On gaps between sums of powers and other topics in Number Theory and Combinatorics

Luca Ghidelli

Thesis submitted to the Faculty of Science in partial fulfillment of the requirements
for the degree of
Doctorate in Philosophy Mathematics and Statistics[1]

Department of Mathematics and Statistics
Faculty of Science
University of Ottawa

[1]The Ph.D. program is a joint program with Carleton University, administered by the Ottawa-Carleton Institute of Mathematics and Statistics

Abstract

One main goal of this thesis is to show that for every K it is possible to find K consecutive natural numbers that cannot be written as sums of three nonnegative cubes. Since it is believed that approximately 10% of all natural numbers can be written in this way, this result indicates that the sums of three cubes distribute unevenly on the real line. These sums have been studied for almost a century, in relation with Waring's problem, but the existence of "arbitrarily long gaps" between them was not known. We will provide two proofs for this theorem. The first is relatively elementary and is based on the observation that the sums of three cubes have a positive bias towards being cubic residues modulo primes of the form $p = 1 + 3k$. Thus, our first method to find consecutive non-sums of three cubes consists in searching them among the natural numbers that are non-cubic residues modulo "many" primes congruent to 1 modulo 3. Our second proof is more technical: it involves the computation of the Sato-Tate distribution of the underlying cubic Fermat variety $\{x^3 + y^3 + z^3 = 0\}$, via Jacobi sums of cubic characters and equidistribution theorems for Hecke L-functions of the Eisenstein quadratic number field $\mathbf{Q}(\sqrt{-3})$. The advantage of the second approach is that it provides a nearly optimal quantitative estimate for the size of gaps: if N is large, there are $\gg \sqrt{\log N}/(\log\log N)^4$ consecutive non-sums of three cubes that are less than N. According to probabilistic models, an optimal estimate would be of the order of $\log N/\log\log N$. In this thesis we also study other gap problems, e.g. between sums of four fourth powers, and we give an application to the arithmetic of cubic and biquadratic theta series. We also provide the following additional contributions to Number Theory and Combinatorics: a derivation of cubic identities from a parameterization of the pseudo-automorphisms of binary quadratic forms; a multiplicity estimate for multiprojective Chow forms, with applications to Transcendental Number Theory; a complete solution of a problem on planar graphs with everywhere positive combinatorial curvature.

Résumé

Un des objectifs principaux de cette thèse est de montrer que pour chaque K il existe K nombres naturels consécutifs qui ne peuvent pas s'écrire comme sommes de trois cubes non négatifs. Comme on conjecture qu'environ 10% des nombres naturels peuvent s'écrire de cette façon, ce résultat indique que les sommes de trois cubes ne sont pas réparties de manière uniforme. Ces sommes ont été étudiées pendant près d'un siècle, en relation avec le problème de Waring, mais l'existence d'écarts arbitrairement longs entre elles n'était pas connue. Nous donnons deux preuves de ce théorème. La première est relativement élémentaire et repose sur le fait que les sommes de trois cubes ont un biais positif pour être congrues à un cube modulo les nombres premiers de la forme $p = 1 + 3k$. Ainsi, notre première méthode pour trouver des nombres consécutifs qui ne sont pas sommes de trois cubes, consiste à les rechercher parmi les nombres qui sont des résidus non-cubiques modulo "beaucoup" de premiers congruents à 1 modulo 3. Notre deuxième preuve est plus complexe: elle requiert un calcul de la distribution de Sato-Tate de la variété de Fermat cubique sous-jacente $\{x^3 + y^3 + z^3 = 0\}$, via les sommes de Jacobi des caractères cubiques et les théorèmes d'équidistribution pour les fonctions L de Hecke du corps de nombres d'Eisenstein $\mathbf{Q}(\sqrt{-3})$. L'avantage de cette seconde approche est qu'elle fournit une estimation quantitative de la taille des écarts: si N est grand, il y a $\gg \sqrt{\log N}/(\log\log N)^4$ nombres consécutifs inférieurs à N qui ne sont pas des sommes de trois cubes. Selon les modèles probabilistiques, une estimation optimale serait de l'ordre de $\log N/\log\log N$. Dans cette thèse, nous étudions également d'autres problèmes d'écart, par exemple entre les sommes de quatre puissances quatrièmes, et nous donnons une application de ces résultats à l'arithmétique de séries thêta cubiques et biquadratiques. Cette thèse apporte aussi les contributions suivantes en théorie des nombres et en combinatoire: une dérivation d'identités cubiques à partir d'une paramétrisation des pseudo-automorphismes de formes binaires quadratiques; une estimation de la multiplicité des formes de Chow multiprojectives avec applications à la théorie des nombres transcendantaux; une solution complète d'un problème sur les graphes planaires ayant partout une courbure combinatoire positive.

Dedications and Acknowldgements

Dedico questa tesi alla mia mamma

I would like to express my gratitude to the many people that, directly or indirectly, supported me in the process of writing this thesis.

The biggest thanks go to Damien, my doctoral advisor. You welcomed me very well in Canada and guided my professional development. From you I always received wholehearted support, and I treasure your precious advices. You are also the person that more than anyone was capable to take me by the hand during a difficult period in the last year. Without you this thesis would not have been possible. Thank you.

My work has been influenced by the positive atmosphere I experienced in the various places I lived in, both in Ottawa and in Gatineau. Thank you Gaspard, David, Saad. Thank you José, Saruul, Anuujin, Campbell. Thank you Shu, Kelly. Thank you Jean-Jacques and the others. Thank you Mr.Dee, Grace, Shea, Aalok. Thank you Prem, Claudius. Merci Paul, Saul, merci Lina, John, Mark, Marc, merci tout le monde. Merci Lisette, Christian, Normand, thank you Scott.

Thanks to the people that I met at the University of Ottawa. Thanks to Benoit, Diane, Mayada, Carolynne, Janick and everyone that makes the Department a nice place. Thanks to the people of the research group in Number Theory. Thanks Martin for proposing the problem I discuss in Chapters 3 and 8. Thanks to all my friends among the fellow students in Mathematics and Statistics. Thanks to the MSGSA/AÉDMS. Thanks to my grad mentor Irene Xia Zhou. Thanks to my professors. Thanks to my students. Thanks to everyone.

Grazie agli amici della Banda di Nese. Thanks to the musicians of the Ottawa Pops Orchestra and the UOPO. Grazie Philip. Grazie a tutti i miei amici in Italia e in giro per il mondo. Grazie Ermanno. Grazie alla famiglia Paolini. Спасибо семье Барановых. Thanks to the Al-Shbeil and Al-Shgoor family.

Grazie alla mia famiglia, che rappresenta da sempre un punto fermo della mia esistenza, una certezza che porto dentro di me, dovunque io vada. Grazie ai sette intrepidi che hanno rischiato il congelamento per farmi visita in un freddo Aprile canadese. Grazie a Sara e a papà che hanno nuovamente attraversato l'oceano per assistere alla discussione della mia tesi.

Анюта, я хочу закончить этот абзац, сказав спасибо тебе. Возможно, вы не помогли мне написать этот тезис, но наверняка с вами я написал страницы жизни. Некоторые сны написаны на листьях, которые несёт ветер. Это мы должны поймать. Но самые большие мечты написаны в виде блестящих маленьких звездочек на огромном полотне небосвода. Это часть судьбы.

This work was supported in part by the full International Admission Scholarship of the University of Ottawa and the FGPS, in part by the International Doctoral Scholarship 712230205087, and in part by NSERC.

Contents

Preface

This thesis is divided in 3 parts and 10 chapters including the Introduction. The first two parts cover different aspects of a common topic in Number Theory, relative to the representability of natural numbers as sums of powers of low degree. In the third part instead I discuss two more results in Commutative Algebra and Combinatorics, which I obtained during my PhD but originated from previous research. For the sake of brevity the other research problems I investigated as a doctoral student at uOttawa (for instance `arXiv:1903.03881v1`) are omitted from this thesis.

In Parts 1 and 2 the main focus is the existence of long runs of consecutive nonnegative integers that cannot be written as sums of two squares, three nonnegative cubes or four fourth powers. Part 1, which consists of Chapters 2 to 5, is mostly concerned with the qualitative results that can be proved by elementary techniques. In Part 2, I use instead analytic methods to derive quantitative estimates on the size of gaps between sums of powers, and I provide an application to the study of generalized theta values.

Parts 2 and 3 consist of the following single-authored preprints:

Chapter 6 `arXiv:1910.05070`

Chapter 7 `arXiv:1910.05079`

Chapter 8 `arXiv:1910.05076`

Chapter 9 `arXiv:1912.04047`

Chapter 10 `arXiv:1708.08502v2`

This thesis was written in LaTeX with the template provided by the Department of Mathematics and Statistics of the University of Ottawa. The numerical experiments of Section 3.1 and for the tables of Section 3.3 were conducted by means of ad-hoc programs written in C and Gnumeric spreadsheets.

All images contained in the Introduction or in Chapter 10 are obtained with tikz or hand-drawn, with the exception of Fig. 1.3.2. The colors of Fig. 1.3.5 and Fig. 1.3.6 were added with Paint. The images for Chapter 2 were obtained with Geogebra.

We now describe briefly the content of all the chapters. A more complete description of the problems, methods and results will be given in the Introduction.

Chapter 2 is mostly a survey of results on sums of two squares. There we showcase theorems of Erdős and Richards on the gaps between these numbers, and we describe geometrically and diophantine-theoretically a construction of Bambah, Chowlah, Huxley, Green and Lindqvist.

Chapter 3 is the starting point of the thesis. There we prove that for every K there exist K consecutive natural numbers none of which can be written as a sum of three cubes.

In Chapter 4 we streamline and generalize some computations of the preceding chapter with Gauss-Jacobi sums, and then we prove that there are arbitrarily many consecutive numbers that cannot be written as sums of four fourth powers.

In Chapter 5 we describe the set of linear changes of variables that leave a binary quadratic form invlariant up to scalar multiplication, and then we derive cubic identities such as $9^3 + 10^3 - 12^3 = 1$ or $3^3 + 4^3 + 5^3 = 6^3$. This is joint work with A. Granville.

In Chapter 6 we estimate the size of gaps between the values of polynomials of the form $F(\mathbf{x}) = a_1x_1^k + \cdots + a_kx_k^k$ with $k \in \{3, 4\}$. The proof requires an analysis of the number of solutions to the congruence $F(\mathbf{x}) \equiv 0$ modulo p prime.

In Chapter 7 we use the circle method to prove that in almost all every interval of the form $(N - N^\gamma, N]$ with $\gamma > 4059/16384$ and N large enough there is a sum of four fourth powers.

In Chapter 8 we prove that a number of the form $\sum_{n=1}^{\infty} q^{-n^k}$ with integer $q \geq 2$ and $k \in \{3, 4\}$ is never an algebraic number of degree $\leq k$. The proof uses a "nested gap" technique of Bradshaw and the previous results on sums of cubes and fourth powers.

In Chapter 9 we prove that if $r + 1$ polynomials have N common roots on a multiprojective variety V, then the Chow form of V vanishes with multiplicity at least N at that $(r + 1)$-tuple of polynomials. We derive a multiplicity estimate for resultants for polynomials on commutative algebraic groups that could be used to prove Transcendence and Algebraic Independence results.

In Chapter 10 we prove that every planar graph with everywhere positive combinatorial curvature, which is not a prism or an antiprism, has at most 208 vertices. This result is sharp as there are examples of such graphs with 208 vertices.

Notation and conventions

The following notation will be sometimes used tacitly throughout the thesis.

We denote the set of nonnegative integers by $\mathbb{N} := \{0, 1, \ldots\}$ and the set of positive integers by $\mathbb{N}_+ := \mathbb{N} \setminus \{0\}$. Given a finite set S we denote its cardinality by $\#S$. If f, g are two composable maps, we let $(f \circ g)(t) := f(g(t))$ and we say it is the composition of f and g, or that f is precomposed with g. The notation log denotes the natural logarithm.

The big-O notation has the usual meaning: given two quantities $A(t)$, $B(t)$ parametrized by a variable t we say that $A = O(B)$ if there exists a constant $c > 0$ such that $|A(t)| \leq cB(t)$ for all t for which this makes sense. Similarly, the little-o notation $A = o(B)$ for $t \to \infty$ means that for all $c > 0$ the inequality $|A(t)| \leq cB(t)$ is eventually true in the limit as $t \to \infty$. We use Vinogradov's notation $A \ll B$ or $B \gg A$ as an equivalent of $A = O(B)$. Finally, we use $A \asymp B$ to mean that we have both $A = O(B)$ and $B = O(A)$. We do not give $A \approx B$ a precise meaning and use it only in informal discussion.

Chapter 1

Introduction

This thesis solves three main problems which are presented in reverse chronological order in the form of five papers which are complemented by four additional chapters.

- The first problem is of combinatorial nature. It is treated in Chapter 10 and deals with planar graphs that have positive combinatorial curvature everywhere. This is a research that I started as an undergraduate student in Pisa and which I completed during my PhD years.

- The second problem belongs to commutative algebra. This research is in continuation with my MSc thesis in Pisa, which itself was dealing with transcendental number theory. In this thesis, I developed a generalization of results of Roy concerning the multiplicity of the resultant which I transposed to the multiprojective setting. In Chapter 9 of the current thesis I further extend this to a more general framework that encompasses results of Chardin as well, together with an application in the context of arbitrary commutative algebraic groups.

- Finally the third problem is strongly connected to Waring's problem in number theory. We show the existence of arbitrarily long gaps between sums of three nonnegative cubes as well as between values of cubic and biquadratic diagonal forms. The main tools are explicit formulas for the number of points of Fermat varieties over finite fields together with estimates for L-functions. This is the main part of the thesis. It occupies Chapters 2 to 7. The main result is presented in Chapter 6. It is the culminating point of a sequence of refinements that require the introduction of more tools at each step. The more elementary approaches are described in Chapters 3 and 4. Chapter 7 is a generalization of a result of Daniel that we need for the application in Chapter 8 on the arithmetic of generalized theta functions.

1.1 Gaps between sums of powers

The most fascinating sequences from the beginning of number theory have been: the integers, their squares, their cubes, their fourth powers and the prime numbers. They are very much interconnected. For example, we know that there are squares that can be written as the sum of two positive squares (Pythagoras) and that the prime numbers which are sums of two positive squares are precisely those which are not congruent to 3 modulo 4 (Fermat-Girard). We further know that each integer can be written as the sum of four squares (Lagrange), nine cubes, nineteen fourth powers, etc. In general Waring's problem asks for the smallest integer $g(k)$ (resp. $G(k)$) such that all natural numbers (resp. all sufficiently large natural numbers) can be written as a sum of $g(k)$ (resp. $G(k)$) nonnegative k-th powers. For example, we know that $g(2) = G(2) = 4$, $g(3) = 9$, $4 \leq G(3) \leq 7$, $g(4) = 19$ and $G(4) = 16$ [158]. Thanks to the work of many mathematicians culminating with Wiles and Taylor we also know that for each $k \geq 3$ there is no k-th power that can be written as a sum of two positive k-th powers. In this thesis we are concerned with the integers that can be written as sums of two squares, three nonnegative cubes and four fourth powers. There are many important problems and conjectures connected with these sets.

1.1.1 Sums of two squares

The case that is most understood is that of squares. We know that the set $\mathcal{S}_{2,2}$ of integers that can be written as sums of two squares is closed under multiplication. It consists of all integers n whose prime factors congruent to 3 modulo 4 appear in the prime factorization of n with an even exponent. Using tools of multiplicative number theory Landau was able to estimate its counting function $S_{2,2}(N) := \#(\mathcal{S}_{2,2} \cap [0, N])$: asymptotically, there are $c_{LR}N/\sqrt{\log N}$ sums of two squares $\leq N$, for some constant $c_{LR} > 0$ (called the Landau-Ramanujan constant). This means that the average gap $n_2 - n_1$ between consecutive sums of squares $n_1, n_2 \leq N$ has size $\asymp \sqrt{\log N}$, and in fact it is known that most gaps have exactly this order of magnitude. However, a nice construction of Richards shows that for arbitrarily large N there is an interval $I \subseteq [0, N]$ of length $\gg \log N$ containing no sum of two squares. It is a long-standing open problem to estimate the maximal size of gaps between sums of two squares $\leq N$: the only known upper bound is $\ll N^{1/4}$, which is easy to prove. We provide an extensive discussion about sums of two squares and their gaps in Chapter 2, including recent results of Ben Green, Lindqvist and Huxley.

1.1.2 Sums of three cubes

Since there are approximately $N^{1/3}$ nonnegative cubes less than N, the set of integers that can be written as a sum of two nonnegative cubes is quite sparse: there are only

$O(N^{2/3})$ such numbers up to N. As a matter of fact, we know [76, Theorem 2] that the exact asymptotic count is $(c + o(1))N^{2/3}$, where $c = (1/12)\Gamma^2(1/3)/\Gamma(2/3) \approx 0.44165$. However, unlike the case of squares, we do not have a nice arithmetic description for the sums of two or three cubes, and so the tools of multiplicative number theory do not apply.

It is expected that the set $\mathcal{S}_{3,3}$ of sums of three nonnegative cubes has positive natural density, which means that the counting function $S_{3,3}(N) := \#(\mathcal{S}_{3,3} \cap [0, N])$ satisfies

$$S_{3,3}(N) = (\delta + o(1))N$$

for some $\delta > 0$. This is supported theoretically in [79, 44, 35]. It is even conjectured [36] that $\delta = 0.0999425\ldots$. In this direction Hooley proved, based on serveral conjectures including the Generalized Riemann Hypothesis, that the counting function of this set grows faster than $N^{1-\epsilon}$ for every $\epsilon > 0$. At present we are far from proving this, and the best unconditional result we have, due to Wooley [166], is that there are at least N^{α} such numbers not exceeding N, for all sufficiently large N, with $\alpha \approx 0.916862$.

Although the above suggests that the average gap between elements of $\mathcal{S}_{3,3}$ is bounded (by 11, according to the most optimistic conjectures), I could prove in this thesis that these gaps are in fact unbounded. This is done in Chapter 3 using elementary methods together with Weil's proof of the Riemann Hypothesis over finite fields. The method that I use provides a lower bound for the largest gap in $\mathcal{S}_{3,3} \cap [1, N]$ as a function on N but I did not compute it explicitly as it would be much weaker than the results that I obtained later by adding more advanced techniques. The careful reader may check that this lower bound has the form $(\log\log N)^A$ for some $A > 0$.

By contrast in Chapter 6 our main result has the following consequence.

Theorem 1.1.1. There is a constant $\kappa > 0$ such that for all sufficiently large integer N there exist gaps of length at least

$$\kappa \frac{\sqrt{\log N}}{(\log\log N)^2}$$

between the elements of $\mathcal{S}_{3,3}$ in $[1, N]$.

The probabilistic models for sums of three pseudocubes [35, 44] predict the existence of gaps of size $\gg \log N/\log\log N$ between elements of $\mathcal{S}_{3,3} \cap [1, N]$. Our estimate, therefore, is short by roughly a square root of $\log N$ with respect to what is expected. Previous to this result very little was known about the gaps in the set $\mathcal{S}_{3,3}$ besides a result of Daniel who proves that almost all gaps in $\mathcal{S}_{3,3} \cap [1, N]$ have size less than $N^{17/108+\epsilon}$, for all $\epsilon > 0$ and all N large enough.

1.1.3 Sums of four fourth powers and generalizations

Like for sums of three cubes, we expect that the set $\mathcal{S}_{4,4}$ of sums of four fourth powers has positive density and so the average gap between its elements should be bounded.

In my thesis I prove that there are arbitrarily large gaps. The precise result from Chapter 6 is the following.

Theorem 1.1.2. There is a constant $\kappa' > 0$ such that for all sufficiently large integer N there exist gaps of length at least

$$\kappa' \frac{\log\log\log N}{\log\log\log\log N}$$

between the elements of $\mathcal{S}_{4,4}$ in $[1, N]$.

In this case the result is much weaker than what is predicted by the probabilistic methods. Indeed it is expected in general that for every $k \geq 3$ the set $\mathcal{S}_{k,k}$ of sums of k nonnegative k-th powers has positive density and that there exist gaps of size $\gg \log N/\log\log N$ between the elements of $\mathcal{S}_{k,k} \cap [1, N]$.

A qualitative proof of Theorem 1.1.2 showing simply the unboundedness of gaps is presented in Chapter 4. The method of proof uses a different strategy than for the case of cubes and involves more sophisticated tools. Unfortunatly, it does not extend to sums of higher powers.

Our main results in Chapter 6 deal with values of cubic and biquadratic diagonal forms

$$F(\mathbf{x}) = a_1 x_1^s + \cdots + a_s x_s^s$$

with $s \in \{3, 4\}$ having positive integer coefficients $a_1, \ldots, a_s$. Here by values of $F(\mathbf{x})$ we mean the natural numbers obtained by evaluating the diagonal form at nonnegative integers $x_1, \ldots, x_s \in \mathbb{N}$. With this notation these results read as follows.

Theorem 1.1.3. Suppose that $s = 3$. Then there is a constant $\kappa_F > 0$ such that for all integers N, K satisfying $N > e^e$, $K \geq 2$ and

$$K < \kappa_F \frac{\sqrt{\log N}}{(\log\log N)^2},$$

there exist gaps of length K between the values of $F(\mathbf{x})$ in $[1, N]$.

Theorem 1.1.4. Suppose that $s = 4$ and that $F(\mathbf{x})$ is not equal to

$$a(c_1 x_1)^4 + b(c_2 x_2)^4 + 4a(c_3 x_3)^4 + 4b(c_4 x_4)^4, \tag{1.1.1}$$

for any $a, b, c_1, c_2, c_3, c_4 \in \mathbb{N}_+$, up to a permutation of the variables. Then there is a constant $\kappa_F > 0$ such that for all integers N, K satisfying $N > e^{e^{e^e}}$, $K \geq 2$ and

$$K < \kappa_F \frac{\log\log\log N}{\log\log\log\log N},$$

there are gaps of length at least K between the values of $F(\mathbf{x})$ in $[1, N]$.

The proof is based on the existence of some $\beta > 0$ and of an infinite set $\mathcal{P}_s$ of prime numbers p for which the congruence $F(\mathbf{x}) \equiv 0 \bmod p$ has at most

$$p^{s-1} - \beta p^{s/2} = (1 - \beta p^{1-s/2})p^{s-1} \tag{1.1.2}$$

solutions. For those primes this shows a negative bias towards the number of representations of the zero class as a value of $F(\mathbf{x})$ since p^{s-1} is the average number of solutions of the congruence $F(\mathbf{x}) \equiv m \bmod p$, as m varies. We can even show that the set $\mathcal{P}_s$ is large enough so that

$$\prod_{p \in \mathcal{P}_s} (1 - \beta p^{1-s/2}) = 0.$$

This cannot happen if $s \geq 5$, and this explains the limitation of the method. Moreover we show, in section 4 of Chapter 6, that for $s = 4$ the bias disappears exactly when $F(\mathbf{x})$ is of the form (1.1.1). A detailed outline of the method is provided in section 2 of Chapter 6.

Briefly, what we do is to compute a closed form for the number of representations of $0 \bmod p$ as a value of $F(\mathbf{x})$ in terms of Jacobi sums involving the coefficients $a_1, \ldots, a_s$ of $F(\mathbf{x})$. This can be written as

$$p^{s-1} + p^{s/2-1}(p-1)(K_p + 2\,\mathbb{R}\mathrm{e}\, H_p),$$

where K_p is a finite sum of values of Dirichlet characters while H_p is a value of a unitary Hecke character of the field $\mathbb{Q}(e^{2\pi i/s})$. The term K_p is equal to zero in the case $s = 3$. Otherwise it is relatively easy to calculate using Kummer theory and class field theory. We restrict to the primes for which K_p is minimal. Then on this set of primes I prove equidistribution results for H_p using the theory of Hecke L-functions.

1.1.4 Complements and applications

In 1925, Hardy and Littlewood proposed their Hypothesis K according to which the number of representations of an integer n as a sum of k nonnegative k-th powers is at most n^ϵ for any given $\epsilon > 0$ provided that n is sufficiently large as a function of ϵ. However, in 1936, Mahler disproved this hypothesis for $k = 3$ by constructing explicit integers n for which the number of representations is at least $\lfloor 9^{1/3} n^{1/12} \rfloor$. His formula is based on a remarkable cubic identity due to Gérardin and rediscovered by Mahler. I found a simple interpretation for the Gérardin-Mahler identity and later, in joint work with Granville, we extended this interpretation to more general cubic identities due to Binet and Euler. This is presented in Chapter 5 of this thesis.

For the applications that I give in Chapter 8, I needed an upper bound for the size of almost all gaps between sums of four fourth powers, extending the result of Daniel for cubes mentioned in Section 1.1.2. The result that I obtained is the following.

Theorem 1.1.5. Define $\gamma_0 := 4059/16384 \approx 0.24774$ and let $\gamma > \gamma_0$. Then for almost all $n \in \mathbb{N}$ (in the sense of natural density) there is a sum of four fourth powers in the interval $(n - n^\gamma, n]$.

What I actually needed was this to hold for $\gamma = 0.25$ and I consider myself fortunate for having reached this by a so small margin. The method of proof follows the general strategy of Daniel and is based on the technique of diminishing ranges. This means that we consider only the sums of fourth powers $x_1^4 + \cdots + x_4^4$ where each x_i is restricted to a prescribed interval $(P_i/2, P_i]$, for powers $P_1 \geq P_2 \geq P_3 \geq P_4$ of some parameter N. These sums are analyzed by means of the circle method. Arguing as Daniel, I obtain the following estimate.

Theorem 1.1.6. Let γ_0 be as in Theorem 1.1.5 and let $\gamma_1 := 4992/16384 \approx 0.3046$. Given $N > 0$ and $\gamma_0 < \gamma \leq \gamma_1$, define

$$Y := N^\gamma, \quad P_1 := \sqrt[4]{N}, \quad \text{and} \quad P_{j+1} = P_j^{13/16} \quad \text{for } 1 \leq j \leq 3.$$

For each integer n, let $R(n)$ denote the number of solutions to the equation

$$n = x_1^4 + x_2^4 + x_3^4 + x_4^4 + y$$

subject to

$$0 < y \leq Y, \quad \tfrac{1}{2}P_i < x_i \leq P_i \quad (1 \leq i \leq 4).$$

Then for each $\epsilon > 0$ we have

$$\sum_{\frac{1}{2}N < n \leq N} \left| R(n) - \bar{R}(n) \right|^2 \ll_\epsilon Y N^{1-\gamma_0+\epsilon}, \tag{1.1.3}$$

where $\bar{R}(n) := \frac{1}{32} Y P_2 P_3 P_4 n^{-3/4}$ and where the implied constant depends only on ϵ.

The relation between Theorem 1.1.5 and Theorem 1.1.6 can be explained as follows. If an interval of the form $[n - N^\gamma, n)$ with $N/2 < n \leq N$ contains no sum of four fourth powers, then $R(n) = 0$ and so it contributes $\bar{R}(n)^2 \asymp N^{2\gamma-2\gamma_0}$ to the sum appearing in (1.1.3). Therefore by (1.1.3) the number of such intervals is at most $N^{1-\gamma+\gamma_0+\epsilon} = o(N)$ by choosing ϵ small enough. From there it is relatively easy to deduce Theorem 1.1.5.

The application that I give in the thesis concerns the values of the generalized theta series

$$\theta_\ell(q) = \sum_{n=0}^{\infty} \frac{1}{q^{n^\ell}},$$

where $\ell \geq 2$ is an integer. These series converge for all complex numbers q with $|q| > 1$. For $\ell = 2$, this is essentially, up to a simple renormalization, the well-known theta function, which is a modular form of weight $1/2$. By a theorem of Nesterenko we know that $\theta_2(q)$ is transcendental for every algebraic number q with $|q| > 1$. For $\ell > 2$ very little is known about the values of θ_ℓ, even at the integers $q > 1$. In this thesis I prove the following.

Theorem 1.1.7. Let $\ell \in \{3, 4\}$, let $q \geq 2$ be an integer and suppose that $\theta_\ell(q)$ is algebraic. Then $\deg \theta_\ell(q) \geq \ell + 1$.

To prove this I follow the strategy of Bradshaw from [16], where it is proved that for any integer $\ell, q \geq 2$ the number $\theta_\ell(q)$, if algebraic, has degree at least ℓ. This used the fact that the series

$$(\theta_\ell(q))^k = \sum_{n=0}^{\infty} r_{\ell,k}(n) q^{-n}$$

are lacunary for each $k = 0, 1, \ldots, \ell - 1$. Here the number $r_{\ell,k}(n)$ denotes the number of representations of n as a sum of k nonnegative ℓ-th powers. It is a real challenge to adapt this method to include the series $(\theta_\ell)^\ell$ because this requires the existence of arbitrarily large gaps between sums of ℓ nonnegative ℓ-th powers. This is why Theorem 1.1.7 is restricted to $\ell \in \{3, 4\}$. In fact the mere existence of arbitrarily large gaps is not sufficient for the diophantine application. It requires also a strenghtening of my estimates of Chapter 6 to prove the existence of what I call "mild gaps". More precisely I prove that the following technical criterion is fulfilled.

Proposition 1.1.8. Let $q \geq 2$ be an integer. Suppose that for every $J > 0$ there are $E, N > 0$, integers $K_1 \leq K_2 \in \mathbb{N}_+$ and $n_1, n_2 \in \mathbb{N}_+$ such that:

(i) $r_{\ell,\ell-1}(n) = 0$ for all $n_1 \leq n < n_2 + K_2$;

(ii) $r_{\ell,\ell}(n) = 0$ for each integer $n \in [n_1, n_1 + K_1) \cup [n_2, n_2 + K_1)$;

(iii) $\sum_{i=0}^{\infty} r_{\ell,\ell}(n_j + K_1 + i) 2^{-i} \leq E$ for $j = 1, 2$;

(iv) there exists $n_3 \in [n_1, n_2)$ with $r_{\ell,\ell}(n_3) > 0$;

(v) $n_1 + K_1 < n_2$ and $n_2 + K_2 \leq N$;

(vi) $q^{K_1} > JE$ and $q^{K_2} > JN$.

Then either $\theta_\ell(q)$ is transcendental or it is algebraic with degree at least $\ell + 1$.

The first item says that there are no sums of $\ell - 1$ nonnegative ℓ-th powers in the large interval $[n_1, n_2 + K_2)$. The second one (together with the fifth) asks for two sub-intervals of lenght K_1 containing no sum of ℓ nonnegative ℓ-th powers. In Chapter 8, the third condition is expressed by saying that these sub-intervals are mild gaps for the series $\theta_\ell^\ell(1/z)$. The fourth condition requires that there exist at least one sum of ℓ nonnegative ℓ-th powers in between these two sub-intervals. This is where I need a control from above on the size of the intervals where $r_{\ell,\ell}$ vanishes. The last item is meant to control the denominators of certain truncations of the series.

1.1.5 Future work

Following suggestions of Wooley, I would like to extend this work in two directions. The first one would be to prove the existence of arbitrarily large gaps between the values of non-homogeneous diagonal forms in three and four variables. I already have some partial results in this direction. For example, I can prove that there exist unbounded intervals that contain no integer of the form $n = x_1^2 + x_2^3 + x_3^7 + x_4^{42}$. The quantitative estimate that I can prove in this case for the size of gaps is similar to the one I obtained for sums of four fourth powers. The main difference with respect to the homogeneous case is that in the explicit formula for the representations of the zero class modulo p, the character H_p is replaced by a linear combination of unitary Hecke characters. The values of these equidistribute on the unit circle as p varies, but the characters themselves may not be independent of each other. The resulting distribution is therefore harder to determine in general. For the example given above, there is in fact only one character. The other direction of research is to consider diagonal forms whose arguments have restricted ranges.

1.2 Multiplicity estimates for the resultant

The classical resultant of a sequence $\underline{f} = (f_0, \dots, f_r)$ of $r+1$ homogeneous polynomials in $r+1$ variables over $\mathbb{C}$ is an irreducible polynomial in the set of coefficients of $f_0, \dots, f_r$ that vanishes if and only if the polynomials have a common zero in projective r-space $\mathbb{P}^r_{\mathbb{C}}$ over $\mathbb{C}$. More generally, if the polynomials vanish at finitely many points together with their first few derivatives, then we may expect that the resultant vanishes with some multiplicity at the tuple of coefficients of the polynomials. The problem is to estimate to which order this vanishing occurs.

This is important for applications to transcendental number theory. In [132] my supervisor Roy proposed a conjecture that is equivalent to Schanuel's conjecture. This new conjecture assumes that a sequence of polynomials in two variables with integer coefficients take small values, together with some of their derivatives, at many points of a finitely generated subgroup of $\mathbb{C} \times \mathbb{C}^*$. The conclusion is an upper bound on the transcendence degree of the field generated over $\mathbb{Q}$ by the coordinates of the points of this subgroup. A general approach to the conjecture is to show that if P, Q, R are homogenized versions of such polynomials then their resultant is very small. Since it is an integer, it should then vanish.

This approach raises two main problems. The first one is that we need to take into account the degrees of the polynomials in each variable separately, which amounts to working with bihomogeneous polynomials. So we need multiplicity estimates for resultants of multihomogeneous polynomials. This is the problem that I solve in Chapter 9 of my thesis. The second problem, which is not addressed in the thesis, is to deal with a number of points and derivatives which exceeds the number of

unknown coefficients of the polynomials. In fact I already addressed the first problem in my MSc thesis and I used it to extend the results of my supervisor [133] dealing with polynomials taking small values at a single point of $\mathbb{C} \times \mathbb{C}^*$ together with their derivatives along a natural one-dimensional subgroup. This was based on Rémond's multihomogeneous elimination theory.

In my PhD thesis, I extend the validity of the multiplicity estimate by replacing the ambient space by an arbitrary multiprojective variety of dimension r. In order to state the result I need to introduce some notation.

We fix a field k, an integer $q \in \mathbb{N}_+$ and positive natural numbers $n_1, \dots, n_q \in \mathbb{N}_+$. We denote the ambient multiprojective space over k by

$$\mathbb{P}^{\mathbf{n}}_k := \mathbb{P}^{n_1}_k \times \cdots \times \mathbb{P}^{n_q}_k.$$

We introduce the set of variables $\mathbf{x} = (x_{p,i})_{p=1,\dots,q,\, i=0,\dots,n_p}$ so that the coordinate ring of this space is $k[\mathbf{x}]$. For each $\mathbf{d} \in \mathbb{N}^q$ we denote by $k[\mathbf{x}]_{\mathbf{d}}$ its multihomogenous part of multidegree $\mathbf{d}$. Within this multiprojective space we fix a multiprojective scheme $Z \subseteq \mathbb{P}^{\mathbf{n}}_k$ of dimension r, and denote by $I \subseteq k[\mathbf{x}]$ the multihomogeneous ideal of definition of Z. For any subset $J \subseteq k[\mathbf{x}]$ we denote by $\mathcal{Z}(J)$ the zero subscheme of J. For an ideal $J \subseteq k[\mathbf{x}]$ we define $J_{\mathbf{d}} := J \cap k[\mathbf{x}]_{\mathbf{d}}$ for each $\mathbf{d} \in \mathbb{N}^q$ and we say that J is irrelevant if $J_{\mathbf{d}} = k[\mathbf{x}]_{\mathbf{d}}$ for some $\mathbf{d}$. Equivalently, J is irrelevant if $\mathcal{Z}(J)$ is empty. Finally, let $\mathfrak{d} = (\mathbf{d}^{(0)}, \dots, \mathbf{d}^{(r)})$ be a collection of nonzero multidegrees (i.e. a sequence of $r+1$ elements of $\mathbb{N}^q$). We denote by $\text{rés}_{\mathfrak{d}}(I)$ the resultant form attached to Z, for $r+1$ multihomogeneous polynomials $f_0 \in k[\mathbf{x}]_{\mathbf{d}^{(0)}}, \dots, f_r \in k[\mathbf{x}]_{\mathbf{d}^{(r)}}$.

Theorem 1.2.1. Let J be a multihomogeneous ideal of $k[\mathbf{x}]$ such that $I \subseteq J$ and $\dim \mathcal{Z}(J) = 0$. Suppose also that, for every $i = 0, \dots, r-1$, we have $\dim \mathcal{Z}(J_{\mathbf{d}^{(i)}}) = 0$ and that, for every relevant $\mathfrak{p} \in \text{Ass}_{k[\mathbf{x}]}(k[\mathbf{x}]/J_{\mathbf{d}^{(i)}}k[\mathbf{x}])$, the local ring (module) $(k[\mathbf{x}]/I)_{\mathfrak{p}}$ is Cohen-Macaulay of (Krull) dimension r. Then the resultant form $\text{rés}_{\mathfrak{d}}(I)$ vanishes to order at least $\deg(J)$ at each $(r+1)$-tuple $\underline{f} = (f_0, \dots, f_r) \in J_{\mathbf{d}^{(0)}} \times \cdots \times J_{\mathbf{d}^{(r)}}$.

For the application, let $k = \mathbb{C}$ and let $G = G_1 \times \ldots \times G_q$ be a connected commutative algebraic group of dimension n_G embedded in $\mathbb{P}^{\mathbf{n}}_{\mathbb{C}}$. We denote by $\mathfrak{G} \subseteq \mathbb{C}[\mathbf{x}]$ the multihomogeneous ideal of definition of its Zariski closure $\overline{G}$. Let $\Sigma = \{\gamma_1, \dots, \gamma_\ell\} \subseteq G(\mathbb{C})$ be a finite subset of complex points of G and let $\Delta = \{\partial_1, \dots, \partial_d\}$ be a set of linearly independent invariant derivations on G. For every $\sigma \in \mathbb{N}^d$ we define the differential operator $\partial^\sigma = \partial_1^{\sigma_1} \dots \partial_d^{\sigma_d}$ of order $|\sigma| = \sigma_1 + \cdots + \sigma_d$. Then, for each multidegree $\mathbf{d}$ and each positive integer T, we define the evaluation map

$$\begin{array}{rccc} \text{ev}_{\Sigma,T,\mathbf{d}} : & \mathbb{C}[\mathbf{x}]_{\mathbf{d}} & \longrightarrow & \mathbb{C}^{|\Sigma|\binom{T-1+d}{d}} \\ & P & \mapsto & \left(\partial^\sigma \left(\dfrac{P}{x_{1,0}^{d_0} \dots x_{q,0}^{d_q}} \right) (\gamma) : \ |\sigma| < T, \gamma \in \Sigma \right). \end{array}$$

Finally, for every multidegree $\mathbf{d} \in \mathbb{N}^q$ we let

$$I_{\mathbf{d}}^{\Sigma,T} := \ker(\mathrm{ev}_{\Sigma,T,\mathbf{d}}).$$

Then $I^{\Sigma,T} := \bigoplus_{\mathbf{d}\in\mathbb{N}^q} I_{\mathbf{d}}^{\Sigma,T}$ is an ideal called the interpolation ideal for the data (Σ, Δ, T). The following result is my main application of Theorem 1.2.1.

Theorem 1.2.2. Let $\mathfrak{d} = (\mathbf{d}^{(0)}, \ldots, \mathbf{d}^{(n_G)})$ be a collection of multidegrees such that $\mathrm{ev}_{\Sigma,T,\mathbf{d}^{(i)}}$ is surjective for all $i = 0, \ldots, n_G - 1$. Then the resultant $\mathrm{rés}_{\mathfrak{d}}(\mathfrak{G})$ of index $\mathfrak{d}$ attached to the prime ideal $\mathfrak{G}$ vanishes with multiplicity at least $|\Sigma|\binom{T-1+d}{d}$ on every (n_G+1)-uple of polynomials in $I_{\mathbf{d}^{(0)}}^{\Sigma,T} \times \cdots \times I_{\mathbf{d}^{(n_G)}}^{\Sigma,T}$.

In another direction, I noticed a similitude between Theorem 1.2.1 and a result of Chardin saying that if the reduction modulo p of $r+1$ homogeneous polynomials $\underline{f} = (f_0, \ldots, f_r)$ with integer coefficients in $r+1$ variables have N common zeros over the algebraic closure of $\mathbb{F}_p$, then their resultant $\mathrm{Res}(\underline{f})$ is an integer divisible by p^N. In order to bring the two results together I developed a framework for a generalized notion of resultant. This is done over an arbitrary Noetherian Unique Factorization Domain A and deals with sequences $\underline{f}$ of polynomials in $A[\mathbf{x}]$. The role of $k[\mathbf{x}]/I$, which is the coordinate ring of Z, is replaced in this theory by a multihomogeneous $A[\mathbf{x}]$-module M which is free over A. Instead of Rémond's multihomogeneous elimination theory, I use the Cayley determinant of the Koszul complex $\mathbf{K}_\bullet(\underline{f}, M)$ to define the resultant (see Definition 2.9 of Chapter 9). I prove that this generalizes the resultant of Rémond (see Theorem 2.14) and I obtain a formulation that generalizes both Theorem 1.2.1 and Chardin's theorem (see Theorem 3.3).

1.3 Planar graphs with positive combinatorial curvature

In Chapter 10, I study a special class of planar graphs that generalizes the notion of convex polyhedra with regular faces. Such convex polyhedra have been completely classified by Johnson [90] and Zalgaller [169] up to isomorphism: they consist of the 5 Platonic solids, the 14 Archimedean solids, the infinite families of prisms and antiprisms (Figure 1.3.1) and the 92 Johnson solids depicted in Figure 1.3.2.

For a vertex v of a polyhedron we define $\mathrm{anglesum}(v)$ to be the sum of the angles (at v) of faces incident in v. If the polyhedron is convex then $\mathrm{anglesum}(v)$ is less than a full angle, so the quantity

$$\mathrm{curvature}(v) := 1 - \frac{1}{2\pi}\,\mathrm{anglesum}(v),$$

sometimes known as (normalized) *angular defect*, is strictly positive. This function defines, on the vertex set of a polyhedron P, a discrete notion of *curvature* akin to

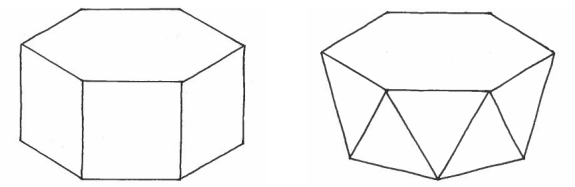


Figure 1.3.1: A prism and an antiprism.

