}TR}_{0,\tau}(x_1, \dots, x_n)$ correctly). Let $\tau = 0.5\alpha n = \sqrt{n}/(2\varepsilon\sqrt{d})$. We denote $D \triangleq \{\mathbf{c} : G(\mathbf{c}) = 1\}$, that is, D contains all vectors of communication for which the curator answers 1. There are two cases. If the probability of D under the distribution $\mathcal{A}$ is small, then the curator has a big error when the inputs are distributed according to $\mathcal{A}$. Otherwise, by Corollary 7.4.9, the probability of D when the inputs are all zero is big, hence the curator has a big error when the inputs are the all-zero vector. Formally, there are two cases:

Case 1: $\Pr_{\mathcal{A}}[D] < 0.99$**.** We consider the event that the sum of the inputs is at least $\tau = 0.5\alpha n$ and the curator returns an answer 0, that is, the curator errs. We next prove that when the inputs are distributed according to $\mathcal{A}$ the probability of the complementary of this event is not too big. By the union bound the probability of the complementary event is at most $\Pr_{\mathcal{A}}\left[\sum_{i=1}^{n} X_i < 0.5\alpha n\right] + \Pr_{\mathcal{A}}[D]$. By Lemma 7.4.10,

$$\Pr_{\mathcal{A}}[D] + \Pr_{\mathcal{A}}\left[\sum_{i=1}^{n} X_i < 0.5\alpha n\right] \leq 0.99 + \exp\left(-0.25\sqrt{n}/(2\varepsilon\sqrt{d})\right) \approx 0.99.$$

Thus, in this case, with probability ≈ 0.01 the curator errs.

Case 2: $\Pr_{\mathcal{A}}[D] \geq 0.99$**.** In this case, we consider the event that the input is the all-zero vector and the curator answers 1, that is, the curator errs. We next prove using Corollary 7.4.9 that when the inputs are all zero, the probability of this event is bounded away from 0 when taking $\nu = \theta(\ell \log \ell)$ and $d = \ell\nu = \theta(\ell^2 \log \ell)$,

$$\Pr[\mathrm{View}_C(\mathbf{0}) \in D] \;\geq\; \frac{\Pr_{\mathcal{A}}[D] - \ell \exp\left(-(\nu - 32)^2/32d\right)}{\exp\left(\ell\nu/d\right)} > \frac{0.99 - 0.5}{\exp\left(1\right)} > 0.01.$$

Thus, in this case, with probability at least 0.01, the curator errs. As $d = \theta(\ell^2 \log \ell)$, we get that $\tau = \sqrt{n}/(2\varepsilon\sqrt{d}) = \theta(\sqrt{n}/(\varepsilon\ell\sqrt{\log \ell}))$. □

We now state the lowerbound for SUM_n which follows from Theorem 7.4.11 by applying the local model variant of Claim 7.2.13.

Corollary 7.4.12. *Let $0 < \varepsilon \leq 1$. There exist constants $\delta > 0$ and $p > 0$ such that in any ℓ-round, ε-differentially private, local protocol for computing SUM_n the curator errs with probability at least p by at least $\frac{\delta\sqrt{n}}{\varepsilon\ell\sqrt{\log \ell}}$.*

Proof. Let Π be an ℓ-round, ε-differentially private, local protocol for computing SUM_n, for which the curator errs by at most $\tau = \frac{\delta\sqrt{n}}{\varepsilon\ell\sqrt{\log \ell}}$ with probability at most p. By Claim 7.2.13 there exists an ℓ-round, ε-differentially private, local protocol for computing $\mathrm{GAP\text{-}TR}_{0,2\tau}$ errs with probability at most p. By Theorem 7.4.11 no such protocol exists. □

7.5 Lowerbounds for Binary Sum and GAP-TR in the Distributed Model

We prove that in any ℓ-round, fixed-communication, (t, ε)-differentially private protocol computing the binary sum with additive error less than $\sqrt{n}/\tilde{O}(\ell)$, the number of messages sent in the protocol is $\Omega(nt)$. We say that two parties are adjacent if they communicate in the fixed-communication protocol. In the heart of our proof is the more general observation that in the information theoretic setting, a party that has at most t adjacent parties must protect its privacy with respect to the set of its adjacent parties, since this set separates it from the rest of the parties. Thus, any such party, is essentially as limited as any party participating in a protocol in the local communication model.

7.5.1 Reduction to the Local Model

We start with the transformation of a distributed protocol, using a small number of messages to a protocol in the local model.

Lemma 7.5.1. *If there exists an ℓ-round, fixed communication, (t, ε)-differentially private protocol that (γ, τ)-approximates SUM_n sending at most $n(t+1)/4$ messages, then there exists an $(\ell + 1)$-round, ε-differentially private protocol in the local model that (γ, τ)-approximates $\mathrm{SUM}_{n/2}$.*

Proof. The intuition behind the proof is that in the information theoretic model if an input of a party affects the output, then the set parties adjacent to this party must learn information on its input. Recall that a party in a protocol Π is lonely if it communicates with at most t other parties and it is popular otherwise (see Section 7.2). If a party p_i is lonely, then it has most t adjacent parties, and from the privacy requirement in (t, ε)-differentially private protocols, they are not allowed to learn "too much" information on the input of p_i. Therefore, the inputs of lonely parties cannot affect the output of the protocol by too much, thus, since there are many lonely parties, the protocol must have a large error.

Formally, assume that there is a distributed protocol Π satisfying the conditions in the lemma. As the protocol sends at most $n(t+1)/4$ messages, the protocol uses at most $n(t+1)/4$ channels. Since each channel connects two parties, there are at least $n/2$ lonely parties. We will construct a protocol in the local model which (γ,τ)-approximates $\mathrm{SUM}_{n/2}$ in two stages: We first construct a protocol $\mathcal{P}$ in the local model which (γ,τ)-approximates SUM_n and only protects the privacy of the lonely parties. We next fix the inputs of the popular parties and obtain a protocol $\mathcal{P}'$ for $n/2$ parties that protects the privacy of all parties.