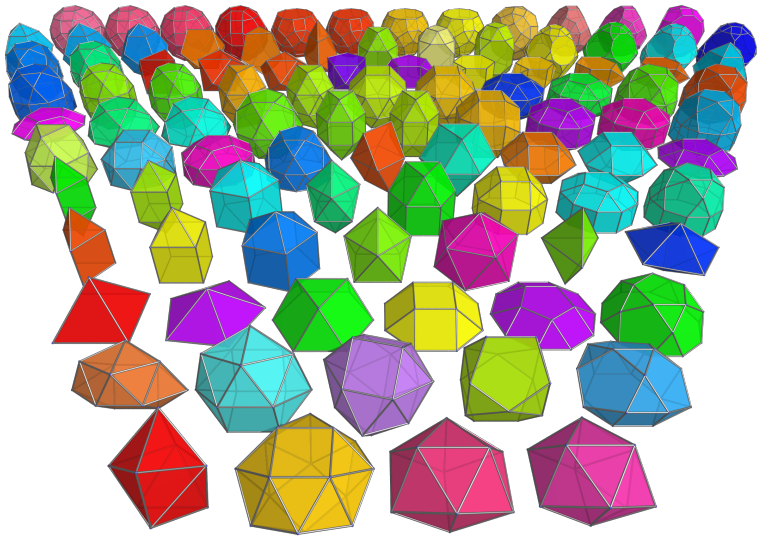


Figure 1.3.2: The 92 Johnson solids (eusbeia.dydns.org).

the notion of Gaussian curvature for 2-dimensional manifolds. For instance, we have the following discrete version of the Gauss-Bonnet theorem:

$$\sum_{v \text{ vertex}} \text{curvature}(v) = \chi(P), \tag{1.3.1}$$

where the sum runs over all vertices of the polyhedron P and $\chi(P)$ is the Euler characteristic of its surface boundary. This theorem is also known as *Descartes' total angular defect* formula because it was discovered by Descartes in the case of polyhedra homeomorphic to the sphere [34]. The general version (1.3.1), proved by Euler, is historically very important because it led to the discovery of the Euler characteristic, which is now one of the most important invariants of topology [130].

The internal angles of a regular n-gon are equal to $\pi - 2\pi/n$, so for a polyhedron with regular faces the discrete curvature can be computed as

$$\text{curvature}(v) = 1 - \sum_{f \sim v} \left(\frac{1}{2} - \frac{1}{|f|} \right), \tag{1.3.2}$$

where the sum runs over all faces f that are adjacent to v and $|f|$ denotes the number

of sides of f. This last formula is used to define the discrete curvature for abstract finite connected planar graphs.

We are interested in classifying the connected finite planar graphs that have positive curvature at all vertices. Of course this includes all graphs coming from convex polyhedra with regular faces, in particular the infinite families of prisms and antiprisms. To avoid trivialities we consider only the graphs which have no vertex of degree less than 3 and no repeated edges. Then, if we exclude the prisms and the antiprisms, a theorem of DeVos and Mohar [37] asserts that there remain only finitely many planar graphs up to isomorphism. We call them the planar PCC graphs (from Positive Combinatorial Curvature). DeVos and Mohar further provide an upper bound of 3444 for the number of vertices for PCC graphs, and ask for the best possible bound.

The first conjectured answer was 120, corresponding to the great rhombic icosidodecahedron (Figure 1.3.3), but in 2005 the lower bound was improved to 138 in [127] (Figure 1.3.4), showing that there are large PCC graphs that do not come from convex polyhedra. In 2011, Nicholson and Sneddon found the first example of a PCC graph with 208 vertices [118] (Figure 1.3.5). On the other hand, more effort is required to improve the upper bound on the number of vertices. In his MSc thesis, Oldridge [120] used linear programming to lower the bound to 244, conditionally on the hypothesis that there do not exist PCC graphs with a face having 42 edges or more. Recently an unconditional upper bound of 380 vertices was given by Oh [119].

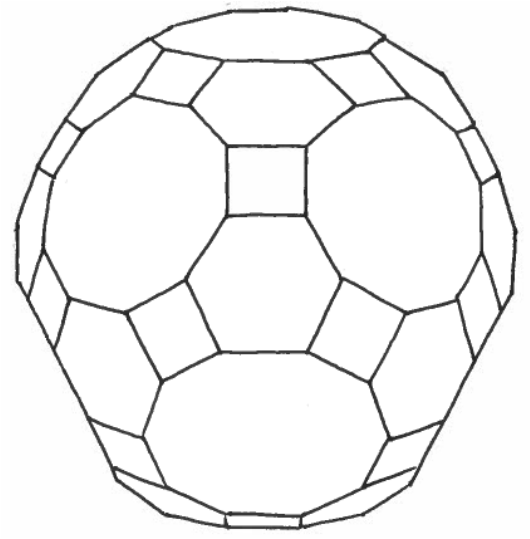


Figure 1.3.3: A 3D view of the rhombicosidodecahedron.

When I started working on this problem in 2011 I was only aware of the example of a PCC graph with 138 vertices. Trying to improve that construction I found several examples of large PCC graphs, including one with 208 vertices (Figure 1.3.6).

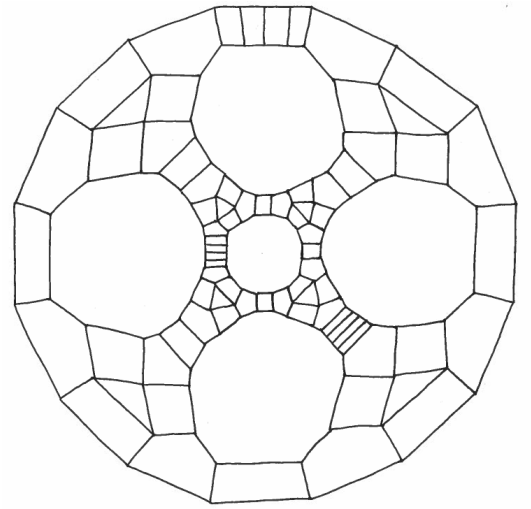


Figure 1.3.4: A PCC graph with 138 vertices due to Réti-Bitái-Kosztol'anyi.

It happened to be non-isomorphic to the one of Nicholson and Sneddon, and was rediscovered independently in 2017 by Oldridge [120]. In the subsequent years, I was able to sharpen the upper bound on the number of vertices of PCC graphs from 3444 to 264 (in 2011) and then to approximately 220 (in 2013). I was guided by the observation that the total curvature of a planar graph is equal to 2 (by Bonnet's formula) and so a large PCC graph necessarily has a very small average curvature. My approach at that time, to estimate from below the average curvature, was to prove the existence of a vertex with large curvature next to each vertex with small curvature. However, the task of lowering the bound all the way down to 208, which I expected to be best-possible, was becoming computationally unfeasible.

During my PhD, I returned to the problem with a new idea coming from analysis (more precisely, transportation theory). This consisted in redistributing the curvature from the vertices to the faces of the graph. The technical implementation of this idea is based on the so-called discharging method. This is a general-purpose technique in structural graph theory (used for example in the proof of the Four Color Map theorem) that reduces a global statement to a number of local verifications.

The discharging method requires a choice of weights that is essentially the result of a linear optimization problem. In future work, I would like to implement my discharging argument in a computer system. The framework that I envision, involving Integer Linear Programming, is similar to the one set up by Oldridge for a simpler set of local linear constraints [120]. This would provide a powerful tool to study the possible subgraphs of PCC graphs, especially of those with a large number of vertices. It is my hope that this approach would in turn lead to a complete classification of PCC graphs.

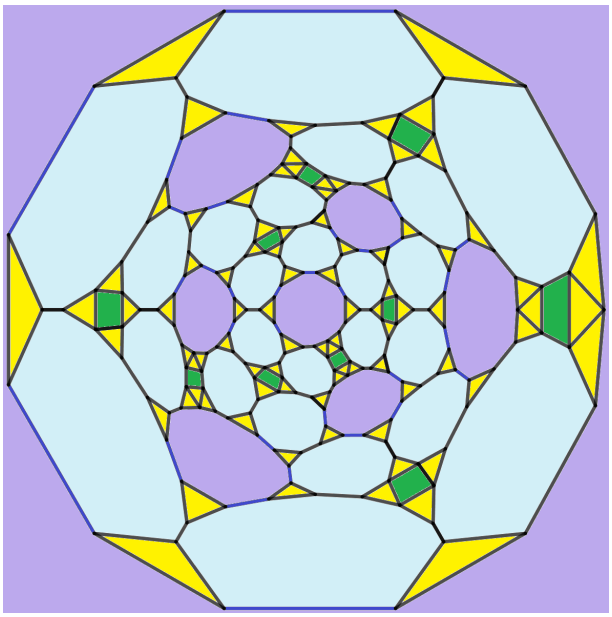


Figure 1.3.5: A PCC graph with 208 vertices due to Nicholson and Sneddon.

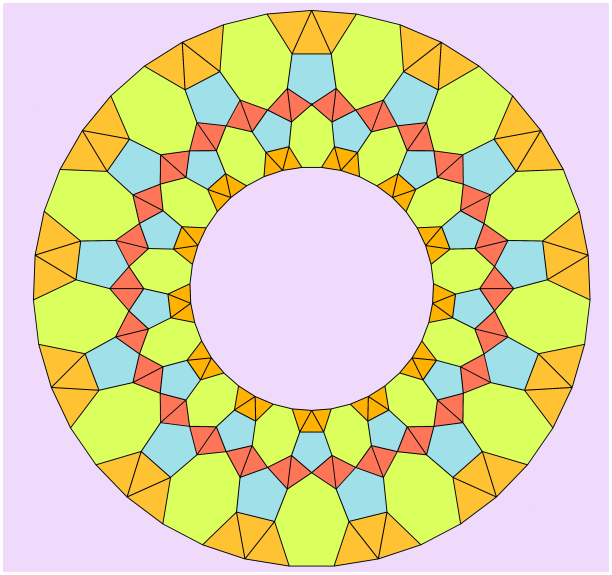


Figure 1.3.6: A PCC graph with 208 vertices found by Oldridge and myself.

Part I

Sums of powers - Elementary methods

Chapter 2

Gaps between numbers that are sums of two squares: theorems and conjectures

Let $\mathcal{S}_{2,2} = \{0, 1, 2, 4, 5, 8, 9, \dots\}$ be the set of integers that can be written as a sum of two squares. In this chapter we collect both old and new results about this set, in particular concerning the size of gaps between its elements. We organize the content in three sections. In the first one we recall the well-known arithmetic description of the elements of $\mathcal{S}_{2,2}$. In the second we feature a theorem of Richards: between the elements of $\mathcal{S}_{2,2} \cap [0, N]$ there are gaps of size $\gg \log N$. Then in the third section we discuss a long-standing open problem: are there in $\mathcal{S}_{2,2} \cap [0, N]$ gaps of size $\gg N^{1/4}$, or at least $\gg N^{\epsilon}$ for some $\epsilon > 0$?

2.1 A characterization of the sum of two squares

It is clear that every perfect square is in $\mathcal{S}_{2,2}$ because we have the trivial decomposition $n^2 = n^2 + 0^2$. It is also easy to see that a sum of two squares cannot be congruent to 3 modulo 4. In particular only the *even* powers p^{2k} of a prime $p \equiv 3 \bmod 4$ are elements of $\mathcal{S}_{2,2}$. By contrast, any prime number $p \equiv 1 \bmod 4$ can be written as a sum of two squares, a fact that was first stated by Girard in 1625 [149, Q.XII, p.622]. This result is also known as Fermat's theorem on sums of two squares, or *Fermat's Christmas Theorem* because it appears also in a letter of Fermat to Mersenne dated December 25th, 1640 [150]. The first written proof was given by Euler [46, 47, 48] using the method of infinite descent. Since then, several alternative proofs have been proposed by many authors: a proof by Lagrange [94], then simplified by Gauss [57, art. 182] that uses the theory of binary quadratic forms; two proofs by Dedekind [32, 33] based on the algebraic properties of the Gaussian integers; a constructive proof of Smith [145, 146, 24] that use palindromic continuants; a proof of Ewell [51] that uses

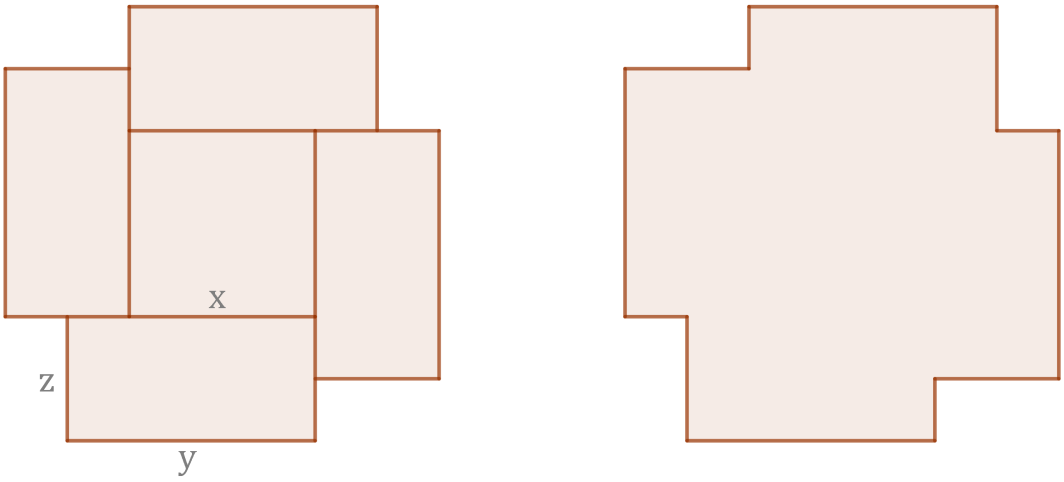


Figure 2.1.1: A winged square and its shape.

Jacobi's triple product formula to simplify an argument of Uspensky and Heaslet [156, pp.446-448]; a one-line involution-theoretic proof by Zagier [167], that refines ideas of Heath-Brown [74] and Liouville (see the survey by Elsholtz [42]); a proof of Lucas [104] with "regular satins" (see the exposition of Decaillot [31]); a lattice-theoretic proof by Grace [64], and some that use ideas of Minkowski [155, 154]; a proof given by Larson [98] coming from the existence of a symmetric arrangements of queens on a chessboard [93], and based on an intuition of Pólya [125]; a partition-theoretic proof of Christopher [22], etc. Here we propose the beautiful visual *proof from THE BOOK* [1, sec.4] due to A.Spivak [147].

Theorem 2.1.1. For each prime number $p \equiv 1 \bmod 4$ we have $p \in \mathcal{S}_{2,2}$.

Proof: We define a *winged square* to be a configuration of four rectangles arranged around a square with $\pi/2$-rotational symmetry as in Figure 2.1.1. We require that each rectangle shares a vertex and a portion of an edge with the square. If the square has size x and the rectangles have sizes y and z (the edge with length y is the one adjacent to the square), then we say that the winged square has type (x, y, z). Its area is equal to $x^2 + 4yz$. The *shape* of a winged square is simply the union of its five components. We now consider the set $\mathcal{W}$ of the winged squares with area equal to p considered up to isometry, where $p \equiv 1 \bmod 4$ is a prime number. The set $\mathcal{W}$ is naturally in bijection with the set

$$\mathcal{W}' := \{(x, y, z) \in \mathbb{N}_+^3 : \ x^2 + 4yz = p\}.$$

We are going to consider two natural involutions on this set. The first one is geometric and is exemplified in Figure 2.1.2. For every winged square W_1 there is at most

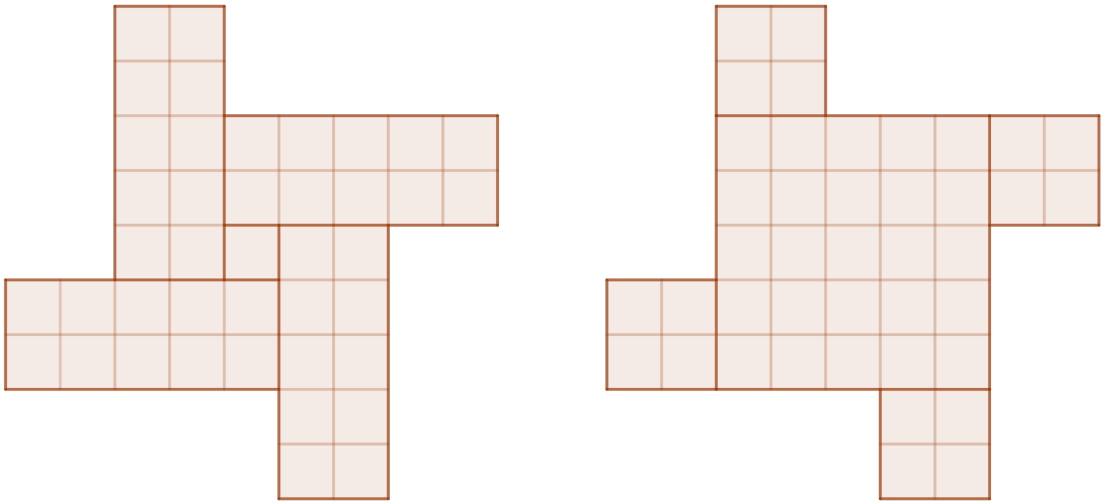


Figure 2.1.2: A winged square of type $(1, 5, 2)$ and its dual, of type $(5, 2, 2)$.

one other winged square W_2 that has a congruent shape (possibly with opposite orientation). If such $W_2 \neq W_1$ exists, we say that W_2 is the *dual* of W_1 (and viceversa); otherwise we say that W_1 is self-dual. For example, $(1, 5, 2)$ is the dual of $(5, 2, 2)$ as we see in the picture. A winged square is self-dual only if $x = y$, that is, if it has the shape of a Greek cross. However the equality

$$x^2 + 4xz = x(x + 4z) = p$$

forces $x = 1$. In other words, in $\mathcal{W}$ there is only one self-dual winged square, namely the one with type $(1, 1, \frac{p-1}{4})$. In particular, we get:

Lemma 2.1.2. The set $\mathcal{W}$ (and so also $\mathcal{W}'$) has odd cardinality.

We now consider the simple arithmetic involution

$$\text{swap} : (x, y, z) \mapsto (x, z, y)$$

on $\mathcal{W}'$ that swaps the last two coordinates of the triple. Since $\mathcal{W}'$ is a set with odd cardinality, the involution swap must have some fixed point. This is a triple of the form (x, y, y), which corresponds to the decomposition

$$x^2 + (2y)^2 = p.$$

∎

However beautiful, the proof above is not constructive. An explicit procedure to decompose a prime number $p \equiv 1 \bmod 4$ as a sum $p = x^2 + y^2$ of two squares was given by Serret [138] and Hermite [75] with arguments similar to Smith's. Their algorithm was then improved and simplified by Brillhart [18]. His algorithm consists of two parts: the first is the computation of $\sqrt{-1} \bmod p$; the second is better understood as an Euclidean long division performed in the ring of the Gaussian integers $\mathbb{Z}[i]$.

```
    input prime p
output int x, int y
begin
     find z such that p divides z^2 + 1
     compute x + iy = gcd(p, z − i)
     return (x, y)
end
```

Other algorithms such as Cornacchia's [27, 112, 5], are used to compute more general representations of primes as values of binary forms, or solving the Diophantine equation $x^2 + dy^2 = m$ for given m and d, see also [25, p.35], [72] and [157].

It is known since the antiquity that the set $\mathcal{S}_{2,2}$ is a multiplicative monoid, i.e. the product of two sums of two squares is still a sum of two squares. More precisely, we have the following identity of the Alexandrian Hellenistic mathematician Diophantus [40, III, 19 and 22]

$$(a^2 + b^2)(c^2 + d^2) = (ac - bd)^2 + (ad + bc)^2. \tag{2.1.1}$$

This formula is sometimes known also as the Brahmagupta-Fibonacci identity, because it was rediscovered and generalized in the 7th century by the Indian mathematician Brahmagupta [17, XVIII.65-66] (see [26] for a translation into English of the chapters XVII and XVIII of the Brāhmasphuṭasiddhānta.), and then it reappeared in 1225 in the Liber Quadratorum of Fibonacci [52]. Since we also have $2 = 1^2 + 1^2 \in \mathcal{S}_{2,2}$, the previous discussion implies that a number n is expressible as a sum of two squares, if all the prime factors of n congruent to 3 modulo 4 occur to an even exponent. The converse also hold, and this is known as the *sum of two squares theorem*. We now prove it using one of the approaches of Dedekind.

Theorem 2.1.3. A nonnegative integer n satisfies $n \in \mathcal{S}_{2,2}$ if and only if it can be written as $n = 2^\alpha P Q^2$, where $\alpha \in \mathbb{N}$ and P (resp. Q) is an integer divisible only by primes congruent to 1 modulo 4 (resp. 3 modulo 4).

Proof: It suffices to prove that $n \in \mathcal{S}_{2,2}$ cannot be divisible exactly by p^k, if $p \equiv 3 \bmod 4$ and k is odd. We recall that the ring of the Gaussian integers $\mathbb{Z}[i]$ is an Euclidean domain, and so in particular it is a Principal Ideal Domain and a Unique Factorization Domain. Moreover we have the following

Lemma 2.1.4. Each prime number $p \equiv 3 \bmod 4$ is a prime element in $\mathbb{Z}[i]$.

Indeed, if we have a nontrivial factorization $(a+bi)(c+di) = p$ then taking norms we get the nontrivial factorization in integers $(a^2+b^2)(c^2+d^2) = p^2$. But this forces $a^2+b^2 = c^2+d^2 = p$, which is impossible since $p \equiv 3 \bmod 4$.

Now, if $n \in \mathcal{S}_{2,2}$ then in the Gaussian integers we have a factorization

$$n = x^2 + y^2 = (x - iy)(x + iy)$$

for some $x, y \in \mathbb{Z}$. If $n = p^k m$ for some prime number $p \equiv 3 \bmod 4$ and some integer m coprime with p, then by the lemma we have that p^{k_1} divides $x - iy$ and p^{k_2} divides $x + iy$ for some k_1, k_2 with $k_1 + k_2 = k$. However, using the conjugation isomorphism we see that $k_1 = k_2$ and so k must be even. ∎

From the decomposition in Theorem 2.1.3 it is possible to read the *sum of two squares function* $r_2(n)$, that is the number of ways the number $n = 2^\alpha PQ^2$ can be decomposed as $n = x^2 + y^2$. We define

$$r_2(n) := \#\{(x, y) \in \mathbb{N}^2 : \ n = x^2 + y^2\} \tag{2.1.2}$$
$$r_2^*(n) := \#\{(x, y) \in \mathbb{Z}^2 : \ n = x^2 + y^2\} \tag{2.1.3}$$

so that $r_2^*(n) = 4r_2(n)$ if n is not a perfect square and $r_2^*(n) = 4r_2(n) - 4$ if it is. Then

$$r_2^*(n) = 4 \cdot d(P), \tag{2.1.4}$$

where $n = 2^\alpha PQ^2$ as in Theorem 2.1.3 and $d(P)$ is the number of the divisors of P. Again, the formula (2.1.4) was stated by Fermat in his letter to Mersenne.

2.2 Gaps of logarithmic size

We now write the elements of $\mathcal{S}_{2,2} = \{s_1 < s_2 < s_3 < \dots\}$ as an increasing sequence and we turn to the following question: how big can be the quantity $s_{n+1} - s_n$, say, compared to the element s_n? From the characterization given in Theorem 2.1.3, Shiu [141] was able to prove that for every $k \in \mathbb{N}_+$ there exists some $n = n(k)$ such that $s_{n+1} - s_n = k$. However, this precision on the size of the gap goes at the expense of the size of n: it can be very large compared to k. From a completely different perspective, Landau [96, 95] proved an asymptotic formula for the counting function $S(x)$ of $\mathcal{S}_{2,2}$, as $x \to \infty$:

$$S(x) := \sum_{\substack{n \in \mathcal{S}_{2,2} \\ n \le x}} 1 = (c_{LR} + o(1)) \frac{x}{\sqrt{\log x}}, \tag{2.2.1}$$

where

$$c_{LR} := \frac{1}{\sqrt{2}} \prod_{\substack{q \text{ prime} \\ q \equiv 3 \bmod 4}} \left(1 - \frac{1}{q^2}\right)^{-1/2} = 0.76422305\ldots.$$

The constant c_{LR} is known as the *Landau-Ramanujan constant* [99] because a statement equivalent to (2.2.1) appeared in the first letter of Ramanujan to Hardy [10, p.25], see also [8, pp.60-66] or [113]. The growth $S(x) \asymp x/\sqrt{\log x}$ can be heuristically inferred from Theorem 2.1.3 with a simple sieve-theoretic argument [102], and in fact Iwaniec [87] has proposed a proof of the asymptotic (2.2.1) using the so-called half-dimensional sieve, see also the exposition [55, sec. 14.3]. Instead, the classical proof of Landau's asymptotics is analytic and is obtained from a study of the Dirichlet series $B(s) := \sum_{n \in \mathcal{S}_{2,2}} n^{-s}$ for $s \to 1$. First, $B(s)$ can be expanded as an Euler product

$$B(s) = \frac{1}{1-2^{-s}} \prod_{\substack{p \text{ prime} \\ p \equiv 1 \bmod 4}} \frac{1}{1-p^{-s}} \prod_{\substack{q \text{ prime} \\ q \equiv 3 \bmod 4}} \frac{1}{1-q^{-2s}} \tag{2.2.2}$$

and so its square can be factored as $(B(s))^2 = \zeta(s) L(s, \chi_{2,4}) \psi(s)$, where $\zeta(s) = \prod_{p \text{ prime}} (1-p^{-s})^{-1}$ is the Riemann zeta function,

$$L(s, \chi_{2,4}) := \prod_{\substack{p \text{ prime} \\ p \neq 2}} \frac{1}{1 - (-1)^{\frac{p-1}{2}} p^{-s}} \tag{2.2.3}$$

is the L-function associated to the primitive Dirichlet character modulo 4 (which is of order 2) and $\psi(s)$ is a Dirichlet series that is absolutely convergent on the half plane $\mathbb{R}\mathrm{e}(s) > 1/2$. Since $\zeta(s)$ has a simple pole at $s = 1$ and $L(s, \chi_{2,4})$ extends to an entire function, we see that $B(s) \sim \beta \cdot (s-1)^{-1/2}$ as $s \to 1$, for some $\beta \neq 0$. One computes that $\beta = c_{LR} \cdot \Gamma(1/2)$, hence Equation (2.2.1) follows from a generalized Wiener-Ikehara Tauberian theorem ([6, Sec. 7.4] or [152, Sec. 7.5]), see also the thesis [129].

A direct consequence of Equation (2.2.1) is that the set $\mathcal{S}_{2,2}$ has zero density in the natural numbers. Moreover, it says that on "average" the gap between its elements has size $s_{n+1} - s_n \approx \sqrt{\log s_n}$. However, the set $\mathcal{S}_{2,2}$ exhibits some irregularities and it is actually easy to show the existence of larger gaps. For example, if $p_1 < \cdots < p_k$ are the first k prime numbers congruent to 3 modulo 4, and $M_k := \prod_{i=1}^{k} p_i^2$ the product of their squares, then by the Chinese Remainder Theorem there is a natural number $m < M_k$ such that

$$m + j \equiv p_j \bmod p_j^2$$

for all $j = 1, \ldots, k$. Then by the characterization in Theorem 2.1.3 none of the numbers $m+1, \ldots, m+k$ is in $\mathcal{S}_{2,2}$. If $n = S(m)$ then we have $s_{n+1} - s_n > k$ and,

if we work out the estimates coming from Dirichlet's prime number theorem for the arithmetic progression $3 + 4\mathbb{N}$, then we get

$$s_{n+1} - s_n \gg \frac{\log s_n}{\log\log s_n}. \tag{2.2.4}$$

In a letter to Erdős, Turàn asked if one could do better than (2.2.4) by some more sophisticated argument. Erdős replied with the following ingenious construction: he chooses some $t \approx k\sqrt{\log k}$ and calls $a_1, \dots, a_z$ the numbers $\leq t$ for which the implication $p_i \,|a_j \Rightarrow p_i^2 \,|a_j$ holds for all $i \leq \lfloor k/2 \rfloor$ and all $j \leq z$; then he selects $m < M_k$ so that

$$\begin{array}{lll} m \equiv 0 & \mod p_i^2 & \text{for } i \leq \lfloor k/2 \rfloor, \\ m + a_j \equiv p_{\lfloor k/2 \rfloor + j} & \mod p_{\lfloor k/2 \rfloor + j}^2 & \text{for } i \leq z. \end{array}$$

In this way it turns out that none of the numbers $m+1, \dots, m+t$ is in $\mathcal{S}_{2,2}$ and this is sufficient to conclude that

$$s_{n+1} - s_n \gg \frac{\log s_n}{(\log\log s_n)^{1/2}}. \tag{2.2.5}$$

In 1982 Richards [128] found a construction that has the effect of removing the factor $(\log\log s_n)^{1/2}$ from (2.2.5). His argument is simple and short, so we reproduce it entirely in the proof below.

Theorem 2.2.1. For every $\epsilon > 0$ there are infinitely many $n \in \mathbb{N}_+$ such that

$$\boxed{s_{n+1} - s_n \geq \left(\frac{1}{4} - \epsilon\right) \log s_n.}$$

Proof: Given $T > 0$, we let $\mathcal{P}_T$ be the set of prime numbers $p \leq T$ congruent to 3 modulo 4, and we define M_T by the product

$$M_T := \prod_{p \in \mathcal{P}_T} p^{\beta(p,T)+1},$$

where $\beta = \beta(p, T)$ is the largest exponent such that $p^\beta \leq T$. Then, we let $m = m(T)$ be unique natural number $m < M_T$ such that

$$4m \equiv -1 \mod M_T.$$

Claim 2.2.2. None of the numbers $m + 1, \cdots, m + \lfloor T/4 \rfloor$ is in $\mathcal{S}_{2,2}$.

Indeed, for every $j \le T/4$ we have

$$4(m+j) \equiv 4j-1 \bmod M_T$$

and the number $4j-1$ is and integer $\le T$ congruent to 3 modulo 4. So, $4j-1$ must be divisible by some $p \in \mathcal{P}_T$ to some odd power $\alpha \le \beta(p,T)$. Since $p^{\beta(p,T)+1} \,|M_T$, this means that p also divides $m+j$ exactly to the power α, hence $m+j$ is not a sum of two squares.

Notice that for all but $O(\sqrt{T})$ elements of $\mathcal{P}_T$ we have $\beta(p,T)=2$, while the $p \in \mathcal{P}_T$ are roughly half of all the primes $\le T$. Then the prime number theorem in arithmetic progressions implies that

$$\log m \le \log M_T = (1+o(1))T$$

as $T \to \infty$. If we choose $n = S(m)$ then we have $s_n \le m$ and $s_{n+1} \ge m + T/4$, so the theorem follows making T grow to infinity. ∎

Dietmann and Elsholts [39] observed that some large primes can be excluded in the above argument. For example, it is not necessary that $p^2 \,|M_T$ for primes $p \equiv 11-4m \bmod 16$ with $T/5 < p \le T$: if $4j-1 \le T$ is composite, then it is divisible to an odd power by some prime $p \le T/5$ congruent to 3 modulo 4; if instead $q := 4j-1$ is prime and $q \equiv 11-4m \bmod 16$, then $m+j \equiv 3 \bmod 4$ and so the assertion $m+j \notin \mathcal{S}_{2,2}$ is trivial. With this and similar arguments the constant $1/4$ in Richard's theorem can be replaced by $195/449 = 0.434\ldots$. A paper of Balog and Wooley [2] enriches Richard's argument with more sophisticated tools from sieve theory in the following way: similarly as with M_T above, they consider

$$M_{T,S} := \prod_{p \in \mathcal{P}_T} p^{\alpha(p,S)}$$

where $\alpha = \alpha(p,S)$ is the smallest *odd* exponent such that $p^\alpha > S$, then they pick $m_\pm < M_{T,S}$ so that $4m_\pm \equiv \pm 1 \bmod M_{T,S}$. The class of $m = m_-$ modulo $M_{S,T}$ is chosen as in Richard's proof, so that the numbers $m+j$ with $j < S/4$ have reduced chance of being in $\mathcal{S}_{2,2}$; symmetrically, $m = m_+$ increases this chance. Balog and Wooley consider the "rectangle" $R^\pm$ of integers (see the Maier matrix method in Chapters 3, 4 and 6) given by

$$R^\pm := \{n \le N : \ n \equiv m_\pm + j \text{ for some } 1 \le j < S/4\}$$

and they estimate the cardinality of $R^\pm \cap \mathcal{S}_{2,2}$. They do so using the half-dimensional sieve and the choice of parameters $S \asymp (\log N)^A$ and $T = \log N / \log\log N$. In this way they prove that for all $A > 0$ there are constants $0 < \delta_-(A), \delta_+(A)$ such that both the following inequalities

$$S(x+y) - S(x) > (1+\delta_+(A)) \frac{c_{LR} y}{\sqrt{\log x}} \tag{2.2.6}$$

$$S(x+y) - S(x) < (1 - \delta_-(A)) \frac{c_{LR} y}{\sqrt{\log x}} \tag{2.2.7}$$

have infinitely many integer solutions with $y = (\log x)^A$. This statement should be compared with the following result of Hooley [80]: if $y/\sqrt{\log x}$ grows to infinity as $x \to \infty$, then there exist constants $0 < A_1 \leq A_2$ such that

$$A_1 \frac{y}{\sqrt{\log x}} \leq S(x+y) - S(x) \leq A_2 \frac{y}{\sqrt{\log x}}. \tag{2.2.8}$$

for "almost all x". See also Friedlander [54] for the upper bound; Plaskin [122], [123] and Harman [73] for the lower bound. In other words, Landau's estimate (2.2.1) persists, up to a constant and for almost all x, when the set $\mathcal{S}_{2,2}$ is restricted on intervals $(x, x+y]$ with $\sqrt{\log x} = o(y)$, but it fails to hold as an asymptotic for all x when $\log y \ll \log\log x$. Also relevant to this discussion is a recent result of Maynard [108], proved with an adaptation of the GPY sieve [63]: there are short intervals $(x, x+y]$ containing $\gg y^{1/10}$ elements of $\mathcal{S}_{2,2}$. This is of an higher order of magnitude than the "expected" $\approx y/\sqrt{\log x}$ when the interval is "very short", namely when $y = o((\log x)^{5/9})$. Another way of measuring the gaps of $\mathcal{S}_{2,2}$ is by estimating their "moments". In this sense Hooley [77] proves that

$$\sum_{n \leq x} (s_{n+1} - s_n)^\gamma \ll \frac{x}{\sqrt{\log x}} \cdot (\log x)^{\gamma/2}$$

for every $\gamma \leq 5/3$. This result is consistent with the estimate $\sqrt{\log x}$ for the average gap. A similar estimate was given by Kalminin [91]

$$\sum_{n \leq x} (s_{n+1} - s_n)^\gamma \ll x (\log x)^{3(\gamma-1)/2} \tag{2.2.9}$$

for the range $\gamma < 2$, see also [122, 73, 123]. An essentially logarithmic size for the gaps of $\mathcal{S}_{2,2}$ is also predicted by some probabilistic models, see [35].

2.3 Points in circles and folklore conjectures

In the previous section we learned that the average gap of $\mathcal{S}_{2,2}$ satisfies $s_{n+1} - s_n \asymp \sqrt{\log s_n}$ but also that $\mathcal{S}_{2,2}$ is not uniformily distributed and there exist gaps with $s_{n+1} - s_n \asymp \log s_n$. A natural question is then: what could be a reasonably sharp upper bound for the size of gaps? The following simple estimate gives $s_{n+1} - s_n \ll s_n^{1/4}$. For convenience we phrase it in terms the existence of elements of $\mathcal{S}_{2,2}$ in a *left* short interval of the form $(n-k, n]$, but n should not be confused with the index of some s_n.

Theorem 2.3.1. For every $n \in \mathbb{N}_+$ there exists a sum of two squares $s = x^2 + y^2$ with

$$n - 2\sqrt{2} n^{1/4} < x^2 + y^2 \leq n.$$

Proof: We let $x = \lfloor \sqrt{n} \rfloor$, so $x^2 \leq n < (x+1)^2$ and we get

$$0 \leq n - x^2 \leq 2x \leq 2\sqrt{n}.$$

Now we let $y = \lfloor \sqrt{n - x^2} \rfloor$, so we have

$$0 \leq n - x^2 - y^2 \leq 2\sqrt{n - x^2} \leq 2\sqrt{2} \cdot n^{1/4}.$$

∎

This construction is known as the *greedy argument* because we first take x as large as we can and then we do the same for y. This result was published by Bambah and Chowla in 1952 [3] but it is reasonable to think that this trick was known by earlier mathematicians, such as Hardy, Littlewood or even Euler, Gauss or Landau, who worked on related problems. In fact the greedy argument is a motivation for the *diminishing ranges* technique of Hardy and Littlewood, see the chapter on the cirle method in this thesis. Surprisingly, currently there is essentially no result that improves on Theorem 2.3.1! The following problem, attributed to Littlewood in [111, Appendix, Problem 64], has become folklore:

Problem 2.3.2. Prove that if $f(n)$ tends to 0 sufficiently slowly, then every interval $(n - f(n)n^{1/4}, n]$ contains a sum of two squares.

In fact, it is reasonable to expect that the same holds for intervals of length $O(n^\epsilon)$ for every given $0 < \epsilon$. It is possible to give a geometric interpretation of this problem, as follows. Let

$$\mathcal{C}_{\sqrt{n}} := \{x^2 + y^2 = n\}$$
$$\mathcal{D}_{\sqrt{n}} := \{x^2 + y^2 \leq n\}$$

be respectively the circumference and the disk of radius $r = \sqrt{n}$, then an element $s = x^2 + y^2$ of $\mathcal{S}_{2,2} \cap (n-k, n]$ corresponds to a lattice point $(x, y) \in \mathbb{Z}^2$ in the annulus

$$\mathcal{A}_{r,r-d} := \mathcal{D}_r - \mathcal{D}_{r-d}$$

of width

$$d = \sqrt{n} - \sqrt{n-k} = \left(\frac{1}{2} + o(1)\right) \frac{k}{\sqrt{n}}. \tag{2.3.1}$$

In other words, sums of squares satisfying $n - x^2 - y^2 = o(n^{1/4})$ as in Littlewood's problem correspond to lattice points $P = (x, y)$ with distance

$$\operatorname{dist}(P, \mathcal{C}_{\sqrt{n}}) = o(n^{-1/4}).$$

The greedy argument may be formulated geometrically as in Figure 2.3.1: we first choose a lattice point $A = (x_A, 0)$ on the x-axis, so that it is as close as possible

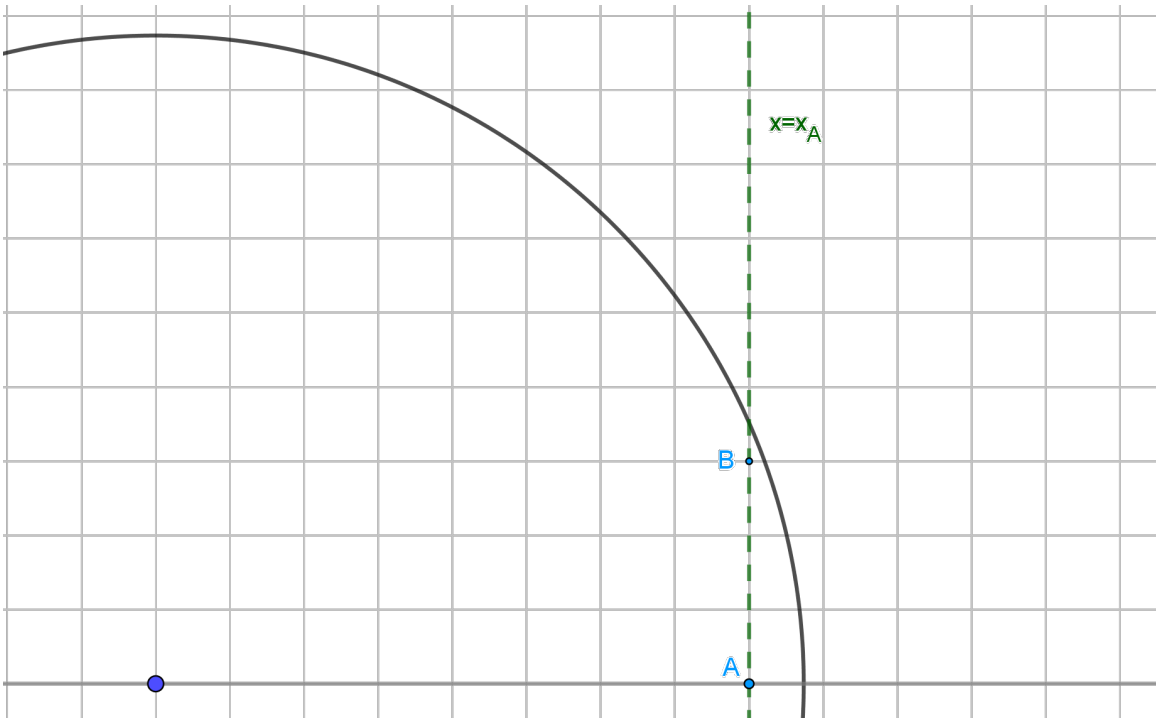


Figure 2.3.1: A geometric representation of the greedy argument

to the circumference $\mathcal{C}_{\sqrt{n}}$, i.e. $x_A = \lfloor \sqrt{n} \rfloor$; then we find the lattice point $B = (x_A, \lfloor \sqrt{n - x_A^2} \rfloor)$, vertically above A and inside the circle, but again as close as possible to the circumference. In fact,

$$\mathrm{dist}(B, \mathcal{C}_{\sqrt{n}}) \ll n^{-1/4}.$$

The greedy argument produces a lattice point B whose second coordinate is of a smaller order of magnitude than the first. In other words, B is located in vicinity of the point $A' = (\sqrt{n}, 0)$. We now describe a modification of the greedy argument that produces lattice approximations to the circumference with almost the same precision, but distributed in other parts of the circle. This is a geometric construction, illustrated in Figure 2.3.2, which we call the *almost-tangent method*.

The almost-tangent method

Choose some integers a, b, not both zero, and let A' be a point of intersection between the circumference $\mathcal{C}_{\sqrt{n}}$ and the line $\ell_1 : ax - by = 0$, so that the tangent at A' to the circle has the same direction as

$$\vec{v} := (-a, b).$$

Let $A = (x_A, y_A)$ be any lattice point inside the circle and "close" to A' and draw the line ℓ_2 passing through A with direction $\vec{v}$. See Figure 2.3.2 for an example with $a = 2$ and $b = 1$. Let B' be an intersection between $\mathcal{C}_{\sqrt{n}}$ and the line ℓ_2. Finally, choose a lattice point B inside the circle and on the line ℓ_2, so that B is as close as possible to B'. Then B is "very" close to the circumference, namely

$$\operatorname{dist}(B, \mathcal{C}_{\sqrt{n}}) \ll_{a,b} n^{-1/4},$$

where the constant depends on a and b.

Indeed, we can choose the lattice point A inside the circle so that $\overline{AA'} \leq 1$ and B' on the same side of A with respect to ℓ_1. Then we have $\operatorname{dist}(A, \mathcal{C}_{\sqrt{n}}) \leq 1$ and

$$\overline{AB'} \leq \sqrt{n - (\sqrt{n} - 1)^2} \leq \sqrt{2} \cdot n^{1/4}.$$

In particular we see from simple geometry that the angle between ℓ_2 and the tangent at B' to the circle measures $\alpha \ll n^{-1/4}$. Moreover $\overline{BB'} \ll \max\{|a|, |b|\}$ and so

$$\operatorname{dist}(B, \mathcal{C}_{\sqrt{n}}) \leq \overline{BB'} \sin(\alpha) \ll \max\{|a|, |b|\} \cdot n^{-1/4}. \tag{2.3.2}$$

Similar constructions appeared (independently) in a work of Green and Lindqvist [67] about the Ramsey theory of the equation $x + y = z^2$, and in a study of Huxley [83] about the lattice points close to smooth planar curves. In fact on close inspection the estimate (2.3.2) is essentially equivalent to one found in [67, sec.6].