First Stage. We convert the distributed protocol Π to a protocol $\mathcal{P}$ in the local model as follows: Recall that in the local model each round consists of two phases where in the first phase the curator sends queries to the parties and in the second phase parties send the appropriate responses. We hence have a single round in $\mathcal{P}$ for every round of Π such that every message m that Party p_j sends to Party p_k in round i in protocol Π, Party p_j sends m to the curator in round i and the curator sends m to Party p_k in the first phase of round $i+1$. Finally, at the end of the protocol Party p_1 sends the output to the curator.

We next prove that $\mathcal{P}$ protects the privacy of lonely parties. Without loss of generality, let p_1 be a lonely party, let T be the set of size at most t of parties that are adjacent to p_1, and let $R=\{p_1,\ldots,p_n\}\setminus(T\cup\{p_1\})$. See Figure 7.1 for a description of these sets. Fix any neighboring vectors of inputs $\mathbf{x}$ and $\mathbf{x}'$ which differ on x_1. The view of the curator in $\mathcal{P}$ contains all messages sent in the protocol. It suffices to prove that for every view v,

$$\Pr[\mathrm{View}_{\mathcal{C}}^{\mathcal{P}}(\mathbf{x})=v] \quad\leq\quad e^{\varepsilon}\cdot\Pr[\mathrm{View}_{\mathcal{C}}^{\mathcal{P}}(\mathbf{x}')=v] \tag{7.8}$$

(by simple summation it will follow for every set of views $\mathcal{V}$).

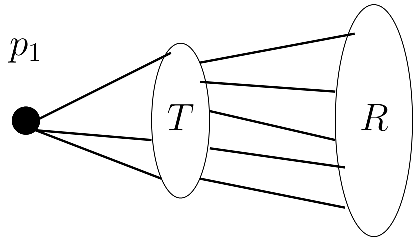

Figure 7.1: The partition of the parties to sets.

Fix a view v of the curator. For a set A, define α_A and α'_A as the probabilities in Π that in each round the set A with inputs from $\mathbf{x}$ and $\mathbf{x}'$ respectively sends messages according to v if it gets messages according to v in previous rounds (these probabilities are taken over the random inputs of the parties in A). Observe that if $p_1 \notin A$, then $\alpha_A = \alpha'_A$ (since $\mathbf{x}$ and $\mathbf{x}'$ only differ on x_1). By simulating p_1, T, R by three parties and applying Lemma 7.2.7, and by the construction of $\mathcal{P}$ from Π

$$\begin{aligned}\Pr\left[\mathrm{View}_{\mathcal{C}}^{\mathcal{P}}(\mathbf{x}) = v\right] &= \alpha_{\{p_1\}} \cdot \alpha_T \cdot \alpha_R, \quad \text{and} \\ \Pr\left[\mathrm{View}_{\mathcal{C}}^{\mathcal{P}}(\mathbf{x}') = v\right] &= \alpha'_{\{p_1\}} \cdot \alpha'_T \cdot \alpha'_R = \alpha'_{\{p_1\}} \cdot \alpha_T \cdot \alpha_R.\end{aligned}$$

Thus, proving (7.8) is equivalent to proving that

$$\alpha_{\{p_1\}} \leq e^{\varepsilon} \alpha'_{\{p_1\}}. \tag{7.9}$$

We use the (t, ε)-privacy of protocol Π to prove (7.9). Let v_T be the messages sent and received by the parties in T in v. As T separates p_1 from R, the messages in v_T are all messages in v except for the messages exchanged between parties in R. The view of T includes the inputs of T in $\mathbf{x}$, the messages v_T, and the random inputs $\mathbf{r_T} = \{r_i : p_i \in T\}$. For a set A, define β_A and β'_A as the probabilities that in Π in each round the set A with inputs from $\mathbf{x}$ and $\mathbf{x}'$ respectively sends messages according to v_T if it gets messages according to v_T in previous rounds. Note that $\beta_{\{p_1\}} = \alpha_{\{p_1\}}$ and $\beta'_{\{p_1\}} = \alpha'_{\{p_1\}}$ by the definition of $\mathcal{P}$. By simulating p_1, T, R by three parties, where the random inputs of T are fixed to $\mathbf{r_T}$, and by Lemma 7.2.7,

$$\Pr[\mathrm{View}_T^{\Pi}(\mathbf{x}) = (\mathbf{x_T}, \mathbf{r_T}, v_T)] = \alpha_{\{p_1\}} \cdot \beta_R, \tag{7.10}$$

and

$$\Pr[\mathrm{View}_T^{\Pi}(\mathbf{x}') = (\mathbf{x_T}, \mathbf{r_T}, v_T)] = \beta'_{\{p_1\}} \cdot \beta'_R = \alpha'_{\{p_1\}} \cdot \beta_R. \tag{7.11}$$

(recalling that $\mathbf{x_T} = \mathbf{x'_T}$). The above probabilities are taken over the random strings of R and p_1 when the random strings of T are fixed to $\mathbf{r_T}$. The (t, ε)-differential privacy of Π implies that

$$\Pr[\mathrm{View}_T^{\Pi}(\mathbf{x}) = (\mathbf{x_T}, \mathbf{r_T}, v_T)] \leq e^{\varepsilon} \Pr[\mathrm{View}_T^{\Pi}(\mathbf{x}') = (\mathbf{x_T}, \mathbf{r_T}, v_T)]. \tag{7.12}$$

Thus, by Equations 7.10, 7.11, and 7.12, $\alpha_{\{p_1\}} \leq e^{\varepsilon} \alpha'_{\{p_1\}}$ and therefore $\mathcal{P}$ is ε-differentially private with respect to inputs of lonely parties.