We can rephrase the almost-tangent method as follows: *if a lattice point A is somewhat close to the circumference $\mathcal{C}_{\sqrt{n}}$ and the ratio y_A/x_A is close to some rational number a/b of small height, then visiting the lattice points in the direction $\vec{v} = (-a, b)$ we find a second lattice point B that better approaches $\mathcal{C}_{\sqrt{n}}$.* Notice also that "A is close to $\mathcal{C}_{\sqrt{n}}$" means that the line $\ell_2 : A + t\vec{v}$ is "almost tangent" to the circumference. It also means that the point $A'' := \ell_1 \cap \ell_2$ is a point on $\ell_1 : ax - by = 0$, close to $\mathcal{C}_{\sqrt{n}}$, that has rational coordinates with denominator $a^2 + b^2$. This suggests the following Diophantine-approximation-theoretic reformulation of the almost-tangent method, which is essentially equivalent to [83, Lemma 2]. Notice that, comparing with (2.3.2), we improve on the dependency in a, b by choosing $\delta < 2$.

Proposition 2.3.3. Let a, b, n be integers, with $n > 0$ and a, b coprime. Let $H := \sqrt{a^2 + b^2}$ and suppose that

$$\sqrt{(a^2 + b^2)n} = m + \delta$$

for some $m \in \mathbb{N}$ and some $\delta_0 \leq \delta < 1$, where

$$\delta_0 := H^3/(2\sqrt{n}). \tag{2.3.3}$$

Then there exist integers x_B, y_B with $ax_B - by_A \leq \sqrt{2\delta} n^{1/4}$ and

$$n - \kappa < x_B^2 + y_B^2 \leq n,$$

where $\kappa := 2\sqrt{2\delta H} \cdot n^{1/4}$.

Proof: Let $\vec{u} := (b, a)/H^2$ be a vector of length $\|\vec{u}\| = 1/H$ in a direction parallel to the line $\ell_1 : ax - by = 0$. Then the points

$$A' := H\sqrt{n}\vec{u}$$
$$A'' := m\vec{u}$$

satisfy respectively $A' \in \ell_1 \cap \mathcal{C}_{\sqrt{n}}$ and $A'' \in \ell_1 \cap \ell_2$, where $\ell_2 : \ bx + ay = m$. Notice that

$$\overline{A'A''} = \delta \|\vec{u}\| = \frac{\delta}{H}.$$

We let B' be an intersection between ℓ_2 and $\mathcal{C}_{\sqrt{n}}$, then the Pythagorean theorem, applied on the right triangle with vertices $O = (0, 0)$, A'' and B' gives

$$\overline{A''B'} = \sqrt{n - (\sqrt{n} - \delta/H)^2} \leq \sqrt{2\delta/H} \cdot n^{1/4}. \tag{2.3.4}$$

We let $\alpha = \angle A''OB'$ and we notice that α is equal to the angle between ℓ_2 and the tangent at B' to the circle. Finally, we let $B = (x_B, y_B)$ be a lattice point $B \in \mathbb{Z}^2 \cap \ell_2$ with

$$\overline{BB'} \leq H.$$

Such point exists because a, b are coprime and so the Diophantine equation $bx_B + ay_B = m$ has solutions. We notice that $H \leq 2\overline{A''B'}$ because $\delta \geq \delta_0$, therefore we can take B to be inside the circle. Moreover

$$\sin\alpha = \frac{\overline{A''B'}}{\overline{OA'}} \leq \sqrt{\frac{2\delta}{H}} n^{-1/4}$$

and so

$$\delta_B := \operatorname{dist}(B, \mathcal{C}_{\sqrt{n}}) \leq \overline{BB'} \sin\alpha \leq \sqrt{2\delta H} \cdot n^{-1/4}$$

Therefore

$$n - x_B^2 - y_B^2 = n - (\sqrt{n} - \delta_B)^2 \leq 2\sqrt{n}\delta_B \leq 2\sqrt{2\delta H} \cdot n^{1/4}.$$

Notice also that $y_B/x_B \approx a/b$, more precisely:

$$ax_B - by_B = \overline{A''B} \cdot H \leq \sqrt{2\delta H} n^{1/4}$$

by (2.3.4) and the inequality $\overline{A''B} \leq \overline{A''B'}$. ∎

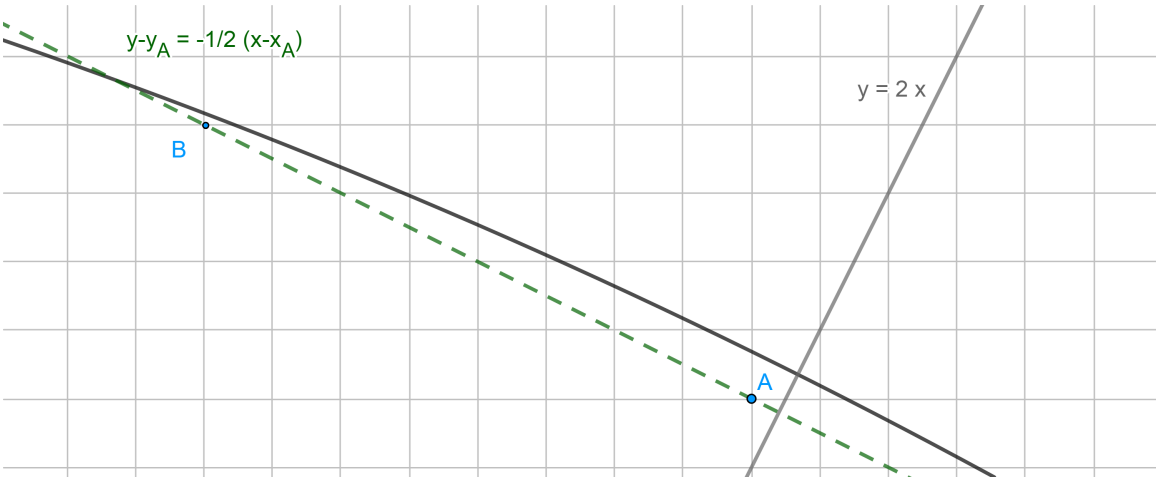


Figure 2.3.2: An illustration of the almost-tangent method

We already remarked that the existence of elements of $\mathcal{S}_{2,2}$ in short intervals is equivalent to the existence of lattice points in thin annuli $\mathcal{A}_{r,r-d}$. Then a natural approach is to estimate the cardinality of $\mathbb{Z}^2 \cap \mathcal{A}_{r,r-d}$ by analytic means. We may write

$$\#(\mathbb{Z}^2 \cap \mathcal{A}_{\sqrt{n},\sqrt{n-k}}) = R(n) - R(n-k)$$

where

$$R(n) := \#(\mathbb{Z} \cap \mathcal{D}_{\sqrt{n}})$$

counts the number of lattice points inside the circle of radius $\sqrt{n}$. One may naively hope to address Littlewood's problem (Problem 2.3.2) by computing the value of $R(n)$ with enough precision. Since the area of the circle is πn, we expect $R(n)$ to have more or less this value.

Problem 2.3.4. Find $\theta > 0$ as small as possible such that the asymptotic

$$R(n) = \pi n + O(n^{\theta}) \tag{2.3.5}$$

holds for $n \to \infty$.

This is known as the *Gauss circle problem* because around 1800 Gauss [58, p.277] [69, p.67] proved (2.3.5) with $\theta = 1/2$. This problem is fundamentally number-theoretic because $R(n)$ can be written as

$$R(n) = \sum_{i=0}^{n} r_2^*(i),$$

where $r_2^*(\cdot)$ is the sum of two squares function defined in Equation (2.1.3). Because of Equation (2.1.4) the Gauss circle problem is related to the *Dirichlet divisor problem* of estimating the quantity

$$D(x) := \sum_{n \leq x} d(n) = \sum_{m \leq x} \left\lfloor \frac{x}{m} \right\rfloor$$

and it turns out that it is also related to the estimates for $|\zeta(1/2 + it)|$, that is the values of the Riemann zeta function on the critical line. The Gauss circle problem has a long history and a complete treatment of this subject would take us too afar. The interested reader is encouraged to consult the well-documented survey articles [9, 86] for more bibliographical details. The first progress was recorded by Sierpiński [142], who proved (2.3.4) with $\theta = 1/3$,
(1)

shortly after Voronoï [159] had made a similar progress on the Dirichlet divisor problem. In the last decades this estimate has been refined by many authors [89, 84, 15] using an ingenious and widely applicable method of Bombieri and Iwaniec [14]. The current record $\theta = 517/1648 \approx 0.31371$ is due to Bourgain and Watts [15]. However already in 1915 both Hardy [70] and Landau [97] observed that θ cannot be too small. More precisely Hardy proves that

$$R(n) - \pi n = \omega((n \log n)^{1/4}),$$

where the small-ω notation $A = \omega(B)$ means that the ratio $|A|/B$ is not bounded from above. In particular, we see that even an optimal solution to Problem 2.3.4 would not be enough to address Problem 2.3.2. Nevertheless, we would like to spend a few words on the ideas that are exploited in classical approaches to the Gauss circle problem, since they might find a use in attacking Littlewood's problem as well. As stated in [86], much of the progress after Sierpiński's has been obtained using formulas involving Bessel functions $J_\nu(x)$ of the first kind. For integer $\nu \in \mathbb{Z}$ these functions can be defined as

$$J_\nu(x) := \frac{1}{\pi} \int_0^\pi \cos(\nu\alpha - x \sin\alpha)\, d\alpha$$

and they come out naturally in problems with bidimensional rotational symmetry. For example the 2D Fourier transform of the characteristic function $\mathbb{1}_{\mathcal{D}_1}$ of the unit disk is

$$\widehat{\mathbb{1}_{\mathcal{D}_1}}(u, v) = \frac{1}{\sqrt{u^2 + v^2}} J_1(2\pi\sqrt{u^2 + v^2}).$$

(1) Sierpiński's proved this result when competing for a scholarship as an undergraduate student of the University of Warszaw. His proof, 41 pages long, is ingenious but essentially elementary. A French translation can be found in his Œuvres [143]. Later authors, starting with Landau, shortened and simplified the argument. The author's favourite treatment of Sierpiński's result is via Fourier analysis as in [153].

The role of the Bessel functions in the Gauss circle problem is demonstrated by formulas such as Hardy's identity

$$R(n) - \frac{r_2^*(n)}{2} = \pi n + \sqrt{n} \sum_{m=0}^{\infty} \frac{r_2^*(m)}{\sqrt{m}} J_1(2\pi\sqrt{mn})$$

and other formulas due to Ramanujan, Landau, Voronoï, Chandrasekharan, Narasimhan and others, see [9]. One hopes to get an asymptotic estimate from this formula because the functions $J_\nu(x)$ exhibit oscillatory behaviour and slowly decay to zero, namely

$$J_\nu(x) = \left(\frac{2}{\pi x}\right)^{1/2} \cos\left(x - \frac{\nu}{2}\pi - \frac{\pi}{4}\right) + O(x^{-3/2})$$

as $x \to \infty$. Bessel functions have also been used in the study of the size of gaps of $\mathcal{S}_{2,2}$ by e.g. Kalmynin, in his proof of (2.2.9).

Another folklore conjecture, related to Littlewood's problem, is that Landau's asymptotic should hold for intervals of length x^ϵ:

Problem 2.3.5. Fix $0 < \epsilon < 1$. Is it true that

$$S(x + x^\epsilon) - S(x) = (c_{LR} + o(1)) \frac{x^\epsilon}{\sqrt{\log x}}$$

as $x \to \infty$?

In [78] Hooley writes that it is possible to confirm this conjecture for $\epsilon > 1/2$ assuming the Riemann Hypothesis for both the Riemann zeta function $\zeta(s)$ and the Dirichlet L-function $L(s, \chi_{2,4})$, where $\chi_{2,4}$ is the only non-principal (quadratic) character modulo 4 (compare the proof of Landau's asymptotics at the beginning of Section 2.2). He also states that one can prove it unconditionally for $\epsilon > 7/12$ using methods of Ingham, Montgomery and Huxley for primes [82]. Recently, this has been improved to $\epsilon > 0.55$ by Matomäki and Teräväinen [107]. The function field analogue of Problem 2.3.5 has been studied in [4].

Chapter 3

Gaps between sums of three cubes

In this chapter we consider the set $\mathcal{S}_{3,3}$ of natural numbers that can be written as a sum of three nonnegative cubes. Our main goal here is to show that its complement $\mathbb{N} \setminus \mathcal{S}_{3,3}$ is a *thick* set. In other words:

Theorem 3.0.1. There are arbitrarily long sequences of consecutive numbers, none of which is a sum of three nonnegative cubes.

The set $\mathcal{S}_{3,3}$ is conjectured to have positive natural density [44, 79, 35, 36]. Therefore Theorem 3.0.1 suggests that the elements of $\mathcal{S}_{3,3}$ distribute unevenly in the set of all natural numbers.

We begin the chapter with the discussion of a naive strategy of proof that was successful for the analogous problem for sums of two squares but does not work here. We then propose a modification of that approach which leads to a proof of Theorem 3.0.1.

3.1 Arithmetic progressions without sums of three cubes

If p is a prime number congruent to 3 modulo 4, we know from Fermat's Christmas theorem that the congruence $x^2 + y^2 \equiv p \bmod p^2$ does not admit any solution. This observation was sufficient, in the previous chapter, to conclude that there are arbitrarily long gaps between sums of two squares. By the Chinese reminder theorem, an analogous result would follow for sums of three cubes if we were able to find infinitely many pairwise coprime moduli M_i and residue classes m_i modulo M_i such that the congruence $x^3 + y^3 + z^3 \equiv m_i \bmod M_i$ has no integer solutions. For example, it is easy to show that there are no sums of three cubes that are congruent to 4 or 5 modulo 9. This is because the only cubic residues modulo 9 are -1, 0 and 1. However, it turns out that there is no other modular constraint on the set $\mathcal{S}_{3,3}$.

Proposition 3.1.1. Let M and m be any integers with $3 \nmid M$. Then the congruence

$$x^3 + y^3 + z^3 \equiv m \pmod{M} \tag{3.1.1}$$

is solvable in integers x, y, z.

Proof: By the Chinese remainder theorem, it suffices to prove the proposition in case M is equal to a power of a prime $p \neq 3$. We notice that the number 9 is invertible modulo M, so there exists some integer T such that $9T \equiv m \bmod M$. We are going to use the following polynomial identity

$$(T^3 \mp 1)^3 + (-T^3 + 3T \pm 1)^3 + (\pm 3T^2 + 3T)^3 = 9T(T^2 \pm T + 1)^3. \tag{3.1.2}$$

This is a special case of some more general formulas from the book [106]. We now observe that, for every integer T, there is a choice of sign such that p does not divide $T^2 \pm T + 1$. Then if we divide both sides of (3.1.2) by $T^2 \pm T + 1$ we find a solution in $\mathbb{Z}/M\mathbb{Z}$ to the equation $\bar{x}^3 + \bar{y}^3 + \bar{z}^3 = m$, which lifts to an integer solution of (3.1.1). ∎

As a consequence, we see that the strategy outlined above is not practicable. As an aside, I take a few lines to answer the following question, which was asked to me by James Dowdall during a seminar: are there 7 consecutive numbers that can be written as sums of three cubes? The question arises because, since in $\mathcal{S}_{3,3}$ there are no elements congruent to 4 or 5 modulo 9, it is not possible to have more than 7 consecutive numbers all of which are in $\mathcal{S}_{3,3}$. We answer the question in the affirmative:

$$\begin{aligned}
47420214 &= 302^3 + 263^3 + 119^3 \\
47420215 &= 359^3 + 92^3 + 72^3 \\
47420216 &= 348^3 + 174^3 + 20^3 \\
47420217 &= 290^3 + 234^3 + 217^3 \\
47420218 &= 333^3 + 213^3 + 94^3 \\
47420219 &= 360^3 + 91^3 + 22^3 \\
47420220 &= 348^3 + 142^3 + 127^3
\end{aligned}$$

If we trust our computer search, this should be the smallest 7-tuple with the required property. Moreover each of these 7 integers is representable as a sum of three nonnegative cubes in only one way up to permutation.

3.2 Arithmetic progressions with few sums of three cubes

The game-changing modification to the approach examined in the previous section is summarized in the following sentence.

What if we consider congruences as in (3.1.1) *that have "few" solutions modulo M?*

The idea is that such congruence gives rise to an arithmetic progression $m + \mathbb{N}M$ in which the elements of $\mathcal{S}_{3,3}$ are relatively scarce. Then if we are able to package K consecutive arithmetic progressions with the above property, we obtain an arithmetic progression of intervals $\{[m+1, m+K] + hM\}_{h\in\mathbb{N}}$, each of which has a high chance of being in the complement of $\mathcal{S}_{3,3}$. This heuristic argument is formalized in the following proposition, in which we use an elementary double-counting technique known as the Maier matrix method [66]. First, we set some notation. For all $n \in \mathbb{N}$ we define

$$r_3(n) := \#\{(x, y, z) \in \mathbb{N}^3 : \ x^3 + y^3 + z^3 = n\}$$

to be the number of representations of n as a sum of three cubes, so that $\mathcal{S}_{3,3} = \{n : r_3(n) \geq 1\}$. In the same spirit, for $M \in \mathbb{N}_+$ and $m \in \mathbb{Z}/M\mathbb{Z}$ we define

$$r_3(m, M) := \#\{(\bar{x}, \bar{y}, \bar{z}) \in (\mathbb{Z}/M\mathbb{Z})^3 : \ \bar{x}^3 + \bar{y}^3 + \bar{z}^3 \equiv m \bmod M\}$$

be the number of solutions modulo M to (3.1.1). Since there are M^3 choices for the values of $\bar{x}, \bar{y}, \bar{z}$ modulo M, and there are M possible values for the sum $\bar{x}^3 + \bar{y}^3 + \bar{z}^3$, the average value of $r_3(m, M)$ is M^2, as m varies through the residue classes modulo M.

Proposition 3.2.1. Let K be a positive integer and suppose there exist $m, M \in \mathbb{N}$ with $0 \leq m$ and $m + K < M$ such that

$$r_3(m + i, M) < \frac{1}{K}M^2 \tag{3.2.1}$$

for $i = 1, \ldots, K$. Then there exist K consecutive natural numbers in the complement of $\mathcal{S}_{3,3}$. More precisely, there is an integer h with $1 \leq h < M^2$ such that none of the numbers $m + hM + 1, \ldots, m + hM + K$ belongs to $\mathcal{S}_{3,3}$.

Proof: Suppose the contrary. So for every $h \in \mathbb{N}$ there is some $i_h \in \{1, \ldots, K\}$ and there are $x_h, y_h, z_h \in \mathbb{N}$ such that

$$x_h^3 + y_h^3 + z_h^3 = m + hM + i_h.$$

Notice that if $h < M^2$ then $x_h^3 + y_h^3 + z_h^3 < M^3$ and so $x_h, y_h, z_h < M$. Therefore we see that the set

$$\mathcal{A} := \{(\overline{x}_h, \overline{y}_h, \overline{z}_h) \in (\mathbb{Z}/M\mathbb{Z})^3 : \ 0 \leq h < M^2\}$$

given by their residues $\bmod M$ has cardinality exactly M^2. However, we also have $\overline{x}_h^3 + \overline{y}_h^3 + \overline{z}_h^3 \equiv m + i_h \bmod M$ and so the set $\mathcal{A}$ has cardinality bounded above by

$$\sum_{i=1}^{K} r_3(m + i, M) < M^2$$

which is a contradiction. ∎

Our next objective is to show that the hypotheses of Proposition 3.2.1 are fulfilled for every K. Equivalently, we need to show that for every $K \in \mathbb{N}_+$ and all $\epsilon > 0$ there are K consecutive nonzero congruence classes $m+1, \dots, m+K$ modulo some natural integer M such that

$$r_3(m+i, M) \leq \epsilon M^2$$

for every $i = 1, \dots, K$. We start by solving this problem for a single class m mod M and after we will have understood this case, we will explain how to produce K consecutive classes $m+1, \dots, m+K$ with the required property. We observe that our goal cannot be reached with M equal to a prime number. In fact it is possible to prove (see e.g. the Weil bound for Jacobi sums in the next chapter) that for a prime p we have $r_3(m,p) = (1+o(1))p^2$ as $p \to \infty$, uniformly in m. However, if we take M that factors as $M = p_1 \dots p_\ell$ for distinct primes p_j and use the Chinese remainder theorem, we see that a weaker estimate for prime moduli turns out to be sufficient. Namely, it is enough to find a set of prime numbers $\mathcal{P}$ and congruence classes m mod p such that

$$r_3(m,p) \leq (1-\epsilon_p)p^2 \tag{3.2.2}$$

for some $\epsilon_p > 0$ such that

$$\prod_{p \in \mathcal{P}} (1-\epsilon_p) = 0$$

or, equivalently, such that $\sum_{p\in\mathcal{P}} \epsilon_p = \infty$. In this chapter we consider the primes p congruent to 1 modulo 3 and the classes m mod p that are not cubic residues, for which we can prove (3.2.2) with $\epsilon_p \asymp p^{-1}$. In Chapter 6 we will discuss some refinements based on a more elaborate argument. There we construct a more complicated set of primes and use the zero class $m = 0$, for which we can prove a stronger estimate $\epsilon_p \asymp p^{-1/2}$.

3.3 Solutions count modulo p and noncubic residues

For a prime number p we let $\mathbb{F}_p := \mathbb{Z}/p\mathbb{Z}$ be the finite field with p elements and we denote by $\mathbb{F}_p^\times$ its multiplicative group, that is a cyclic group with $p-1$ elements. If p is congruent to 2 modulo 3, then the map $x \mapsto x^3$ is a bijection in $\mathbb{F}_p$. Therefore for every $m \in \mathbb{F}_p$ we have

$$\#\{(x,y,z) \in \mathbb{F}_p^3 : \ x^3+y^3+z^3 = m\} = \#\{(x',y',z') \in \mathbb{F}_p^3 : \ x'+y'+z' = m\},$$

so $r_3(m,p) = p^2$ for every m. If $p \equiv 1 \bmod 3$ the situation is more interesting. In this case the image of $x \mapsto x^3$ is equal to $\{0\} \cup H$, where H is a subgroup of $\mathbb{F}_p^\times$

with index 3. If g is a generator of the multiplicative group $\mathbb{F}_p^\times$, we have that the residue classes modulo p decompose into four sets $\{0\}, H, gH, g^2H$, that are the orbits of the multiplicative action $H \curvearrowright \mathbb{F}_p$. We observe that the function $r_3(\cdot, p)$ is constant on these four H-orbits because for every $u \in \mathbb{F}_p^\times$ the multiplication by u induces a bijection

$$\{(x,y,z) \in \mathbb{F}_p^3 :\ x^3+y^3+z^3 = m\} = \{(x',y',z') \in \mathbb{F}_p^3 :\ x'^3+y'^3+z'^3 = mu^3\}.$$

p	p^2	$r_3(0,p)$	$r_3(1,p)$	$r_3(g,p)$	$r_3(g^2,p)$	p^2-3p
7	49	55	90	27	27	28
31	961	1081	1143	864	864	868
37	1369	973	1602	1269	1269	1258
163	26569	30619	27522	26055	26055	26080
313	97969	87049	99882	97065	97065	97030
997	994009	1003969	999981	991008	991008	991018

Table 3.3.1: Values of $r_3(-,p)$ on the four H-orbits for $p \equiv 1 \bmod 3$.

In Table 3.3.1 we list some values of $r(m,p)$ for $p \equiv 1 \bmod 3$ and we observe some interesting patterns. First, we notice that $r_3(1,p)$ seems to be always larger than the average value p^2 of $r_3(\cdot,p)$. Symmetrically, the values of $r_3(g,p)$ and $r_3(g^2,p)$ seem to be consistently smaller than this value. Moreover, we observe that the equality $r_3(g,p) = r_3(g^2,p)$ holds for all tabulated values. This phenomenon is nontrivial and is special to the cubic polynomial $x^3+y^3+z^3$. In Table 3.3.2 we compute the number of solutions to $F(\mathbf{x}) = m$ modulo $p = 31$ using diagonal polynomials $F(\mathbf{x})$ with a different number of variables or a different degree (notice that 3 is a multiplicative generator modulo 31), and we do not see an analogous equality anymore. Since we

$F(\mathbf{x})$	$r_F(0,31)$	$r_F(1,31)$	$r_F(3,31)$	$r_F(3^2,31)$	$r_F(3^3,31)$	$r_F(3^4,31)$
$x_1^3+x_2^3$	91	33	36	18	–	–
$x_1^3+\cdots+x_4^3$	35371	30225	30132	28458	–	–
$x_1^5+x_2^5+x_3^5$	1951	2040	825	500	900	375
$x_1^5+\cdots+x_5^5$	1422751	1288025	876500	733125	942375	694375

Table 3.3.2: Number of solutions $r_F(m,31)$ to $F(\mathbf{x}) = m$ in $\mathbb{F}_{31}$.

are interested in the classes m for which $r_3(m,p) < p^2$, we now try to estimate $r_3(g,p)$ or $r_3(g^2,p)$ as p varies. Notice that the residue classes in $gH \cup g^2H$ are exactly the noncubic residues modulo p. The last columns of Table 3.3.1 show that the value of $r_3(g,p)$ oscillates around p^2-3p. The key to find a neat exact formula is to add a contribution coming from $r_3(0,p)$, as follows.

Proposition 3.3.1. Let p be a prime number congruent to 1 modulo 3 and let m be any noncubic residue modulo p. Then

$$r_3(0,p) + (p-1)r_3(m,p) = p^3 - 3p(p-1). \tag{3.3.1}$$

We will prove Proposition 3.3.1 in the next section. For the moment we remark that one may prove the estimate $r_3(0,p) = p^2 + O(p^{3/2})$, which implies by (3.3.1) that $r_3(m,p) = p^2 - 3p + O(\sqrt{p})$ if m is a noncubic residue. However for our goal we may content ourselves with a simple estimate from above.

Corollary 3.3.2. Let $p \equiv 1 \bmod 3$ be a prime number and let m be any noncubic residue modulo p. Then

$$r_3(m,p) \leq p^2 - 2p. \tag{3.3.2}$$

Proof: Among the triples counted by $r_3(0,p)$ we find the p triples $(x,-x,0)$ for $x \in \mathbb{F}_p$. Thus $r_3(0,p) \geq p$ and so (3.3.1) implies (3.3.2). ∎

This is a promising result with respect to the strategy outlined in the previous section, because

$$\sum_{\substack{p \text{ prime} \\ p \equiv 1 \bmod 3}} \frac{2}{p} = \infty \tag{3.3.3}$$

by the prime number theorem in arithmetic progressions.

3.4 Multiplicative characters and Fermat cubics

In order to detect the cubic and noncubic residues modulo p we are going to use a multiplicative character modulo p. The results of this section hold for an arbitrary finite field $\mathbb{F}_q$, where q is a power of a prime, as long as $q \equiv 1 \bmod 3$. Let $\zeta = e^{2\pi i/3}$ be a primitive cube root of unity in $\mathbb{C}$ and denote by $\mu_3 := \{1, \zeta, \zeta^{-1}\}$ the cyclic multiplicative group of order 3 generated by ζ. Since the multiplicative group $\mathbb{F}_q^\times$ is a cyclic group with $q-1$ elements and $3 \,|\, q-1$, there exists some nontrivial group homomorphism $\chi_q^\times : \mathbb{F}_q^\times \to \mu_3$, which we extend to a multiplicative monoid morphism $\chi_q : \mathbb{F}_q \to \mu_3 \cup \{0\}$ by the condition $\chi_q(0) = 0$. We call such χ_q a *cubic character* of $\mathbb{F}_q$. A *noncubic residue* in $\mathbb{F}_q$ is an element $a \in \mathbb{F}_q$ for which $x^3 = a$ has no solution in $\mathbb{F}_q$. Notice that the noncubic residues are precisely the elements for which $\chi_q(a) \in \{\zeta, \zeta^2\}$. We also have the following useful lemma.

Lemma 3.4.1. For every $t \in \mathbb{F}_q$ the number of solutions of the equation $x^3 = t$ in $\mathbb{F}_q$ is equal to

$$1 + \chi_q(t) + \chi_q(t^2),$$

if $q \equiv 1 \bmod 3$ and χ_q is a cubic character as above.

Proof: Since $3 \,|\, q-1$, the field $\mathbb{F}_q$ contains third roots of unity. If t is a nonzero cubic residue, then $\chi_q(t) = 1$ and there are three solutions to the equation $x^3 = t$. If t is a noncubic residue, then $\chi_q(t) \in \{\zeta, \zeta^2\}$ and $1 + \zeta + \zeta^2 = 0$. Finally, if $t = 0$ the displayed expression is equal to 1 and the only solution to $x^3 = 0$ is $x = 0$. ∎

The left-hand side of (3.3.1) can be rewritten as

$$r_3(0,p) + (p-1)r_3(m,p) = \#\{(x,y,z,t) \in \mathbb{F}_p^4 : \ x^3 + y^3 + z^3 + mt^3 = 0\}$$

because we have $r_3(m(-t)^3, p) = r_3(m,p)$ for every $t \in \mathbb{F}_p^\times$. This expression has therefore a natural geometric interpretation: it is (in the case $q = p$) the number of $\mathbb{F}_q$-points of the twisted Fermat affine cubic threefold

$$\tilde{\mathcal{F}}[m] := \mathcal{V}(x^3 + y^3 + z^3 + mt^3 = 0) \subseteq \mathbb{A}^4_{\operatorname{Spec} \mathbb{F}_q}.$$

Indeed, for all prime power q and all $m \in \mathbb{F}_q$ we have

$$\#\tilde{\mathcal{F}}[m](\mathbb{F}_q) = \#\{(x,y,z,t) \in \mathbb{F}_q^4 : \ x^3 + y^3 + z^3 + mt^3 = 0\}.$$

By this discussion, we therefore see that Proposition 3.3.1 is a special case of the following proposition. This result follows from an exercise in [137, Chapter 2, Exercise (d)], whose Hint suggests to use Galois descent on the (projectivization of) the Fermat cubic. The Galois descent theory of Fermat varieties is intimately related with their so-called inductive structure [19] and so to Jacobi sums. We will show how to use the theory of Jacobi sums for similar statements in later chapters. In the following proof instead we reproduce the first computation done by the author when working on this problem, which is based on a more direct handling of the character sums involved.

Proposition 3.4.2. If $q \equiv 1 \bmod 3$ and $m \in \mathbb{F}_q^\times$ is a noncubic residue in $\mathbb{F}_q$, we have

$$\#\tilde{\mathcal{F}}[m](\mathbb{F}_q) = q^3 - 3q^2 + 3q.$$

Proof: Let $\chi_q : \mathbb{F}_q \to \mu_3 \cup \{0\}$ be a cubic character as above and let $u = m^{-1} \in \mathbb{F}_q^\times$ be the multiplicative inverse of m. Then $\chi_q(u) = \zeta$ or $\chi_q(u) = \zeta^2$.

By Lemma 3.4.1, for every $c \in \mathbb{F}_q$ the (integer!) numbers

$$1 + \chi_q(c) + \chi_q(c^2) \quad \text{and} \quad 1 + \chi_q(u)\chi_q(c) + \chi_q(u^2)\chi_q(c^2),$$

are respectively equal to the number of solutions x, in $\mathbb{F}_q$, of the equations $c = x^3$ and $c = mx^3$. Then we reformulate our counting problem as follows

$$\#\tilde{\mathcal{F}}[m](\mathbb{F}_q) = \sum_{x_1+x_2+x_3+x_4=0} (1 + \chi_q(u)\chi_q(x_4) + \chi_q(u^2)\chi_q(x_4^2)) \prod_{i=1}^{3} (1 + \chi_q(x_i) + \chi_q(x_i^2)).$$

We expand this product using the multiplicativity of the character χ_q and, according to the exponents appearing in the monomials, we collect the terms into $3^4 = 81$ sums

$$\#\tilde{\mathcal{F}}[m](\mathbb{F}_q) = \sum_{a_1,a_2,a_3,a_4\in\{0,1,2\}} \chi_q(u^{a_4}) \sum_{x_1+x_2+x_3+x_4=0} \chi_q(x_1^{a_1}x_2^{a_2}x_3^{a_3}x_4^{a_4}), \tag{3.4.1}$$

where we formally define $0^0 = 1$. The terms with $a_1 = a_2 = a_3 = a_4 = 0$ give the main contribution

$$\sum_{x_1+x_2+x_3+x_4=0} 1 = q^3.$$

If some but not all of the exponents a_i vanish, the corresponding contribution is zero. Indeed, if for example $a_4 = 0$ but $a_3 \in \{1, 2\}$, we have

$$\sum_{x_1+x_2+x_3+x_4=0} \chi_q(x_1^{a_1}x_2^{a_2}x_3^{a_3}) = \sum_{x_1,x_2,x_3\in\mathbb{F}_q} \chi_q(x_1^{a_1}x_2^{a_2})\chi_q(x_3^{a_3}) = 0,$$

because $\sum_{x_3\in\mathbb{F}_q} \chi_q(x_3^{a_3}) = 0$. This argument applies to all the 64 sums of this form. Now we see that also the 2 sums corresponding to the cases $a_1 = a_2 = a_3 = a_4 \in \{1, 2\}$, where all exponents are equal, vanish. Indeed if $g \in \mathbb{F}_q^\times$ is any element with $\chi_q(g) = \zeta$ we have

$$\begin{aligned}\sum_{x_1+x_2+x_3+x_4=0} \chi_q(x_1^{a_1}x_2^{a_1}x_3^{a_1}x_4^{a_1}) = &\sum_{x_1+x_2+x_3+x_4=0} \frac{1}{3}\chi_q(x_1^{a_1}x_2^{a_1}x_3^{a_1}x_4^{a_1}) + \\ &+ \frac{1}{3}\chi_q^{-4a_1}(g)\chi_q((gx_1)^{a_1}(gx_2)^{a_1}(gx_3)^{a_1}(gx_4)^{a_1}) + \\ &+ \frac{1}{3}\chi_q^{-8a_1}(g)\chi_q((g^2x_1)^{a_1}(g^2x_2)^{a_1}(g^2x_3)^{a_1}(g^2x_4)^{a_1}) = \\ = &\left(\sum_{x_1+x_2+x_3+x_4=0} \chi_q(x_1^{a_1}x_2^{a_1}x_3^{a_1}x_4^{a_1})\right)\frac{1+\zeta+\zeta^2}{3} = \\ = &\ 0,\end{aligned}$$

because the multiplication by g (or g^2) induces a linear automorphism of the hyperplane in $\mathbb{F}_q^4$ given by $x_1 + \cdots + x_4 = 0$. To deal with the remaining 14 sums we notice the following identity:

Lemma 3.4.3. For every for every $c \in \mathbb{F}_q$ we have

$$\sum_{b\in\mathbb{F}_q} \chi_q(b^2(c-b)) = \begin{cases} q-1 \text{ if } c = 0, \\ -1 \text{ if } c \neq 0. \end{cases}$$

Proof: If $c = 0$ the above sum reduces to

$$\sum_{b\in\mathbb{F}_q} \chi_q((-b)^3) = q - 1,$$

because $\chi_q((-b)^3)$ is equal to 1 if $b \neq 0$ and is equal to 0 otherwise. For $c \neq 0$ we have

$$\sum_{b \in \mathbb{F}_q} \chi_q(b^2(c-b)) = \sum_{b \in \mathbb{F}_q^\times} \chi_q(b^2)\chi_q(c-b)$$

and for $b \neq 0$ we have $\chi_q(b^2) = \chi_q(b^{-1})$, so the above sum becomes

$$\sum_{b \in \mathbb{F}_q^\times} \chi_q(cb^{-1}-1) = \sum_{\substack{d \in \mathbb{F}_q \\ d \neq -1}} \chi_q(d) = 0 - \chi_q(-1) = -1.$$

∎

With this identity in hand, we can compute the contribution of the sums corresponding to the cases in which $a_i \in \{1,2\}$ for $i = 1,2,3,4$ are not all equal. For example, if $a_4 = 1$ and $a_3 = 2$ we get

$$\begin{aligned}\sum_{x_1+x_2+x_3+x_4=0} \chi_q(x_1^{a_1}x_2^{a_2}x_3^2x_4^1) &= \sum_{s \in \mathbb{F}_q} \sum_{x_1+x_2=s} \chi_q(x_1^{a_1}x_2^{a_2}) \sum_{x_3 \in \mathbb{F}_q} \chi_q(x_3^2(s-x_3)) \\ &= \sum_{s \in \mathbb{F}_q} \sum_{x_1+x_2=s} \chi_q(x_1^{a_1}x_2^{a_2})(q[s=0]-1) \\ &= q \sum_{x_1+x_2=0} \chi_q(x_1^{a_1}x_2^{a_2}) - \sum_{s \in \mathbb{F}_q} \sum_{x_1+x_2=s} \chi_q(x_1^{a_1}x_2^{a_2}),\end{aligned}$$

where the Yverson bracket notation $[s=0]$ means

$$[s=0] = \begin{cases} 1 \text{ if } s = 0 \\ 0 \text{ if } s \neq 0. \end{cases}$$

We see that the second part of the last expression always vanishes, since $a_1, a_2 \neq 0$:

$$\sum_{s \in \mathbb{F}_q} \sum_{x_1+x_2=s} \chi_q(x_1^{a_1}x_2^{a_2}) = \left(\sum_{x_1 \in \mathbb{F}_q} \chi_q(x_1^{a_1})\right)\left(\sum_{x_2 \in \mathbb{F}_q} \chi_q(x_2^{a_2})\right) = 0.$$

On the other hand the first part simplifies to

$$q \sum_{x_1+x_2=0} \chi_q(x_1^{a_1}x_2^{a_2}) = q \sum_{x_1 \in \mathbb{F}_q} \chi_q(x_1^{a_1+a_2}) \underbrace{\chi_q^{a_2}(-1)}_{=1},$$

which vanishes if $a_1 + a_2 \neq 3$ and is equal to $q(q-1)$ otherwise. Therefore, with this argument we see that there are exactly 6 sums other than the one corresponding to $a_1 = a_2 = a_3 = a_4 = 0$ which contribute nontrivially to (3.4.1), namely the sums corresponding to

$$(a_1,a_2,a_3,a_4) \in \{(1,1,2,2),(1,2,1,2),(1,2,2,1),(2,1,1,2),(2,1,2,1),(2,2,1,1)\}.$$

Their contribution was shown to be $q(q-1)$, and each of them is multiplied by $\chi_q(u^{a_4})$. The final computation thus gives

$$\#\tilde{\mathcal{F}}[m](\mathbb{F}_q) = q^3 + 3(\chi_q(u) + \chi_q(u^2))q(q-1) = q^3 - 3q(q-1)$$

as we claimed, since $\chi_q(u) + \chi_q(u^2) = \zeta + \zeta^2 = -1$. ∎

3.5 Existence of consecutive noncubic residue classes

At this point in order to finish the proof of Theorem 3.0.1 it is enough to show that for every positive integer K, and every prime number $p \equiv 1 \bmod 3$ large enough, there exist K consecutive noncubic residue classes modulo p. The Chinese remainder theorem and the estimate of Corollary 3.3.2 would imply that the hypothesis of Proposition 3.2.1 is fulfilled for every K and so our claimed result would follow.

In order to prove the existence of consecutive residue classes with a prescribed multiplicative pattern there is a general approach [139] based on the classical Weil's bound for character sums [23, 161]. We state this bound for cubic characters, but it is true in general for characters of any order at least 2.

Lemma 3.5.1. Let $q \equiv 1 \bmod 3$ be a power of a prime, let χ_q be a cubic character of $\mathbb{F}_q$ and let $f \in \mathbb{F}_q[x]$ be a polynomial of positive degree that is not a constant multiple of the 3rd power of a polynomial. Let d be the number of distinct roots of f in its splitting field over $\mathbb{F}_q$ then

$$\left| \sum_{x \in \mathbb{F}_q} \chi_q(f(x)) \right| \leq (d-1)\sqrt{q}.$$

We are now ready for the last proposition of the chapter.

Proposition 3.5.2. Given a prime number $p \equiv 1 \bmod 3$ and a positive number K let $\mathcal{C}_{p,K}^{nr}$ be the set of residue classes $m \in \mathbb{Z}/p\mathbb{Z}$ such that $m+i$ is a noncubic residue for $i = 1, \ldots, K$. Then we have

$$\#\mathcal{C}_{p,K}^{nr} = (2/3)^K p + O(2^K K\sqrt{p})$$

for some absolute implied constant. In particular $\mathcal{C}_{p,K}^{nr}$ is not empty for all p large enough.

Proof: First we observe that every element $m \in \mathcal{C}_{p,K}^{nr}$ admits an integer representative with $m < p - K$ because the zero class is not a noncubic residue. Let χ_p be a

(nontrivial) cubic character of $\mathbb{F}_p$ as in the previous section and observe that for every $t \in \mathbb{F}_p^\times$ we have

$$2 - \chi_p(t) - \chi_p(t^2) = \begin{cases} 3 & \text{if } \chi_p(t) \in \{\zeta, \zeta^2\} \\ 0 & \text{if } \chi_p(t) = 1. \end{cases}$$

Therefore we have

$$\begin{aligned} \#\mathcal{C}_{p,K}^{nr} &= 3^{-K} \sum_{m=0}^{p-K-1} \prod_{i=1}^{K} \Big(2 - \chi_p(m+i) + \chi_p((m+i)^2)\Big) \\ &= 3^{-K} \sum_{x=0}^{p-1} \prod_{i=1}^{K} \Big(2 - \chi_p(x+i) + \chi_p((x+i)^2)\Big) - \theta_1 K, \quad 0 \le \theta_1 \le 1, \end{aligned}$$

since the contribution of each $x \in [p-K, p-1]$ to the whole sum is at most $3^{-K} \cdot 3^K = 1$.