Second Stage. There are at least $n/2$ lonely parties in Π; w.l.o.g., parties $p_1, \ldots, p_{n/2}$ are lonely. We construct a protocol $\mathcal{P}'$ for (γ, τ)-approximating $\mathrm{SUM}_{n/2}$ by executing Protocol $\mathcal{P}$ where (i) Party p_i, where $1 \leq i \leq n/2$, with input x_i sends messages in $\mathcal{P}'$ as Party p_i with input x_i sends them in $\mathcal{P}$; and (ii) Party p_1 in $\mathcal{P}'$ simulates all other $n/2$ parties in $\mathcal{P}$, that is, for every $n/2 < i \leq n$, it chooses a random input r_i for p_i and in every round it sends to the curator the same messages as p_i would send with $x_i = 0$ and r_i. Since the curator sees the same view in $\mathcal{P}$ and $\mathcal{P}'$ and since the privacy of lonely parties is protected in $\mathcal{P}$, the privacy of each of the $n/2$ parties in $\mathcal{P}'$ is protected. Protocol $\mathcal{P}'$, therefore, (γ, τ)-approximates $\mathrm{SUM}_{n/2}$ (since we fixed $x_i = 0$ for $n/2 < i \leq n$ and $\mathcal{P}'$ returns the same output distribution of Π, which (γ, τ)-approximates SUM_n for all possible inputs). □

We are now ready to state the main theorem of this section.

Theorem 7.5.2. *Let* $0 < \varepsilon \leq 1$. *There exist constants* $\delta > 0$ *and* $\gamma > 0$ *such that in any* ℓ*-round, fixed-communication,* (t, ε)*-differentially private protocol for approximating* SUM_n *that sends at most* $n(t+1)/4$ *messages the protocol errs with probability at least* γ *by at least* $\frac{\delta\sqrt{n}}{\varepsilon(\ell+1)\sqrt{\log(\ell+1)}}$.

Proof. Assume to the contrary, that there is an ℓ-round, (t, ε)-differentially private protocol Π for computing SUM_n, which sends at most $n(t+1)/4$ messages and errs by at most $\tau = \frac{\delta\sqrt{n}}{\varepsilon(\ell+1)\sqrt{\log(\ell+1)}}$ with probability at least $1 - \gamma$. By Lemma 7.5.1 there exists an $\ell + 1$-round, ε-differentially private, local protocol $\mathcal{P}$ for computing $\mathrm{SUM}_{n/2}$ which errs by at most $\tau = \frac{\delta\sqrt{n}}{\varepsilon(\ell+1)\sqrt{\log(\ell+1)}} = \frac{\sqrt{2}\delta\sqrt{n/2}}{\varepsilon(\ell+1)\sqrt{\log(\ell+1)}}$ with probability at least $1 - \gamma$. This contradicts Corollary 7.4.12. □

Theorem 7.5.2 suggests that whenever we require that the error of a differentially private protocol for approximating SUM to be of magnitude smaller than $\sqrt{n/\varepsilon}$, there is no reason to relinquish the simplicity and modularity suggested by the natural paradigm for constructing protocols. In this case, it is possible to construct relatively simple efficient SFE protocols, which use $O(nt)$ messages, and compute an additive $(O(1/\varepsilon), O(1))$-approximation of SUM. We thus observe a phase transition threshold at $\theta(\sqrt{n/\varepsilon})$ magnitude of error.

Remark 7.5.3. It can also be shown that in any ℓ-round, fixed-communication, (t, ε)-differentially private protocol computing the $\mathrm{GAP\text{-}TR}_{\kappa,\tau}$, for any $0 \leq \kappa \leq n - \tau$ the number of messages sent in the protocol is $\Omega(nt)$, for $\tau = \sqrt{n}/\tilde{O}(\ell)$. To show this, use the ideas similar to those of Lemma 7.5.1 and apply Theorem 7.4.11 to assert that any ℓ-round, fixed-communication, (t, ε)-differentially private protocol computing the $\mathrm{GAP\text{-}TR}_{0,\tau}$, the number of messages sent in the protocol is $\Omega(nt)$, for $\tau = \sqrt{n}/\tilde{O}(\ell)$. Then, using Claim 7.2.14, infer that the same is true for general κ.

7.6 SFE for Symmetric Approximations of Binary-Sum

In this section we show the advantage of using the alternative paradigm for constructing distributed differentially private protocols whenever we allow an $O(\sqrt{n}/\varepsilon)$ approximation. Recall that it is possible to construct differentially private protocols for such approximations that use $2n$ messages and are secure against coalitions of size up to $t = n-1$ (see Section 7.3.1). We next prove, using ideas from Chor and Kushilevitz [20], that any SFE protocol for computing a symmetric approximation for SUM_n, using less than $nt/4$ messages, has error magnitude $\Omega(n)$.

Remark 7.6.1. We remark that allowing $O(nt)$ messages, it is fairly straightforward to construct a (t, ε)-differentially private protocol with constant additive error, using the natural paradigm. That is, first selecting an ε-private approximation, say the one described in Example 6.3.1, and then constructing a t-secure protocol for computing it.

We first give the definition of SFE protocols computing a given randomized function $\hat{f}(\cdot)$. Here, again, we only consider protocols where all parties are honest-but-curious and compute the *same* output. The definition is given in the information-theoretic model, a definition of SFE in the computational model can be found in [45].

Definition 7.6.2 (SFE). *Let $\hat{f} : (\{0,1\}^*)^n \to \{0,1\}^*$ be an n-ary randomized function. Let Π be an n-party protocol for computing $\hat{f}$. For a coalition $T \subseteq \{1,\ldots,n\}$, the view of T during an execution of Π on $\mathbf{x} = (x_1 \ldots x_n)$, denoted $\mathrm{View}_T(\mathbf{x})$, is defined as in Definition 7.2.1, i.e., $\mathrm{View}_T(x_1,\ldots,x_n)$ is the random variable containing the inputs of the parties in T (that is, $\{x_i\}_{i\in T}$), the random inputs of the parties in T, and the messages that the parties in T received during the execution of the protocol with inputs $\mathbf{x} = (x_1,\ldots,x_n)$.*

We say that Π is a t-secure protocol for $\hat{f}$ if there exist a randomized function, denoted S, such that for every $t' \leq t$ and for every $T = \{i_1,\ldots,i_{t'}\}$ as above and for every input vector $\mathbf{x} = (x_1 \ldots x_n)$, the following two random variables are identically distributed,