Expanding the product, and using the multiplicativity of the character χ_p, one can write $\#\mathcal{C}_{p,K}^{nr}$ as a sum of the main term $3^{-K} \sum_x 2^K = (2/3)^K p$ and $3^K - 1$ additional terms, each of the form $2^\alpha 3^{-K} \sum_x \chi_p(Q(x))$, where $\alpha \le K$ and $Q(x)$ is a nonconstant polynomial with coefficients in $\mathbb{F}_p$ and degree at most $2K$. Moreover, these polynomials all have the shape $Q(x) = \prod_{i=1}^K (x+i)^{a_i}$ for some $a_1, \dots, a_K \in \{0, 1, 2\}$. Hence we see that they are not constant multiples of cubes of polynomials. By Weil's bound (Lemma 3.5.1) we get that each of these remaining terms does not exceed $(2/3)^K \cdot (K-1)\sqrt{p}$ in absolute value. As a result, since $2^K \sqrt{p} > \theta_1 K$, we finally obtain

$$\#\mathcal{C}_{p,K}^{nr} = (2/3)^K p + \theta_2 2^K K \sqrt{p}, \qquad |\theta_2| < 1.$$

∎

Chapter 4

Gauss-Jacobi sums and gaps between sums of four fourth powers

In this chapter we prove the existence of arbitrarily long gaps between the numbers that can be written as sums of four fourth powers. In the strategy of the previous chapter a crucial step was to count the number of solutions modulo p to a congruence of the form $x_1^3 + x_2^3 + x_3^3 \equiv m \bmod p$. This is a goal that we achieved through computations that involved sums of cubic characters. In Sections 4.1 and 4.2 we review the theory of Gauss and Jacobi sums. We establish their main properties and we show how to use them to count the number of solutions to a congruence $F(\mathbf{x}) \equiv m \bmod p$, modulo a prime p, when $F(\mathbf{x})$ is a diagonal polynomial.

The computations with Gauss-Jacobi sums suggest that an adaptation of the strategy of the previous section might be successful for proving the existence of arbitrarily long gaps between the values of $F(\mathbf{x})$ in $\mathbb{N}$, if $F(\mathbf{x})$ has at most four variables. In Section 4.3 we show that this is indeed the case for the special polynomial $F_4(\mathbf{x}) := x_1^4 + x_2^4 + x_3^4 + x_4^4$. The proof is relatively elementary and it takes into account the primes that are congruent to 5 modulo 8. For more general diagonal polynomials instead one needs also to use equidistribution results of Sato-Tate type coming from the theory of L-functions [61, 114, 115]. The case of twisted homogeneous cubic and biquadratic diagonal forms is examined in a Chapter 6, while the case of unequal degrees will be the objective of a future publication.

4.1 Diagonal polynomials and Jacobi sums

If p is a prime number, we recall that by $\mathbb{F}_p := \mathbb{Z}/p\mathbb{Z}$ we denote the field with p elements and by $\mathbb{F}_p^\times := \mathbb{F}_p \setminus \{0\}$ the multiplicative group of its nonzero elements. A multiplicative character of $\mathbb{F}_p$ is by definition a group homomorphism $\chi \in \mathrm{Hom}(\mathbb{F}_p^\times, \mathbb{C}^\times)$. We denote

by $\mathbf{1}$ the trivial character, i.e. the one satisfying $\mathbf{1}(t) = 1$ for all $t \in \mathbb{F}_p^\times$. The multiplication between characters is defined pointwise on $\mathbb{F}_p^\times$, and the trivial character is a unit element with respect to this operation. With a slight abuse of notation, we extend every multiplicative character $\chi \in \mathrm{Hom}(\mathbb{F}_p^\times, \mathbb{C}^\times)$ to a map $\chi : \mathbb{F}_p \to \mathbb{C}$ with the following convention: $\mathbf{1}(0) := 1$ and $\chi(0) := 0$ if $\chi \neq \mathbf{1}$. The characters of order dividing k can be used to detect the k-th power residue classes modulo p, by the following elementary lemma.

Lemma 4.1.1. For every $k \in \mathbb{N}_+$ and every $t \in \mathbb{F}_p$ we have

$$\sum_{\chi^k = \mathbf{1}} \chi(t) = \#\{x \in \mathbb{F}_p : \ x^k = t\},$$

where the sum ranges over the multiplicative characters of order dividing k.

Let now $F(\mathbf{x}) := x_1^{k_1} + \cdots + x_\ell^{k_\ell}$ be a diagonal polynomial with trivial coefficients and define

$$r_F(m, p) := \#\{\mathbf{x} \in \mathbb{F}_p^\ell : \ F(\mathbf{x}) = m\} \tag{4.1.1}$$

for every $m \in \mathbb{F}_p$. The theory that we expose in this section is valid also if we consider twisted diagonal polynomials of the form

$$F_{\mathbf{a}}(\mathbf{x}) = a_1 x_1^{k_1} + \cdots + a_\ell x_\ell^{k_\ell}$$

with $a_i \in \mathbb{F}_p^\times$. Here we avoid covering the general case, for which we refer to [85], in order to keep the notation a little simpler. Using Lemma 4.1.1 we may compute $r_F(m, p)$ as follows:

$$\begin{aligned} r_F(m,p) &= \sum_{\substack{t_1,\dots,t_\ell \in \mathbb{F}_p \\ t_1+\cdots+t_\ell = m}} \prod_{i=1}^{\ell} \#\{x \in \mathbb{F}_p : \ x^{k_i} = t_i\} \\ &= \sum_{\substack{t_1,\dots,t_\ell \in \mathbb{F}_p \\ t_1+\cdots+t_\ell = m}} \prod_{i=1}^{\ell} \sum_{\chi^{k_i} = \mathbf{1}} \chi(t_i) \\ &= \sum_{\substack{\chi_1,\dots,\chi_\ell \in \mathrm{Hom}(\mathbb{F}_p^\times, \mathbb{C}^\times) \\ \chi_1^{k_1} = \cdots = \chi_\ell^{k_\ell} = \mathbf{1}}} \sum_{\substack{t_1,\dots,t_\ell \in \mathbb{F}_p \\ t_1+\cdots+t_\ell = m}} \chi_1(t_1) \dots \chi_\ell(t_\ell). \end{aligned}$$

Motivated by the above computation, given $m \in \mathbb{F}_p$ and ℓ multiplicative characters $\chi_1, \dots, \chi_\ell$ of $\mathbb{F}_p$, we introduce the *generalized Jacobi sum*

$$J_m(\chi_1, \dots, \chi_\ell) := \sum_{\substack{t_1,\dots,t_\ell \in \mathbb{F}_p \\ t_1+\cdots+t_\ell = m}} \chi_1(t_1) \dots \chi_\ell(t_\ell).$$

In the literature some authors [85, 88] use $J(\chi_1, \dots, \chi_\ell)$ to denote the Jacobi sum $J_1(\chi_1, \dots, \chi_\ell)$, while other authors [162, 11] prefer to use J to denote the Jacobi sum J_{-1}. The following lemma explicits the relation between distinct Jacobi sums.

Lemma 4.1.2. Let $\chi_1, \dots, \chi_\ell$ be arbitrary multiplicative characters of $\mathbb{F}_p$ and take $\chi_{\ell+1}$ so that the product $\chi_1 \dots \chi_{\ell+1} = \mathbf{1}$ is the trivial character. Then for every $m \in \mathbb{F}_p^\times$ we have the following formulas

$$J_m(\chi_1, \dots, \chi_\ell) = \chi_{\ell+1}^{-1}(m) J_1(\chi_1, \dots, \chi_\ell), \tag{4.1.2}$$

$$J_m(\chi_1, \dots, \chi_\ell) = \frac{\chi_{\ell+1}^{-1}(-m)}{p-1}\Big(J_0(\chi_1, \dots, \chi_{\ell+1}) - \chi_{\ell+1}(0) J_0(\chi_1, \dots, \chi_\ell)\Big). \tag{4.1.3}$$

Proof: Using the substitution $t_i' = m t_i$ and the multiplicativity of the characters we see that

$$\sum_{t_1'+\dots+t_\ell'=m} \chi_1(t_1') \dots \chi_\ell(t_\ell') = \prod_{i=1}^{\ell} \chi_i(m) \sum_{t_1+\dots+t_\ell=1} \chi_1(t_1) \dots \chi_\ell(t_\ell),$$

which is equivalent to (4.1.2). To prove the second formula, we divide the summation in the following expression according to the possible values $u = t_1 + \dots + t_\ell$, and we get

$$\sum_{t_1+\dots+t_{\ell+1}=0} \chi_1(t_1) \dots \chi_{\ell+1}(t_{\ell+1}) = \sum_{u \in \mathbb{F}_p} J_u(\chi_1, \dots, \chi_\ell) \chi_{\ell+1}(-u).$$

Using (4.1.2) on the elements $u \neq 0$ we get

$$J_0(\chi_1, \dots, \chi_{\ell+1}) = (p-1)\chi_{\ell+1}(-1) J_1(\chi_1, \dots, \chi_\ell) + \chi_{\ell+1}(0) J_0(\chi_1, \dots, \chi_\ell),$$

which is equivalent to (4.1.3). ∎

Moreover we have that $J_0(\chi_1, \dots, \chi_\ell)$ is nonzero only for specific choices of characters.

Lemma 4.1.3. Let $\chi_1, \dots, \chi_{\ell-1}$ be multiplicative characters of $\mathbb{F}_p$. If either (1) some but not all χ_i are trivial, or (2) the product $\chi_1 \dots \chi_\ell \neq \mathbf{1}$ is nontrivial, then $J_0(\chi_1, \dots, \chi_\ell) = 0$.

Proof: If at least one of the characters is trivial, we may suppose it is χ_ℓ because the Jacobi sum is invariant under permutation of the characters. In this case we have that

$$J_0(\chi_1, \dots, \chi_\ell) = \sum_{t_1, \dots, t_{\ell-1} \in \mathbb{F}_p} \prod_{i=1}^{\ell-1} \chi_i(t_i) = \prod_{i=1}^{\ell-1} \sum_{t \in \mathbb{F}_p} \chi_i(t),$$

which is equal to zero if any of the characters $\chi_1, \dots, \chi_{\ell-1}$ is nontrivial. For the second assertion, suppose there exists $g \in \mathbb{F}_p^\times$ such that $\prod_{i=1}^{\ell} \chi_i(g) \neq 1$. Then the change of variables $t_i' = g t_i$ in the definition of the 0-th Jacobi sum gives

$$J_0(\chi_1, \dots, \chi_\ell) = \prod_{i=1}^{\ell} \chi_i(g) J_0(\chi_1, \dots, \chi_\ell)$$

and so $J_0(\chi_1, \dots, \chi_\ell) = 0$. ∎

By (4.1.3) and Lemma 4.1.3 (1) we also have that $J_m(\chi_1, \dots, \chi_\ell) = 0$, for any $m \in \mathbb{F}_p$, if some but not all the characters χ_i are trivial. On the other hand, if all the characters are trivial we get directly from the definition that

$$J_m(\underbrace{\mathbf{1}, \dots, \mathbf{1}}_{\ell \text{ times}}) = p^{\ell-1}$$

for each $m \in \mathbb{F}_p$. We can summarize the above discussion in the following proposition.

Proposition 4.1.4. Let $F(\mathbf{x}) = x_1^{k_1} + \cdots + x_\ell^{k_\ell}$ and let $r_F(m,p)$ for $m \in \mathbb{F}_p$ be as in (4.1.1). Then

$$r_F(m,p) = p^{\ell-1} + \sum_{\substack{\chi_1,\dots,\chi_\ell \neq \mathbf{1} \\ \chi_i^{k_i} = \mathbf{1}}} J_m(\chi_1, \dots, \chi_\ell). \tag{4.1.4}$$

Moreover if $m = 0$ we can restrict the sum to the ℓ-tuples of nontrivial characters, respectively with order dividing k_i, whose product is trivial.

In order to estimate the "error term" $r_F(m,p) - p^{\ell-1}$ the following result is very useful.

Proposition 4.1.5. Let $\chi_1, \dots, \chi_\ell$ be multiplicative characters with $\chi_1 \dots \chi_\ell = \mathbf{1}$. Then

$$|J_0(\chi_1, \dots, \chi_\ell)| = (1 - 1/p)p^{\ell/2}.$$

This is known as the *Weil bound* for Jacobi sums, and it is a special case of the Riemann Hypothesis for function fields of positive characteristic [92, 160, 161]. We will prove this proposition in the next section, whereas now we focus on its consequences. By (4.1.3), Lemma 4.1.3 and Proposition 4.1.5, we also get that

$$|J_m(\chi_1, \dots, \chi_\ell)| = p^{\ell/2-1/2} \qquad \text{if } \chi_1 \dots \chi_\ell \neq \mathbf{1}, \tag{4.1.5}$$

$$|J_m(\chi_1, \dots, \chi_\ell)| = p^{\ell/2-1} \qquad \text{if } \chi_1 \dots \chi_\ell = \mathbf{1}, \tag{4.1.6}$$

for each $m \in \mathbb{F}_p^\times$ and any nontrivial characters $\chi_1, \dots, \chi_\ell$. Using these inequalities into Proposition 4.1.4, we get the following estimate for $r_F(m,p)$, also known as a Weil bound.

Corollary 4.1.6. Let $F(\mathbf{x})$ be as in Proposition 4.1.4. Then

$$\left| r_F(m,p) - p^{\ell-1} \right| \leq C_0 p^{\ell/2} \qquad \text{if } m = 0, \tag{4.1.7}$$

$$\left| r_F(m,p) - p^{\ell-1} \right| \leq C_1 p^{(\ell-1)/2} \qquad \text{if } m \neq 0, \tag{4.1.8}$$

where C_0 and C_1 are the respective numbers of ℓ-tuples of characters that are involved in the sums of (4.1.4).

At this point, we pause to record some observations on the relevance of Corollary 4.1.6 with respect to our basic plan.

Remark 4.1.7. A crucial part of the strategy of the previous chapter was to show that for all primes p in an infinite set $\mathcal{P}$ and for some residue classes $m \bmod p$ the quantity $r_F(m,p)$ is smaller than the "average" value of $r_F(\cdot,p)$ by a factor of at least $1-\epsilon_p$, with

$$\sum_{p\in\mathcal{P}} \epsilon_p = \infty.$$

Corollary 4.1.6 shows that it is not possible to take ϵ_p larger than $O(p^{1/2-\ell/2})$ if $m \neq 0$, or larger than $O(p^{1-\ell/2})$ if $m \neq 0$. Since it is known that the infinite sum

$$\sum_{p \text{ prime}} p^{-3/2}$$

is convergent, we see that the residue classes $m \neq 0$ are of no use to us if $\ell \geq 4$. Similarly, we see that the zero class $m = 0$ becomes useless when $\ell \geq 5$. However if $\ell = 4$ the Weil bound for the zero class reads as follows:

$$r_F(0,p) = (1 + O(p^{-1}))p^3.$$

If only we could prove that

$$r_F(0,p) < (1 + \beta p^{-1}))p^3$$

for a positive proportion of the primes and for a fixed *negative* parameter β, we could then try to leverage on this inequality and hopefully prove the existence of arbitrarily large gaps between the integer values of $F(\mathbf{x})$. In Section 4.3 we will show how to do this in the case where $F(\mathbf{x}) = x_1^4 + x_2^4 + x_3^4 + x_4^4$.

4.2 Gauss sums and square-root cancellation

On $\mathbb{F}_p$ we also have an additive group operation and so a notion of additive characters $\psi :\in \mathrm{Hom}(\mathbb{F}_p, \mathbb{C}^\times)$. Since the additive group of $\mathbb{F}_p$ is cyclic, we have exactly p additive characters with values in the p-th roots of unity, given by

$$\psi_a(t) := e(at/p),$$

where $a \in \mathbb{Z}/p\mathbb{Z}$ and $e(z) := e^{2\pi iz}$ is the normalized exponential function. The space of all functions $f : \mathbb{F}_p \to \mathbb{C}$ is a p-dimensional complex vector space that can be given a Hermitian structure via the discrete L^2-pairing

$$\langle f, g\rangle := \sum_{t\in\mathbb{F}_p} f(t)\overline{g(t)}.$$

With respect to this Hermitian product, the additive characters form an orthogonal basis:

$$\frac{1}{p}\langle \psi_a, \psi_b \rangle = \begin{cases} 1 & \text{if } a = b, \\ 0 & \text{if } a \neq b. \end{cases} \tag{4.2.1}$$

If χ is a nontrivial multiplicative character of $\mathbb{F}_p$, its Gauss sum is defined by

$$G(\chi) := \sum_{t \in \mathbb{F}_p^\times} \chi(t) e(t/p).$$

In other words $G(\chi)$ is the L^2-pairing between the multiplicative character χ and the additive character $t \mapsto e(-t/p)$. More generally, given an additive character ψ and a multiplicative character χ one may consider the modified Gauss sums

$$g(\chi, \psi) := \langle \chi, \psi \rangle,$$

so that $G(\chi) = g(\chi, \psi_{-1})$. The various Gauss sums are related as follows.

Lemma 4.2.1. If $\chi \neq \mathbf{1}$ then $g(\chi, \psi_0) = 0$ and

$$g(\chi, \psi_a) = \chi^{-1}(-a) G(\chi) \tag{4.2.2}$$

for every $a = 1, \ldots, p-1$.

Proof: If $a, b \in \mathbb{Z}/p\mathbb{Z}$ and $a \neq 0$, we compute

$$\begin{aligned} \chi(a) g(\chi, \psi_b) &= \sum_{t \in \mathbb{F}_p} \chi(at) e(-bt/p) \\ &= \sum_{s \in \mathbb{F}_p} \chi(s) e(-a^{-1} bs/p) \\ &= g(\chi, \psi_{a^{-1}b}) \end{aligned}$$

with the substitution $s = at$. Since $G(\chi) = g(\chi, \psi_{-1})$, the formula (4.2.2) follows. Moreover, if $\chi \neq \mathbf{1}$, then there exists $a \neq 0$ such that $\chi(a) \neq 1$. Then the formula $\chi(a) g(\chi, \psi_0) = g(\chi, \psi_0)$ implies that $g(\chi, \psi_0) = 0$. ∎

In particular we observe that $G(\overline{\chi}) = \overline{g(\chi, \psi_1)}$, which together with (4.2.2) gives

$$G(\overline{\chi}) = \chi(-1) \overline{G(\chi)} \tag{4.2.3}$$

because $\chi(-1) = \pm 1$ for each nontrivial multiplicative character χ. The relation between Gauss sums and Jacobi sums is given by the following formula.

Proposition 4.2.2. Let $\chi_1, \dots, \chi_\ell$ be nontrivial multiplicative characters of $\mathbb{F}_p$ such that $\chi_1 \dots \chi_\ell = \mathbf{1}$. Then

$$J_0(\chi_1, \dots, \chi_\ell) = \frac{p-1}{p} G(\chi_1) \cdots G(\chi_\ell).$$

Proof: We have

$$\begin{aligned} G(\chi_1) \dots G(\chi_\ell) &= \sum_{t_1,\dots,t_\ell \in \mathbb{F}_p} \chi_1(t_1) \dots \chi_\ell(t_\ell) e((t_1 + \cdots + t_\ell)/p) \\ &= \sum_{m \in \mathbb{F}_p} J_m(\chi_1, \dots, \chi_\ell) e(m/p) \\ &= J_0(\chi_1, \dots, \chi_\ell) \left(1 - \frac{1}{p-1} \sum_{m \in \mathbb{F}_p^\times} e(m/p)\right), \end{aligned}$$

by (4.1.3) and Lemma 4.1.3 (1). The proposition follows because the term in parenthesis is $1 + 1/(p-1)$. ∎

An important property of the Gauss sums is that they exhibit "square-root cancellation". That is, despite the fact that $G(\chi)$ is defined as a sum of $p-1$ terms with absolute value 1, the norm of the Gauss sum is only the square root of p. Here is a conceptual way to explain this phenomenon: the normalized Gauss sums $p^{-1/2} g(\chi, \psi_a)$ are the projections of χ onto the L^2-orthonormal basis given by the normalized additive characters $p^{-1/2}\psi_a$; because of the action by $\mathbb{F}_p^\times$ and the multiplicative nature of χ, the multiplicative character χ does not correlate preferentially with any of the nontrivial additive characters; therefore the Gauss sums $g(\chi, \psi_a)$ are "as small as they can get".

Proposition 4.2.3. For every multiplicative character $\chi \neq \mathbf{1}$ we have

$$|G(\chi)| = \sqrt{p}.$$

Proof: By orthogonality (4.2.1), we have the Pontryagin-Fourier inverse formula

$$\chi(t) = \frac{1}{p} \sum_{a=0}^{p-1} g(\chi, \psi_a) \psi_a$$

for the character χ. Taking L^2 norms on both sides and noticing that $\|\psi_a\|_2 = \sqrt{p}$ for all a, we get the Pontryagin-Parseval-Plancherel formula

$$\sum_{t \in \mathbb{F}_p} |\chi(t)|^2 = \frac{1}{p} \sum_{a=0}^{p-1} |g(\chi, \psi_a)|^2 .$$

The left-hand side is equal to $p-1$ because $\chi(0)=0$ by definition and $|\chi(t)|=1$ for all $t\in\mathbb{F}_p^\times$. On the other hand $g(\chi,\psi_0)=0$ and $|g(\chi,\psi_a)|=|G(\chi)|$ for all $a\neq 0$. The claimed equality $|G(\chi)|=\sqrt{p}$ follows. ∎

Together with Proposition 4.2.2, the square-root cancellation for Gauss sums implies the Weil bound for Jacobi sums of Proposition 4.1.5.

4.3 Gaps between sums of four fourth powers

We now specialize the theory of Section 4.1 to the diagonal form

$$F_4(\mathbf{x}) := x_1^4+x_2^4+x_3^4+x_4^4$$

with the objective of proving the existence of arbitrarily large gaps between the integers that can be written as sums of four fourth powers. The idea is to use Jacobi sums to compute $r_4(0,p)$ when p is a prime number congruent to 5 modulo 8, where for every $m \bmod M$ we denote

$$r_4(m,M) := \#\{\mathbf{x}\in\mathbb{Z}/M\mathbb{Z}:\ F_4(\mathbf{x})\equiv m \pmod M\}.$$

If p is a prime number congruent to 1 modulo 4 there exists a nontrivial multiplicative character

$$\chi_p : \mathbb{F}_p^\times \to \mu_4 := \{1,-1,i,-i\}$$

of order exactly four. Thus there are exactly three nontrivial multiplicative characters of $\mathbb{F}_p$ whose order divides 4, namely χ_p, ${\chi_p}^2$ and ${\chi_p}^3$. Since the real part of a Jacobi sum $J_0(\chi_1,\dots,\chi_4)$ does not change if we permute the χ_i or if we replace all of them with their conjugates, we see from Proposition 4.1.4 that

$$\begin{aligned} r_4(0,p) = p^3 &+ 2\,\mathrm{Re}\, J_0(\chi_p,\chi_p,\chi_p,\chi_p) + \mathrm{Re}\, J_0({\chi_p}^2,{\chi_p}^2,{\chi_p}^2,{\chi_p}^2) \\ &+ 6\,\mathrm{Re}\, J_0(\chi_p,\chi_p,{\chi_p}^3,{\chi_p}^3) + 12\,\mathrm{Re}\, J_0(\chi_p,{\chi_p}^2,{\chi_p}^2,{\chi_p}^3). \end{aligned} \tag{4.3.1}$$

An interesting observation can be made about the algebraic number

$$\pi_p := J_1(\chi_p,\chi_p) = \frac{1}{p}G(\chi_p)G(\chi_p)G({\chi_p}^2), \tag{4.3.2}$$

which is a Gaussian integer $\pi_p\in\mathbb{Z}[i]$ since by definition it is a sum of elements of μ_4. In fact, by (4.1.5) we see that π_p has norm equal to $\sqrt{p}$ and this implies that it can be written as

$$\pi_p = A+iB \tag{4.3.3}$$

for some integers $A,B\in\mathbb{Z}$ such that $A^2+B^2=p$. We have the following explicit result.

Proposition 4.3.1. Let $p \equiv 5 \bmod 8$ be a prime number and let $H := (A^2 - B^2)/p$ where A, B are as in Equation (4.3.3). Then

$$r_4(0,p) = p^3 + p(p-1)\,(2H-5)\,. \tag{4.3.4}$$

Proof: Since $p \equiv 5 \bmod 8$, we have no eight roots of unity modulo p. Hence there are no solutions to the equation $x^4 = -1$ in $\mathbb{F}_p$ and so $\chi_p(-1) = -1$. Thus by (4.2.3) and Proposition 4.2.3 we have

$$G(\chi_p)G(\chi_p{}^3) = -p \qquad \text{and} \qquad G(\chi_p{}^2)^2 = p. \tag{4.3.5}$$

Using (4.3.5) and Proposition 4.2.2 we get

$$\begin{aligned} J_0(\chi_p{}^2, \chi_p{}^2, \chi_p{}^2, \chi_p{}^2) &= p(p-1), \\ J_0(\chi_p\,, \chi_p\,, \chi_p{}^3, \chi_p{}^3) &= p(p-1), \\ J_0(\chi_p\,, \chi_p{}^2, \chi_p{}^2, \chi_p{}^3) &= -p(p-1). \end{aligned}$$

Moreover by (4.3.2) and (4.3.5) we get

$$J_0(\chi_p, \chi_p, \chi_p, \chi_p) = \frac{p-1}{p} G(\chi_p)^4 G(\chi_p{}^2)^2 \cdot G(\chi_p{}^2)^{-2} = (p-1)\pi_p^2,$$

and so using (4.3.3) we obtain

$$\mathbb{R}\mathrm{e}\, J_0(\chi_p, \chi_p, \chi_p, \chi_p) = (p-1)\,\mathbb{R}\mathrm{e}(A^2 - B^2 + 2Bi) = p(p-1)H.$$

The proposition follows from (4.3.1). ∎

The term $H = (A^2 - B^2)/p$ that appears in the above proposition is a rational number $H \in (-1, 1)$ of absolute value at most one, because we know that $A^2 + B^2 = p$. Therefore (4.3.4) yields

$$r_4(0,p) \le p^3 - 3p(p-1) < p^3 - 2p^2,$$

because $p \ge 5$. Since moreover there are exactly $3^4 = 81$ quadruples of nontrivial characters in $\mathbb{F}_p$ with order dividing 4, we get from Proposition 4.3.1 and (4.1.8) the following estimate.

Corollary 4.3.2. Let p be a prime number congruent to 5 modulo 8. Then

$$r_4(0,p) < (1 - 2p^{-1})p^3, \tag{4.3.6}$$

$$r_4(m,p) \le (1 + 81p^{-3/2})p^3 \qquad \text{if } m \ne 0. \tag{4.3.7}$$

This last statement is good enough to imply the following key result. Notice that the following argument is different than the one we used in the previous chapter to prove the analogous result for sums of three cubes.

Theorem 4.3.3. For every integer $K \geq 2$ there exist natural numbers m, M with $m + K < M$ such that

$$r_4(m+i, M) < \frac{1}{K} M^3 \tag{4.3.8}$$

for all $i = 1, \ldots, K$.

Proof: The sum $\sum_{p=5+8k} p^{-3/2}$ performed over all prime numbers congruent to 5 modulo 8 converges and in fact

$$\sum_{p=5+8k} p^{-3/2} < \int_4^\infty t^{-3/2} dt = 1.$$

On the contrary, the sum $\sum_{p=5+8k} p^{-1}$ diverges. More precisely, the prime number theorem in arithmetic progressions shows that

$$\sum_{\substack{p=5+8k \\ p \leq T}} p^{-1} = \frac{1}{4} \log\log T + O(1) \tag{4.3.9}$$

as $T \to \infty$.In particular it is possible to choose K non-empty, finite and disjoint sets $\mathcal{P}_1, \ldots, \mathcal{P}_K$ of prime numbers congruent to 5 modulo 8 such that

$$\sum_{p \in \mathcal{P}_i} \frac{2}{p} > \log K + 81. \tag{4.3.10}$$

Then, let M be the (squarefree) product of all $p \in \mathcal{P}_1 \cup \cdots \cup \mathcal{P}_K$, and let m denote the unique integer with $0 \leq m < M$ such that $m + i \equiv 0 \pmod{p}$ for all $1 \leq i \leq K$ and all $p \in \mathcal{P}_i$. Notice that in fact $m < M - K$. By the Chinese Remainder Theorem and the estimates of Corollary 4.3.2, we get

$$\begin{aligned} \log\left(\frac{r_4(m+i, M)}{M^3}\right) &\leq \log\left(\prod_{p \in \mathcal{P} \setminus \mathcal{P}_i} \left(1 + \frac{81}{p^{3/2}}\right) \prod_{p \in \mathcal{P}_i} \left(1 - \frac{2}{p}\right)\right) \\ &\leq 81 \sum_{p \in \mathcal{P} \setminus \mathcal{P}_i} \frac{1}{p^{3/2}} - 2 \sum_{p \in \mathcal{P}_i} \frac{1}{p} \end{aligned}$$

for all $1 \leq i \leq K$. This gives (4.3.8) because of (4.3.10). ∎

The existence of arbitrarily long gaps between sums of four fourth powers now follows from Theorem 4.3.3 and a Maier matrix argument, like in the previous chapter.

Corollary 4.3.4. Let $\mathcal{S}_{4,4}$ be the set of natural numbers that can be written as a sum of four fourth powers. Then for every positive integer K there are K consecutive numbers in $\mathbb{N} \setminus \mathcal{S}_{4,4}$.

Proof: Suppose $K \geq 2$, let m, M be as in Theorem 4.3.3 and consider the $K \times M^3$ "matrix of numbers"

$$\mathcal{R} = \{m + i + (h-1)M : \ 1 \leq i \leq K, \ 1 \leq h \leq M^3\}.$$

If we suppose that among every K consecutive natural numbers there is an element of $\mathcal{S}_{4,4}$, we get that

$$\#\mathcal{S}_{4,4} \cap \mathcal{R} \geq M^3.$$

Since $x_1^4 + x_2^4 + x_3^4 + x_4^4 < M^4$ implies $x_1, x_2, x_3, x_4 \in \{0, \ldots, M-1\}$, we also have:

$$\begin{aligned} \#\mathcal{S}_{4,4} \cap \mathcal{R} &\leq \sum_{i=1}^{K} \sum_{h=1}^{M^3} \#\{\mathbf{x} \in \mathbb{N}^4 : \ F(\mathbf{x}) = m + i + (h-1)M\} \\ &\leq \sum_{k=1}^{K} \#\{\mathbf{x} \in (\mathbb{Z}/M\mathbb{Z})^4 : \ F(\mathbf{x}) \equiv m + i \pmod{M}\} \\ &= \sum_{i=1}^{K} r_4(m+i, M). \end{aligned}$$

This is strictly less than M^3 by Theorem 4.3.3, so we have a contradiction. ∎

Remark 4.3.5. We repeat that the purpose of this chapter is to give an elementary proof for the existence of arbitrarily long gaps between sums of four fourth powers. A more general theorem, proved with more technical and sophisticated tools from both algebraic and analytic number theory, will be presented in Chapter 6. We have described the proof of this chapter in full detail, possibly with the only exception of the estimate (4.3.9), which we deduced as a consequence of the prime number theorem in arithmetic progressions.

The prime number theorem states that the number of primes $p \leq T$ is asymptotic to $T/\log T$ as $T \to \infty$. This was conjectured by Gauss [60, 59, 62] and Legendre [100, 101] at the end of the 18th century but it was only in 1896 that Hadamard and de la Valleé-Poussin [68, 30] independently succeeded in proving it using ideas of Riemann. The proof amounts to showing that the analytic continuation of the Riemann zeta function $\zeta(s) = \sum_{n=1}^{\infty} n^{-s}$ has a pole at $s = 1$ and does not have any zero on the line $s = 1 + it$, where $t \in \mathbb{R}$. Later authors have greatly simplified the original proofs, but the result remains fundamentally nontrivial. See [117, 168] for simple analytic proofs, [136, 43] for difficult elementary proofs, or [29, Chapter 5] for more on this topic.

The prime number theorem in arithmetic progressions further states that for every $m \in \mathbb{N}_+$ the prime numbers equidistribute among the $\phi(m)$ congruence classes $\ell \bmod m$ with $(\ell, m) = 1$. The fact that there are infinitely many primes $p \equiv \ell \bmod m$ was demonstrated in 1837 by Dirichlet [41]. This is a difficult result, proved via analytic methods, which nevertheless predates the fundamental memoir of Riemann [131], published in 1859. For some congruence classes, such as $p \equiv 5 \bmod 8$, the infinitude of primes can be proved with a more elementary Euclid-like argument [7]. In fact such proof can be given whenever the congruence $p \equiv \ell \bmod m$ satisfies $\ell^2 \equiv 1 \bmod m$ [116].

The divergence of $\sum p^{-1}$ over all primes was already known to Euler [45] in the 18th century. A more precise statement is that for each $m \in \mathbb{N}_+$ and each ℓ coprime with m the estimate

$$\sum_{\substack{p \le T \\ p \equiv \ell \bmod m}} p^{-1} = \left(\frac{1}{\phi(m)} + o(1) \right) \log \log T \tag{4.3.11}$$

holds as $T \to \infty$. This result was first proved by Mertens [109, 110] in 1874 as a refinement of the work of Dirichlet. See for instance [165] or [28, Chapter 7] for modern presentations.

Mertens' theorem (4.3.11) follows from the prime number theorem in arithmetic progressions, but it is not equivalent to it. At the level of the analytic proofs, one sees that the prime number theorem fundamentally requires an analysis of the Riemann zeta function $\zeta(s)$ on the whole vertical line $s = 1 + it$, while Mertens' theorem simply follows from the behaviour of $\zeta(s)$ in a neighborhood of $s = 1$, see [151].

The most difficult part of most proofs of Mertens' theorem in arithmetic progressions is related to the nonvanishing of Dirichlet L-series $L(s, \chi)$ at $s = 1$. Nevertheless it is also possible to prove (4.3.11) in just a few pages by purely elementary methods, using ideas of Selberg [135, 140, 65]. This means that our proof for the existence of arbitrarily long gaps between sums of four fourth powers could indeed be considered as fully elementary.

Remark 4.3.6. The estimate $\sum_{p \equiv 1 \bmod 3} p^{-1} = \infty$ which we stated in (3.3.3) in the previous chapter can be treated in the same way as (4.3.9). However our proof of the existence of arbitrarily long gaps between sums of three cubes makes use of Weil bounds for multiplicative character sums. This is another difficult result, which is a consequence of the Riemann Hypothesis for curves over finite fields, proved by Weil around 1941 [160]. A relatively elementary approach to prove this statement is the so-called Stepanov method, conceived by Stepanov in 1969 [148] and later generalized and simplified by Bombieri [13].

Chapter 5

Pseudo-automorphisms of binary quadratic forms and cubic identities

In Chapter 3 we discussed the irregularity of the distribution of the set $\mathcal{S}_{3,3}$ by showing that in certain locations of the number line we have an "abundance" of natural numbers that admit no representations as a sum of three cubes. Conversely, one may ask if there are single numbers $n \in \mathbb{N}$ that admit several representations $x^3 + y^3 + z^3 = n$ with x, y, z nonnegative. Such numbers were studied by Mahler [105] using a polynomial parametrization of the solutions to the homogeneous cubic equation

$$X^3 + Y^3 + Z^3 = U^3. \tag{5.0.1}$$

In Section 5.1 we introduce the formula used by Mahler and we propose a simple and conceptual way to derive it, by looking at how the quadratic form $g(x, y) = x^2 + 3xy + 3y^2$ transforms via the linear change of variable $(x, y) \mapsto (3y, x)$. In fact, we establish a close connection between more general cubic polynomial identities and the pseudo-automorphisms $g(x, y)$, i.e. those matrices $\left(\begin{smallmatrix} \alpha & \beta \\ \gamma & \delta \end{smallmatrix}\right) \in GL_2(\mathbb{Q})$ whose natural action on binary forms leaves $g(x, y)$ invariant up to multiplication by a scalar:

$$g(\alpha x + \beta y, \gamma x + \delta y) = \lambda g(x, y).$$

In Sections 5.2 and 5.3 we investigate and describe the set of pseudo-automorphisms of a generic binary quadratic form $f(x, y)$ without repeated factors. Then in Section 5.5 we specialize the results to the binary form $g(x, y)$ and we derive the cubic identities. This is joint work with Andrew Granville.

5.1 The Mahler-Gérardin identity

Given an integer $k \geq 2$ we say that the diagonal form $F_k(\mathbf{x}) := x_1^k + \cdots + x_k^k$ with homogeneous degree k in k variables satisfies the *Hypothesis K* of Hardy and Littlewood [71] if $r_k(n) \ll n^\epsilon$ for every $\epsilon > 0$, where

$$r_k(n) := \#\{\mathbf{x} \in \mathbb{N}^k : \ F_k(\mathbf{x}) = n\}.$$

It is known, by (2.1.4) that the Hypothesis K holds for F_2, but it is not known if it is true for F_k with $k \geq 4$ [81]. For the case of three cubes, i.e. $k = 3$, the Hypothesis K was disproved by Mahler in [105], who constructed natural numbers for which $r_3(n) > \left\lfloor 9^{1/3} n^{1/12} \right\rfloor$. Mahler used the following polynomial identity in two variables [105]:

$$\boxed{(9x^4)^3 + (3xy^3 - 9x^4)^3 + (y^4 - 9x^3y)^3 = y^{12}} \tag{5.1.1}$$

and he considered the natural numbers that are perfect 12th powers. Indeed, (5.1.1) gives a representation of $n = y^{12}$ as a sum of three nonnegative cubes for every $0 \leq x \leq 9^{-1/3}y$. Notice that for each y this identity also shows that the number y^{12} can be written in infinitely many ways as a sum of three cubes of signed integers. Mahler derived (5.1.1) by setting (using his notation)

$$f = \frac{3}{2}x, \ g = \frac{1}{2}x, \ f' = y, \ g' = 0$$

into the following more general identity of Euler and Binet [49, 50, 12], see [38, pp.554-555]:

$$(\rho^2 - \sigma\rho')^3 + (\sigma'\rho' - \rho^2)^3 + (\rho'^2 - \rho\sigma')^3 = (\rho'^2 - \rho\sigma)^3, \tag{5.1.2}$$

where

$$\begin{aligned} \rho &= f^2 + 3g^2, \quad \sigma = ff' + 3gg' + 3fg' - 3f'g, \\ \rho' &= f'^2 + 3g'^2, \quad \sigma' = ff' + 3gg' - 3fg' + 3f'g, \end{aligned} \tag{5.1.3}$$

so that $\rho = 3x^2$, $\sigma = 0$, $\rho' = y^2$ and $\sigma' = 3xy$.

We observe that (5.1.1) was in fact already found by Gérardin in 1911, see [38, p.559]. Of course one could dispense with the appeal to the Euler-Binet identity by checking the Mahler-Gérardin identity directly, but that would not be illuminating. Instead, we propose the following simple alternative derivation. Consider

$$g(A, B) := \frac{(A + B)^3 - B^3}{A} = A^2 + 3AB + 3B^2 \tag{5.1.4}$$

and notice that $g(A, B)$ is an eigenfunction for the transformation $(A, B) \mapsto (3B, A)$:

$$g(3B, A) = 3g(A, B). \tag{5.1.5}$$

In other words:

$$\boxed{A\left[(A+3B)^3 - A^3\right] = 9B\left[(A+B)^3 - B^3\right]}. \tag{5.1.6}$$

Now, there is an obvious change of variables that makes (5.1.6) into an identity between sums of cubes, namely $A = y^3$ and $B = -3x^3$. This gives exactly (5.1.1).

The polynomial identity of Mahler and Gérardin has been used to equate a cube with a sum of three cubes, but it can equally be used to find numbers that can be written as sums of two cubes in more than one way. We show this for the well-known taxicab identity $1728 + 1 = 1000 + 729$ [144]. Indeed if we plug in $x = 1$ and $y = -1$ into

$$(9x^4 - 3xy^3)^3 + y^{12} = (y^4 - 9xy^3)^3 + (9x^4)^3$$

we get precisely the equality

$$12^3 + 1^3 = 10^3 + 9^3$$

found by Frenicle [53] [38, p.552] and made famous by an anecdote about Ramanujan [126, p.387] [69, p.12]. However, the identity (5.1.1) does not account for all integer solutions of (5.0.1). For example, it does not include the well-known representation

$$3^3 + 4^3 + 5^3 = 6^3 \tag{5.1.7}$$

of Plato's number 216 [163, p.144] [103, p.66] [124] as the sum of the three cubes of the Pythagorean triple $(3, 4, 5)$.

5.2 Pseudo-automorphisms of binary quadratic forms

To formalize the observation in (5.1.5) we introduce the following definition. Given a binary quadratic form $f(x, y) = ax^2 + bxy + cy^2$ with rational coefficients, we say that $M = \left(\begin{smallmatrix} \alpha & \beta \\ \gamma & \delta \end{smallmatrix}\right) \in GL_2(\mathbb{Q})$ is a pseudo-automorphism of $f(x, y)$ if there exists $\lambda \in \mathbb{Q}^\times$ such that

$$f(\alpha x + \beta y, \gamma x + \delta y) = \lambda f(x, y). \tag{5.2.1}$$

We denote the set of pseudo-automorphisms of $f(x, y)$ by $\mathrm{pAut}(f)$. Thus, (5.1.5) means that $\left(\begin{smallmatrix} 0 & 3 \\ 1 & 0 \end{smallmatrix}\right) \in \mathrm{pAut}(g)$, where $g(x, y) = x^2 + 3xy + 3y^2$. We remark that the left-hand side of (5.2.1) defines the usual contravariant action of $GL_2(\mathbb{Q})$ on the set of binary forms, so we have that $\mathrm{pAut}(f)$ is a subgroup of $GL_2(\mathbb{Q})$. If we write $f(x, y) = \mathbf{x}^t A \mathbf{x}$ where

$$\mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix} \qquad \text{and} \qquad A = \begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix},$$

then (5.2.1) is equivalent to the quadratic matrix equation

$$M^t A M = \lambda A. \tag{5.2.2}$$

We now give a complete and explicit description of $\mathrm{pAut}(f)$ for a generic $f(x,y) \in \mathbb{Q}[x,y]_2$.

Proposition 5.2.1. Let $f(x,y) = ax^2 + bxy + cy^2$ with no repeated linear factor and $a \neq 0$. Then $M = \left(\begin{smallmatrix} \alpha & \beta \\ \gamma & \delta \end{smallmatrix}\right)$ is a pseudo-automorphism of $f(x,y)$ if and only if either

$$aM = \alpha \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix} + \gamma \begin{pmatrix} 0 & -c \\ a & b \end{pmatrix} \tag{5.2.3}$$

or

$$aM = \alpha \begin{pmatrix} a & b \\ 0 & -a \end{pmatrix} + \gamma \begin{pmatrix} 0 & c \\ a & 0 \end{pmatrix}. \tag{5.2.4}$$

Proof: First, suppose that M is a pseudo-automorphism of $f(x,y)$ and set

$$Mt := \frac{\alpha t + \beta}{\gamma t + \delta}$$

for each $t \in \mathbb{C}$. The polynomial $F(t) := f(t,1)$ factors as $F(t) = a(t-\rho_1)(t-\rho_2)$ for some (distinct) $\rho_1, \rho_2 \in \mathbb{C}$ and for $i = 1, 2$ we have

$$f(\alpha\rho_i + \beta, \gamma\rho_i + \delta) = \lambda F(\rho_i) = 0. \tag{5.2.5}$$

We notice that $\gamma\rho_i + \delta$ and $\alpha\rho_i + \beta$ cannot both vanish because $M \in GL_2(\mathbb{Q})$. In fact we see that $\gamma\rho_i + \delta \neq 0$ by (5.2.5) and the hypothesis $a \neq 0$. Thus (5.2.5) yields $F(M\rho_i) = 0$. Since $\rho_1 \neq \rho_2$ and $M \in GL_2(\mathbb{Q})$ we deduce that $\{M\rho_1, M\rho_2\} = \{\rho_1, \rho_2\}$.

If $M\rho_1 = \rho_1$ then both ρ_1 and ρ_2 satisfy the polynomial equation derived from $Mt = t$, namely

$$\gamma t^2 + (\delta - \alpha)t - \beta = 0,$$

which must then be a multiple of $at^2 + bt + c = 0$. Therefore (5.2.3) holds because, if $\kappa = \gamma/a$ denotes the proportionality ratio, we get

$$\beta = -\kappa c \qquad \text{and} \qquad \delta = \alpha + b\kappa.$$

If $M\rho_1 = \rho_2$ and $M\rho_2 = \rho_1$ then $\gamma\rho_1\rho_2 + \delta\rho_2 = \alpha\rho_1 + \beta$ and $\gamma\rho_1\rho_2 + \delta\rho_1 = \alpha\rho_2 + \beta$. Subtracting the two equations we obtain that $(\alpha + \delta)(\rho_1 - \rho_2) = 0$ and so $\alpha + \delta = 0$. Therefore (5.2.4) holds because

$$\gamma c + \alpha b = \gamma a \rho_1\rho_2 - \alpha a(\rho_1 + \rho_2) = a\beta.$$

The above arguments show that every $M \in \mathrm{pAut}(f)$ satisfies (5.2.3) or (5.2.4). Conversely, it is straightforward to verify that in both cases (5.2.1) holds with

$$\lambda = a^{-1} f(\alpha, \gamma). \tag{5.2.6}$$

See also Corollary 5.4.1 for a neat proof of this last claim. ∎

Remark 5.2.2. If $a = 0$ but $c \neq 0$ then we interchange x and y and the above result applies. If $a = c = 0$ then $f(x, y) = bxy$ and so there are again two families:

$$M = \alpha \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \delta \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad M = \beta \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} + \gamma \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}.$$

5.3 The structure of the set of pseudo-automorphisms

It is clear from the definitions that the set $\mathrm{pAut}(f)$ of pseudo-automorphisms of $f(x, y)$ is a subgroup of $GL_2(\mathbb{Q})$, so it is closed under composition, i.e. matrix multiplication. However, it is amusing to observe from Proposition 5.2.1 that its Zariski closure $\overline{\mathrm{pAut}(f)}$ in $\mathrm{Mat}_{2\times 2}(\mathbb{Q})$ is the union of two vector subspaces of $\mathrm{Mat}_{2\times 2}(\mathbb{Q})$. This fact is a little surprising at first, but it becomes clear under the following considerations. First, we recall the matricial formulation given in (5.2.2): an invertible matrix M belongs to $\mathrm{pAut}(f)$ if and only if

$$M^t A M = \lambda A \quad \text{where } A = \begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix}. \tag{5.3.1}$$

The hypothesis that the quadratic form $f(x, y)$ does not have a repeated linear factor translates into the nonvanishing of $\det A = \Delta/4$, where Δ is the discriminant of f. Taking determinants on both sides of (5.3.1) we obtain $(\det M)^2 = \lambda^2$, or

$$\lambda = \pm \det M.$$

If we multiply by the inverse M^{-1} on the right and we denote by $\mathrm{Adj}\, M := (\det M)M^{-1}$ the adjoint of M, we see that (5.3.1) amounts to two separate linear problems

$$M^t A = \pm A \,\mathrm{Adj}\, M. \tag{5.3.2}$$

We denote by $\mathrm{pAut}^+(f)$ and $\mathrm{pAut}^-(f)$ the solutions to (5.3.2) in $\mathrm{Mat}_{2\times 2}(\mathbb{Q})$ corresponding to the choice of sign $+$ and $-$ respectively. Thus, we have

$$\overline{\mathrm{pAut}(f)} = \mathrm{pAut}^+(f) \cup \mathrm{pAut}^-(f).$$

We are now going to explore the algebraic structure of $\overline{\mathrm{pAut}(f)}$. For simplicity of exposition, we restrict ourselves to the case $a = 1$, but similar considerations are valid also when $f(x, y)$ is not monic. By Proposition 5.2.1 the solutions to (5.3.2) are parametrized respectively by

$$M^+_{x,y} := \begin{pmatrix} x & -cy \\ y & x + by \end{pmatrix} \quad \text{and} \quad M^-_{x,y} := \begin{pmatrix} x & bx + cy \\ y & -x \end{pmatrix}. \tag{5.3.3}$$

It is clear that $M^+_{x,y}$ and $M^-_{x,y}$ are linear matrix-valued functions in the two variables x, y. Therefore the additive structure on $\mathrm{pAut}^+(f)$ and $\mathrm{pAut}^-(f)$ is simply given by

$$M^\pm_{x,y} + M^\pm_{x',y'} = M^\pm_{x+x',y+y'}. \tag{5.3.4}$$

The multiplicative structure on $\overline{\mathrm{pAut}(f)}$ is instead described as follows.

Theorem 5.3.1. For every $x, y, \alpha, \gamma \in \mathbb{Q}$ we have

$$M^+_{\alpha,\gamma} M^\pm_{x,y} = M^\pm_{x',y'},$$
$$M^-_{\alpha,\gamma} M^\pm_{x,y} = M^\mp_{x'',y''},$$

where $\left(\begin{smallmatrix} x' \\ y' \end{smallmatrix}\right) = M^+_{\alpha,\gamma} \left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} x'' \\ y'' \end{smallmatrix}\right) = M^-_{\alpha,\gamma} \left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)$.

Proof: First we observe that the multiplication in $\mathrm{pAut}^+(f)$ is commutative, because $\mathrm{pAut}^+(f)$ is linearly generated by $\{M^+_{1,0}, M^+_{0,1}\}$, and $M^+_{1,0}$ is the identity matrix. Moreover we note the following interesting property: for every $x, y, v, w \in \mathbb{Q}$ we have

$$M^+_{x,y} \begin{pmatrix} v \\ w \end{pmatrix} = M^+_{v,w} \begin{pmatrix} x \\ y \end{pmatrix}.$$

This fact is easily proved by noticing that the first column of $M^+_{x,y}$ is equal to $M^+_{1,0} \left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)$ and that its second column is equal to $M^+_{0,1} \left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)$. For a more compact notation, we let $\boldsymbol{\alpha}, \mathbf{x}, \mathbf{v}$ be respectively the vectors $(\alpha, \gamma)^t, (x, y)^t$ and $(v, w)^t$. Then the multiplication formula in $\mathrm{pAut}^+(f)$ follows from this computation:

$$M^+_{\boldsymbol{\alpha}} M^+_{\mathbf{x}} \mathbf{v} = M^+_{\boldsymbol{\alpha}} M^+_{\mathbf{v}} \mathbf{x} = M^+_{\mathbf{v}} M^+_{\boldsymbol{\alpha}} \mathbf{x} = M^+_{M^+_{\boldsymbol{\alpha}} \mathbf{x}} \mathbf{v}.$$

Since $\mathbf{v}$ is arbitrary, we deduce that $M^+_{\alpha,\gamma} M^+_{x,y} = M^+_{x',y'}$. To extend the result to $\overline{\mathrm{pAut}(f)}$ we consider

$$J := M^-_{1,0} = \begin{pmatrix} 1 & b \\ 0 & -1 \end{pmatrix} \quad \text{and} \quad R := M^+_{0,1} = \begin{pmatrix} 0 & -c \\ 1 & b \end{pmatrix} \tag{5.3.5}$$

and we notice the following equalities:

$$R \cdot J = \begin{pmatrix} 0 & c \\ 1 & 0 \end{pmatrix} = M^-_{0,1}, \qquad J \cdot R = \begin{pmatrix} b & -c + b^2 \\ -1 & -b \end{pmatrix} = M^-_{b,-1}, \tag{5.3.6}$$

and

$$J^2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

By linearity one gets from (5.3.6) the following formulas that express the elements of $\mathrm{pAut}^-(f)$ as the result of left or right multiplication by J on $\mathrm{pAut}^+(f)$:

$$M^-_{\mathbf{x}} = M^+_{\mathbf{x}} \cdot J,$$

$$M^-_{J\mathbf{x}} = J \cdot M^+_{\mathbf{x}}.$$

The conclusion follows formally:

$$\begin{array}{llll} M^+_{\boldsymbol{\alpha}} M^-_{\mathbf{x}} & = M^+_{\boldsymbol{\alpha}} M^+_{\mathbf{x}} J & = M^+_{\mathbf{x}'} J & = M^-_{\mathbf{x}'}, \\ M^-_{\boldsymbol{\alpha}} M^+_{\mathbf{x}} & = M^+_{\boldsymbol{\alpha}} J M^+_{\mathbf{x}} & = M^+_{\boldsymbol{\alpha}} M^-_{J\mathbf{x}} & = M^-_{\mathbf{x}''}, \\ M^-_{\boldsymbol{\alpha}} M^-_{\mathbf{x}} & = M^+_{\boldsymbol{\alpha}} J^2 M^+_{J^{-1}\mathbf{x}} & = M^+_{\boldsymbol{\alpha}} M^+_{J^{-1}\mathbf{x}} & = M^+_{\mathbf{x}''}, \end{array}$$

where $\mathbf{x}' := M^+_{\boldsymbol{\alpha}}\mathbf{x}$ and $\mathbf{x}'' := M^-_{\boldsymbol{\alpha}}\mathbf{x}$. ∎

The previous result can be interpreted abstractly as an equivariance property of the parametrization (5.3.3), as follows. The group $GL_2(\mathbb{Q})$ acts naturally on $\mathrm{Mat}_{2\times 2}(\mathbb{Q})$ by left multiplication, and this action restricts to

$$\mathrm{pAut}(f) \curvearrowright \overline{\mathrm{pAut}(f)}.$$

On the other hand $GL_2(\mathbb{Q})$ also acts by left multiplication on the set $\mathbb{Q}^2$ of column vectors, and we may restrict this action to $\mathrm{pAut}(f) \curvearrowright \mathbb{Q}^2$. Since the elements of $\mathrm{pAut}(f)$ are either in $\mathrm{pAut}^+(f)$ or $\mathrm{pAut}^-(f)$ but not in both, we actually have a well-defined action

$$\mathrm{pAut}(f) \curvearrowright \mathbb{Q}^2 \times \{+,-\}$$

given by

$$M^{\pm}_{\alpha,\gamma} \cdot \left(\begin{pmatrix} x \\ y \end{pmatrix}, \epsilon \right) = \left(M^{\pm}_{\alpha,\gamma} \cdot \begin{pmatrix} x \\ y \end{pmatrix}, \pm\epsilon \right).$$

Furthermore, we have that (5.3.3) determines a surjective parametrization

$$M : \ \mathbb{Q}^2 \times \{\pm 1\} \to \overline{\mathrm{pAut}(f)} \tag{5.3.7}$$

Then Theorem 5.3.1 is equivalent to the following statement.

Corollary 5.3.2. The map (5.3.7) defined by (5.3.3) is $\mathrm{pAut}(f)$-equivariant.

5.4 Determinants of pseudo-automorphisms

Another interesting property of the matrices $M^{\pm}_{x,y}$ is that their determinant is a scalar multiple of the binary quadratic form $f(x,y) = x^2 + bxy + cy^2$:

$$\det M^{\pm}_{x,y} = \pm f(x,y). \tag{5.4.1}$$

Therefore, if we take determinants in Theorem 5.3.1, we get

Corollary 5.4.1. For every $x, y, \alpha, \gamma \in \mathbb{Q}$ we have

$$\boxed{f(\alpha,\gamma)f(x,y) = f(x',y')}, \tag{5.4.2}$$

where $\left(\begin{smallmatrix} x' \\ y' \end{smallmatrix}\right) = M^{\pm}_{\alpha,\gamma} \left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)$.

We notice that the formula in Corollary 5.4.1 resembles the Diophantus-Brahmagupta-Fibonacci identity (2.1.1) for the product of sums of two squares. Motivated by the usual complex-theoretic proof of this identity, we now present one more interpretation for the matrices $M^{\pm}_{x,y}$. Let $\mathbb{K}$ denote the quotient ring $\mathbb{Q}[t]/(F(t))$, with $F(t) := t^2 + bt + c$ being a de-homogeneization of $f(x, y)$. Let

$$\rho := t \bmod F(t) \qquad \text{and} \qquad \overline{\rho} := -b - t \bmod F(t),$$

and notice that $\mathbb{K} = \mathbb{Q} \oplus \rho\mathbb{Q}$ is a two-dimensional $\mathbb{Q}$-algebra together with a $\mathbb{Q}$-linear involution given by $1 \mapsto 1$ and $\rho \mapsto \overline{\rho}$. We consider on $\mathbb{K}$ the basis $\{1, -\rho\}$, so that every element of $\mathbb{K}$ is written uniquely as $x - \rho y$ for some rational numbers x, y.

Proposition 5.4.2. Let $\rho, \overline{\rho} \in \mathbb{K}$ as above. Then for every $x, y, v, w \in \mathbb{Q}$ we have

$$(x - \rho y)(v - \rho w) = (v' - \rho w'),$$
$$(x - \rho y)(v - \overline{\rho} w) = (v'' - \rho w''),$$

where $\left(\begin{smallmatrix} v' \\ w' \end{smallmatrix}\right) = M^{+}_{x,y} \left(\begin{smallmatrix} v \\ w \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} v'' \\ w'' \end{smallmatrix}\right) = M^{-}_{x,y} \left(\begin{smallmatrix} v \\ w \end{smallmatrix}\right)$.

In other words, $M^{+}_{x,y}$ is the matrix of the multiplication by $x - \rho y$, while $M^{-}_{x,y}$ is the matrix of this multiplication precomposed with the conjugation involution. In particular the matrices R and J defined in (5.3.5) correspond to the multiplication by $-\rho$ and to the conjugation, respectively. With this interpretation, the equivariance discussed in Corollary 5.3.2 (or equivalently, displayed in Theorem 5.3.1) simply encodes of the basic properties of associativity of multiplication in $\mathbb{K}$ and of distribution of conjugation with respect to multiplication. For example, using the notation in the proof of Theorem 5.3.1, we see that the identity

$$M^{-}_{\boldsymbol{\alpha}} M^{-}_{\mathbf{x}} \mathbf{v} = M^{+}_{M^{-}_{\boldsymbol{\alpha}} \mathbf{x}} \mathbf{v}$$

is just a translation of the formula $\boldsymbol{\alpha}\overline{\overline{\mathbf{x}\mathbf{v}}} = (\boldsymbol{\alpha}\overline{\mathbf{x}})\mathbf{v}$. On $\mathbb{K}$ we have a multiplicative "signed norm" defined formally by

$$\|x - \rho y\|_f := (x - \rho y)(x - \overline{\rho} y) = f(x, y),$$

where the product lies in the copy of $\mathbb{Q}$ inside $\mathbb{K}$ spanned by $1 \in \mathbb{K}$. Hence Corollary 5.4.1 is a consequence of the multiplicativity of $\|\cdot\|_f$.

For completeness, we may write the identity (5.4.2) in the two expanded forms

$$f(\alpha,\gamma)f(x,y) = f(\alpha x - c\gamma y, \gamma x + \alpha y + b\gamma y)$$
$$f(\alpha,\gamma)f(x,y) = f(\alpha x + c\gamma y + b\alpha y, \gamma x - \alpha y)$$

corresponding to $M^{+}_{\alpha,\gamma}$ and $M^{-}_{\alpha,\gamma}$ respectively. However, the compact form (5.4.2) is more neat and intelligible.

5.5 Cubic identities

In this last section we apply the previous results to the quadratic form

$$g(x,y) := x^2 + 3xy + 3y^2 = \frac{(x+y)^3 - y^3}{x}. \tag{5.5.1}$$

By Proposition 5.2.1 every pseudo-automorphism of g can be written as

$$M^+_{\alpha,\gamma} = \begin{pmatrix} \alpha & -3\gamma \\ \gamma & \alpha + 3\gamma \end{pmatrix} \quad \text{or} \quad M^-_{\alpha,\gamma} = \begin{pmatrix} \alpha & 3\alpha + 3\gamma \\ \gamma & -\alpha \end{pmatrix}.$$

For example, the transformation $(x,y) \mapsto (3y,x)$ that we used in Section 5.1 corresponds to $M^-_{0,1}$. The pseudo-automorphism formula in Corollary 5.4.1 reads as follows:

$$g(x',y') = \lambda g(x,y) \tag{5.5.2}$$

where

$$\begin{cases} \lambda &= g(\alpha,\gamma) \\ x' &= \alpha x - 3\gamma y \\ y' &= \gamma x + (\alpha + 3\gamma) y \end{cases} \quad \text{or} \quad \begin{cases} \lambda &= g(\alpha,\gamma) \\ x' &= \alpha x + (3\alpha + 3\gamma) y \\ y' &= \gamma x - \alpha y \end{cases}, \tag{5.5.3}$$

depending on whether we use $M^\pm_{\alpha,\gamma}$ or $M^-_{\alpha,\gamma}$. If we substitute these formulas in the expression for $g(x,y)$ given to the right of (5.5.1), we obtain identities relating linear combinations of cubes. In view of the denominators implicit in (5.5.2), coming from (5.5.1), it is better to write the resulting formulas in terms of x and x'.

Proposition 5.5.1. For every $x, x', \alpha, \delta, \lambda$ such that $\alpha^2 + \alpha\delta + \delta^2 = 3\lambda$ we have

$$\boxed{x(\lambda x - \alpha x')^3 - x(\lambda x - \delta x')^3 = \lambda x'(x' - \alpha x)^3 - \lambda x'(x' - \delta x)^3}. \tag{5.5.4}$$

Proof: We use the formula (5.5.2) corresponding to the transformation displayed to the left of (5.5.3). We define $\delta := \alpha + 3\gamma$, so that

$$3\lambda = 3g(\alpha,\gamma) = \alpha^2 + \alpha\delta + \delta^2$$

and then we compute:

$$\begin{aligned} y &= (\alpha x - x')/(3\gamma), \\ x + y &= (\delta x - x')/(3\gamma), \\ y' &= (\lambda x - \delta x')/(3\gamma), \\ x' + y' &= (\lambda x - \alpha x')/(3\gamma). \end{aligned}$$

Thus we get (5.5.4) from (5.5.2) and (5.5.1), after simplifying the common denominator $(3\gamma)^3 xx'$. ∎

We observe that using the relations to the right of (5.5.3) we get

$$\begin{aligned} y &= (\alpha x - x')/(-\beta), \\ x + y &= (\xi x - x')/(-\beta), \\ y' &= (\lambda x - \alpha x')/\beta, \\ x' + y' &= (\lambda x - \xi x')/\beta, \end{aligned}$$

where $\xi := -2\alpha - 3\gamma$ and $\beta := 3\alpha + 3\gamma$. Moreover we have

$$3\lambda = \alpha^2 + \alpha\xi + \xi^2$$

and so we get an identity that is equivalent to (5.5.4). Using the substitutions $x' = n\lambda^2 x$ and $(\alpha, \delta) = (p + 3q, p - 3q)$ we recover the following identity of Piezas [121], which generalizes to all n a formula of Binet [12].

Corollary 5.5.2. If $\lambda = p^2 + 3q^2$, we have

$$\frac{(1 - \lambda n(p + 3q))^3 - (1 - \lambda n(p - 3q))^3}{(\lambda^2 n - (p + 3q))^3 - (\lambda^2 n - (p - 3q))^3} = n.$$

It is known that the Binet-Piezas formula accounts for all rational solutions of $A^3 - B^3 = n(C^3 - D^3)$ up to multiplication by a scalar [121, Ch. 6.I.2, Link 10]. If instead we operate the substitutions $(x, x') = (m, -n^2)$ and $\lambda = mn$ we obtain the following complete parametrization of the rational solutions of $A^3 - B^3 = C^3 - D^3$ due to Werebrusov [164] and Schwering [134], and found also in the third notebook of Ramanujan [126, p.387].

Corollary 5.5.3. If $\alpha^2 + \alpha\delta + \delta^2 = 3mn$, then

$$(m^2 + \alpha n)^3 - (m^2 + \delta n)^3 = (n^2 + \alpha m)^3 - (n^2 + \delta m)^3.$$

The formula of Euler and Binet (5.1.2) mentioned in the beginning of the chapter follows from that of Werebrusov and Schwering, because if $\rho, \rho', \sigma, \sigma'$ are defined as in (5.1.3), we have

$$\sigma^2 + \sigma\sigma' + \sigma'^2 = 3\rho\rho'.$$

In fact, also the formula of Euler and Binet gives a complete parametrization of the primitive rational solutions of the cubic equation $X^3 + Y^3 + Z^3 = U^3$ [20]. For some discussions of the integer solutions to this equation, see [21, 56]. For extensive lists of other algebraic identities, we refer to [38] and [121].

Part II

Sums of powers: analytic methods

Chapter 6

Arbitrarily long gaps between the values of positive-definite cubic and biquadratic diagonal forms

ARBITRARILY LONG GAPS BETWEEN THE VALUES OF POSITIVE-DEFINITE CUBIC AND BIQUADRATIC DIAGONAL FORMS

LUCA GHIDELLI

Abstract. For $s = 3, 4$, we prove the existence of arbitrarily long sequences of consecutive integers none of which is a sum of s nonnegative s-th powers. More generally, we study the existence of gaps between the values $\leq N$ of diagonal forms of degree s in s variables with positive integer coefficients. We find: (1) gaps of size $\gg \frac{\sqrt{\log N}}{(\log\log N)^2}$ when $s = 3$; (2) gaps of size $\gg \frac{\log\log\log N}{\log\log\log\log N}$ if $s = 4$ and the form, up to permutation of the variables, is not equal to $a(c_1x_1)^4 + b(c_2x_2)^4 + 4a(c_3x_3)^4 + 4b(c_4x_4)^4$.

Contents

1. Introduction

Let $s \in \mathbb{N}_+$ and let $F(\mathbf{x}) = a_1x_1^s + \cdots + a_sx_s^s$ be a diagonal form of degree s in s variables with positive integer coefficients $a_1, \ldots, a_s \in \mathbb{N}_+$. In this article by *values of* $F(\mathbf{x})$ we mean the natural numbers obtained by evaluating the diagonal form at nonnegative integers $x_1, \ldots, x_s \in \mathbb{N}$. A *gap* of length K between these values is a sequence of consecutive nonnegative integers $n + 1, \ldots, n + K$ that are not values of $F(\mathbf{x})$. When $s = 2$ the polynomial $F(\mathbf{x})$ is a multiple of a norm form and so the values of $F(\mathbf{x})$ form a set with natural density 0 in $\mathbb{N}$ (see Landau [21] for the prototypical case $F(\mathbf{x}) = x_1^2 + x_2^2$ and Odoni [28] for general norm forms). In particular if $s = 2$ there are arbitrarily long gaps between the values of $F(\mathbf{x})$. When $s \geq 3$ the polynomial $F(\mathbf{x})$ is irreducible over $\mathbb{C}$ and so it is not a norm form. In fact very little is known unconditionally about the distribution of the values of $F(\mathbf{x})$ if $s \geq 3$ (see [15] for some results conditional on GRH) but it is reasonable to expect, on the basis of probabilistic models [6] [7], that the set of values of $F(\mathbf{x})$ has positive density. Nevertheless, we may ask if there are arbitrarily long gaps between the values of $F(\mathbf{x})$, when $s \geq 3$. In this

Date: December 10, 2019.
2010 *Mathematics Subject Classification.* Primary 11B05, Secondary 11R37, 11R45, 11T06, 11T24.

article we give a positive answer in two cases. First, for all trinomial positive-definite cubic diagonal forms:

Theorem 1.1. Let $F(\mathbf{x})$ be as above, with $s = 3$. Then there is a constant $\kappa_F > 0$ such that for all integers N, K satisfying $N > e^e$, $K \geq 2$ and $K < \kappa_F \frac{\sqrt{\log N}}{(\log\log N)^2}$ there exist gaps of length K between the values of $F(\mathbf{x})$ less than N.

Second, for almost all quadrinomial positive-definite biquadratic diagonal forms:

Theorem 1.2. Let $F(\mathbf{x})$ be as above, with $s = 4$, and suppose that $F(\mathbf{x})$ is not equal to $a(c_1x_1)^4 + b(c_2x_2)^4 + 4a(c_3x_3)^4 + 4b(c_4x_4)^4$, for some $a, b, c_1, c_2, c_3, c_4 \in \mathbb{N}_+$, up to a permutation of the variables. Then there is a constant $\kappa_F > 0$ such that for all integers N, K satisfying $N > e^{e^{e^e}}$, $K \geq 2$ and $K < \kappa_F \frac{\log\log\log N}{\log\log\log\log N}$ there are gaps of length at least K between the values of $F(\mathbf{x})$ less than N.

Notice that in both theorems the upper bound on K goes to infinity with N, but the growth is much faster when $s = 3$. We refer to Remark 2.3 for some explanation. In Theorem 8.8 we show more precisely that, for a suitable $\kappa_F > 0$ and the same hypotheses, there exist at least $c(F, K)N$ gaps of length K between the values of $F(\mathbf{x})$ less than N, where $c(F, K) > 0$ is independent of N.
The above theorems include the important special cases $F(\mathbf{x}) = x_1^3 + x_2^3 + x_3^3$ and $F(\mathbf{x}) = x_1^4 + x_2^4 + x_3^4 + x_4^4$. The values of these forms are often studied in connection with Waring's problem [32], which more generally concerns the representability of natural numbers as sums of perfect powers. Moreover, the results of the present paper concerning these two special cases have been used in a crucial way to improve some results of Bradshaw [3] in regard to values of cubic and biquadratic theta series [9].
On the other hand Theorem 1.2 doesn't apply to some biquadratic forms such as $F(\mathbf{x}) = x_1^4 + x_2^4 + 4x_3^4 + 4x_4^4$. We show that these exceptions are characterized among all biquadratic diagonal forms by a local property (see Theorem 6.2). This is further discussed in Remark 2.4.
We now compare the above results with the literature. When $s = 2$ Richards [30] proved, with an ingenious elementary proof, that there are gaps of length at least $\gamma_F \log N$ between the values of $F(\mathbf{x})$, for some constant $\gamma_F > 0$. It is an important open-problem to estimate sharply the order of growth of the gaps between the values of $F(\mathbf{x}) = x_1^2 + x_2^2$. However when $s \geq 3$ our knowledge is even weaker. For example, if $F(\mathbf{x}) = x_1^3 + x_2^3 + x_3^3$ we only know by an elementary greedy argument [5] that for N large enough there are no gaps of size greater than $3^{19/9}N^{8/27}(1 + o(1))$, among the values of $F(\mathbf{x})$ less than N. On the other hand, working out the predictions of the probabilistic models, we should expect the existence of gaps of length as large as $O(\log N/\log\log N)$, for all $s \geq 3$.
In the following section we expose our strategy towards the proofs of Theorems 1.1 and 1.2. As it will be clear, the same method can be used to prove the existence of arbitrarily long gaps between the values of other polynomials, provided they satisfy a certain local property (see "Step 2" below). Following a suggestion of Wooley, we are going to treat in a future publication the case of non-homogeneous diagonal forms such as $x_1^2 + x_2^3 + x_3^7 + x_4^{42}$.

2. Detecting the existence of long gaps - the method

Let $s, F(\mathbf{x})$ be as in Theorems 1.1 and 1.2. Let $\mathcal{S}_F \subseteq \mathbb{N}$ be the set of values of $F(\mathbf{x})$ and for all $n \in \mathbb{N}$ let $r_F(n) := \#\{\mathbf{x} \in \mathbb{N}^s : \ F(\mathbf{x}) = n\}$ be the number of representations of n as a value of $F(\mathbf{x})$. Moreover, for all $M \in \mathbb{N}_+$ and $m \in \mathbb{Z}$ let $r_F(m, M)$ denote the number of solutions $\mathbf{x} \in (\mathbb{Z}/M\mathbb{Z})^s$ to the congruence $F(\mathbf{x}) \equiv m \pmod M$. Our strategy to find gaps between the values of $F(\mathbf{x})$ consists of three parts:

Step 1: Estimate $r_F(m,p)$ for prime numbers p, with special attention to the case $m=0$. In particular we find a set $\mathcal{P}_F$ of primes and positive real numbers $\{\epsilon_p\}_{p\in\mathcal{P}_F}$ with the following properties: $r_F(0,p)\leq p^{s-1}(1-\epsilon_p)$ for all $p\in\mathcal{P}_F$, and $\sum_{p\in\mathcal{P}_F}\epsilon_p=+\infty$.

Step 2: Show that for every $\epsilon>0$ and $K\in\mathbb{N}_+$ there are $m,M\in\mathbb{N}$ with $0\leq m<M-K$ such that $r_F(m+k,M)<\epsilon M^{s-1}$ for all $k=1,\dots,K$.

Step 3: Form the intersection of $\mathcal{S}_F$ with a set of the form

$$\mathcal{R}=\{m+k+(h-1)M:\ 1\leq k\leq K,\ 1\leq h\leq H\}.$$

If M and m are obtained from Step 2 with $\epsilon<\frac{1}{K}$, and H is suitably chosen, we find that the cardinality of the intersection is strictly less than H. This implies that $m+(h_0-1)M+[1,K]$ is a gap between the values of $F(\mathbf{x})$, for some $h_0\leq H$.

The underlying idea is the following: suppose that the number of solutions to the congruence $F(\mathbf{x})\equiv m\pmod M$ is significantly smaller than the "expected" number M^{s-1}; then a number of the form $m+(h-1)M$ has a low chance to be a value of $F(\mathbf{x})$, if h is randomly chosen. In other words, these numbers are likely to be in a gap of $F(\mathbf{x})$. To make this observation rigorous in Step 3, we require that the form $F(\mathbf{x})$ is positive-definite.

The first step constitutes the bulk of this article, and occupies all the sections from 3 to 7. Steps 2 and 3 are performed in section 8, together with the derivations of the quantitative estimates announced in section 1. We now give more details about the strategy outlined above, in the case of biquadratic diagonal forms. The case of cubic forms is analogous: it is only slightly more delicate in Step 2, and overall considerably easier in Step 1. See also Remark 8.9 for some variants of our proof.

2.1. **Step 1.** Let $s=4$, then fix $F(\mathbf{x})$ as in Theorem 1.2, and let Σ_F be the set of primes that divide some coefficient of $F(\mathbf{x})$. The outcome of Step 1 is the following.

Proposition 2.1. For all $m\in\mathbb{Z}$ and all prime $p\equiv 1\pmod 4$ with $p\notin\Sigma_F$, we have

$$r_F(m,p)\leq p^3(1+81p^{-3/2}).$$

Moreover for all $\beta\in(0,1)$ there is a set of primes $\mathcal{P}_F$ with positive relative density $\delta>0$ such that for all $p\in\mathcal{P}_F$ we have $p\equiv 1\pmod 4$ and:

$$r_F(0,p)\leq p^3\left(1-\beta p^{-1}\right). \tag{2.1}$$

The first upper estimate for $r_F(m,p)$ is a consequence of the Deligne-Weil bounds [31, Chapter 4.5] (see Proposition 3.2 below). The second result for $r_F(0,p)$ comes from an exact formula of the form

$$r_F(0,p)=p^3+p(p-1)(2\,\mathbb{R}\mathrm{e}\,H_{F,p}+K_{F,p}) \tag{2.2}$$

which is established in sections 3 and 4 using the theory of cyclotomy, more precisely with Gauss and Jacobi sums [2] [16, Sec. 8]. Here $H_{F,p}$ and $K_{F,p}$ denote explicit character sums modulo p, where $p\equiv 1\pmod 4$ is prime and $p\notin\Sigma_F$. Moreover, $H_{F,p}$ is a complex number of absolute value 1 well-defined up to conjugation, and $K_{F,p}$ is an integer satisfying $-7\leq K_{F,p}\leq 19$. The formula (2.2) is related to the Sato-Tate distribution [31, Chapter 8] of the affine scheme associated to $F(\mathbf{x})$: the continuous part of the Sato-Tate distribution corresponds to $H_{F,p}$, and the discrete part to $K_{F,p}$. By a theorem of Weil, we are able to interpret $H_{F,p}$ as a Hecke character of infinite order and absolute value 1. Using the theory of Hecke L-functions, we prove in section 7 that $H_{F,p}$ equidistributes on the unit circle (up to conjugation) as $p\to\infty$. In particular, for all $\beta\in(0,1)$ we have $2\,\mathbb{R}\mathrm{e}\,H_{F,p}<-1-\beta$ for a positive proportion of the primes.

On the other hand, in section 6.2 we relate $K_{F,p}$ to the Kummer extension L/K, where $K := \mathbb{Q}(i)$, and $L = K(\sqrt[4]{a_1}, \dots, \sqrt[4]{a_4}, \sqrt[4]{-1})$ is generated by the fourth roots of -1 and of the coefficients of $F(\mathbf{x})$. By Chebotarev's density theorem and Kummer's theory, we are able to compute the possible values of $K_{F,p}$ explicitly from the characters of $\mathrm{Gal}(L/K)$, which is a finite abelian group of order at most 512. In particular we can show that $K_{F,p} \leq 1$ for a positive proportion of the primes, if $F(\mathbf{x}) \neq a(c_1x_1)^4 + b(c_2x_2)^4 + 4a(c_3x_3)^4 + 4b(c_4x_4)^4$ up to a permutation of the variables. In fact, this hypothesis on $F(\mathbf{x})$ is necessary to have $K_{F,p} \leq 1$, as we show in section 6.1 by an elementary argument.

2.2. **Step 2.** In order to construct M and m, we start by choosing suitable disjoint finite subsets $\mathcal{P}_1, \dots, \mathcal{P}_K$ of $\mathcal{P}_F$ and we form their union $\mathcal{P} := \mathcal{P}_1 \cup \dots \cup \mathcal{P}_K$. Then, we let M be the (squarefree) product of all $p \in \mathcal{P}$, and we take m so that $m + k \equiv 0 \pmod{p}$ for all $k \leq K$ and all $p \in \mathcal{P}_k$. In this way, by the Chinese Remainder Theorem and the estimates of Step 1, we have

$$\begin{aligned}\log\left(\frac{r_F(m+k,M)}{M^3}\right) &\leq \log\left(\prod_{p\in\mathcal{P}\setminus\mathcal{P}_k}\left(1+\frac{81}{p^{3/2}}\right)\prod_{p\in\mathcal{P}_k}\left(1-\frac{\beta}{p}\right)\right)\\ &\leq 81\sum_{p\in\mathcal{P}\setminus\mathcal{P}_k}\frac{1}{p^{3/2}} - \beta\sum_{p\in\mathcal{P}_k}\frac{1}{p}\end{aligned}$$

for all $1 \leq k \leq K$. We notice that the series $\sum_p p^{-3/2}$ ranging over all primes is bounded above by an absolute constant C_1. On the other hand, since $\mathcal{P}_F$ has positive density, we have that $\sum_{p\in\mathcal{P}_F} p^{-1}$ diverges, and therefore it is possible to choose $\mathcal{P}_1, \dots, \mathcal{P}_K$ so that $r_F(m+k, M) \leq \epsilon M^3$, for all $k \leq K$ and for any given $\epsilon > 0$.

2.3. **Step 3.** The conclusion is now obtained by a simple double-counting technique that is sometimes known as the Maier matrix method [10]. Fix $K \in \mathbb{N}_+$ and $0 < \epsilon < \frac{1}{K}$, and construct M, m as in Step 2, with $0 \leq m < M - K$. Let $\mathcal{R} = \{m + k + (h-1)M : 1 \leq k \leq K,\ 1 \leq h \leq M^3\}$. Since $F(\mathbf{x}) < M^4$ implies $x_1, x_2, x_3, x_4 \in \{0, \dots, M-1\}$, we have:

$$\begin{aligned}\#\mathcal{S}_F \cap \mathcal{R} &\leq \sum_{k=1}^{K}\sum_{h=1}^{M^3} \#\{\mathbf{x} \in \mathbb{N}^4 :\ F(\mathbf{x}) = m + k + (h-1)M\}\\ &\leq \sum_{k=1}^{K} \#\{\mathbf{x} \in (\mathbb{Z}/M\mathbb{Z})^4 :\ F(\mathbf{x}) \equiv m + k \pmod{M}\},\end{aligned}$$

which is equal to $\sum_{k=1}^{K} r_F(m+k, M)$, and so it is at most $K\epsilon M^3 < M^3$ by Step 2. On the other hand, suppose by contradiction that for all $h \leq M^3$ the interval $m + [1, K] + (h-1)M$ contains a value of $F(\mathbf{x})$. Then $\#\mathcal{S}_F \cap \mathcal{R}$ contains at least M^3 elements, and this is a contradiction.

Remark 2.2. A modification of Steps 1 and 2 proves the existence of residue classes $m \pmod{M}$ that satisfy $r_F(m, M) > cM^{s-1}$ for arbitrarily large $c > 0$. This can be used to show (see [14, Chapter IV.1]) that for any given $A > 0$ there exists $n \in \mathbb{N}_+$ such that the equation $F(\mathbf{x}) = n$ has at least A solutions $\mathbf{x} \in \mathbb{N}^s$.

Remark 2.3. If $s = 3$ we have an analog of (2.1) of the form

$$r_F(0, p) \leq p^2\left(1 - \beta p^{-1/2}\right),$$

so Step 1 is fulfilled with $\epsilon_p \asymp p^{-1/2}$. Then the series $\sum_{p\in\mathcal{P}_F} \epsilon_p \asymp \sum_{p\in\mathcal{P}_F} p^{-1/2}$ diverges to infinity much faster than the series $\sum_{p\in\mathcal{P}_F} p^{-1}$ which appears in Step 2

above, in the case $s = 4$. This is the technical reason that explains why the estimate on K in our main result Theorem 1.1 for cubic forms is much better than the one for biquadratic forms in Theorem 1.2.

Remark 2.4. When $s \geq 5$, it is well known [6] that

$$r_F(m, q) = q^{s-1}\left(1 + O(q^{-3/2})\right)$$

for every power of a prime $q = p^\nu$ and every residue class $m \bmod q$. Reasoning as in Step 2, since the series $\sum_{q \in \mathbb{N}_+} q^{-3/2}$ converges, we see that there exist positive constants c_0, c_1 such that

$$c_0 M^{s-1} \leq r_F(m, M) \leq c_1 M^{s-1}$$

for all $M \in \mathbb{N}_+$ and all $m \bmod M$. This explains why our approach does not yield arbitrarily long gaps between the values of diagonal forms in 5 or more variables. Step 2 also fails when $s = 4$ and

$$F(\mathbf{x}) = a(c_1x_1)^4 + b(c_2x_2)^4 + 4a(c_3x_3)^4 + 4b(c_4x_4)^4$$

for some $a, b, c_1, c_2, c_3, c_4 \in \mathbb{N}_+$, because any such form satisfies $M^3 \leq r_F(0, M)$ for all odd squarefree moduli M (see section 6.1). Taking into account higher powers of primes and the residue classes other than zero, it is in fact possible to prove that $cM^3 \leq r_F(m, M)$ for all m, M with a constant $c = c(F) > 0$.

3. Multiplicative characters and diagonal congruences

3.1. **Characters and character sums.** If $\mathbb{F}$ is a field we denote by $\mathbb{F}^\times := \mathbb{F} \setminus \{0\}$ the multiplicative group of its nonzero elements. A multiplicative character of $\mathbb{F}$ is by definition a group homomorphism $\chi \in \mathrm{Hom}(\mathbb{F}^\times, \mathbb{C}^\times)$. We denote by $\mathbf{1}$ the trivial character, i.e. the one satisfying $\mathbf{1}(t) = 1$ for all $t \in \mathbb{F}^\times$. If χ is a nontrivial multiplicative character of $\mathbb{F}$, it is customary to declare $\chi(0) = 0$, thus extending χ to a map $\chi : \mathbb{F} \to \mathbb{C}$. Given nontrivial multiplicative characters $\chi_1, \ldots, \chi_\ell$ of a finite field $\mathbb{F}$ we consider the generalized Jacobi sum

$$J(\chi_1, \ldots, \chi_\ell) := \sum_{\substack{t_1, \ldots, t_\ell \in \mathbb{F} \\ t_1 + \cdots + t_\ell = 1}} \prod_{i=1}^{\ell} \chi_i(t_i). \tag{3.1}$$

and we let $J_0(\chi_1, \ldots, \chi_\ell)$ be defined analogously, but with the sum performed over the ℓ-tuples satisfying $t_1 + \cdots + t_\ell = 0$. If $\#\mathbb{F} = p$ is a prime number, then the finite field $\mathbb{F}$ is canonically isomorphic to $\mathbb{F}_p := \mathbb{Z}/p\mathbb{Z}$. For every $s \in \mathbb{N}_+$ we define

$$\mathfrak{X}_p^{(s)} := \{\chi \in \mathrm{Hom}(\mathbb{F}_p^\times, \mathbb{C}^\times) : \ \chi^s = \mathbf{1} \text{ and } \chi \neq \mathbf{1}\}$$

to be the set of the nontrivial multiplicative characters of $\mathbb{F}_p$ with order dividing s. We observe that $\mathbb{F}_p^\times$ is a cyclic group of order $p - 1$, so every multiplicative character χ of $\mathbb{F}_p$ is determined by its value at the multiplicative generators modulo p, and $\#\mathfrak{X}_p^{(s)} = \gcd(s, p-1) - 1$. Since the complex exponential function is periodic with period $2\pi i$, the map $x \mapsto e^{\frac{2\pi i x}{p}}$ gives a well-defined additive character of $\mathbb{F}_p$. If χ is a multiplicative character of $\mathbb{F}_p$, its associated Gauss sum is

$$G(\chi) := \sum_{t \in \mathbb{F}_p} \chi(t) e^{\frac{2\pi i t}{p}}.$$

3.2. **Cubic and biquadratic power residue characters.** For $s \in \mathbb{N}_+$, let $\zeta_s := e^{\frac{2\pi i}{s}}$ and let $\mu_s := \{\zeta_s^i : \ 0 \le i < s\} \subseteq \mathbb{C}$. Let K be a number field containing μ_s, let $\mathcal{O}_K$ be its ring of integers, and let $\mathfrak{p}$ be a prime ideal of $\mathcal{O}_K$ not dividing s. The discriminant of $X^s - 1$ is divisible only by the primes dividing s, and so the elements $\zeta_s^i \in \mu_s$ are pairwise incongruent modulo $\mathfrak{p}$. Thus $\mu_s \bmod \mathfrak{p}$ has cardinality s, and is the complete set of s-th roots of unity in the residue field $\mathcal{O}_K/\mathfrak{p}$. This implies that $s \,|N\mathfrak{p} - 1$, where $N\mathfrak{p} := \#(\mathcal{O}_K/\mathfrak{p})$ is the norm of $\mathfrak{p}$. From this we conclude that for every $a \in \mathcal{O}_K$ with $a \notin \mathfrak{p}$ there is a unique $\chi_{s,\mathfrak{p}}(a) \in \mu_s$, also denoted by $\left(\frac{a}{\mathfrak{p}}\right)_s$ (see Definition 5.1 below), such that

$$\chi_{s,\mathfrak{p}}(a) \equiv a^{\frac{N\mathfrak{p}-1}{s}} \pmod{\mathfrak{p}}.$$

The multiplicative character $\chi_{s,\mathfrak{p}}$ of $\mathcal{O}_K/\mathfrak{p}$ is the s-th power residue character modulo $\mathfrak{p}$. Fix now $s \in \{3, 4\}$. The ring $\mathbb{Z}[\zeta_s]$ is an Euclidean domain (it is the ring of Eisenstein integers for $s = 3$ and the ring of Gaussian integers for $s = 4$), and coincides with the ring of integers of the quadratic number field $\mathbb{Q}(\zeta_s)$. If p is a prime number satisfying $p \equiv 1 \pmod{s}$, then it splits in $\mathbb{Z}[\zeta_s]$. We choose an arbitrary prime $\mathfrak{p}$ above p, so that $p\mathbb{Z}[\zeta_s] = \mathfrak{p}\,\overline{\mathfrak{p}}$, and we define $\chi_{s,p} := \chi_{s,\mathfrak{p}}$. We notice that $N\mathfrak{p} = p$, and so we may, as we will, consider $\chi_{s,p}$ as a multiplicative character of $\mathbb{F}_p$. It is easy to see that the order of $\chi_{s,p}$ is exactly s. Finally, we let for brevity

$$\pi_{s,p} := J(\chi_{s,p}, \chi_{s,p}). \tag{3.2}$$

It is well-known [16, Sec.9.4, Lemma 1, Proposition 9.9.4] that $\mathfrak{p} = (\pi_{s,p})$ and that $p = \pi_{s,p}\overline{\pi}_{s,p}$.