- $\left\{S\left(T, \left(x_{i_1},\ldots,x_{i_{t'}}\right), o\right), o\right\}$ *where, o is obtained first by sampling $\hat{f}(\mathbf{x})$ and then S is applied to $\left(T, \left(x_{i_1},\ldots,x_{i_{t'}}\right), o\right)$.*
- $\left\{\mathrm{View}_T(\mathbf{x}), \mathrm{Output}^{\Pi}(\mathrm{View}_T(\mathbf{x}))\right\}$ *where $\mathrm{Output}^{\Pi}(\mathrm{View}_T(\mathbf{x}))$ denotes the output during the execution represented in $\mathrm{View}_T(\mathbf{x})$.*

Definition 7.6.3 (Symmetric randomized function). *We say that a randomized function $\hat{f}$ over domain D with range R is* symmetric if it does not depend on the ordering on the coordinates of the input. I.e., for every $(x_1, \ldots, x_n) \in D^n$ and every permutation $\pi : [n] \to [n]$ the distributions (over R) implied by $\hat{f}(x_1, \ldots, x_n)$ and by $\hat{f}(x_{\pi(1)}, \ldots, x_{\pi(n)})$ are identical.

Lemma 7.6.4. *Let $\hat{f}$ be a symmetric function approximating SUM_n such that for every input vector $\mathbf{x}$, it holds that $\left|\hat{f}(\mathbf{x}) - \mathrm{SUM}(\mathbf{x})\right| < n/4$, and let $t \leq n-2$. Every fixed-communication t-secure protocol Π for computing $\hat{f}$ uses at least $n(t+1)/4$ messages*[8].

Proof. Let Π be a t-secure protocol computing $\hat{f}$ using less than $n(t+1)/4$ messages. There are at least $\frac{n}{2}$ lonely parties in Π. The intuition for the proof is that a lonely party does not affect the computation, since the set of parties adjacent to it, being smaller than $t+1$, would be able to infer information about its input. The proof is given in two steps. In the first step, we show that for any given lonely party p_i, for any fixed inputs for all other parties, and for any transcript c of the protocol, the probability of c being the transcript of the protocol when $x_i = 0$ is exactly the same as the probability of c being the transcript of the protocol when $x_i = 1$. In the second step of the proof, we use this to show that with probability at least $1/2$, the protocol errs by $n/4$.

Without loss of generality, assume p_1 is lonely and assume p_2 is not adjacent to p_1. Let T be the set of parties adjacent to p_1 and let $R = \{p_1, \ldots, p_n\} \backslash (T \cup \{p_1\})$ (e.g., $p_2 \in R$). Recall that for a transcript c we denote by $\alpha_1^c(x_1)$, the probability that p_1 is consistent with c with input x_1, namely, the probability that p_1 with input x_1 sends at each round messages according to c, provided it received all messages according to c in previous rounds. Our goal in the first part of the proof is to prove that for any transcript of the protocol c, it holds that $\alpha_1^c(0) = \alpha_1^c(1)$. Towards this end, we pursue the following proof structure. We first consider two inputs $\mathbf{z}$ and $\mathbf{y}$ such that $\mathrm{SUM}(\mathbf{z}) = \mathrm{SUM}(\mathbf{y})$, but $y_1 = 0$ while $z_1 = 1$ and consider the restriction of c to its intersection c_T with the view of T (in other words, we think of c_T as the protocol with three parties, p_1, T, and R, implied by c). We use Definition 7.6.3 to prove that the probability of c_T is the same with $\mathbf{z}$ and with $\mathbf{y}$. We then use Lemma 7.2.7 to present these probabilities as a product of $\alpha_1^{c_T}(x_1), \alpha_T^{c_T}(\mathbf{x_T})$ and $\alpha_R^{c_T}(\mathbf{x_R})$, where $\mathbf{x_T}$ (respectively, $\mathbf{x_R}$) are the inputs of parties in T (respectively, in R). We then assert, by considering all prefices of c_T, that each factor of these two multiplications is the same in both cases and hence $\alpha_1^c(0) = \alpha_1^{c_T}(0) = \alpha_1^{c_T}(1) = \alpha_1^c(1)$.

Fix any inputs $x_3, \ldots, x_n$ for the parties $p_3, \ldots, p_n$. Let $\mathbf{y}$ be the input vector such that $y_k = x_k$ for $k > 2$, $y_1 = 0$, and $y_2 = 1$ and let $\mathbf{z}$ be the input vector

[8] We note that the lemma does not hold for non-symmetric functions. For example, we can modify the bit flip protocol described in Section 7.3.1 to an SFE protocol for a non-symmetric function, retaining the number of messages sent (but not their length): in Step (2) p_1 (acting as the curator) also sends $\mathbf{z} = (z_1, \ldots, z_n)$, and in Step (3) each p_i locally outputs $\hat{f} + \mathbf{z}2^{-n}$, treating $\mathbf{z}$ as an n-bit binary number.

such that $z_k = x_k$ for $k > 2$, $z_1 = 1$, and $z_2 = 0$. We first claim that the distribution over the views of T when the protocol is executed with $\mathbf{y}$ is the same as when the protocol is executed with $\mathbf{z}$. I.e., we claim that for any possible view v_T of the set T, it holds that, $\Pr\left[\mathrm{View}_T(\mathbf{y}) = v_T\right] = \Pr\left[\mathrm{View}_T(\mathbf{z}) = v_T\right]$. This is true since the two random variables $\left\{S\left(T, \left(y_{i_1}, \ldots, y_{i_{t'}}\right), o\right), o\right\}$ and $\left\{S\left(T, \left(z_{i_1}, \ldots, z_{i_{t'}}\right), o\right), o\right\}$ (as defined in Definition 7.6.2) are identically distributed, since $\mathrm{SUM}(\mathbf{y}) = \mathrm{SUM}(\mathbf{z})$ and since $\hat{f}$ is symmetric. Hence, it holds by the t-security of the protocol that both $\left\{\mathrm{View}_T(\mathbf{y}), \mathrm{Output}^{\Pi}(\mathrm{View}_T(\mathbf{y}))\right\}$ and $\left\{\mathrm{View}_T(\mathbf{z}), \mathrm{Output}^{\Pi}(\mathrm{View}_T(\mathbf{z}))\right\}$ are also identically distributed. Thus, since the view of T contains the transcript c_T of messages sent between parties in T and parties in $\{p_1\} \cup R$, we have that for any such possible transcript c_T, the probability that parties send messages according to c_T is the same when the protocol is executed with $\mathbf{y}$ and when the protocol is executed with $\mathbf{z}$. Furthermore, for any possible prefix c'_T of any transcript c_T of T, the probability of messages sent according to c'_T when executing Π with input $\mathbf{y}$ is the same as when executing Π with input $\mathbf{z}$. This is true as this probability is merely the sum over the probabilities of all transcripts completing c'_T.