3.3. **The number of solutions of diagonal congruences.** Let $s, k \in \mathbb{N}_+$ and fix a diagonal form $F(\mathbf{x}) = a_1x_1^k + \ldots + a_sx_s^k$ of degree k in s variables, with nonzero integer coefficients $a_1, \ldots, a_s \in \mathbb{Z} \setminus \{0\}$. Let Σ_F be the (finite) set of primes dividing $a_1 \cdots a_s$. For all $M \in \mathbb{N}_+$ and $m \in \mathbb{Z}$ we define $r_F(m, M) := \#\mathcal{R}_F(m, M)$, where

$$\mathcal{R}_F(m, M) := \{\mathbf{x} \in (\mathbb{Z}/M\mathbb{Z})^s : \ F(\mathbf{x}) \equiv m \pmod{M}\}.$$

In other words, we count the solutions of the congruence $F(\mathbf{x}) \equiv m \bmod M$. A classical application of the Chinese Remainder Theorem shows that the function $r_F(m, M)$ is multiplicative in its second variable. This allows us to reduce the computation of $r_F(m, M)$ to the case of prime moduli, if M is squarefree. When p is prime and m is not divisible by p we content ourselves with classical estimates for $r_F(m, p)$. On the other hand, for $r_F(0, p)$ we will use an explicit computation in terms of modified Jacobi sums, which in turn can be computed via Gauss sums.

Lemma 3.1. Let $m \in \mathbb{Z}$ and $M = \prod_{i=1}^{\ell} p_i$ for distinct primes $p_1, \ldots, p_\ell$. Then

$$r_F(m, M) = \prod_{i=1}^{\ell} r_F(m, p_i). \tag{3.3}$$

Proof. We have (3.3) because the Chinese Remainder Theorem provides a bijection

$$\psi : \ \mathcal{R}_F(m, M) \to \mathcal{R}_F(m, p_1) \times \cdots \times \mathcal{R}_F(m, p_\ell),$$

sending an s-tuple $(x_1, \ldots, x_s) \in (\mathbb{Z}/M\mathbb{Z})^s$ to the sequence of s-tuples $(x_1^{(i)}, \ldots, x_s^{(i)}) \in (\mathbb{Z}/p_i\mathbb{Z})^s$ with $1 \le i \le \ell$ obtained by reducing modulo p_i each component. □

Proposition 3.2. Let p be a prime number with $p \notin \Sigma_F$, and let $m \in \mathbb{Z}$ witht $p \nmid m$. Then

$$\left|r_F(m, p) - p^{s-1}\right| \le (k-1)^s p^{\frac{s-1}{2}},$$

Proof. This follows from the case $b \neq 0$ of [16, Sec. 8.7, Theorem 5], since we have $p^{\frac{s}{2}-1} \leq p^{\frac{s-1}{2}}$ and $\#\mathfrak{X}_p^{(k)} \leq k-1$. □

Proposition 3.3 ([16, Sec. 8.7, Theorem 5])**.** Let p be a prime number with $p \notin \Sigma_F$. Then

$$r_F(0,p) = p^{s-1} + \sum_{\chi_1,\dots,\chi_s} \overline{\chi}_1(a_1)\cdots\overline{\chi}_s(a_s) J_0(\chi_1,\dots,\chi_s)$$

where the sum ranges over the s-tuples of characters $\chi_i \in \mathfrak{X}_p^{(k)}$ that satisfy $\chi_1\cdots\chi_s = \mathbf{1}$, and where $\overline{\chi}_i$ denotes the complex conjugate of χ_i.

Proposition 3.4 ([16, Sec. 8.5, Prop 8.5.1 & Cor. 1])**.** Let p be a prime number and let $\chi_1,\dots,\chi_\ell$ be nontrivial multiplicative characters of $\mathbb{F}_p$ such that $\chi_1\cdots\chi_\ell = \mathbf{1}$. Then

$$J(\chi_1,\dots,\chi_{\ell-1}) = \frac{\chi_\ell(-1)}{p} G(\chi_1)\cdots G(\chi_\ell); \tag{3.4}$$

$$J_0(\chi_1,\dots,\chi_\ell) = \frac{p-1}{p} G(\chi_1)\cdots G(\chi_\ell). \tag{3.5}$$

4. The zero residue class in the cubic and biquadratic cases

4.1. **Evaluation of the Jacobi sums.** We now specialize to the case $s = k \in \{3,4\}$. We first compute the modified Jacobi sums appearing in Proposition 3.3, using the notation introduced in section 3.2. We will then get an explicit formula for $r_F(0,p)$. In the next sections we will use it to deduce good upper bounds on $r_F(0,p)$ for special choices of p. Recall from section 3.2 and (3.2) the definition of $\chi_{s,p}$ and $\pi_{s,p}$ for $s \in \{3,4\}$ and $p \equiv 1 \bmod s$.

Lemma 4.1. Let p be a prime number with $p \equiv 1 \pmod 3$. Then

$$J_0(\chi_{3,p},\chi_{3,p},\chi_{3,p}) = (p-1)\pi_{3,p}. \tag{4.1}$$

Analogously, let q be a prime number with $q \equiv 1 \pmod 4$. Then

$$J_0(\chi_{4,q},\chi_{4,q},\chi_{4,q},\chi_{4,q}) = (q-1)\pi_{4,q}^2; \tag{4.2}$$

$$J_0(\chi_{4,q},\chi_{4,q},\chi_{4,q}^3,\chi_{4,q}^3) = q(q-1); \tag{4.3}$$

$$J_0(\chi_{4,q}^2,\chi_{4,q}^2,\chi_{4,q},\chi_{4,q}^3) = q(q-1)\chi_{4,q}(-1); \tag{4.4}$$

$$J_0(\chi_{4,q}^2,\chi_{4,q}^2,\chi_{4,q}^2,\chi_{4,q}^2) = q(q-1). \tag{4.5}$$

Proof. Since $\chi_{3,p}(-1) = \chi_{3,p}((-1)^3) = 1$, equation (4.1) is a direct consequence of (3.4) and (3.5) applied to the triple of characters $(\chi_{3,p},\chi_{3,p},\chi_{3,p})$. It is immediate to see from the definitions that $J_0(\chi_{4,q},\chi_{4,q}^3)\chi_{4,q}(-1) = J_0(\chi_{4,q}^2,\chi_{4,q}^2) = q-1$ and that $\chi_{4,q}^2(-1) = \chi_{4,q}((-1)^2) = 1$. Then (3.5) applied to the tuples of characters $(\chi_{4,q},\chi_{4,q}^3)$, $(\chi_{4,q}^2,\chi_{4,q}^2)$ and (3.4) applied to $(\chi_{4,q},\chi_{4,q},\chi_{4,q}^2)$ give respectively

$$G(\chi_{4,q})G(\chi_{4,q}^3) = \chi_{4,q}(-1)q; \tag{4.6}$$

$$G(\chi_{4,q}^2)G(\chi_{4,q}^2) = q; \tag{4.7}$$

$$G(\chi_{4,q})G(\chi_{4,q})G(\chi_{4,q}^2) = q\,\pi_{4,q}. \tag{4.8}$$

Combining (4.7) and (4.8) we get

$$G(\chi_{4,q})^4 = q\,\pi_{4,q}^2. \tag{4.9}$$

Now, (4.2)-(4.5) follow at once from (4.6)-(4.9) and (3.5). □

	$[\chi_{4,q}(\underline{a})] \in \mu_4^4/\sim$	$b_{F,q}$	$c_{F,q}$	$b_{F,q}+c_{F,q}$	$b_{F,q}-c_{F,q}$
U_1	[(1, 1, 1, 1)]	7	12	19	-5
U_2	[(1, 1, 1,-1)]	-5	0	-5	-5
U_3	[(1, 1, 1, i)]	-1	-6	-7	5
U_4	[(1, 1,-1,-1)]	7	-4	3	11
U_5	[(1, 1,-1, i)]	-1	2	1	-3
U_6	[(1, 1, i, i)]	3	4	7	-1
U_7	[(1, 1, i,-i)]	-1	0	-1	-1
U_8	[(1,-1, i,-i)]	3	-4	-1	7

TABLE 4.1. Table displaying the quantities appearing in Proposition 4.3.

4.2. **Cubic and biquadratic diagonal congruences.** The required estimate in the case of cubic diagonal forms in 3 variables is readily obtained.

Proposition 4.2. Let $F(\mathbf{x}) = a_1x_1^3 + a_2x_2^3 + a_3x_3^3$ with $a_1, a_2, a_3 \in \mathbb{Z} \setminus \{0\}$ and let p be a prime number with $p \notin \Sigma_F$ and $p \equiv 1 \pmod 3$. Then

$$r_F(0,p) = p^2 + 2\,\mathbb{R}\mathrm{e}\, H_{F,p}(p\sqrt{p} - \sqrt{p}), \tag{4.10}$$

with $H_{F,p} := \overline{\chi}_{3,p}(a_1a_2a_3)\pi_{3,p}/\sqrt{p}$.

Proof. There are only two nontrivial cubic characters of $\mathbb{F}_p$: $\mathfrak{X}_p^{(3)} = \{\chi_{3,p}, \overline{\chi}_{3,p}\}$. Notice that $\chi_{3,p}^{-1} = \chi_{3,p}^2 = \overline{\chi}_{3,p}$. Therefore by Proposition 3.3, Lemma 4.1 and the multiplicativity of characters, we get

$$\begin{aligned} r_F(0,p) &= p^2 + \overline{\chi}_{3,p}(a_1a_2a_3)(p-1)\pi_{3,p} + \chi_{3,p}(a_1a_2a_3)(p-1)\overline{\pi}_{3,p} \\ &= p^2 + 2(p-1)\,\mathbb{R}\mathrm{e}(\overline{\chi}_{3,p}(a_1a_2a_3)\pi_{3,p}). \end{aligned}$$

□

The case of biquadratic diagonal forms comes with some extra complication, so we introduce some notation. Let q be a prime number with $q \equiv 1 \pmod 4$ and let $\underline{a} = (a_1, a_2, a_3, a_4) \in \mathbb{Z}^4$ with $q \nmid a_1a_2a_3a_4$. We denote by $\chi_{4,q}(\underline{a})$ the quadruple

$$\chi_{4,q}(\underline{a}) := (\chi_{4,q}(a_1), \chi_{4,q}(a_2), \chi_{4,q}(a_3), \chi_{4,q}(a_4)) \in \mu_4^4,$$

where $\mu_4 = \{1, -1, i, -i\}$. We say that two quadruples $\mathbf{u}_1, \mathbf{u}_2 \in \mu_4^4$ are equivalent if $\mathbf{u}_2$ can be obtained from $\mathbf{u}_1$ by performing some or all of the following operations: (1) permutation of the components; (2) componentwise multiplication by an element of μ_4; (3) componentwise complex conjugation. The quotient $\mu_4^4/\sim$ obtained by this equivalence relation has 8 elements, displayed in table 4.1. For later reference, we label these 8 elements with the names $U_1, \dots, U_8$. We denote the equivalence class of an element $\mathbf{u} \in \mu_4^4$ by $[\mathbf{u}]$.

Proposition 4.3. Let $F(\mathbf{x}) = a_1x_1^4 + a_2x_2^4 + a_3x_3^4 + a_4x_4^4$ with $\underline{a} \in (\mathbb{Z} \setminus \{0\})^4$ as above. Let q be a prime number with $q \notin \Sigma_F$ and $q \equiv 1 \pmod 4$. Then

$$r_F(0,q) = q^3 + (2\,\mathbb{R}\mathrm{e}\, H_{F,q} + K_{F,q})\, q(q-1), \tag{4.11}$$

where

$$\begin{aligned} H_{F,q} &:= \overline{\chi}_{4,q}(a_1a_2a_3a_4)\pi_{4,q}^2/q, \\ K_{F,q} &:= b_{F,q} + \chi_{4,q}(-1)c_{F,q}, \end{aligned}$$

and $b_{F,q}, c_{F,q} \in \mathbb{Z}$ depend on $[\chi_{4,q}(\underline{a})]$ as indicated in table 4.1.

Proof. We have $\mathfrak{X}_q^{(4)} = \{\chi_{4,q}, \chi_{4,q}^2, \chi_{4,q}^3\}$, i.e. there are only three nontrivial biquadratic characters of $\mathbb{F}_q$. Thus Proposition 3.3 and Lemma 4.1 give

$$r_F(0,q) = q^3 + q(q-1)(b_{F,q} + \chi_{4,q}(-1)c_{F,q}) + (q-1)d_{F,q},$$

where

$$b_{F,q} = \chi_{4,q}^2(a_1a_2a_3a_4) + \frac{1}{4}\sum_{\sigma\in\mathfrak{S}_4} \chi_{4,q}(a_{\sigma(1)}a_{\sigma(2)})\chi_{4,q}^3(a_{\sigma(3)}a_{\sigma(4)});$$

$$c_{F,q} = \frac{1}{2}\sum_{\sigma\in\mathfrak{S}_4} \chi_{4,q}^2(a_{\sigma(1)})\chi_{4,q}^2(a_{\sigma(2)})\chi_{4,q}(a_{\sigma(3)})\chi_{4,q}^3(a_{\sigma(4)});$$

$$d_{F,q} = \overline{\chi}_{4,q}(a_1a_2a_3a_4)\pi_{4,q}^2 + \chi_{4,q}(a_1a_2a_3a_4)\overline{\pi}_{4,q}^2.$$

Here $\mathfrak{S}_4$ denotes the set of permutations of $\{1,2,3,4\}$. We observe that both $b_{F,q}$ and $c_{F,q}$ are symmetric polynomial combinations of the components of $\chi_{4,q}(\underline{a})$. They are both homogeneous of degree 8, so they are invariant with respect to multiplying the entries of $\chi_{4,q}(\underline{a})$ by some $\lambda \in \mu_4$. Moreover, we notice that both $b_{F,q}$ and $c_{F,q}$ are invariant under conjugation. Therefore $b_{F,q}$ and $c_{F,q}$ depend only on the class $[\chi_{4,q}(\underline{a})] \in \mu_4^4/\sim$. Now, a straightforward computation gives the values listed in table 4.1 in all the 8 cases. The proposition follows, since moreover $d_{F,q} = 2\,\mathrm{Re}(\overline{\chi}_{4,q}(a_1a_2a_3a_4)\pi_{4,q}^2)$. □

We remark that in the above statements we have $\mathrm{Re}\, H_{F,p}, \mathrm{Re}\, H_{F,q} \in [-1,1]$ for all $p \equiv 1 \pmod 3$ and $q \equiv 1 \pmod 4$, because $|\pi_{3,p}| = \sqrt{p}$ and $\left|\pi_{4,q}^2\right| = q$. In fact, in the next sections we are going to use the fact that for all $\rho \in (-1,1)$ the inequalities $\mathrm{Re}\, H_{F,p} < \rho$ and $\mathrm{Re}\, H_{F,q} < \rho$ are satisfied for a positive proportion of the primes. Notice moreover that $\chi_{4,q}(-1) = 1$ if $q \equiv 1 \pmod 8$ and $\chi_{4,q}(-1) = -1$ if $q \equiv 5 \pmod 8$. Therefore a necessary condition to have $r_F(0,q) < q^3$ in the case $s = k = 4$, is that $K_{F,q} = b_{F,q} \pm c_{F,q} < 2$ for some choice of sign $\pm$. Compare this with table 4.1.

5. Hecke L-functions and asymptotic estimates

There is a universal strategy, which we will implement later, to study the range of values of $H_{F,p}$ and $H_{F,q}$ from the previous section, or more generally quantities likewise computed from Jacobi sums. In this section we collect the main ingredients of the method: following Weil the Jacobi sums can be interpreted as Hecke characters; the theory of Hecke L-functions provides "generalized prime number theorem"-type estimates; finally these estimates are feeded into equidistribution lemmas. This game plan is inspired by Moreno [25], even though in detail we follow more closely an approach of Heath-Brown and Patterson [12, p.115] by using the generalized prime number theorem of Kubilyus and the equidistribution lemma of Erdős and Turán.

5.1. **Hecke characters.** Let K be a number field of degree $d := [K:\mathbb{Q}]$. A *Hecke character* (also named Grössencharakter) of K is a character of the idèle class group $\mathbb{A}_K^\times/K^\times$. More down to earth, let $\mathcal{O}_K$ be the ring of integers of K, let $\mathfrak{m} \subseteq \mathcal{O}_K$ be a nonzero ideal and let $\mathcal{I}_\mathfrak{m}$ be the set of the ideals of $\mathcal{O}_K$ that are coprime to $\mathfrak{m}$. Since $\mathcal{O}_K$ is a Dedekind domain, $\mathcal{I}_\mathfrak{m}$ is a multiplicative monoid generated by the prime ideals of $\mathcal{O}_K$ that don't divide $\mathfrak{m}$. A multiplicative homomorphism

$$H:\ \mathcal{I}_\mathfrak{m} \to \mathbb{C}^\times$$

is a Hecke character of K if there is a continuous group homomorphism $\chi_\infty : (K \otimes_\mathbb{Q} \mathbb{R})^\times \to \mathbb{C}^\times$ such that $H((\alpha)) = \chi_\infty(\alpha \otimes 1)$ for all $\alpha \in \mathcal{O}_K$ satisfying $\alpha \equiv 1 \pmod{\mathfrak{m}}$. In other words, H is a Hecke character if, for the same α,

$$H((\alpha)) = \prod_{\sigma: K\to\mathbb{C}} \sigma(\alpha)^{k_\sigma}\,|\sigma(\alpha)|^{c_\sigma} \tag{5.1}$$

for some integers $(k_\sigma)_\sigma$ and complex numbers $(c_\sigma)_\sigma$. We say that the $2n$-tuple $(k_\sigma, c_\sigma)_\sigma$ is a *vector of exponents* of H. The ideal $\mathfrak{m}$ is a *defining ideal* of H and χ_∞ is the *infinity type* of H. A Hecke character H is unitary if $|H(\mathfrak{a})| = 1$ for all $\mathfrak{a} \in \mathcal{I}_\mathfrak{m}$.
As a word of caution, we mention the fact that some authors define χ_∞^{-1} to be the infinity type of H. Moreover, sometimes in the literature the Hecke characters are required to be unitary by definition, while those that are not unitary are called quasicharacters. For more details on the basic facts and properties of Hecke characters, we refer to the fundational article of Hecke [13] or to the first chapter of Kubilyus [19]. According to the general theory, we know that the unitary Hecke characters of K with defining ideal $\mathfrak{m}$ form a finitely generated abelian group $G(K, \mathfrak{m})$. This group contains a natural free subgroup $G^{(1)}(K, \mathfrak{m})$ of order $d - 1$ whose elements are called *Hecke characters of the first kind*. Then the group $G(K, \mathfrak{m})$ of all unitary Hecke characters (which in the literature are sometimes called Hecke characters *of the second kind* [19]) splits into a direct product

$$G(K, \mathfrak{m}) = G^{(1)}(K, \mathfrak{m}) \times T(K, \mathfrak{m}),$$

where $T(K, \mathfrak{m})$ is the subgroup of Hecke characters of finite order, which are sometimes called *abelian characters*. This is a finite subgroup with cardinality $h(\mathfrak{m}) := 2^{r_1} h(K)\varphi(\mathfrak{m})$, where r_1 is the number of real embeddings of K, $h(K)$ is the ideal class number and $\varphi(\mathfrak{m}) := \#(\mathcal{O}_K/\mathfrak{m})^\times$ is the Euler function. In other words, every unitary Hecke character H can be written uniquely as a product $H = \chi\xi$ of an abelian character and a Hecke character of the first kind. Every Hecke character can be normalized to a unitary one through multiplication by a real power of the norm character.
In this paper we are concerned with two important examples of Hecke characters.

Definition 5.1. Let $s \in \mathbb{N}_+$, let K be a number field containing all s-th roots of unity, with ring of integers $\mathcal{O}_K$, and let $a \in \mathcal{O}_K \setminus \{0\}$. Let also $\mathfrak{m}_1 := (as) \subseteq \mathcal{O}_K$ and $\mathfrak{m}_2 := (s) \subseteq \mathcal{O}_K$. Recall the definition of $\chi_{s,\mathfrak{p}}(\cdot)$ from section 3.2 and that $\chi_{s,\mathfrak{p}}$ can be seen as a character of the finite field $\mathcal{O}_K/\mathfrak{p}$. Then we define the power residue symbol and the (normalized) Jacobi sum symbol

$$\left(\frac{a}{\cdot}\right)_s : \mathcal{I}_{\mathfrak{m}_1} \to \mathbb{C}^\times , \quad \mathfrak{J}_s(\cdot) : \mathcal{I}_{\mathfrak{m}_2} \to \mathbb{C}^\times$$

by setting $\left(\frac{a}{\mathfrak{p}}\right)_s := \chi_{s,\mathfrak{p}}(a)$ and $\mathfrak{J}_s(\mathfrak{p}) := -J(\chi_{s,\mathfrak{p}}, \chi_{s,\mathfrak{p}})(N\mathfrak{p})^{-1/2}$ for all prime ideal $\mathfrak{p}$ coprime to $\mathfrak{m}_1$ for the first, to $\mathfrak{m}_2$ for the second, and then extending by multiplicativity. Here $N\mathfrak{p} := \#(\mathcal{O}_K/\mathfrak{p})$ denotes the norm of $\mathfrak{p}$.

Proposition 5.2. Keep the notation of Definition 5.1.

(i) The power residue symbol $\left(\frac{a}{\cdot}\right)_s$ is a unitary abelian character of K with trivial infinity type and with $\mathfrak{m}_{a,s} := (as)^{f_{a,s}}$ as a defining ideal, for some $f_{a,s} \in \mathbb{N}_+$.
(ii) The Jacobi sum symbol $\mathfrak{J}_s(\cdot)$ is a Hecke character of K with defining ideal $\mathfrak{m}_{\mathfrak{J}_s} := (s^2)$. It is unitary if $s \geq 3$. Moreover if $s \in \{3, 4\}$ then the infinity type of $\mathfrak{J}_s(\cdot)$ satisfies $\chi_\infty(\alpha \otimes 1) = \alpha/|\alpha|$ for all $\alpha \in K^\times$.

Proof. Statement (i) is a consequence of Class Field Theory [18, Theorem 1.13(8) in Ch. 2.§ 1.8, and Example 36 in Ch. 1.§ 6.3]. The assertions in (ii) follow instead from the work of Weil [33], as follows. The fact that $\mathfrak{J}_s(\cdot)$ is a Hecke character is the main theorem of that paper: notice in particular that the minus sign in the definition of $\mathfrak{J}_s(\cdot)$ reflects the different sign convention for Jacobi sums in Weil's paper [33, eq. (I)] and in ours (eq. (3.1)). The fact that $\mathfrak{J}_s(\cdot)$ is unitary for $s \geq 3$ follows from [33, eq. (10)]. To compute the infinity type, Weil gives explicit general formulas in [33, eq.(9)

and the bottom of p.491]. According to these formulas, if $s \in \{3, 4\}$, we get that

$$\mathfrak{J}_s((\alpha)) = \bar{\alpha}^{\omega_1} \alpha^{\omega_2} N((\alpha))^{-1/2}$$

for every $\alpha \in \mathcal{O}_K$ such that $\alpha \equiv 1 \bmod s^2$, where

$$\omega_i = \left\lfloor \frac{i}{s} + \frac{i}{s} \right\rfloor.$$

Then, we have $\omega_1 = 0$ and $\omega_2 = 1$. Since $N((\alpha)) = \alpha\bar{\alpha} = |\alpha|^2$, the claim is proved. □

5.2. **Hecke L-functions.** Given a Hecke character H of K, one considers the attached Hecke L-function

$$L(H, s) = \sum_{\mathfrak{a} \in \mathcal{I}_\mathfrak{m}} H(\mathfrak{a})(N\mathfrak{a})^{-s} = \prod_{\mathfrak{p} \in \mathcal{I}_\mathfrak{m} \cap \operatorname{Spec} \mathcal{O}_K} (1 - H(\mathfrak{p})(N\mathfrak{p})^{-s})^{-1},$$

where $\operatorname{Spec} \mathcal{O}_K$ is the set of prime ideals of $\mathcal{O}_K$. Hecke L-functions form a class of relatively well-behaved L-functions. If H is unitary then both the Dirichlet series and the infinite Euler product above converge absolutely on the right half plane $\mathbb{R}\mathrm{e}(s) \geq 1$. Moreover $L(H, s)$ has a meromorphic analytic continuation on all the complex plane, which is entire if H is nontrivial.
Analytic estimates for $L(H, s)$ can be given in terms of the "size" of the character H. Following Kubilyus [19] we fix arbitrarily a basis $\boldsymbol{\xi} = (\xi_1, \dots, \xi_{d-1})$ of the group of Hecke characters of the first kind, so that every Hecke character H can be written uniquely as

$$H = \chi \xi_1^{m_1}, \dots, \xi_{d-1}^{m_{d-1}}, \tag{5.2}$$

for some abelian character χ and some integers $m_i \in \mathbb{Z}$. Then we define the *size of H with respect to $\boldsymbol{\xi}$* by

$$v_{\boldsymbol{\xi}}(H) := \prod_{i=1}^{d-1} (|m_i| + 3). \tag{5.3}$$

With this notation, a classical result concerning zero-free regions of Hecke L-functions states that $L(H, s) \neq 0$ if $s = \sigma + it$ satisfies

$$\sigma > 1 - \frac{c(K, \boldsymbol{\xi})}{\log(|t| + 3) + \log v_{\boldsymbol{\xi}}(H)}, \tag{5.4}$$

for some constant $c(K, \boldsymbol{\xi})$ independent of H [19, Lemma 2]. In fact, more recent results for zero-free regions of Hecke L-functions are available, which provide more precise estimates than (5.4) both in the $v_{\boldsymbol{\xi}}$ and t aspects [4, 1] (see also Remark 5.3). For more about the theory of (Hecke) L-functions, see [13], [17, Chapter 5.10], [22] or [26].

Remark 5.3. In the literature there is no universally accepted notation for the "size" of an Hecke character. For example, Coleman [4] defines it as the L^2-norm of some suitable vector of exponents $(k_\sigma, c_\sigma)_\sigma$ of the Hecke character, while Mitsui [24] uses a quantity related to the L^1-norm of this vector of exponents. In the book of Iwaniec and Kowalski [17], instead, the role of $v_{\boldsymbol{\xi}}(H)$ is played by the analytic conductor $\mathfrak{q}(H)$. The analytic conductor is a quantity that is computed in terms of the norm of the algebraic conductor of H (which is the largest defining ideal of H, with respect to inclusion) and the γ-factors of the functional equation of the $L(H, s)$. In fact, all these notions are related. For example, Hecke [13] describes the Hecke characters of the first kind by means of explicit formulas, which themselves are given in terms of the choice of a basis for the units of $\mathcal{O}_K$. From such description one can explicit a choice of a basis $\boldsymbol{\xi}$ for the set of Hecke characters of the first kind. Furthermore, Hecke provides

explicit formulas for the γ-factors of $L(H,s)$. From these formulas it is possible to verify that

$$\log v_{\boldsymbol{\xi}}(H) \asymp \log \mathfrak{q}(H),$$

where the implied constant may depend on K and $\mathfrak{m}$, but is independent of H. Similar considerations apply to the "sizes" defined by Coleman and Mitsui.

5.3. **Asymptotic estimates.** We are interested in Hecke characters primarily because they give access to the following version of the prime number theorem [19, Lemma 4].

Lemma 5.4. Let $\mathfrak{m}$ be an ideal of $\mathcal{O}_K$ and let $\boldsymbol{\xi}$ be a basis of $G^{(1)}(K,\mathfrak{m})$. Then there are effective constants $c_1(K,\boldsymbol{\xi}), c_2(K,\boldsymbol{\xi}) > 0$ such that for each nontrivial unitary Hecke character $H \in G(K,\mathfrak{m})$ and every $T \geq 2$ we have

$$\left| \sum_{\substack{\mathfrak{p} \in \mathcal{I}_{\mathfrak{m}} \cap \operatorname{Spec} \mathcal{O}_K \\ N\mathfrak{p} \leq T}} H(\mathfrak{p}) \right| \leq c_1(K,\boldsymbol{\xi})\, T \, \exp\left(\frac{-2c_2(K,\boldsymbol{\xi}) \log T}{\log v_{\boldsymbol{\xi}}(H) + \sqrt{\log T}} \right). \tag{5.5}$$

Lemma 5.4 is proved via standard arguments concerning zero-free regions of Hecke L-functions [13] [17, Thm 5.13], using (5.4). Refinements can be given using the more precise estimates for zero-free regions due to Coleman et al. [4, 1].
Assuming that $v_{\boldsymbol{\xi}}(H) \leq \sqrt{\log T}$, the expression on the right-hand side of (5.5) simplifies to

$$c_1 T e^{-c_2 \sqrt{\log T}}.$$

The strength of Lemma 5.4 is appreciated by noticing that the number of summands in the left-hand side of (5.5) is asymptotic to $T/\log T$ by a classical theorem of Landau. We now collect some estimates that can be easily checked by partial integration-summation (e.g. [11, Thm 421, 22.5.2]). It is useful, in order to simplify the calculations and the final estimates, to use the fact that

$$e^{-\alpha\sqrt{\log T}} = o((\log T)^{-A}), \tag{5.6}$$

for any fixed $A, \alpha > 0$ and for $T \to \infty$.

Lemma 5.5. Let $\mathcal{A} \subseteq \mathbb{N}$ be a set of positive integers such that for $T \to \infty$ the following estimate holds, for some $c, d > 0$, and where $\operatorname{Li}(T) := \int_2^T dx/\log x$:

$$\#\mathcal{A} \cap [1,T] = c \operatorname{Li}(T) + O(Te^{-d\sqrt{\log T}}).$$

Then:

$$\sum_{p \in \mathcal{A} \cap [1,T]} \log p \;\; = (c + o(1))T; \tag{5.7}$$

$$\sum_{p \in \mathcal{A} \cap [1,T]} p^{-1/2} = (2c + o(1)) \frac{\sqrt{T}}{\log T}; \tag{5.8}$$

$$\sum_{p \in \mathcal{A} \cap [1,T]} p^{-1} \;\; = c \log\log T + O(1); \tag{5.9}$$

$$\sum_{p \in \mathcal{A} \cap [1,T]} p^{-3/2} = O(1); \tag{5.10}$$

$$\sum_{p \in \mathcal{A} \cap [1,T]} p^{-2} \;\; = O(1). \tag{5.11}$$

5.4. **Equidistribution.** The estimates coming from Lemma 5.4 will be used to show that the values of some Hecke characters equidistribute on the unit circle. Classical tools to prove such results are Weyl's equidistribution lemma or its quantitative version due to Erdős and Turán [8, Theorem III]. The following proposition is a direct consequence of the Erdős-Turán equidistribution lemma.

Lemma 5.6. Let $\{h_a\}_{a\in\mathcal{A}}$ be a sequence of complex numbers of modulus 1 indexed by a finite set $\mathcal{A}$ and for every $n \in \mathbb{N}_+$ let $S_n := \sum_{a\in\mathcal{A}} \mathbb{R}\mathrm{e}(h_a^n)$. Let ϕ_1, ϕ_2 be real numbers satisfying $0 \le \phi_1 < \phi_2 \le \pi$. Then

$$\#\{a \in \mathcal{A} :\ \mathbb{R}\mathrm{e}\, h_a \in [\cos\phi_2, \cos\phi_1]\} = \frac{\phi_2 - \phi_1}{\pi}\#\mathcal{A} + E$$

with

$$|E| \le C\left(\frac{\#\mathcal{A}}{N} + \sum_{n=1}^{N}\frac{1}{n}\,|S_n|\right)$$

for every $N \in \mathbb{N}_+$ and for an absolute constant $C > 0$.

See [17, Chapter 5] for a general discussion on L-functions and equidistribution and [26, Exercise 3.2] for more precise versions of the Erdős-Turán inequality. In this article, the above results will be used to show the equidistribution of $H_{F,p}$ and $H_{F,q}$ of Propositions 4.2 and 4.3 as p, q vary. In other words, equidistribution of Jacobi sum symbols at prime elements. We remark that there are also equidistribution results for Gauss sums, which in turn are related to a famous problem of Kummer [25, 12, 29].

6. Exceptional forms and the term $K_{F,q}$

6.1. **Exceptional biquadratic diagonal forms.** In this paragraph we study a special family of biquadratic diagonal forms.

Definition 6.1. We say that a biquadratic diagonal form $F(\mathbf{x})$ is exceptional if there are positive integers a, b, c_1, c_2, c_3, c_4 and a permutation $\sigma \in \mathfrak{S}_4$ such that

$$F(\mathbf{x}) = a(c_1x_{\sigma(1)})^4 + b(c_2x_{\sigma(2)})^4 + 4a(c_3x_{\sigma(3)})^4 + 4b(c_4x_{\sigma(4)})^4.$$

We will prove the following characterization of exceptional forms.

Theorem 6.2. A biquadratic diagonal form $F(\mathbf{x})$ is exceptional if and only if for all prime numbers $q \notin \Sigma_F$ we have $r_F(0, q) \ge q^3$.

We first show that the condition is necessary through the following two lemmas which treat separately the cases $q \equiv 1 \pmod 4$ and $q \equiv 3 \pmod 4$. The proof of sufficiency is posponed to section 8.

Lemma 6.3. Let $F(\mathbf{x}) = a(c_1x_{\sigma(1)})^4 + b(c_2x_{\sigma(2)})^4 + 4a(c_3x_{\sigma(3)})^4 + 4b(c_4x_{\sigma(4)})^4$ be an exceptional form and let q be a prime number with $q \equiv 1 \pmod 4$ and $q \nmid abc_1c_2c_3c_4$. Then $r_F(0, q) \ge q^3$.

Proof. Since $\#\mathbb{F}_q^\times$ is divisible by four, $\mathbb{F}_q$ contains a fourth root of unity $\omega \in \mathbb{F}_q$ with $\omega^2 = -1$. Let $\lambda := 1 + \omega$, and notice that $\lambda^4 = -4$. Consider now $F'(\mathbf{x}) = ax_1^4 + bx_2^4 - ax_3^4 - bx_4^4$. Then the map $(x_1, \dots, x_4) \mapsto (c_1x_{\sigma(1)}, c_2x_{\sigma(2)}, \lambda c_3x_{\sigma(3)}, \lambda c_4x_{\sigma(4)})$ gives a bijection between $\mathcal{R}_F(0, q)$ and $\mathcal{R}_{F'}(0, q)$.
For every $t \in \mathbb{F}_q$ define $n_t := \#\{(y, z) \in \mathbb{F}_q^2 : ay^4 + bz^4 = t\}$. Then we deduce that

$$r_F(0, q) = r_{F'}(0, q) = \sum_{t\in\mathbb{F}_q} n_t^2.$$

However, it is clear that $\sum_{t\in\mathbb{F}_q} n_t = q^2$. Hence, from the quadratic-arithmetic mean inequality (or Cauchy-Schwartz) we get $r_F(0, q) \ge q^3$. □

Lemma 6.4. Let $F(\mathbf{x}) = a_1x_1^4+\ldots+a_4x_4^4$ with $a_1,\ldots,a_4 \in \mathbb{Z}\setminus\{0\}$ and let q be a prime number with $q \equiv 3 \pmod 4$ and $q \nmid a_1a_2a_3a_4$. Then $r_F(0,q) = q^3+\left(\frac{a_1a_2a_3a_4}{q}\right) q(q-1)$, where $\left(\frac{\cdot}{q}\right)$ denotes the Legendre symbol.

Proof. Recall that the Legendre symbol $\chi_{2,q}(\cdot \bmod q) := \left(\frac{\cdot}{q}\right)$ is the only nontrivial quadratic character of $\mathbb{F}_q$, so: $\mathfrak{X}_q^{(2)} = \{\chi_{2,q}\}$. For $a \in \mathbb{F}_q^\times$ we have $\chi_{2,q}(a) = 1$ if and only if a is a quadratic residue modulo q, and we have $\chi_{2,q}(a) = -1$ otherwise.
Let $F'(\mathbf{x}) : a_1x_1^2 + \ldots + a_4x_4^2$ be a quadratic form with the same coefficients as $F(\mathbf{x})$. Since $J(\chi_{2,q}) = 1$, we get $G(\chi_{2,q})^2 = \chi_{2,q}(-1)q$ by (3.1). Then by (3.2) we get $J_0(\chi_{2,q},\chi_{2,q},\chi_{2,q},\chi_{2,q}) = (q-1)q$, and since $\mathfrak{X}_q^{(2)} = \{\chi_{2,q}\}$ we see that

$$r_{F'}(0,q) = q^3 + \chi_{2,q}(a_1a_2a_3a_4)(q-1)q,$$

by Proposition 3.3 and multiplicativity of $\chi_{2,q}$. Finally, we notice that $r_F(0,q) = r_{F'}(0,q)$, because, since $q \equiv 3 \pmod 4$, we have $\#\{x \in \mathbb{F}_q :\ x^4 = y\} = \#\{x \in \mathbb{F}_q :\ x^2 = y\}$ for all $y \in \mathbb{F}_q$. □

If $F(\mathbf{x}) = a(c_1x_{\sigma(1)})^4 + b(c_2x_{\sigma(2)})^4 + 4a(c_3x_{\sigma(3)})^4 + 4b(c_4x_{\sigma(4)})^4$ is an exceptional form, the product of its coefficients is a perfect square. Then Lemma 6.4 implies that $r_F(0,q) \geq q^3$ if $q \equiv 3 \pmod 4$ and $q \notin \Sigma_F$. Together with Lemma 6.3 we conclude that $r_F(0,q) \geq q^3$ for every $q \notin \Sigma_F$, as claimed in Theorem 6.2.

6.2. **Computing $K_{F,q}$ via Kummer's theory.** In order to prove that the condition in Theorem 6.2 is sufficient, we need to analyze in more detail the formula given in Proposition 4.3. Fix $F(\mathbf{x}) = a_1x_1^4 + a_2x_2^4 + a_3x_3^4 + a_4x_4^4$ with $a_1,a_2,a_3,a_4 \in \mathbb{Z}\setminus\{0\}$ and recall that $\mu_4 = \{1,-1,i,-i\}$. We notice that the term $K_{F,q}$ in (4.11) depends only on $\chi_{4,q}(-1)$ and $\chi_{4,q}(a_1),\ldots,\chi_{4,q}(a_4)$, and that the character $\chi_{4,q}$ depends on the choice of a prime ideal $\mathfrak{q}$ of $\mathbb{Z}[i]$ above q. A prime $q \equiv 1 \bmod 4$ splits in $\mathbb{Z}[i]$ as $q = \mathfrak{q}\overline{\mathfrak{q}}$ and we have $\chi_{4,\overline{\mathfrak{q}}} = \overline{\chi_{4,\mathfrak{q}}}$.
Let $\mathcal{P}_{F,1}$ denote the set of prime numbers q that satisfy $q \equiv 1 \bmod 4$ and $q \notin \Sigma_F$. If $q \in \mathcal{P}_{F,1}$, let $\chi_{4,q}(\underline{a},-1) \in \mu_4^4\times\{\pm1\}$ be a shorthand for $((\chi_{4,q}(a_1),\ldots,\chi_{4,q}(a_4)),\chi_{4,q}(-1))$. For all $\mathbf{u} \in \mu_4^4\times\{\pm1\}$ let $\overline{\mathbf{u}} \in \mu_4^4\times\{\pm1\}$ be obtained from $\mathbf{u}$ by componentwise complex conjugation and let

$$\mathcal{P}_{F,\mathbf{u}} := \{q \in \mathcal{P}_{F,1} :\ \chi_{4,q}(\underline{a},-1) \in \{\mathbf{u},\overline{\mathbf{u}}\}\}.$$

The natural setting to study these sets is over the Gaussian quadratic field, via Kummer's theory. Let $K = \mathbb{Q}(i)$ and let $\Delta_F \subseteq K^\times/(K^\times)^4$ be the (finite abelian) subgroup multiplicatively generated by $a_1,a_2,a_3,a_4,-1$. Notice that $-1 \bmod (K^\times)^4 = 4 \bmod (K^\times)^4$, because $(1+i)^4 = -4$. Moreover, observe that $(K^\times)^4 \cap \mathbb{Q}_+ = (\mathbb{Q}^\times)^4$, where $\mathbb{Q}_+$ denotes the multiplictive group of strictly positive rational numbers. Therefore we can view Δ_F as the subgroup of $\mathbb{Q}_+/(\mathbb{Q}^\times)^4 \subseteq K^\times/(K^\times)^4$ multiplicatively generated by $a_1,a_2,a_3,a_4,4$. Notice that $\mathbb{Q}_+/(\mathbb{Q}^\times)^4 \cong \bigoplus_{\ell \text{ prime}} \mathbb{Z}/4\mathbb{Z}$ as an abelian group.
Let $L = K(\sqrt[4]{\Delta_F})$. By Kummer's theory [27, Ch. I.§ 5] we have that L/K is a finite abelian extension of exponent 4 with Galois group $G := \mathrm{Gal}(L/K) \cong \mathrm{Hom}(\Delta_F,\mu_4)$. The isomorphism $\psi : G \to \mathrm{Hom}(\Delta_F,\mu_4)$ and the dual $\hat{\psi} : \Delta_F \to \mathrm{Hom}(G,\mu_4)$ are induced by the perfect pairing $G\times\Delta_F \to \mu_4$ given by $(\sigma,a) \mapsto \frac{\sigma(\sqrt[4]{a})}{\sqrt[4]{a}}$). The link with the power residue characters is given by the fact that

$$\left(\frac{a}{\mathfrak{p}}\right)_4 = \frac{(\mathfrak{p},L/K)(\sqrt[4]{a})}{\sqrt[4]{a}}$$

for all $a \in \mathcal{O}_K \cap (L^\times)^4$ and all prime ideal $\mathfrak{p} \subseteq \mathcal{O}_K$ coprime with ma where $m = 2a_1 \dots a_4$. Here $(\mathfrak{p}, L/K) \in \mathrm{Gal}(L/K)$ denotes the Frobenius element of $\mathfrak{p}$, which is well-defined because L/K is abelian. In other words, the values of $\chi_{4,\mathfrak{p}}$ on $a_1, a_2, a_3, a_4, -1$ are obtained by applying $\hat{\psi}(a_1), \dots, \hat{\psi}(-1) \in \mathrm{Hom}(G, \mu_4)$ to the Frobenius element $(\mathfrak{p}, L/K) \in G$. Or dually, by applying $\psi((\mathfrak{p}, L/K)) \in \mathrm{Hom}(\Delta_F, \mu_4)$ to $a_1, a_2, a_3, a_4, -1 \in \Delta_F$.