Without loss of generality, we can analyze the execution of the protocol as if at each round only a single message is sent by a single party. Let j be such that p_1 sends a message in round j and denote by $h_j = h_{j-1}, m_j$ the, prefix of c_T also viewed by p_1 (messages sent or received by p_1), in the first j rounds, where h_{j-1} is the history of messages viewed by p_1 in the first $j-1$ rounds, and m_j is the message p_1 sends in round j, according to c_T. By the argument above, the probabilities of h_{j-1} being seen by p_1 are the same when the protocol is executed with $\mathbf{y}$ and when the protocol is executed with $\mathbf{z}$ and the probabilities of h_j being seen by p_1 are the same when the protocol is executed with $\mathbf{y}$ and when the protocol is executed with $\mathbf{z}$.

Therefore, for any given history of messages h_{j-1} viewed by p_1, and for any possible message m_j of p_1, the probabilities of p_1 sending m_j having seen message history h_{j-1} are the same when $x_1 = 0$ and when $x_1 = 1$. Thus, since the probability of p_1 being consistent with a view c_T (of T) is the product of the probabilities that it is consistent at each round, we have $\alpha_1^{c_T}(0) = \alpha_1^{c_T}(1)$. Let c be a full transcript of the protocol, and c_T be its restriction to messages sent between parties in T and parties in $\{p_1\} \cup R$. Since p_1 does not see any message in c that is not in c_T, it holds for every x_1 that $\alpha_1^{c}(x_1) = \alpha_1^{c_T}(x_1)$. Thus, $\alpha_1^{c}(0) = \alpha_1^{c}(1)$.

Hence, we proved that for any lonely party p_i, and any full transcript of the protocol c, it holds that $\alpha_i^{c}(0) = \alpha_i^{c}(1)$. Consider the all zero input vector and the input vector $\mathbf{x}$ such that $x_i = 1$ if and only if p_i is lonely. By Lemma 7.2.7 we have that for any given full transcript c, the probability of c being exchanged with $\mathbf{0}$ is exactly the probability of c being exchanged with $\mathbf{x}$. Thus, if with probability at least $1/2$, when executing the protocol with $\mathbf{0}$, the exchanged transcript implies a value less than $n/4$, then with probability at least $1/2$, the protocol errs by at least $n/4$ when executed with $\mathbf{x}$. Otherwise, with probability at least $1/2$, the protocol errs by at least $n/4$ when executed with $\mathbf{0}$. $\square$

Chapter 8

Conclusions and Future Work

8.1 Phase Transition for Ackermannian Ramsey Numbers

We have proved sharp phase transition thresholds for the regressive and Paris-Harrington Ramsey numbers for pair colorings, similar to the thresholds obtained in [81] for provability in PA. Although the proofs for these results are quite different, it might be interesting to see that they can be motivated by a unifying underlying phase transition principle. As it turned out, finite combinatorics provides bounds (on finite Ramsey numbers) which also provide good bounds on regressive and Paris-Harrington Ramsey numbers below the threshold. Indeed these calculations provide a priori guesses of where the desired thresholds might be located.

In our examples it turned out that the guesses were good, since for parameter functions growing faster than the threshold function, a suitable iteration argument shows that the induced Ramsey functions have extraordinary growth. In vague analogy with dynamic systems, one might consider the threshold region as an unstable fixed point of a renormalization operator given by the bounds on finitary Ramsey numbers. It is interesting to consider more settings in which this paradigm for locating threshold points can be applied.

In Chapter 3 we also use our construction to obtain an incomprehensibly large lowerbound of $A_{53}(2^{2^{274}})$ on the Id-regressive Ramsey number of $k = 82$, where A_{53} is the 53-rd approximation of Ackermann's function.

Finding explicit lower and upper bounds for Ramsey numbers is not just a matter of aesthetics and intellectual challenge. These numbers are so hard to grasp that even finding lower and upper bounds usually requires some profound understanding of their behavior. Thus, looking for more explicit bounds for regressive and Paris-Harrington Ramsey numbers is an important open problem.

Caicedo explores this direction in [16], where he improves our lowerbound.

He also gives simpler arguments for the fact that $R^{\text{reg}}_{\text{Id}}$ grows exactly as the Ackermann function (note that our result only states that it grows at least with Ackermannian rate). His results are indeed obtained by introducing a new approach for the investigation of regressive Ramsey numbers. However, Caicedo only considers the case of $g = \text{Id}$. Applying Caicedo's approach with other functions g may prove beneficial to better understanding the threshold behavior of g-regressive Ramsey numbers.

8.2 Iteration Hierarchies

We have proved phase transition behaviors of functions defined via diagonalization from an iteration hierarchy (of Grzegorczyk type). These transitions are obtained by parameterizing both the start function g and the modulus function h, by which the diagonalization is defined. Specifically, fixing $g(x) = x+1$ (as in the Ackermann hierarchy), we showed a sharp threshold on h at which the resulting hierarchy stops being primitive recursive and becomes Ackermannian. Furthermore, we have shown this threshold to be intrinsically related to g-regressive Ramsey numbers.

For a class of start functions g (starting with $g(x) = x + \varepsilon$ for $0 < \varepsilon \leq 1$ and constantly increasing them), we showed appropriate classes of iteration modulus h for which the resulting classes of hierarchies are slow-growing and very close classes of iteration modulus h for which the resulting classes of hierarchies are fast-growing.

In general we expect that sharp phase transition thresholds can be obtained for any start function $g(x) = A_d(x)$, and we expect that the resulting thresholds are all different. It would be interesting to cover this material and structural stability of resulting phase transitions.

Another natural generalization of our work would be considering phase transition thresholds for the transfinite extensions of the Ackermann hierarchy, which is also known as Schwichtenberg-Wainer hierarchy [15, 73, 76]. We expect essentially that stepping up in the ordinals by one power of ω will allow for one additional iteration of the binary logarithm function in the threshold function.

The functions g_l considered in Chapter 5 seem to appear frequently in weak arithmetic. It seems to be of general interest to explore possible connections.

8.3 Distributed Differential Privacy

We initiated an examination of the paradigm where an analysis and the protocol for computing it are chosen simultaneously. We showed examples that present the potential benefits of using this paradigm: it can lead to simpler protocols and, more importantly, it can lead to more efficient protocols. We examined this paradigm with respect approximations to the binary SUM, and observed a phase transition depending on the magnitude of additive error we allow.

For the upper bound, we observed that for approximations with additive error $\approx \sqrt{n}$ there is a gain: it is possible to construct differentially private protocols that are much more efficient than any SFE protocol for a function in this class. Moreover, these differentially private constant-round protocols are secure against coalitions of size up to $t = n - 1$, and need not rely on secure channels. In the process we proved a generalization of the result of Chor and Kushilevitz [20], who showed a lower bound on the communication complexity of an SFE protocol for computing SUM. We show a similar lowerbound for any symmetric approximation of SUM (with sub-linear error).

The main result presented in Chapter 7 is a lowerbound on the communication complexity of any low-communication protocol for computing differentially private analyses approximating SUM. Our lowerbound is first proved for protocols in the local model using a small number of rounds. We then extend the result to the distributed model, by showing that low-communication protocols in the distributed model for computing SUM are not more powerful than protocols in the simple local model.

The upper and lower bounds are tight and indicate a sharp threshold at additive error $\tilde{\theta}(\sqrt{n})$ for protocols computing $\mathrm{SUM}(\cdot)$ using at most a logarithmic number of rounds. These results also yield a separation between the local model and the global model (where it is possible to compute analyses approximating SUM within constant additive error). They also yield a separation between the computational and the information theoretic settings, since under computational assumptions on the parties, it is possible to construct protocols for computing analyses approximating SUM within constant additive error, which use $2n$ messages (even in the local model).

We view our results as part of the common effort towards a full characterization of *what* can be privately computed in each communication model. Our work leaves open the question of whether interaction (for more than a constant number of rounds) can help in approximating SUM in the local model. We believe it would be very interesting to extend the discussion to the malicious and/or computational settings.

Bibliography

[1] H. L. Abbott. A note on Ramsey's theorem. *Canad. Math. Bull.*, 15:9–10, 1972.

[2] D. Achlioptas and C. Moore. Random k-sat: Two moments suffice to cross a sharp threshold. *SIAM J. Comput.*, 36(3):740–762, 2006.

[3] N. Alon and J. Spencer. *The Probabilistic Method*. Wiley, New York, 1992.

[4] T. Arai. On the slowly well orderedness of ε_0. *Math. Log. Q.*, 48:125–130, 2002.

[5] B. Barak, K. Chaudhuri, C. Dwork, S. Kale, F. McSherry, and K. Talwar. Privacy, accuracy, and consistency too: a holistic solution to contingency table release. In *Proc. of the twenty-sixth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, pages 273–282, New York, NY, USA, 2007. ACM.

[6] A. Beimel, K. Nissim, and E. Omri. Distributed private data analysis: Simultaneously solving how and what. In D. Wagner, editor, *Advances in Cryptology – CRYPTO 2008*, volume 5157 of *Lecture Notes in Computer Science*, pages 451–468. Springer-Verlag, 2008.

[7] M. Ben-Or, S. Goldwasser, and A. Wigderson. Completeness theorems for noncryptographic fault-tolerant distributed computations. In *Proc. of the 20th ACM Symp. on the Theory of Computing*, pages 1–10, 1988.

[8] P. F. Blanchard. On regressive Ramsey numbers. *J. Combin. Theory Ser. A*, 100(1):189–195, 2002.

[9] A. Blum, C. Dwork, F. McSherry, and K. Nissim. Practical privacy: the SuLQ framework. In *Proc. of the twenty-fourth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, pages 128–138, 2005.

[10] A. Blum, K. Ligett, and A. Roth. A learning theory approach to non-interactive database privacy. In *Proc. of the 40th ACM Symp. on the Theory of Computing*, pages 609–618, New York, NY, USA, 2008. ACM.

[11] B. Bollobás, S. Janson, and O. Riordan. The phase transition in inhomogeneous random graphs. *Random Struct. Algorithms*, 31(1):3–122, 2007.

[12] B. Bollobás and A. Thomason. Threshold functions. *Combinatorica*, 7(1):35–38, 1987.

[13] J. Bourgain and G. Kalai. Influences of variables and threshold intervals under group symmetries. *Funct. Anal*, 7:438–461, 1997.

[14] A. Bovykin. Brief introduction to unprovability. Manuscript, 2005.

[15] W. Buchholz, A. Cichon, and A. Weiermann. A uniform approach to fundamental sequences and hierarchies. *Math. Log. Q.*, 40:273–286, 1994.

[16] A. E. Caicedo. Regressive functions on pairs. Preprint, 2007.

[17] C. Calude. *Theories of computational complexity*, volume 35. Annals of Discrete Mathematics, North-Holland, Amsterdam, 1988.

[18] L. Carlucci, G. Lee, and A. Weiermann. Classifying the phase transition threshold for regressive Ramsey functions. Preprint, 2006.

[19] D. Chaum, C. Crépeau, and I. Damgård. Multiparty unconditionally secure protocols. In *Proc. of the 20th ACM Symp. on the Theory of Computing*, pages 11–19, 1988.

[20] B. Chor and E. Kushilevitz. A communication-privacy tradeoff for modular addition. *Inform. Process. Lett.*, 45(4):205–210, 1993.

[21] C. Cooper and A. Frieze. Multi-coloured Hamilton cycles in random edge-coloured graphs. *Comb. Probab. Comput.*, 11(2):129–133, 2002.