6.3. **The sets $\mathcal{P}_{F,\mathbf{u}}$ and Chebotarev's theorem.** Following the discussion in section 6.2, we consider the map

$$\begin{array}{rccc} \varphi_F : & \mathrm{Hom}(\Delta_F, \mu_4) & \longrightarrow & \mu_4^4 \times \{\pm 1\} \\ & \chi & \longmapsto & ((\chi(a_1), \dots, \chi(a_4)), \chi(-1)) \end{array}$$

Proposition 6.5. Let $F(\mathbf{x}) = a_1x_1^4 + a_2x_2^4 + a_3x_3^4 + a_4x_4^4$ with $a_1, \dots, a_4 \in \mathbb{Z} \setminus \{0\}$ and let $\mathbf{u} \in \mu_4^4 \times \{\pm 1\}$ be in the image of φ_F. Then $\mathcal{P}_{F,\mathbf{u}} \neq \emptyset$ and moreover for $T \to \infty$ we have

$$\#\mathcal{P}_{F,\mathbf{u}} \cap [1, T] = \delta \,\mathrm{Li}(T) + O(Te^{-\alpha\sqrt{\log T}}) \tag{6.1}$$

for some $\delta \geq \frac{1}{1024}$ and some effectively computable absolute constant $\alpha > 0$.

Proof. Denote for brevity $\varphi = \varphi_F$ and recall that we described an isomorphism $\psi : \mathrm{Gal}(L/K) \to \mathrm{Hom}(\Delta_F, \mu_4)$ in section 6.2. Then by Chebotarev's theorem [31, Thm. 3.4] the set

$$\mathcal{P} := \{\mathfrak{p} \in \mathrm{Spec}\, \mathcal{O}_K : \ \psi((\mathfrak{p}, L/K)) \in \varphi^{-1}(\{\mathbf{u}, \overline{\mathbf{u}}\})\}$$

satisfies

$$\#\{\mathfrak{p} \in \mathcal{P} : \ N\mathfrak{p} \leq T\} = \delta' \,\mathrm{Li}(T) + O(Te^{-\alpha\sqrt{\log T}}), \tag{6.2}$$

for some $\alpha > 0$ and $\delta' = \frac{\#\varphi^{-1}(\{\mathbf{u}, \overline{\mathbf{u}}\})}{\#\Delta_F}$. Since Δ_F is an abelian group generated by 4 elements of order at most 4, and an element of order 2, we have $\#\Delta_F \leq 512$. In particular the degree of $L = \mathbb{Q}(i, \sqrt[4]{\Delta_F})$ over $\mathbb{Q}$ is at most 1024, and so we can take α to be an effectively computable absolute constant by [20]. For the sake of completeness, we remark that also the constant implied in the O-notation can be computed effectively, and it is an absolute constant if the Dirichlet zeta function of L has no real zero, while it may depend on the discriminant of L otherwise. Since $\mathbf{u}$ is in the image of φ, we have $\#\varphi^{-1}(\{\mathbf{u}, \overline{\mathbf{u}}\}) \geq 1$ and so $\delta' \geq \frac{1}{512}$. Notice that for every T there are at most $\sqrt{T}$ primes $\mathfrak{p} \in \mathcal{O}_K$ of degree two with $N\mathfrak{p} \leq T$. Indeed, these are the primes of the form $\mathfrak{p} = p\mathcal{O}_K$ where p is a (rational) prime number with $p \equiv 3 \pmod 4$, and $N\mathfrak{p} = p^2$. Therefore the estimate in (6.2) is also valid when we restrict to the primes of degree 1 which are coprime with $2a_1a_2a_3a_4$. These come in conjugate pairs, which correspond bijectively to rational primes $q \in \mathcal{P}_{F,1}$ via $q = \pi_{4,q}\overline{\pi}_{4,q}$. For such q we have

$$\pi_{4,q} \in \mathcal{P} \iff \overline{\pi}_{4,q} \in \mathcal{P} \iff q \in \mathcal{P}_{F,\mathbf{u}},$$

therefore we get (6.1) with $\delta = \delta'/2 \geq \frac{1}{1024}$. □

6.4. **Characters of Δ_F, exceptional forms and the inequality $K_{F,q} \leq 1$.** In this paragraph we finally compute the term $K_{F,q}$ of Proposition 4.3 when $F(\mathbf{x})$ is not exceptional and we deduce, together with Proposition 6.5, that $K_{F,q} \leq 1$ for a positive proportion of the primes $q \nmid \Sigma_F$. Let $\bar{\varphi}_F : \mathrm{Hom}(\Delta_F, \mu_4) \longrightarrow (\mu_4^4/\sim) \times \{\pm 1\}$ be the composition of φ_F with the natural projection $\pi : \mu_4^4 \times \{\pm 1\} \to (\mu_4^4/\sim) \times \{\pm 1\}$. See section 4.2 for the definition of $\mu_4^4/\sim$. For brevity, we denote the elements of $\mu_4^4/\sim$ by $U_1, \dots, U_8$ as shown in table 4.1.

Lemma 6.6. Let $F(\mathbf{x}) = a_1x_1^4+a_2x_2^4+a_3x_3^4+a_4x_4^4$ with $a_1,\dots,a_4 \in \mathbb{Z}\setminus\{0\}$. Assume that, in the image of $\bar{\varphi}_F$, there is no element (U,u_5) with

$$(U,u_5) \in \{(U_i,1) :\ i \in \{2,3,5,7,8\}\} \cup \{(U_i,-1) :\ i \in \{1,2,5,6,7\}\}. \tag{6.3}$$

Then $F(\mathbf{x})$ is exceptional.

Proof. We notice that the image of $\bar{\varphi}_F$ doesn't change, if we multiply one coefficient of $F(\mathbf{x})$ by the fourth power of an integer, or if we multiply all its coefficients by the same nonzero integer, or if we permute its cofficients. Therefore we may assume without loss of generality that $\gcd(a_1,a_2,a_3,a_4) = 1$ and that none of $a_1,\dots,a_4$ is divisible by nontrivial fourth powers.
For every prime ℓ we consider the group homomorphism $\chi'_\ell :\ \mathbb{Q}_+/(\mathbb{Q}^\times)^4 \to \mu_4$ given by $r \mapsto i^{v_\ell(r)}$, where $v_\ell(\cdot)$ is the ℓ-adic valuation. Let $\chi_\ell = (\chi'_\ell)|_{\Delta_F} \in \mathrm{Hom}(\Delta_F,\mu_4)$ be the restriction of χ'_ℓ with respect to the inclusion $\Delta_F \hookrightarrow \mathbb{Q}_+/(\mathbb{Q}^\times)^4$.
We cannot have $\bar{\varphi}_F(\chi_2) = (U_3,-1)$, otherwise $\bar{\varphi}_F(\chi_2^2) = (U_2,1)$ is in the image of $\bar{\varphi}_F$. Since $\chi_2(-1) = \chi'_2(4) = -1$, we must have $\bar{\varphi}_F(\chi_2) \in \{(U_4,-1),(U_8,-1)\}$. By the remarks made at the beginning of the proof, we may therefore assume that either

(a) $F(\mathbf{x}) = d_1x_1^4 + d_2x_2^4 + 4d_3x_3^4 + 4d_4x_4^4$, or
(b) $F(\mathbf{x}) = d_1x_1^4 + 2d_2x_2^4 + 4d_3x_3^4 + 8d_4x_4^4$,

for some odd integers d_1,d_2,d_3,d_4 with $\gcd(d_1,d_2,d_3,d_4) = 1$, none of which is divisible by nontrivial fourth powers.
Notice that for a prime number $\ell \neq 2$ we must have $\bar{\varphi}_F(\chi_\ell) \in \{(U_1,1),(U_4,1),(U_6,1)\}$. This means that ℓ doesn't divide $d_1d_2d_3d_4$ or else there are exactly two indices $i,j \in \{1,\dots,4\}$ such that $\ell\,|d_i$ and $\ell\,|d_j$, and moreover $v_\ell(d_i) = v_\ell(d_j) \in \{1,2,3\}$.
Suppose that $F(\mathbf{x})$ is not exceptional. Then observe that one of the following cases must occur, for some distinct odd prime numbers p_1,p_2:

(i) p_1 divides both d_1 and d_2;
(ii) p_1 divides both d_3 and d_4;
(iii) p_1 divides d_j and d_3, p_2 divides d_j and d_4 for some $j \in \{1,2\}$;
(iv) p_1 divides d_1 and d_j, p_2 divides d_2 and d_j for some $j \in \{3,4\}$.

We define auxiliary values $\alpha(1) = \alpha(3) = 2$ and $\alpha(2) = 1$. Now for each case (i)-(iv) we consider the following auxiliary character $\chi_{aux} \in \mathrm{Hom}(\Delta_F,\mu_4)$: in (i) $\chi_{aux} = \chi_{p_1}^{\alpha(v_{p_1}(d_1))}$; in (ii) $\chi_{aux} = \chi_{p_1}^{\alpha(v_{p_1}(d_3))}$; in (iii) and (iv) $\chi_{aux} = \chi_{p_1}^{\alpha(v_{p_1}(d_j))}\chi_{p_2}^{\alpha(v_{p_2}(d_j))}$. In each case we get that $\varphi_F(\chi_{aux})$ is equal to $((-1,-1,1,1),1)$ or $((1,1,-1,-1),1)$. But then we see that $\bar{\varphi}_F(\chi_2\chi_{aux}) = (U_1,-1)$ in case (a) above, and $\bar{\varphi}_F(\chi_2\chi_{aux}) = (U_6,-1)$ in case (b). Both are contrary to our assumptions, so $F(\mathbf{x})$ is exceptional. □

Proposition 6.7. Let $F(\mathbf{x})$ be a biquadratic diagonal form that is not exceptional. Choose $\mathbf{u} \in \mu_4^4 \times \{\pm1\}$ in the image of φ_F such that $\pi(\mathbf{u}) \in (\mu_4^4/\sim) \times \{\pm1\}$ satisfies (6.3). Then $K_{F,q} \leq 1$ for all $q \in \mathcal{P}_{F,\mathbf{u}}$.

Proof. Notice that $\pi(\mathbf{u}) = \pi(\overline{\mathbf{u}})$, so for all $q \in \mathcal{P}_{F,\mathbf{u}}$ we verify from (6.3) and table 4.1 that $K_{F,q} \leq 1$. □

7. Equidistribution and the terms $H_{F,p}$ and $H_{F,q}$

In this section we investigate the remaining terms $\mathbb{R}\mathrm{e}\,H_{F,p}, \mathbb{R}\mathrm{e}\,H_{F,q}$ in Proposition 4.2 and Proposition 4.3. The main fact that we exploit is that $H_{F,p}$ and $H_{F,q}$ essentially take the values of infinite order unitary Hecke characters. This enables us to show that they equidistribute on the unit circle as $p,q \to \infty$, using Lemma 5.6 and the estimates given by Lemma 5.4.

7.1. **Equidistribution of $H_{F,p}$.** The case of cubic forms is almost straightforward.

Proposition 7.1. Let $F(\mathbf{x}) = a_1x_1^3 + a_2x_2^3 + a_3x_3^3$ be a cubic diagonal form with $a_1, a_2, a_3 \in \mathbb{Z} \setminus \{0\}$. For all $\beta \in (-1, 1]$ let

$$\mathcal{P}_{F,\beta} := \{p \text{ prime} : \ p \equiv 1 \bmod 3, \ p \notin \Sigma_F, \text{ and } \mathbb{R}\mathrm{e}\, H_{F,p} \leq \beta\}.$$

Then $\mathcal{P}_{F,\beta}$ is nonempty, and for $T \to \infty$ we have

$$\#\mathcal{P}_{F,\beta} \cap [1, T] = \delta \operatorname{Li}(T) + O(Te^{-\alpha\sqrt{\log T}}) \tag{7.1}$$

for some absolute constant $\alpha > 0$ and with $\delta = \frac{1}{2\pi}\arccos(-\beta)$.

Proof. Notice that $\mathcal{P}_{F,1}$ is just the set of all primes $p \equiv 1 \pmod 3$ with $p \notin \Sigma_F$, because $\mathbb{R}\mathrm{e}\, H_{F,p} \leq 1$ is always satisfied. Now, recall from Proposition 5.2 that the Jacobi sum symbol and the power residue symbols are Hecke characters of cyclotomic fields. For every $n \in \mathbb{N}$ we consider the unitary Hecke character

$$H_n(\cdot) := \mathfrak{J}_3(\cdot)^n \left(\frac{a_1a_2a_3}{\cdot}\right)_3^{-n}$$

of the number field $K = \mathbb{Q}(e^{2\pi i/3})$. We have that $\mathfrak{m} = \mathfrak{m}_{a_1a_2a_3,3} \cap \mathfrak{m}_{\mathfrak{J}_3} \subseteq \mathcal{O}_K$ is a defining ideal of H_n for every $n \in \mathbb{N}$ and the infinity type of H_n is $\alpha \mapsto (\alpha/|\alpha|)^n$. Since the field K has degree $d = 2$, the size of H_n satisfies $v_{\boldsymbol{\xi}}(H_n) \leq c_3 n$ for some $c_3 > 0$ independent of n. Moreover for $n \neq 0$ the character H_n is nontrivial, because, since $\mathfrak{m}$ is a lattice in $\mathbb{C}$, there exists $\alpha \in \mathbb{Z}[e^{2\pi i/3}]$ such that $\alpha \equiv 1 \pmod{\mathfrak{m}}$ and $\alpha^n \notin \mathbb{R}$.
The primes $\mathfrak{p} \in \mathcal{I}_{\mathfrak{m}}$ above a prime $p \equiv 1 \pmod 3$ come in conjugate pairs, they satisfy $N\mathfrak{p} = p$ and we have either $\mathfrak{p} = \pi_{3,p}$ or $\overline{\mathfrak{p}} = \pi_{3,p}$. Therefore from the definitions we have

$$H_n(\mathfrak{p}) + H_n(\overline{\mathfrak{p}}) = 2\,\mathbb{R}\mathrm{e}((H_{F,p})^n).$$

On the other hand the primes $\mathfrak{p} \in \mathcal{I}_{\mathfrak{m}}$ above a prime $p \equiv 2 \pmod 3$ satisfy $N\mathfrak{p} = p^2$, and so there are at most $\sqrt{T}$ of them satisfying $N\mathfrak{p} \leq T$, for every given $T > 0$. By these remarks, and by Lemma 5.4 applied to $H_n(\cdot)$ we get, for every $T \geq 2$ and every positive integer $n \leq c_3^{-1}\exp(\sqrt{\log T})$:

$$\left|\sum_{p \in \mathcal{P}_{F,1} \cap [1,T]} \mathbb{R}\mathrm{e}((H_{F,p})^n)\right| \leq c_1\, T\, e^{-c_2\sqrt{\log T}} + \sqrt{T}, \tag{7.2}$$

for some absolute constants $c_1, c_2 > 0$. We observe that $H_{F,p}$ belongs to the unit circle for all $p \in \mathcal{P}_{F,1}$ and that $\#\mathcal{P}_{F,1} \cap [1, T] = \frac{1}{2}\operatorname{Li}(T) + O(Te^{-c_3\sqrt{\log T}})$ for some effective absolute constant $c_3 > 0$, by the prime number theorem in arithmetic progressions. Now by Lemma 5.6 applied with $\phi_1 = \arccos(\beta)$, $\phi_1 = \pi$ and $N = \lfloor c_3^{-1}e^{\sqrt{\log T}}\rfloor$ we get the asymptotics displayed in (7.1), for any $\alpha < \min\{1, c_2, c_3\}$. □

7.2. **Equidistribution of $H_{F,q}$.** For a biquadratic form $F(\mathbf{x})$ we need that $H_{F,q}$ equidistributes when $q \to \infty$ ranges in the set $q \in \mathcal{P}_{F,\mathbf{u}}$ that we defined in section 6.2, for a fixed $\mathbf{u} \in \mu_4^4 \times \{\pm 1\}$. To detect those primes among the primes in $\mathcal{P}_{F,1}$ (and so to handle sums indexed by them) we use character sums, as follows. We define the auxiliary polynomial $f_{aux}(x) := 1 + x + x^2 + x^3$ and the auxiliary sum

$$S_{aux}(\mathbf{u}, \mathbf{v}) = \sum_{\mathbf{k} \in (\mathbb{Z}/4\mathbb{Z})^5} C(\mathbf{u}, \mathbf{k}) \prod_{i=1}^{5} v_i^{k_i} \tag{7.3}$$

for all $\mathbf{u}, \mathbf{v} \in \mu_4^4 \times \{\pm 1\}$, where $C(\mathbf{u}, \mathbf{k}) = 2^{1-\epsilon_{\mathbf{u}}} 4^{-5} \operatorname{Re}(u_1^{k_1} u_2^{k_2} u_3^{k_3} u_4^{k_4} u_5^{k_5})$, $\epsilon_{\mathbf{u}} = 1$ if $\mathbf{u} = \overline{\mathbf{u}}$ and $\epsilon_{\mathbf{u}} = 0$ otherwise. Observe that

$$S_{aux}(\mathbf{u}, \mathbf{v}) = \frac{2^{-\epsilon_{\mathbf{u}}}}{4^5} \left(\prod_{i=1}^{5} f_{aux}(u_i v_i) + \prod_{i=1}^{5} f_{aux}(u_i^{-1} v_i) \right),$$

from which we see that $S_{aux}(\mathbf{u}, \mathbf{v}) = 1$ if $\mathbf{v} \in \{\mathbf{u}, \overline{\mathbf{u}}\}$ and $S_{aux}(\mathbf{u}, \mathbf{v}) = 0$ otherwise. Therefore, for $q \in \mathcal{P}_{F,1}$ we have $q \in \mathcal{P}_{F,\mathbf{u}}$ if and only if $S_{aux}(\mathbf{u}, \chi_{4,q}(\underline{a}, -1)) = 1$, where $\mathcal{P}_{F,\mathbf{u}}$ and $\chi_{4,q}(\underline{a}, -1)$ are as in section 6.2. In particular, for all $T \geq 1$, all $\mathbf{u} \in \mu_4^4 \times \{\pm 1\}$ and every function $h : \mathcal{P}_{F,1} \to \mathbb{C}$ we have

$$\sum_{q \in \mathcal{P}_{F,\mathbf{u}} \cap [1,T]} h(q) = \sum_{q \in \mathcal{P}_{F,1} \cap [1,T]} S_{aux}(\mathbf{u}, \chi_{4,q}(\underline{a}, -1))\ h(q). \tag{7.4}$$

This is a common technique in analytical number theory, see e.g. [19, Lemma 4] for an application of this trick in a similar context.

Proposition 7.2. Let $F(\mathbf{x}) = a_1 x_1^4 + \ldots + a_4 x_4^4$ be a biquadratic diagonal form with $a_1, \ldots, a_4 \in \mathbb{Z} \setminus \{0\}$. For all $\mathbf{u} \in \mu_4^4 \times \{\pm 1\}$ and all $\beta \in (-1, 1]$ let

$$\mathcal{P}_{F,\mathbf{u},\beta} := \{q \in \mathcal{P}_{F,\mathbf{u}} :\ \operatorname{Re} H_{F,q} \leq \beta\}.$$

If $\mathbf{u}$ is in the image of φ_F, then $\mathcal{P}_{F,\mathbf{u},\beta}$ is nonempty, and for $T \to \infty$ we have

$$\#\mathcal{P}_{F,\mathbf{u},\beta} \cap [1, T] = \delta \operatorname{Li}(T) + O(T e^{-\alpha\sqrt{\log T}}) \tag{7.5}$$

for some effective absolute constant $\alpha > 0$ and with $\delta \geq \frac{1}{1024\pi} \arccos(-\beta)$.

Proof. For all $n \in \mathbb{N}_+$ and all $\mathbf{k} = (k_1, \ldots, k_5) \in (\mathbb{Z}/4\mathbb{Z})^5$ we define the unitary Hecke character

$$H_{n,\mathbf{k}}(\cdot) := \mathfrak{J}_4(\cdot)^{2n} \Big(\frac{a_1 a_2 a_3 a_4}{\cdot}\Big)_4^{-n} \Big(\frac{a_1}{\cdot}\Big)_4^{k_1} \Big(\frac{a_2}{\cdot}\Big)_4^{k_2} \Big(\frac{a_3}{\cdot}\Big)_4^{k_3} \Big(\frac{a_4}{\cdot}\Big)_4^{k_4} \Big(\frac{-1}{\cdot}\Big)_4^{k_5}$$

of the number field $K = \mathbb{Q}(i)$. For every n and $\mathbf{k}$ as above we have that

$$\mathfrak{m} = \mathfrak{m}_{a_1,4} \cap \mathfrak{m}_{a_2,4} \cap \mathfrak{m}_{a_3,4} \cap \mathfrak{m}_{a_4,4} \cap \mathfrak{m}_{-1,4} \cap \mathfrak{m}_{\mathfrak{J}_4} \subseteq \mathcal{O}_K$$

is a defining ideal of $H_{n,\mathbf{k}}$ and $\alpha \mapsto (\alpha/|\alpha|)^{2n}$ is its infinity type. Observe that $H_{n,\mathbf{k}}$ is nontrivial for $n \neq 0$ because, $\mathfrak{m}$ being a lattice in $\mathbb{C}$, there exists $\alpha \in \mathbb{Z}[i]$ such that $\alpha \equiv 1 \pmod{\mathfrak{m}}$ and $\alpha^{2n} \notin \mathbb{R}$. Moreover, the size of $H_{2n,\mathbf{k}}$ satisfies $v_{\boldsymbol{\xi}}(H_{n,\mathbf{k}}) \leq c_3 n$ for some $c_3 > 0$ independent of n.
The primes $\mathfrak{q} \in \mathcal{I}_{\mathfrak{m}}$ with degree $\deg(\mathfrak{q}) \neq 1$ are precisely those above a prime $q \equiv 3 \pmod 4$. These primes satisfy $N\mathfrak{q} = q^2$, and so there are at most $\sqrt{T}$ of them satisfying $N\mathfrak{q} \leq T$, for every given $T > 0$. Therefore, by Lemma 5.4 applied to $H_{n,\mathbf{k}}(\cdot)$ we get, for every $T \geq 2$, every positive integer $n \leq c_3^{-1} \exp(\sqrt{\log T}) \in \mathbb{N}_+$ and every $\mathbf{k} \in (\mathbb{Z}/4\mathbb{Z})^5$:

$$\left| \sum_{\substack{\mathfrak{q} \in \operatorname{Spec} \mathcal{O}_K \cap \mathcal{I}_{\mathfrak{m}} \\ N\mathfrak{q} \leq T,\ \deg(\mathfrak{q}) = 1}} H_{n,\mathbf{k}}(\mathfrak{q}) \right| \leq c_1\, T\, e^{-c_2\sqrt{\log T}}, \tag{7.6}$$

for some constants $c_1, c_2 > 0$ independent of $n \in \mathbb{N}_+$ and $\mathbf{k} \in (\mathbb{Z}/4\mathbb{Z})^5$. The primes $\mathfrak{q} \in \mathcal{I}_{\mathfrak{m}}$ with $\deg(\mathfrak{q})$ come in conjugate pairs, they satisfy $N\mathfrak{p} = q$ for some $q \equiv 1 \pmod 4$ and we have either $\mathfrak{q} = \pi_{4,q}$ or $\overline{\mathfrak{q}} = \pi_{4,q}$. In particular, given such $\mathfrak{q}$ and $\mathbf{v} = \chi_{4,q}(\underline{a}, -1)$ we have:

$$H_{n,\mathbf{k}}(\mathfrak{q}) + H_{n,-\mathbf{k}}(\overline{\mathfrak{q}}) = 2 \operatorname{Re}((H_{F,q})^n) \prod_{i=1}^{5} v_i^{k_i} \tag{7.7}$$

for all $n \in \mathbb{N}_+$ and all $\mathbf{k} \in (\mathbb{Z}/4\mathbb{Z})^5$. Then (7.3), (7.4) and (7.7) imply

$$\sum_{q \in \mathcal{P}_{F,\mathbf{u}} \cap [1,T]} 2\,\mathrm{Re}((H_{F,q})^n) = \sum_{\mathbf{k} \in (\mathbb{Z}/4\mathbb{Z})^5} C(\mathbf{u},\mathbf{k}) \sum_{\substack{\mathfrak{q} \in \mathrm{Spec}\,\mathcal{O}_K \cap \mathcal{I}_{\mathfrak{m}} \\ N\mathfrak{q} \le T,\, \deg(\mathfrak{q})=1}} H_{n,\mathbf{k}}(\mathfrak{q}). \tag{7.8}$$

for some real numbers $C(\mathbf{u},\mathbf{k})$ satisfying $C(\mathbf{u},\mathbf{k}) = C(\mathbf{u},-\mathbf{k})$ and $|C(\mathbf{u},\mathbf{k})| \le \frac{1}{512}$. Finally, by (7.6) and (7.8) we deduce that

$$\left| \sum_{q \in \mathcal{P}_{F,\mathbf{u}} \cap [1,T]} \mathrm{Re}((H_{F,q})^n) \right| \le c_1\, T\, e^{-c_2\sqrt{\log T}}, \tag{7.9}$$

for all $T \ge 2$ and all positive $n \le c_3^{-1} \exp\sqrt{\log T}$. We observe that $H_{F,q}$ belongs to the unit circle for all $p \in \mathcal{P}_{F,\mathbf{u}} \cap [1,T]$ and that $\#\mathcal{P}_{F,\mathbf{u}} \cap [1,T]$ is estimated in Proposition 6.5. Now by Lemma 5.6 applied with $\phi_1 = \arccos(\beta)$, $\phi_1 = \pi$ and $N = \lfloor c_3^{-1} \exp\sqrt{\log T} \rfloor$ we get the asymptotics displayed in (7.1), for any $\alpha < \min\{1, c_2, \alpha'\}$. □

8. Detecting the existence of long gaps - the proof

8.1. **Congruences with few solutions.** For the remaining part of the article let $s \in \{3,4\}$ and let $F(\mathbf{x}) = a_1 x_1^s + \ldots + a_s x_s^s$, with $a_1, \ldots, a_s \in \mathbb{N}_+$ be either a cubic diagonal form or a biquadratic diagonal form that is not exceptional according to Definition 6.1.

Proposition 8.1. Let $s, F(\mathbf{x})$ be as above. Then we can choose a set $\mathcal{P}_F$ of prime numbers and effectively computable absolute constants $\alpha, \beta, \delta_0 > 0$ such that for all $p \in \mathcal{P}_F$ and all $m \in \mathbb{Z}$ we have

$$r_F(0,p) \le p^{s-1}\left(1 - \beta(p^{1-\frac{s}{2}} - p^{-\frac{s}{2}})\right), \tag{8.1}$$

$$r_F(m,p) \le p^{s-1}\left(1 + (s-1)^s p^{\frac{1}{2}-\frac{s}{2}}\right), \tag{8.2}$$

and for $T \to \infty$ we have, for some $\delta \ge \delta_0$:

$$\#\mathcal{P}_F \cap [1,T] = \delta\,\mathrm{Li}(T) + O(Te^{-\alpha\sqrt{\log T}}). \tag{8.3}$$

Proof. If $s = 3$ the inequality (8.1) and the asymptotics (8.3) follow from Proposition 4.2 and Proposition 7.1 by choosing any $\beta \in (0,2)$ and letting $\mathcal{P}_F := \mathcal{P}_{F,-\beta/2}$. If $s = 4$ we may choose any $\beta \in (0,1)$, and let $\mathbf{u}$ be any element in the image of φ_F such that $\pi(\mathbf{u})$ satisfies (6.3). Then (8.1) and (8.3) follow from Proposition 4.3, Proposition 6.5 and Proposition 7.2 with $\mathcal{P}_F := \mathcal{P}_{F,\mathbf{u},-\beta/2}$. For both $s \in \{3,4\}$ and for the same choice of $\mathcal{P}_F$, (8.2) follows from (8.1) when $p \,|\, m$ and it follows from Proposition 3.2 otherwise.

□

From Proposition 8.1 we deduce that a biquadratic diagonal form satisfying $r_F(0,p) \ge p^3$ for all but finitely many primes p must be exceptional in the sense of Definition 6.1. This observation, together with the arguments of section 6.1, completes the proof of Theorem 6.2. For non-exceptional diagonal forms, Proposition 8.1 implies that the ratio $r_F(m,M)/M^{s-1}$ can be made strictly less than 1 for suitable m and $M = p$ prime. In the next proposition we make this ratio arbitrarily small by using products of primes.

Proposition 8.2. Let $s, F(\mathbf{x}), \beta, \mathcal{P}_F$ be as in Proposition 8.1. Let $\mathcal{P}_1 \subseteq \mathcal{P}_2 \subset \mathcal{P}_F$ with $\#\mathcal{P}_2 < \infty$ and let $m \in \mathbb{Z}$ with $m \equiv 0 \pmod p$ for all $p \in \mathcal{P}_1$. Then we have

$r_F(m,M) \le \varepsilon M^{s-1}$ for $M := \prod_{p\in\mathcal{P}_2} p$ and all $\varepsilon > 0$ that satisfy

$$\log\varepsilon \ge -\sum_{p\in\mathcal{P}_1} \beta(p^{1-\frac{s}{2}} - p^{-\frac{s}{2}}) + \sum_{p\in\mathcal{P}_2\setminus\mathcal{P}_1} (s-1)^s p^{\frac{1}{2}-\frac{s}{2}}. \tag{8.4}$$

Proof. By Lemma 3.1 we have that

$$\frac{r_F(m,M)}{M^{s-1}} = \prod_{p\in\mathcal{P}_1} \frac{r_F(0,p)}{p^{s-1}} \prod_{p\in\mathcal{P}_2\setminus\mathcal{P}_1} \frac{r_F(m,p)}{p^{s-1}}.$$

By Proposition 8.1 and the inequality $\log(1+x) \le x$, valid for all $x > -1$, we have that

$$\log\left(\prod_{p\in\mathcal{P}_1} \frac{r_F(0,p)}{p^{s-1}}\right) \le \beta \sum_{p\in\mathcal{P}_1} (-p^{1-\frac{s}{2}} + p^{-\frac{s}{2}}),$$

and

$$\log\left(\prod_{p\in\mathcal{P}_2\setminus\mathcal{P}_1} \frac{r_F(m,p)}{p^{s-1}}\right) \le (s-1)^s \sum_{p\in\mathcal{P}_2\setminus\mathcal{P}_1} p^{\frac{1}{2}-\frac{s}{2}},$$

so the proposition follows. □

8.2. **Low density along arithmetic progressions.** Let $s, F(\mathbf{x})$ be as in section 8.1.

Definition 8.3. For $n \in \mathbb{N}$ we define $r_F(n) := \#\mathcal{R}_F(n)$, where

$$\mathcal{R}_F(n) := \{\mathbf{x} \in \mathbb{N}^s : \ F(\mathbf{x}) = n\}.$$

In other words, $r_F(n)$ counts the number of representations of n via the form $F(\mathbf{x})$. Then the image $\mathcal{S}_F$ of $F(\mathbf{x})$ can be described as

$$\mathcal{S}_F := \{n \in \mathbb{N} : \ r_F(n) \neq 0\}.$$

The relative density of $\mathcal{S}_F$ along an arithmetic progression of the form $m + M\mathbb{N}$ is related to $r_F(m,M)$. A trivial inequality relating the two is sufficient for our purpose.

Proposition 8.4. Let $s, F(\mathbf{x}), \mathcal{S}_F$ be as above. Let $L, M, m \in \mathbb{N}_+$ with $m < M$. Then

$$\#\left(\mathcal{S}_F \cap (m + M\mathbb{N}) \cap [0, L^s M^s)\right) \le r_F(m,M) L^s. \tag{8.5}$$

Proof. We consider the map

$$\begin{array}{rccc} \phi : & \bigcup_{k=1}^{L^s M^{s-1}} \mathcal{R}_F(m+(k-1)M) & \longrightarrow & \mathcal{R}_F(m,M) \\ & (x_1,\dots,x_s) & \longmapsto & (x_1 \bmod M, \dots, x_s \bmod M) \end{array}$$

and for every $k \le L^s M^{s-1}$ we notice that $m + (k-1)M < L^s M^s$. This implies that for every $\mathbf{x} \in \mathcal{R}_F(m+(k-1)M)$ and all $j \in \{1,\dots,s\}$ we have $0 \le x_j < LM$. For every residue class $\overline{x}$ modulo M there are only L integers x satisfying $x \equiv \overline{x} \pmod M$ and $0 \le x < LM$. We deduce that every element in the image of ϕ can have at most L^s preimages. Therefore

$$\sum_{k=1}^{L^s M^{s-1}} r_F(m+(k-1)M) \le L^s\, r_F(m,M).$$

Since

$$\#\left(\mathcal{S}_F \cap (m + M\mathbb{N}) \cap [0, L^s M^s)\right) \le \sum_{k=1}^{L^s M^{s-1}} r_F(m+(k-1)M),$$

the proposition follows. □

Recall that we are interested in intervals contained in $\mathbb{N} \setminus \mathcal{S}_F$, so next we consider arithmetic progressions of intervals with fixed length. The union of these intervals in arithmetic progression forms a "rectangle" of integers $\{m + (h-1)M + k : \ h \leq H, k \leq K\}$. A set of this form is sometimes known as a Maier matrix.

Proposition 8.5. Let $s, F(\mathbf{x}), \mathcal{S}_F$ be as above, let $L, M, m, K \in \mathbb{N}_+$ with $m+K < M$ and let $\mathcal{A} = (m + \mathbb{N}M) \cap [0, L^s M^s)$ be a truncated arithmetic progression. Now let

$$\mathcal{B} := \{a \in \mathcal{A} : \ \mathcal{S}_F \cap (a + [1, K]) \neq \emptyset\}$$

and suppose that

$$r_F(m+1, M) + \ldots + r_F(m+K, M) \leq \frac{1}{2} M^{s-1}. \tag{8.6}$$

Then $\#\mathcal{A} = L^s M^{s-1}$ and $\#\mathcal{B} \leq \frac{1}{2} L^s M^{s-1}$.

Proof. Since $m < M$, the inequality $\#\mathcal{A} = L^s M^{s-1}$ is clear. To estimate $\#\mathcal{B}$, first notice that

$$\#\mathcal{B} \leq \sum_{h=1}^{L^s M^{s-1}} \#\left(\mathcal{S}_F \cap (m + (h-1)M + [1, K])\right) \tag{8.7}$$

because each element of $\mathcal{B}$ contributes at least 1 to the sum in the right hand side of (8.7). Now observe that this sum is equal to

$$\sum_{i=1}^{K} \#\left(\mathcal{S}_F \cap (m + i + M\mathbb{N}) \cap [0, L^s M^s)\right).$$

because $m + K < M$. By Proposition 8.4 we deduce that

$$\#\mathcal{B} \leq \sum_{i=1}^{K} L^s \, r_F(m+i, M) \leq \frac{1}{2} L^s M^{s-1}.$$

□

We remark that $\{a+1, \ldots, a+K\}$ is a gap in the values of $F(\mathbf{x})$ for any $a \in \mathcal{A} \setminus \mathcal{B}$ as in Proposition 8.5. In particular, the existence of such gaps follows from an inequality of the form (8.6).

8.3. **Choice of parameters.** In order to fulfil (8.6), we will use the upper bounds on the summands $r_F(m+i, M)$ coming from Proposition 8.2 and from a suitable choice of sets $\mathcal{P}_1 \subseteq \mathcal{P}_2 \subset \mathcal{P}_F$. We will set $\mathcal{P}_2 = \mathcal{P}_F \cap [1, T]$, for some T large enough, and the next lemma is about finding the appropriate values of T.

Definition 8.6. For all $\gamma \in \mathbb{R}_+$ and $K \in \mathbb{N}_+$ we set

$$\tau_3(\gamma, K) := \gamma K^2 (\log K)^4;$$
$$\tau_4(\gamma, K) := \exp(\exp(\gamma K \log K)).$$

Lemma 8.7. Let $s, F(\mathbf{x}), \mathcal{P}_F, \beta$ be as in Proposition 8.1. Then there is $\gamma_F \geq 1$ such that for all $K \in \mathbb{N}$ with $K \geq 2$, and all $T \geq \tau_s(\gamma_F, K)$, we have

$$\log\left(\frac{1}{2K}\right) \geq \beta - \frac{1}{K} \sum_{p \in \mathcal{P}_F \cap [1,T]} \beta(p^{1-\frac{s}{2}} - p^{-\frac{s}{2}}) + \sum_{p \in \mathcal{P}_F \cap [1,T]} (s-1)^s p^{\frac{1}{2}-\frac{s}{2}}. \tag{8.8}$$

In particular, for this choice of T the set $\mathcal{P}_F \cap [1, T]$ is nonempty.

Please compare (8.8) with (8.4) and notice the extra multiplicative factor $\frac{1}{K}$ in front of the first sum.

Proof. Consider first the case $s = 4$. Let $\gamma \geq 1$ and $T \geq \tau_4(\gamma, K)$. By Proposition 8.1 and Lemma 5.5 we have that:

$$\text{(a)} \qquad \sum_{p \in \mathcal{P}_F \cap [1,T]} (s-1)^s p^{\frac{1}{2}-\frac{s}{2}} \quad \leq C_1;$$

$$\text{(b)} \quad \frac{1}{K} \sum_{p \in \mathcal{P}_F \cap [1,T]} \beta(p^{1-\frac{s}{2}} - p^{-\frac{s}{2}}) \geq \gamma\beta\delta \log K - C_2;$$

for some constants $C_1, C_2 > 0$ independent of K. Then (8.8) holds if $\gamma \geq \gamma_F$ for some γ_F that can be chosen independently of $K \geq 2$. Now we consider the case $s = 3$. Let $\gamma \geq 1$ and $T \geq \tau_3(\gamma, K)$. From Lemma 5.5 and Proposition 8.1 we have that:

$$\text{(a)} \qquad \sum_{p \in \mathcal{P}_F \cap [1,T]} (s-1)^s p^{\frac{1}{2}-\frac{s}{2}} \quad \leq C_3 \log\log \max\{\gamma, K\};$$

$$\text{(b)} \quad \frac{1}{K} \sum_{p \in \mathcal{P}_F \cap [1,T]} \beta(p^{1-\frac{s}{2}} - p^{-\frac{s}{2}}) \geq C_4 \frac{\sqrt{\gamma}(\log K)^2}{\log \max\{\gamma, K\}};$$

for some constants $C_3, C_4 > 0$ independent of K. Again, it is easy to see that (8.8) holds if $\gamma \geq \gamma_F$ for some γ_F that can be chosen independently of K. Finally, we observe that $\beta + \log(2K) > 0$, so (8.8) doesn't hold if $\mathcal{P}_F \cap [1, T] = \emptyset$. □

8.4. **Conclusion.** Let $s, F(\mathbf{x})$ be as in section 8.1. Given $N, K \in \mathbb{N}_+$, we define

$$\mathrm{Gap}_F(N, K) := \{n \in \mathbb{N} : \ n < N \text{ and } \mathcal{S}_F \cap (n + [1, K]) = \emptyset\}.$$

We aim to show that for every K there is $N \in \mathbb{N}_+$ such that $\mathrm{Gap}_F(N, K)$ is nonempty.

Theorem 8.8. Let $s, F(\mathbf{x})$ be as in section 8.1. Then for all $K \geq 2$ there is a constant $C_{F,K} > 0$ such that for all $N \geq e^{sC_{F,K}}$ we have

$$\#\,\mathrm{Gap}_F(N, K) \geq \frac{e^{-C_{F,K}}}{32} N.$$

Moreover we can choose $C_{F,K} = (\delta + o(1))\tau_s(\gamma_F, K)$ as $K \to \infty$, where δ is as in Proposition 8.1, and γ_F is as in Lemma 8.7.

Proof. Fix $K \geq 2$ and let $T \geq \tau_s(\gamma_F, K)$. By Lemma 8.7 we have that $\mathcal{P}_F \cap [1, T] \neq \emptyset$, so let

$$M := \prod_{p \in \mathcal{P}_F \cap [1,T]} p,$$

let $N \in \mathbb{N}$ with $N \geq M^s$, and let $L := \lfloor \sqrt[s]{N}/M \rfloor$. Since $p^{1-\frac{s}{2}} - p^{-\frac{s}{2}} < 1$ for all $p \geq 1$, we can easily construct a partition

$$\mathcal{P}_F \cap [1, T] = \mathcal{P}_F^{(1)} \sqcup \cdots \sqcup \mathcal{P}_F^{(K)}$$

such that for all $i \in \{1, \ldots, K\}$ we have

$$\sum_{p \in \mathcal{P}_F^{(i)}} (p^{1-\frac{s}{2}} - p^{-\frac{s}{2}}) \geq -1 + \frac{1}{K} \sum_{p \in \mathcal{P}_F \cap [1,T]} (p^{1-\frac{s}{2}} - p^{-\frac{s}{2}}). \tag{8.9}$$

By the Chinese Remainder Theorem there is some $m \in \mathbb{N}$ with $m < M$ that satisfies $m \equiv -i \pmod{p}$ for all $i \in \{1, \ldots, K\}$ and all $p \in \mathcal{P}_F^{(i)}$. By (8.9), Lemma 8.7 and Proposition 8.2 with $\varepsilon = \frac{1}{2K}$ we deduce that

$$r_F(m + i, M) \leq \frac{1}{2K} M^{s-1}$$

for all $i \in \{1, \dots, K\}$. Then Proposition 8.5 implies that

$$\#\operatorname{Gap}_F(L^s M^s, K) \geq \frac{1}{2} L^s M^{s-1} = \frac{(L+1)^s M^s}{2M} \left(\frac{L}{L+1}\right)^s \geq \frac{N}{2M} \left(\frac{1}{2}\right)^s \geq \frac{N}{32M}.$$

By Lemma 5.5 we have that $\log M = T(\delta + o(1))$ as $T \to \infty$. Since $\tau_s(\gamma_F, K) \to \infty$ as $K \to \infty$, and since $\operatorname{Gap}_F(L^s M^s, K) \subseteq \operatorname{Gap}_F(N, K)$, the theorem follows. □

We remark that, despite the appearances, in general a larger value of δ corresponds to a smaller value of $C_{F,K}$ in Theorem 8.8. As a corollary of Theorem 8.8 we get the theorems stated in the Introduction.