[22] I. Damgård, M. Fitzi, E. Kiltz, J. B. Nielsen, and T. Toft. Unconditionally secure constant-rounds multi-party computation for equality, comparison, bits and exponentiation. In S. Halevi and T. Rabin, editors, *Proc. of the Third Theory of Cryptography Conference – TCC 2006*, volume 3876 of *Lecture Notes in Computer Science*, pages 285–304. Springer-Verlag, 2006.

[23] I. Dinur and K. Nissim. Revealing information while preserving privacy. In *Proceedings of the Twenty-Second ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*, pages 202–210, 2003.

[24] C. Dwork. Differential privacy. In M. Bugliesi, B. Preneel, V. Sassone, and I. Wegener, editors, *Proc. of the 33rd International Colloquium on Automata, Languages and Programming*, volume 4052 of *Lecture Notes in Computer Science*, pages 1–12. Springer-Verlag, 2006.

[25] C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor. Our data, ourselves: Privacy via distributed noise generation. In *Advances in Cryptology – EUROCRYPT 2006*, pages 486–503. Springer, 2006.

[26] C. Dwork and J. Lei. Differential privacy and robust statistics. In *Proc. of the 41st ACM Symp. on the Theory of Computing*, 2009.

[27] C. Dwork, F. McSherry, K. Nissim, and A. Smith. Calibrating noise to sensitivity in private data analysis. In S. Halevi and T. Rabin, editors, *Proc. of the Third Theory of Cryptography Conference – TCC 2006*, volume 3876 of *Lecture Notes in Computer Science*, pages 265–284. Springer-Verlag, 2006.

[28] C. Dwork, F. McSherry, and K. Talwar. The price of privacy and the limits of LP decoding. In *Proc. of the 39th ACM Symp. on the Theory of Computing*, pages 85–94, 2007.

[29] C. Dwork, M. Naor, O. Reingold, G. Rothblum, and S. Vadhan. When and how can data be efficiently released with privacy? In *Proc. of the 41st ACM Symp. on the Theory of Computing*, 2009.

[30] C. Dwork and K. Nissim. Privacy-preserving datamining on vertically partitioned databases. In M. Franklin, editor, *Advances in Cryptology – CRYPTO 2004*, volume 3152 of *Lecture Notes in Computer Science*, pages 528–544. Springer-Verlag, 2004.

[31] C. Dwork and S. Yekhanin. New efficient attacks on statistical disclosure control mechanisms. In D. Wagner, editor, *Advances in Cryptology – CRYPTO 2008*, volume 5157 of *Lecture Notes in Computer Science*, pages 469–480. Springer-Verlag, 2008.

[32] P. Erdős, A. Hajnal, A. Máté, and R. Rado. *Combinatorial set theory: partition relations for cardinals*, volume 106. Studies in Logic and the Foundations of Mathematics, North-Holland, Amsterdam, 1984.

[33] P. Erdős and G. Mills. Some bounds for the Ramsey-Paris-Harrington numbers. *J. Combin. Theory Ser. A*, 30(1):53–70, 1981.

[34] P. Erdős and R. Rado. A combinatorial theorem. *J. London Math. Soc.*, 25:249–255, 1950.

[35] P. Erdős and R. Rado. Combinatorial theorems on classifications of subsets of a given set. *Proc. London Math. Soc. (3)*, 2:417–439, 1952.

[36] P. Erdős and A. Rényi. On random graphs. I. *Publ. Math. Debrecen*, 6:290–297, 1959.

[37] P. Erdős and A. Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 5:17–61, 1960.

[38] A. Evfimievski, J. Gehrke, and R. Srikant. Limiting privacy breaches in privacy preserving data mining. In *Proceedings of the Twenty-Second ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, pages 211–222, 2003.

[39] D. Feldman, A. Fiat, H. Kaplan, and K. Nissim. Private coresets. In *Proc. of the 41st ACM Symp. on the Theory of Computing*, 2009.

[40] E. Friedgut. Sharp thresholds of graph properties, and the k-sat problem. *J. Amer. Math. Soc*, 12:1017–1054, 1999.

[41] E. Friedgut, G. Kalai, and C. J. N. Kahn. Every monotone graph property has a sharp threshold. *Proc. Amer. Math. Soc*, 124:2993–3002, 1996.

[42] E. Friedgut and M. Krivelevich. Sharp thresholds for certain Ramsey properties of random graphs. *Random Struct. Algorithms*, 17(1):1–19, 2000.

[43] A. Ghosh, T. Roughgarden, and M. Sundararajan. Universally utility-maximizing privacy mechanisms. In *Proc. of the 41st ACM Symp. on the Theory of Computing*, 2009.

[44] K. Gödel. Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme. *Monatshefte f. Math. u. Phys.*, 38:173–198, 1931.

[45] O. Goldreich. *Foundations of Cryptography, Voume II Basic Applications*. Cambridge University Press, 2004.

[46] O. Goldreich, S. Micali, and A. Wigderson. How to play any mental game. In *Proc. of the 19th ACM Symp. on the Theory of Computing*, pages 218–229, 1987.

[47] R. L. Graham, B. L. Rothschild, and J. H. Spencer. *Ramsey theory*. John Wiley & Sons Inc., New York, second edition, 1990.

[48] j. Kahn and G. Kalai. Thresholds and expectation thresholds. *Comb. Probab. Comput.*, 16(3):495–502, 2007.

[49] A. Kanamori and K. McAloon. On Gödel incompleteness and finite combinatorics. *Ann. Pure Appl. Logic*, 33(1):23–41, 1987.

[50] H. Karloff and U. Zwick. A 7/8-approximation algorithm for max 3sat. In *Proc. of the 38th IEEE Symp. on Foundations of Computer Science*, pages 406–415, 1997.

[51] S. Kasiviswanathan, H. K. Lee, K. Nissim, S. Raskhodnikova, and A. Smith. What can we learn privately? In *Proc. of the 49th IEEE Symp. on Foundations of Computer Science*, pages 531–540, 2008.

[52] J. Ketonen and R. Solovay. Rapidly growing Ramsey functions. *Ann. of Math. (2)*, 113(2):267–314, 1981.

[53] D. Kőnig. Sur les correspondances multivoques des ensembles. *Fund. Math.*, 8:114–134, 1926.

[54] L. Kirby and J. B. Paris. Initial segments of models of Peano's axioms. In *Set Theory and Hierarchy Theory V*, volume 619 of *Lecture Notes in Mathematics*, pages 211–226. Springer, Berlin / Heidelberg, 1977.

[55] M. Kojman, G. Lee, E. Omri, and A. Weiermann. Sharp thresholds for the phase transition between primitive recursive and Ackermannian Ramsey numbers. *Journal of Combinatorial Theory*, 115(Series A):1036 – 1055, August 2008.

[56] M. Kojman and S. Shelah. Regressive Ramsey numbers are Ackermannian. *J. Combin. Theory Ser. A*, 86(1):177–181, 1999.

[57] G. Lee. *Phase Transitions in Axiomatic Thought*. PhD thesis, University of Münster, Germany, 2005.

[58] H. Lefmann and V. Rödl. On canonical Ramsey numbers for coloring three-element sets. In *Finite and infinite combinatorics in sets and logic: Proceedings of the NATO Advanced Study Institute on Finite and Infinite Combinatorics in Sets and Logic*, pages 237–247. Springer, Banff, Alberta, Canada, 1991.

[59] F. McSherry and K. Talwar. Mechanism design via differential privacy. In *Proc. of the 48th IEEE Symp. on Foundations of Computer Science*, pages 94–103, 2007.

[60] I. Mironov, O. Pandey, O. Reingold, and S. P. Vadhan. Computational differential privacy. In S. Halevi, editor, *Advances in Cryptology – CRYPTO 2009*, volume 5677 of *Lecture Notes in Computer Science*, pages 126–142. Springer-Verlag, 2009.

[61] D. Moshkovitz and R. Raz. Two query pcp with sub-constant error. In *Proc. of the 49th IEEE Symp. on Foundations of Computer Science*, pages 314–323, Los Alamitos, CA, USA, 2008.

[62] K. Nissim. Private data analysis via output perturbation a rigorous approach to constructing sanitizers and privacy preserving algorithms. In

C. C. Aggarwal and P. S. Yu, editors, *Privacy-Preserving Data Mining: Models and Algorithms*, volume 34 of *Advances in Database Systems*, pages 383–414. Springer Publishing Company, Incorporated, 2008.

[63] K. Nissim, S. Raskhodnikova, and A. Smith. Smooth sensitivity and sampling in private data analysis. In *Proc. of the 39th ACM Symp. on the Theory of Computing*, pages 75–84, 2007.

[64] E. Omri and A. Weiermann. Classifying the phase transition threshold for Ackermannian functions. *Annals of Pure and Applied Logic*, 2008. doi:10.1016/j.apal.2007.02.004.

[65] J. B. Paris. Some independence results for Peano arithmetic. *J. Symbolic Logic*, 43(4):725–731, 1978.

[66] J. B. Paris and L. Harrington. A mathematical incompleteness in Peano arithmetic. In J. Barwise, editor, *Handbook of Mathematical Logic*, volume 90, pages 1133–1142. North-Holland, 1977.

[67] A. Percus, G. Istrate, and C. Moore. *Computational Complexity and Statistical Physics (Santa Fe Institute Studies in the Sciences of Complexity Proceedings).* Oxford University Press, Inc., New York, NY, USA, 2006.

[68] R. Péter. *Recursive functions.* Academic Press, New York, third edition, 1967.

[69] H. J. Prömel, W. Thumser, and B. Voigt. Fast growing functions based on Ramsey theorems. *Discrete Math.*, 95(1-3):341–358, 1991.

[70] P. Pudlák. A bottom-up approach to foundations of mathematics. Manuscript.

[71] F. P. Ramsey. On a problem of formal logic. *Proc. London Math. Soc.*, 30:264–285, 1930.

[72] V. Rastogi, S. Hong, and D. Suciu. The boundary between privacy and utility in data publishing. In *Proc. of the 33rd International Conf. on Very Large Data Bases*, pages 531–542, 2007.

[73] H. Schwichtenberg. Eine klassifikation der ε_0-rekursiven funktionen. *Z. Math. Logik Grundlagen Math.*, 17:61–74, 1971.

[74] A. Smith. Efficient, differentially private point estimators. Technical Report 0809.4794, CoRR, 2008.

[75] W. van Hoof and A. Weiermann. Sharp phase transition thresholds for the paris harrington ramsey numbers for a fixed dimension. *Proceedings of the American Mathematical Society*, 2009. to appear.

[76] S. S. Wainer. A classification of the ordinal recursive functions. *Archiv für Mathematische Logik und Grundlagenforschung*, 13:136–153, 1970.

[77] S. L. Warner. Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60(309):63–69, 1965.

[78] A. Weiermann. A very slow growing hierarchy for γ_0. In *Logic Colloquium '99*, pages 182–199, 1999.

[79] A. Weiermann. An application of graphical enumeration to PA. *J. Symbolic Logic*, 68(1):5–16, 2003.

[80] A. Weiermann. An application of results by Hardy, Ramanujan and Karamata to Ackermannian functions. *Discrete Mathematics and Computer Science*, 6:133–142, 2003.

[81] A. Weiermann. A classification of rapidly growing Ramsey functions. *Proc. Amer. Math. Soc.*, 132(2):553–561., 2004.

[82] A. Weiermann. Analytic combinatorics, proof-theoretic ordinals and phase transitions for independence results. *Ann. Pure Appl. Logic*, 136:189–218, 2005.

[83] A. Weiermann. An extremely sharp phase transition threshold for the slow growing hierarchy. *Math. Structures Comput. Sci.*, 16(5):925–946, 2006.

[84] A. Weiermann. Phase transition thresholds for some natural subclasses of the computable functions. In A. Beckmann, U. Berger, B. Löwe, and J. V. Tucker, editors, *CiE*, volume 3988 of *Lecture Notes in Computer Science*, pages 556–570. Springer, 2006.

[85] A. Weiermann. Phase transitions for some friedman style independence results. *Math. Log. Q.*, 53(1):4–18, 2007.

[86] A. C. Yao. Protocols for secure computations. In *Proc. of the 23th IEEE Symp. on Foundations of Computer Science*, pages 160–164, 1982.