Proof of Theorem 1.1. Let $s = 3$, let $F(\mathbf{x})$ be as in Theorem 1.1 and let $N, K \in \mathbb{N}$ with $K \geq 2$. From Theorem 8.8 (applied to estimate $\#\operatorname{Gap}_F(N-K, K)$) it is possible to compute some constant $\gamma > 0$, independent of N and K, such that whenever the inequality

$$N \geq \exp(\gamma K^2 (\log K)^4) \tag{8.10}$$

holds, there is a gap of length K in the values of $F(\mathbf{x})$ less than N. When $N \geq e^e$ we can write $K = \kappa \frac{\sqrt{\log N}}{(\log\log N)^2}$ for some $\kappa > 0$. If $\kappa \leq 1$ we have $\log K \leq \frac{1}{2}\log\log N$, so (8.10) holds if moreover

$$N \geq \exp\left(\gamma \kappa^2 \frac{1}{2^4} \log N\right),$$

which is satisfied when $\kappa \leq \kappa_F := \min\{1, 4/\sqrt{\gamma}\}$. □

Proof of Theorem 1.2. Let $s = 4$, let $F(\mathbf{x})$ be as in Theorem 1.2 and let $N, K \in \mathbb{N}$ with $K \geq 2$. As in the previous case, we deduce from Theorem 8.8 that there is some constant $\gamma > 0$ independent of N and K such that appropriate gaps of length K esist when the inequality

$$N \geq \exp(\exp(\exp(\gamma K \log K))) \tag{8.11}$$

holds. When $N \geq e^{e^{e^e}}$ we can write $K = \kappa \frac{\log\log\log N}{\log\log\log\log N}$ for some $\kappa > 0$. If $\kappa \leq 1$ we have $\log K \leq \log\log\log\log N$, so (8.11) holds if moreover

$$N \geq \exp(\exp(\exp(\gamma\kappa \log\log\log N))),$$

which is satisfied when $\kappa \leq \kappa_F = \min\{1, 1/\gamma\}$. □

Remark 8.9. For some diagonal forms a more elementary proof can be given, i.e. not involving Hecke characters and Chebotarev's theorem for abelian extensions. For example for the biquadratic diagonal form $F(\mathbf{x}) = x_1^4 + x_2^4 + x_3^4 + x_4^4$ we notice that $K_{F,q} = -5$ for all $q \equiv 5 \pmod 8$. Since $\operatorname{Re} H_{F,q} \leq 2$, we see that Proposition 8.1 holds with $\beta = 3$ and $\mathcal{P}_F = \{q \text{ prime} : \ q \equiv 5 \pmod 8\}$, even without referring to the equidistribution of $H_{F,q}$.
We can avoid the reference to an equidistribution result for cubic forms as well. For example for $F(\mathbf{x}) = x_1^3 + x_2^3 + x_3^3$ we can prove that if $p \equiv 1 \pmod 3$ and m is a nonzero noncubic residue class modulo p (i.e. $\chi_{3,p}(m) \notin \{0, 1\}$), then $r_F(m, p) \leq p^2 - 3p + 2\sqrt{p}$. This is enough to imply the existence of unbounded gaps, though with smaller size compared to Theorem 1.1. With this alternative approach, it helps to observe that for all primes p large enough we can find K consecutive residue classes modulo p at which $\chi_{3,p}$ assumes any given value, see [23].

Acknowledgements

I would like to thank my supervisor Damien Roy for his encouragement and for his many comments on this work. Among the many people to whom I had the pleasure to speak about this project, I am specially grateful to Simon Rydin Myerson and Marc Hindry for their interesting remarks. I also thank Martin Rivard-Cooke for having introduced me to the problem of gaps for $F(\mathbf{x}) = x_1^3 + x_2^3 + x_3^3$ and Daniel Fiorilli for his comments on the content of the paper. For their help in finding references, I thank Daniel Fiorilli, Gerry Myerson and the user EFinat-S from Mathoverflow. I thank Kam Hung Yau for spotting some typos in a previous version of the paper. I thank an anonymous referee for valuable suggestions, especially concerning the introduction and section 5. Finally, I thank Francesco Veneziano for discussing with me the problem of gaps in the case of degree two: the strategy followed in this article was designed as an attempt to generalize our computations to higher degree. This work was supported in part by a full International Scholarship from the Faculty of Graduate and Postdoctoral Studies of the University of Ottawa and by NSERC.

References

[1] J.-H. Ahn and S.-H. Kwon. Some explicit zero-free regions for Hecke L-functions. *Journal of Number Theory*, 145:433–473, 2014.

[2] B. C. Berndt, R. J. Evans, and K. S. Williams. *Gauss and Jacobi sums*. Canadian Mathematical Society Series of Monographs and Advanced Texts. John Wiley & Sons, Inc., New York, 1998. A Wiley-Interscience Publication.

[3] R. Bradshaw. Arithmetic properties of values of lacunary series. Master's thesis, University of Ottawa, 2013.

[4] M. Coleman. A zero-free region for the hecke l-functions. *Mathematika*, 37(02):287–304, 1990.

[5] S. Daniel. On gaps between numbers that are sums of three cubes. *Mathematika*, 44(1):1–13, 1997.

[6] J. Deshouillers, F. Hennecart, and B. Landreau. Sums of powers: an arithmetic refinement to the probabilistic model of Erdős and Rényi. *Acta Arithmetica*, 85(1):13–33, 1998.

[7] J.-M. Deshouillers, F. Hennecart, and B. Landreau. On the density of sums of three cubes. In *Algorithmic number theory*, volume 4076 of *Lecture Notes in Comput. Sci.*, pages 141–155. Springer, Berlin, 2006.

[8] P. Erdős and P. Turán. On a problem in the theory of uniform distribution. I. *Nederl. Akad. Wetensch., Proc.*, 51:1146–1154, 1948. = *Indagationes Math.* 10:370–378, 1948.

[9] L. Ghidelli. Arithmetic properties of cubic and biquadratic theta series. *Preprint*, 2019.

[10] A. Granville. Unexpected irregularities in the distribution of prime numbers. In *Proceedings of the International Congress of Mathematicians*, volume 1, pages 388–399, Basel, 1995. Birkhäuser.

[11] G. H. Hardy and E. M. Wright. *An introduction to the theory of numbers*. The Clarendon Press, Oxford University Press, New York, fifth edition, 1979.

[12] D. Heath-Brown and S. Patterson. The distribution of Kummer sums at prime arguments. *Journal für die reine und angewandte Mathematik*, 310:111–130, 1979.

[13] E. Hecke. Eine neue Art von Zetafunktionen und ihre Beziehungen zur Verteilung der Primzahlen. *Math. Z.*, 6(1-2):11–51, 1920.

[14] C. Hooley. On some topics connected with Waring's problem. *Journal für die reine und angewandte Mathematik*, 369:110–153, 1986.

[15] C. Hooley. On Hypothesis K^* in Waring's problem. In *Sieve methods, exponential sums, and their applications in number theory (Cardiff, 1995)*, volume 237 of *London Math. Soc. Lecture Note Ser.*, pages 175–185. Cambridge Univ. Press, Cambridge, 1997.

[16] K. Ireland and M. Rosen. *A classical introduction to modern number theory*, volume 53 of *Colloquium Publications*. American Mathematical Society, 2004.

[17] H. Iwaniec and E. Kowalski. *Analytic number theory*, volume 53 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2004.

[18] H. Koch. *Algebraic number theory*. Springer-Verlag, Berlin, 1997. Reprint of the 1992 translation.

[19] I. P. Kubilyus. On some problems of the geometry of prime numbers. (russian). *Mat. Sbornik N.S.*, 31(73)(3):507–542, 1952.

[20] J. C. Lagarias and A. M. Odlyzko. Effective versions of the Chebotarev density theorem. In *Algebraic number fields: L-functions and Galois properties (Proc. Sympos., Univ. Durham, Durham, 1975)*, pages 409–464. Academic Press, London, 1977.
[21] E. Landau. Über die Einteilung der positiven ganzen Zahlen in vier Klassen nach der Mindestzahl der zu ihrer additiven Zusammensetzung erforderlichen Quadrate. *Archiv der Mathematik und Physik*, 1908.
[22] S. Lang. *Algebraic Number Theory*, volume 110. Springer Science & Business Media, 1994.
[23] V. Lev (http://mathoverflow.net/users/9924/seva). Consecutive non-quadratic residues. MathOverflow. URL:http://mathoverflow.net/q/161279 (version: 2014-03-28).
[24] T. Mitsui. Generalized prime number theorem. In *Japanese journal of mathematics: transactions and abstracts*, volume 26, pages 1–42. The Mathematical Society of Japan, 1956.
[25] C. J. Moreno. Sur le problème de Kummer. *L'Enseignement Mathématique*, 20(2):45–51, 1974.
[26] M. Murty and V. Murty. *Non-vanishing of L-functions and applications*. Modern Birkhäuser Classics. Springer Basel, 2012.
[27] J. Neukirch. *Class field theory*, volume 280 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1986.
[28] R. Odoni. The Farey density of norm subgroups in global fields (I). *Mathematika*, 20(2):155–169, 1973.
[29] S. Patterson. The distribution of general Gauss sums and similar arithmetic functions at prime arguments. *Proceedings of the London Mathematical Society*, s3-54(2):193–215, 1987.
[30] I. Richards. On the gaps between numbers which are sums of two squares. *Advances in Mathematics*, 46(1):1–2, 1982.
[31] J.-P. Serre. *Lectures on $N_X(p)$*, volume 11 of *Chapman & Hall/CRC Research Notes in Mathematics*. CRC Press, Boca Raton, FL, 2012.
[32] R. C. Vaughan and T. D. Wooley. Waring's problem: a survey. *Number theory for the millennium 3*, pages 301–340, 2002.
[33] A. Weil. Jacobi sums as "Grössencharaktere". *Transactions of the American Mathematical Society*, 73:487–495, 1952.

Chapter 7

On gaps between sums of four fourth powers

ON GAPS BETWEEN SUMS OF FOUR FOURTH POWERS

LUCA GHIDELLI

Abstract. We prove that for almost all N there is a sum of four fourth powers in the interval $(N - N^\gamma, N]$, for all $\gamma > 4059/16384 = 0.24774...$

Contents

1. Introduction

For every $n \in \mathbb{N}$ there is some natural number $x < n^{1/4}$ such that $n - x^4 = O(x^3) = O(n^{3/4})$. If we repeat this procedure we find that for all $n \in \mathbb{N}$ there exist $x_1, x_2, x_3, x_4 \in \mathbb{N}$ such that $x_1^4 + \cdots + x_4^4 = n + O(n^\gamma)$ with $\gamma = (3/4)^4 \approx 0.3164$. In this paper we show that the exponent γ can be reduced if we require the above statement to hold only for *almost all* $n \in \mathbb{N}$. This is motivated by a forthcoming article of the author [9], in which we study arithmetic properties of special values of "cubic" and "biquadratic" theta series. In fact, the arguments of that paper require that almost all intervals of the form $(n - n^\gamma, n]$, for some $\gamma < 0.25$, contain a sum of four fourth powers. Using the circle method, Daniel [3] studied a similar problem in regard to sums of three cubes. Following his approach we are able to prove the following statement.

Theorem 1.1. Define $\gamma_0 := 4059/16384 \approx 0.24774$ and let $\gamma > \gamma_0$. Then for almost all $n \in \mathbb{N}$ (in the sense of natural density) there is a sum of four fourth powers in the interval $(n - n^\gamma, n]$.

To put this theorem in perspective, we now survey the relevant literature on sums of four fourth powers and sums of three cubes. First, we know from a paper of Davenport [4] that there are $\gg N^{\alpha_4}$ distinct sums of four fourth powers up to N, for $\alpha_4 := 331/412 \approx 0.803398$: this means that the average gap between sums of fourth powers is at most of order $\ll N^{1-\alpha_4} \approx N^{0.197}$. However, Davenport's result does not measure how uniformly the sums of four fourth powers distribute on the number line, so it does not imply that almost all gaps have at most this size. In

Date: December 10, 2019.
2010 *Mathematics Subject Classification.* Primary 11P05, 11P55; Secondary 11B05.

fact some probabilistic models [5, 7] suggest that the sums of four fourth powers, and more generally sums of k perfect k-th powers for $k \geq 3$, should have positive natural density. In particular the gaps between these numbers are conjectured to have bounded average size. However, previous work of the author [8] shows that there do exist arbitrarily large gaps between numbers that can be written as sums of four fourth powers. In fact we also showed that a positive proportion of the intervals $(n - \psi(n), n]$ does not contains sums of fourth powers, if $\psi(n)$ grows to infinity sufficiently slowly. If we trust the probabilistic models, we should in fact expect this last statement to hold for $\psi(n) \asymp \log n / \log\log n$.

The situation for sums of three cubes is similar, and has been considered more extensively in the literature. A "greedy argument" as the one in the opening of this introduction shows that for all $n \in \mathbb{N}$ there exist $x_1, x_2, x_3 \in \mathbb{N}$ such that $x_1^3 + x_2^3 + x_3^3 = n + O(n^\gamma)$, with $\gamma = 8/27 \approx 0.296$. The aforementioned paper of Daniel [3] proves instead that almost all gaps between sums of three cubes up to N have length $O(N^\gamma)$, for all $\gamma > 17/108 \approx 0.1574$. For the number of sums of three cubes up to N, the current record is due to Wooley [19], who proves that there are $\gg N^{\alpha_3}$ of them, with $\alpha_3 \approx 0.916862$; this means that on average the gaps between them have order $\ll N^{1-\alpha_3} \approx N^{0.083}$. As we wrote above, it is expected on the basis of probabilistic models that the sums of three cubes have positive density in the set of natural numbers. This expectation is further discussed in [13] and is supported by numerical results [6]. It is also known that there are $\gg N^{1-\epsilon}$ sums of three cubes up to N, for every $\epsilon > 0$, conditionally on analytic conjectures involving certain L-functions [11, 12, 14]. However, if the sums of three cubes have positive natural density, they do not lie uniformly on the number line. In fact, as we prove in [8], there exists a constant $\kappa > 0$ so that, for $\psi(n) := \kappa\sqrt{\log n}(\log\log n)^{-2}$, a positive proportion of the intervals $(n - \psi(n), n]$ does not contain sums of three cubes. More generally, our result belongs to the vast literature on Waring's problem, that is the study of those numbers that can be written as sums of perfect powers. The interested reader is referred to the survey of Vaughan and Wooley [18].

We now provide some details on the basic ideas of this paper. A classical approach known as "diminishing ranges" due to Hardy and Littlewood [10], consists in counting those sums $x_1^4 + \cdots + x_4^4$ in an interval $(n - Y, n]$ whose summands have a prescribed size $x_j^4 \asymp P_j^4$. More precisely, we fix $\mathbf{P} := (P_1, P_2, P_3, P_4, Y) \in \mathbb{R}_+^5$ with

$$P_j^{3/4} \leq P_{j+1} \leq P_j \qquad (1 \leq j \leq 3) \tag{1.1}$$

and let $R(n) = R(n, \mathbf{P})$ denote the number of solutions to the equation

$$n = x_1^4 + x_2^4 + x_3^4 + x_4^4 + y \tag{1.2}$$

subject to

$$0 < y \leq Y, \quad \tfrac{1}{2}P_i < x_i \leq P_i \quad (1 \leq i \leq 4). \tag{1.3}$$

If $n \asymp P_1^4$, say $n \in (N/2, N]$ with $N = P_1^4$, then we expect that, at least on average, $R(n) \asymp Y P_1^{-3} P_2 P_3 P_4$, because there are $\asymp N$ choices for the parameter n and $\asymp Y P_1 P_2 P_3 P_4$ choices for the values of the variables of eq. (1.2). In fact, using the circle method [17] of Hardy and Littlewood we prove the following analog of the main lemma in [3].

Theorem 1.2. Let γ_0 be as in Theorem 1.1 and let $\gamma_1 := 4992/16384 \approx 0.3046$. Given $N > 0$ and $\gamma_0 < \gamma \leq \gamma_1$, we let $Y := N^\gamma$, $P = P_1 := \sqrt[4]{N}$ and $P_{j+1} = P_j^{13/16}$

for $1 \le j \le 3$. Then for each $\epsilon > 0$ we have

$$\sum_{\frac{1}{2}N < n \le N} \left| R(n) - \bar{R}(n) \right|^2 \ll_\epsilon Y N^{1-\gamma_0+\epsilon}, \tag{1.4}$$

where the implied constant depends only on ϵ, and $\bar{R}(n) := \frac{1}{32} Y P_2 P_3 P_4 n^{-3/4}$.

From this quantitative result one may deduce nontrivial moment estimates for the the size of gaps between sums of four fourth powers, as in [3, Corollary 2] or [2, Theorem 1.2]. Moreover, as we will show in the next section, Theorem 1.2 implies Theorem 1.1. We also claim more generally that, with essentially the same strategy and some more work, one may possibly show that in almost every interval of the form $(N - N^\gamma, N]$ there is a number $m = x_1^k + \cdots + x_h^k$ that can be written as the sum of $h \ge 2$ perfect k-th powers, provided that $k \ge 3$ and $\gamma > \gamma_0(h,k)$, where

$$\gamma_0(h,k) := 1 - \frac{1}{k}(1 + \theta_k + \theta_k^2 + \cdots + \theta_k^{h-1}), \tag{1.5}$$

with

$$\theta_k := 1 - \frac{1}{k} + \frac{1}{k 2^{k-2}}.$$

We notice that $\gamma_0(4,4) = 4059/16384$ is the exponent that appears in Theorem 1.1 and that $\theta_4 = 13/16$ is the exponent we use for diminishing the ranges in Theorem 1.2. Therefore our result solves the case $h = k = 4$ while Daniel [3] deals with the case $h = k = 3$. Recently, a paper of Brüdern and Wooley [2] has settled the case $h = 2$ for all $k \ge 3$. Even though the treatment of only two variables simplifies part of the argument (e.g. the final induction on the number of variables becomes trivial), the case treated by Brüdern and Wooley should be considered as the hardest one. In fact their paper introduces some technical modifications to the original strategy of Daniel, which are unnecessary here.

In addition to the results that we have just mentioned, a few more remarks are in order with respect to the general claim enunciated above. The first is that stronger statements are known to be true if h is somewhat larger than k. For example, we know that all natural numbers can be written as a sum of h k-th powers, if h is large enough [18]. Secondly the claim is nontrivial in general: in comparison the greedy argument produces the exponent $\gamma(h,k) = (1 - 1/k)^h$, which is the same as eq. (1.5), with θ_k replaced by the smaller $\theta_k' := 1 - 1/k$. Finally, the recent progress on the Vinogradov mean value theorem [1, 15, 20] should make it possible to replace θ_k with a larger value, if k is large enough; see the note in the introduction of [2] for a more precise remark on this matter.

In closing, let us briefly illustrate the main ingredients in the proof of Theorem 1.2. First the number $R(n)$ is rewritten, by Fourier analysis, as an integral of an exponential sum. Then Bessel's inequality is used to produce an integral formula that estimates from above the left-hand side of (1.4). A characterstic feature of Daniel's approach is that this part of the proof (sections 3 and 4) is performed in conjunction with a triple application of the circle method,[(1)] where only one major arc centered around the origin is considered. The upper bound that results from this preliminary phase is then finally estimated using a more classical application of the circle method and an induction on the number of variables of the underlying

[(1)] Corresponding to the three pairs of integrals $R \sim U$, $S \sim V$ and $T \sim W$ introduced in the proof.

diophantine equations, to produce the expression in the right-hand side of (1.4). Technically, the minor arcs are treated with a version [16, Lemma 1] of the Weyl differencing inequality [17, Lemma 2.4], while the major arcs are treated with classical estimates mostly due to Vaughan [17, Chapter 4]. In conclusion, we express our contentment in noticing the fortuitous happenstance: that this approach produces an exponent $\gamma_0 \approx 0.24774$, that is just barely good enough for our original purpose.

Acknowledgements. I would like to thank my supervisor Damien Roy for his steady encouragement, his careful reading of this manuscript and for his many comments and suggestions. This work was supported in part by a full International Scholarship from the Faculty of Graduate and Postdoctoral Studies of the University of Ottawa and by NSERC.

2. Heuristics and quantitative results

In this section we comment on the statement of Theorem 1.2 and its consequences regarding the size of gaps between sums of four fourth powers.

2.1. **Choice of parameters and notation.** In the remainder of the article we write $N = P^4$ and $Y = P^{4\gamma}$, where

$$\gamma \in (\tfrac{4059}{16384}, \tfrac{4992}{16384}]$$

and P is some parameter that we let grow to infinity. We also let

$$P_1 = P^{\frac{4096}{4096}} \qquad P_2 = P^{\frac{3328}{4096}} \qquad P_3 = P^{\frac{2704}{4096}} \qquad P_4 = P^{\frac{2197}{4096}}$$

as in Theorem 1.2 so that $P_{j+1} = P_j^{13/16}$ for $j = 1, 2, 3$. The inequality $\gamma > \frac{4059}{16384}$ implies that

$$N = o(Y P_1 P_2 P_3 P_4)$$

which is crucial in the approach of this paper. The hypothesis $\gamma \leq \frac{4992}{16384}$ is imposed only for technical reasons, as it ensures that

$$Y^{-2} \geq {P_2}^{-3}. \tag{2.1}$$

In fact the validity of this inequality simplifies some proofs, e.g. that of Proposition 4.7. We denote $\mathbf{P} = (P_1, P_2, P_3, P_4, Y)$ and define $R(n) = R(n, \mathbf{P})$ accordingly, see section 1. Throughout the paper we make various estimates in terms of the parameter P, but we also write the results, when possible, in a way that makes explicit the dependence on the choice of $P_1, \dots, P_4$. As usual, the notation $A \ll B$ means that $|A| \leq cB$ for some absolute $c > 0$. The contributions of terms that are logarithmic in P or anyway asymptotically smaller than any positive power of P will systematically be collected into a "P^ϵ term". We will write $A \ll_\epsilon P^\epsilon B$ to mean that $|A| \leq cP^\epsilon B$, for every $\epsilon > 0$ and for some $c = c(\epsilon) > 0$ depending only on ϵ.

2.2. **The heuristic expected value of $R(n)$.** The diminished ranges (1.3) for the variables of (1.2) reduce the number of sums of fourth powers at our disposal, and so enlarge the gaps between them. However the advantage is that those particular sums of powers are more easily controlled, so that it is possible to estimate $R(n)$ as in Theorem 1.2. The expected average value of $R(n)$, given by the formula

$$\bar{R}(n) := \tfrac{1}{32} Y P_2 P_3 P_4 n^{-3/4}$$

is heuristically obtained as follows. Suppose that P is large and that $n \asymp P^4$ is restricted to an interval $n \in (n_0, n_1]$ with $\Delta n := n_1 - n_0 = o(P_1^4)$ and $Y \le P_2^4 = o(\Delta n)$. Then every solution to eq. (1.2), constrained by (1.3), also satisfies

$$n_1^{1/4} \ge x_1 > (n_0 - 4P_2^4)^{1/4} =: n_1^{1/4} - \Delta x \tag{2.2}$$

with $\Delta x \approx \frac{1}{4}\Delta n \cdot n^{-3/4}$. There are Δn choices for the parameter $n \in (n_0, n_1]$ and $\approx 2^{-3}\Delta x P_2 P_3 P_4 Y$ choices of x_j and y constrained by (1.3) and (2.2), hence we expect that $R(n) \approx \bar{R}(n)$ with $\bar{R}(n)$ as above. We notice en passant that $N^{-3/4}P_2P_3P_4 = N^{-\gamma_0}$, where $\gamma_0 = 4059/16384$, so

$$\bar{R}(n) \asymp Y N^{-\gamma_0}. \tag{2.3}$$

Therefore we also heuristically expect that a typical $n \in (N/2, N]$ satisfies $R(n) \ge 1$, as soon as Y is somewhat larger than N^{γ_0}.

2.3. **Bounding the number of large gaps.** We now show how to prove from Theorem 1.2 that the gaps of size N^γ with $\gamma > \gamma_0 := 4059/16384$ are rare. For every $\gamma > 0$ we denote by $K'(N, N^\gamma)$ the number of $n \in (N/2, N]$ with the property that no element of the interval $(n - N^\gamma, n]$ is a sum of four fourth powers.

Theorem 2.1. Let γ_0 and γ_1 be as in Theorem 1.2. Then

$$K'(N, N^\gamma) \ll N^{1-\xi} \tag{2.4}$$

for every $\gamma > \gamma_0$ and all $\xi < \min\{\gamma_1 - \gamma_0, \gamma - \gamma_0\}$.

Proof. If $\gamma \le \gamma_1$ we may apply Theorem 1.2. Let $K''(N, \mathbf{P})$ denote the number of $n \in (N/2, N]$ for which $R(n) = R(n, \mathbf{P}) = 0$. For each of those n we have $\big|R(n) - \bar{R}(n)\big| = \bar{R}(n) \ge \bar{R}(N)$, hence

$$\sum_{\frac{1}{2}N < n \le N} \big|R(n) - \bar{R}(n)\big|^2 \gg K''(N, \mathbf{P}) \cdot \bar{R}(N)^2. \tag{2.5}$$

It is clear that $K'(N, N^\gamma) \le K''(n, \mathbf{P})$ because whenever the interval $(n - Y, n]$ is empty of sums of four fourth powers, where $Y = N^\gamma$, then $R(n) = 0$. By eqs. (1.4), (2.3) and (2.5) we get

$$K'(N, N^\gamma) \ll \bar{R}(N)^{-2} \sum_{\frac{1}{2}N < n \le N} \big|R(n) - \bar{R}(n)\big|^2 \ll Y^{-1} N^{1+\gamma_0+\epsilon}$$

for every $\epsilon > 0$. This gives eq. (2.4) if $\gamma_0 < \gamma \le \gamma_1$. If $\gamma > \gamma_1 \approx 0.3046$ then we simply use the inequality $K'(N, N^\gamma) \le K'(N, N^{\gamma_1})$. □

We remark that for $\gamma > (3/4)^4 \approx 0.3164$ one in fact has $K'(N, N^\gamma) = 0$ if $N \gg 1$, by the greedy algorithm mentioned in the introduction. We now show that Theorem 1.1 is follows from Theorem 2.1.

Proof of Theorem 1.1. Fix $\gamma > \gamma_0$ and let $K_\gamma(N)$ count the natural numbers $n \le N$ such that no element of the interval $(n - n^\gamma, n]$ is a sum of four fourth powers. Take some $\gamma' \in (\gamma_0, \gamma)$ and let N_0 be such that $N^{\gamma'} \le (N/2)^\gamma$ for all $N \ge N_0$. Then for every real number $N \ge N_0$ we have

$$K_\gamma(N) \le N_0 + \sum_{k=0}^{\lfloor \log_2 N/N_0 \rfloor} K'(N/2^k, (N/2^k)^{\gamma'}).$$

Then by (2.4) we get

$$K_\gamma(N) \ll N^{1-\xi} \sum_{k=0}^{\infty} (2^{-(1-\xi)})^k \ll_\xi N^{1-\xi},$$

where ξ is any positive number with $\xi + \gamma_0 < \min\{\gamma_1, \gamma'\}$. In particular, we have that $K_\gamma(N) = o(N)$ as $N \to \infty$. □

3. On the expected value of $R(n)$

In this section we rewrite the number $R(n)$ in a way that makes it amenable to be studied with analytic methods. Then we give a first estimate of the deviation $R(n) - \bar{R}(n)$ via a partial application of the circle method, with only one major arc centered at zero.

3.1. **Integral representation and Weyl sums.** We denote by $e(\xi) := e^{2\pi i \xi}$ the normalized complex exponential function, considered as an additive character of $\mathbb{R}/\mathbb{Z}$. By the "orthogonality property" we mean the well-known fact that for all $m \in \mathbb{Z}$ we have

$$\int_{\mathbb{R}/\mathbb{Z}} e(m\alpha) d\alpha = \begin{cases} 1 & \text{if } m = 0 \\ 0 & \text{if } m \neq 0. \end{cases}$$

By orthogonality we can rewrite $R(n) = R(n, \mathbf{P})$ as follows

$$R(n) := \int_{\mathbb{R}/\mathbb{Z}} \sum_{\substack{y, x_1, \dots, x_4 \\ \frac{1}{2}P_j < x_j \le P_j \\ 0 \le y < Y}} e((x_1^4 + x_2^4 + x_3^4 + x_4^4 + y - n)\alpha) d\alpha$$

$$= \int_{\mathbb{R}/\mathbb{Z}} f_1 f_2 f_3 f_4 g e(-n\alpha) d\alpha, \tag{3.1}$$

where $f_i = f(\alpha, P_i)$, $g = g(\alpha, Y)$ are given by the following Weyl exponential sums

$$f(\alpha, X) := \sum_{\frac{1}{2}X < x \le X} e(\alpha x^4)$$

$$g(\alpha, Y) := \sum_{0 \le y < Y} e(\alpha y).$$

We observe that $g(\alpha, Y)$ is the sum of a geometric progression, therefore we have

$$g(\alpha, Y) = \frac{e(\alpha(Y + O(1))) - 1}{e(\alpha) - 1}.$$

From this formula, we easily get the following estimates for the function g.

Lemma 3.1.

$$g(\alpha, Y) \le Y \qquad \text{for all } \alpha,$$

$$g(\alpha, Y) \ll \|\alpha\|^{-1} \qquad \text{for all } \alpha,$$

$$g(\alpha, Y) = Y + O(1) \qquad \text{if } \|\alpha\| \le Y^{-2}. \tag{3.2}$$

where $\|\alpha\|$ denotes the distance of $\alpha \in \mathbb{R}/\mathbb{Z}$ from 0.

The estimates contained in Lemma 3.1 imply that the integrand in eq. (3.1) is approximately equal to $e(-\alpha n) f_1 f_2 f_3 f_4 Y$ when α is close to 0, while it becomes "small" when α is bounded away from 0.

3.2. **An approximation.** Under the assumption $\alpha \approx 0$ it is possible to approximate the Weyl sum $f(\alpha, X)$ with its "mollification"

$$\nu(\alpha, X) := \sum_{\frac{1}{16}X^4 < z \le X^4} \tfrac{1}{4} z^{-3/4} e(\alpha z),$$

which is a weighted exponential sum that involves linear phases instead of biquadratic ones. From the book of Vaughan [17] we retrieve the following estimates.

Lemma 3.2.

$$\nu(\alpha, X) \ll X \qquad \text{for all } \alpha, \tag{3.3}$$

$$\nu(\alpha, X) \ll X^{-3} \|\alpha\|^{-1} \qquad \text{for all } \alpha, \tag{3.4}$$

$$f(\alpha, X) \ll X \qquad \text{for all } \alpha, \tag{3.5}$$

$$f(\alpha, X) = \nu(\alpha, X) + O(1) \qquad \text{if } \|\alpha\| \le \tfrac{1}{8} X^{-3}. \tag{3.6}$$

Proof. The estimates (3.3) and (3.4) are a restatement of [17, Lemma 6.2]. The estimate (3.5) is trivial because $f(\alpha, X)$ is a sum of $O(X)$ exponentials. Finally, (3.6) follows from [17, Lemma 6.1] with $q = 1$. □

Then alongside $f_1, \dots, f_4$ we consider the mollified Weyl sums

$$\nu_j := \nu(\alpha, P_j).$$

From (3.6) we have that the approximation $f_j \approx \nu_j$ is admissible, up to an error of $O(1)$, on the interval $\mathfrak{B}_0^{(j)} \subseteq \mathbb{R}/\mathbb{Z}$ given by

$$\mathfrak{B}_0^{(j)} = \{\alpha : \|\alpha\| \le \tfrac{1}{8} P_j^{-3}\}. \tag{3.7}$$

The complement of (3.7) in $\mathbb{R}/\mathbb{Z}$ will be denoted by $\mathfrak{B}_1^{(j)}$. In the range of small $\|\alpha\|$ we also have $g \approx Y$: more precisely by (3.2) and (2.1) we have that $g - Y$ is bounded by an absolute constant on $\mathfrak{B}_0^{(1)}$ and $\mathfrak{B}_0^{(2)}$. Then, we consider the following integral

$$U(n) := Y \int_{\mathbb{R}/\mathbb{Z}} e(-n\alpha) \nu_1 \nu_2 \nu_3 \nu_4 \, d\alpha \tag{3.8}$$

The integrand in eq. (3.8) is approximately equal to $e(-\alpha n) f_1 f_2 f_3 f_4 Y$ when α is close to 0, and it is small when α is bounded away from 0. Thus, by what we said at the end of the previous paragraph, we heuristically expect that $U(n) \sim R(n)$. We now show that $U(n)$ is in fact close to the expected value $\bar{R}(n)$, up to an admissible error.

Proposition 3.3. The following estimate holds uniformly for $n \in (\frac{1}{2}N, N]$:

$$U(n) - \bar{R}(n) \ll Y P_1^{-7} P_2^5 P_3 P_4 = Y P^{-\frac{7131}{4096}}.$$

Proof. By the definitions and by orthogonality, we have

$$U(n) = Y \sum_{\substack{\frac{1}{16}P_j^4 < z_j \le P_j^4 \\ z_1 + z_2 + z_3 + z_4 = n}} \tfrac{1}{256} (z_1 z_2 z_3 z_4)^{-3/4}. \tag{3.9}$$

Since $P_j = o(P_1)$ for each $2 \le j \le 4$, we have the inequality

$$P_2^4 + P_3^4 + P_4^4 < \left(\tfrac{1}{2} - \tfrac{1}{16}\right) P_1^4$$

for all P large enough. Since moreover $\frac{1}{2}P_1^4 < n \leq P_1^4$, we have for every n, z_2, z_3, z_4 in the appropriate range that

$$\tfrac{1}{16}P_1^4 < n - z_2 - z_3 - z_4 \leq P_1^4.$$

In other words in (3.9) we can safely express z_1 in terms of the other variables:

$$U(n) = \tfrac{1}{256} \sum_{\substack{z_2,z_3,z_4 \\ \frac{1}{16}P_j^4 < z_j \leq P_j^4}} (z_2 z_3 z_4)^{-3/4} n^{-3/4} \left(1 - \frac{z_2+z_3+z_4}{n}\right)^{-3/4}.$$

We observe that $z_2 + z_3 + z_4 = O(P_2^4)$ and that

$$\sum_{\frac{1}{16}P_j^4 < z_j \leq P_j^4} \tfrac{1}{4} z_j^{-3/4} = \int_{\frac{1}{16}P_j^4}^{P_j^4} \tfrac{1}{4} t^{-3/4} dt + O(P_j^{-3}),$$

which is equal to $\frac{1}{2}P_j(1 + O(P_j^{-4}))$. Since $P_2^{-4} \ll P_3^{-4} \ll P_4^{-4} \ll P_2^4 P_1^{-4}$ we conclude that

$$U(n) = \tfrac{1}{32} Y P_2 P_3 P_4 n^{-3/4}(1 + O(P_2^4 P_1^{-4})).$$

□

3.3. **First application of the circle method.** For every $n \in \mathbb{N}$ and every measurable set $\mathfrak{B} \subseteq \mathbb{R}/\mathbb{Z}$ (with respect to the natural Lebesgue-Haar measure) we define

$$R(n, \mathbf{P}, \mathfrak{B}) := \int_{\mathfrak{B}} e(-n\alpha) f_1 f_2 f_3 f_4 g d\alpha,$$
$$U(n, \mathbf{P}, \mathfrak{B}) := Y \int_{\mathfrak{B}} e(-n\alpha) \nu_1 \nu_2 \nu_3 \nu_4 d\alpha.$$

Since $Y^{-2} \geq \frac{1}{8}P_2^{-3} \geq \frac{1}{8}P_1^{-3}$ by (2.1), the approximations $g = Y + O(1)$ and $f_j = \nu_j + O(1)$ for $1 \leq j \leq 4$ are valid when $\alpha \in \mathfrak{B}_0^{(1)}$, where

$$\mathfrak{B}_0^{(1)} = [-\tfrac{1}{8}P_1^{-3}, \tfrac{1}{8}P_1^{-3}].$$

We define $\mathfrak{B}_1^{(1)}$ to be its complement so that we have a partition $\mathbb{R}/\mathbb{Z} = \mathfrak{B}_0^{(1)} \sqcup \mathfrak{B}_1^{(1)}$. Then we let $R_i(n) := R(n, \mathbf{P}, \mathfrak{B}_i^{(1)})$ for $i \in \{1, 0\}$ so that $R(n) = R_0(n) + R_1(n)$. In the remaining part of this section, we are going to prove that

$$\left|R(n) - \bar{R}(n)\right| \leq E_R + |R_1(n)| \tag{3.10}$$

where E_R is an error term satisfying the following estimate

$$E_R \ll_\epsilon P^\epsilon Y P^{-\frac{5083}{4096}} \approx Y P^{-1.240967}. \tag{3.11}$$

More precisely, we decompose $U(n) = U_0(n) + U_1(n)$ as we did for $R(n)$ via $U(n, \mathbf{P}, \mathfrak{B})$ and the partition $\mathbb{R}/\mathbb{Z} = \mathfrak{B}_0^{(1)} \sqcup \mathfrak{B}_1^{(1)}$. Then by the triangular inequality (3.10) holds with

$$E_R := |R_0(n) - U_0(n)| + |U_1(n)| + \left|U(n) - \bar{R}(n)\right|.$$

The third absolute value was estimated in Proposition 3.3; the other two terms are treated in the following propositions.

Proposition 3.4.

$$U_1(n) \ll YP_2^{-3}P_3P_4 = YP^{-\frac{5083}{4096}}. \tag{3.12}$$

Proof. By (3.4) applied to ν_1, ν_2 and (3.3) applied to ν_3, ν_4 we have

$$U_1(n) \ll YP_1^{-3}P_2^{-3}P_3P_4 \int_{\mathfrak{B}_1^{(1)}} \|\alpha\|^{-2} d\alpha$$

and so (3.12) follows from an elementary computation. □

Proposition 3.5.

$$R_0(n) - U_0(n) \ll_\epsilon P^\epsilon YP_2^{-3}P_3P_4 = YP^{-\frac{5083}{4096}+\epsilon}. \tag{3.13}$$

Proof. Since $P_j^3 \le P_1^3$ for all j and since $Y^2 \le 8P_1^3$, we have by (3.6) and (3.2)

$$R_0(n) - U_0(n) \ll \int_{\mathfrak{B}_0^{(1)}} (\mu_1\mu_2\mu_3\mu_4 + Y(\mu_1\mu_2\mu_3 + \mu_1\mu_2\mu_4 + \mu_1\mu_3\mu_4 + \mu_2\mu_3\mu_4))\, d\alpha,$$

where $\mu_j := \max\{|\nu_j|, 1\}$. We use (3.3), i.e the trivial estimate $\mu_j \ll P_j$, on the factors with higher indices, to obtain

$$R_0(n) - U_0(n) \ll P_2P_3P_4 \left(1 + \tfrac{Y}{P_2} + \tfrac{Y}{P_3} + \tfrac{Y}{P_4}\right) \int_{\mathfrak{B}_0^{(1)}} \mu_1 d\alpha + YP_3P_4 \int_{\mathfrak{B}_0^{(1)}} \mu_2 d\alpha.$$

Since $Y \ge P_4$ the factor that multiplies the first integral is $\asymp YP_2P_3$. Since $\mu_j \le |\nu_j| + 1$ we can rewrite the last estimate as

$$R_0(n) - U_0(n) \ll YP_2P_3 \cdot P_1^{-3} + YP_2P_3 \int_0^1 |\nu_1|\, d\alpha + YP_3P_4 \int_0^1 |\nu_2|\, d\alpha.$$

Then eq. (3.13) follows from the following lemma, that we state separately for future reference, and the inequality $P_2P_1^{-3} = P^{-\frac{8960}{4096}} < P^{\frac{7787}{4096}} = P_4P_2^{-3}$. □

Lemma 3.6.

$$\int_{\mathbb{R}/\mathbb{Z}} |\nu_j|\, d\alpha \ll P_j^{-3} \log P_j. \tag{3.14}$$

Proof. We estimate ν_j with

$$\nu_j \ll \begin{cases} P_j, & \text{if } \|\alpha\| \le P_j^{-4}, \text{ by (3.3)}, \\ P_j^{-3}/\|\alpha\| & \text{otherwise, by (3.4)}. \end{cases} \tag{3.15}$$

Then the inequality follows from an elementary computation. □

4. On the mean square deviation of $R(n)$

In this section we use Bessel's inequality to find an integral expression that bounds from above the average value of $\left|R(n) - \bar{R}(n)\right|^2$ for $n \in (N/2, N]$. We then perform a change of variables in the underlying arithmetic equation that makes the estimates on the absolute value of the integrand benefit from the restricted ranges $x_2, x_3, x_4 \le P_2 = o(P_1)$. Finally we use again the circle method to estimate the error introduced by this change of variables.

4.1. **Bessel's inequality.** From (3.10) and the inequality $(A+B)^2 \leq 2(A^2+B^2)$ we obtain that

$$\sum_{\frac{1}{2}N<n\leq N} \left|R(n)-\bar{R}(n)\right|^2 \leq NE_R^2 + 2\sum_{\frac{1}{2}N<n\leq N} |R_1(n)|^2 . \tag{4.1}$$

In order to estimate the sum on the right, we use Bessel's inequality, as in [3, eq.(12)], which in this case reveals that

$$\sum_{\frac{1}{2}N<n\leq N} |R_1(n)|^2 \leq \int_{\mathfrak{B}_1^{(1)}} |f_1 f_2 f_3 f_4 g|^2 \, d\alpha. \tag{4.2}$$

It is natural now to consider, for every measurable set $\mathfrak{B} \subseteq \mathbb{R}/\mathbb{Z}$, the integral

$$S(\mathbf{P},\mathfrak{B}) := \int_{\mathfrak{B}} |f_1 f_2 f_3 f_4 g|^2 \, d\alpha \tag{4.3}$$

and to let S, S_0, S_1 denote $S(\mathbf{P},\mathfrak{B})$ respectively for $\mathfrak{B} = \mathbb{R}/\mathbb{Z}, \mathfrak{B}_0^{(1)}, \mathfrak{B}_1^{(1)}$. With this notation, eq. (4.1) and eq. (4.2) can be combined to give the inequality

$$\sum_{\frac{1}{2}N<n\leq N} \left|R-\bar{R}\right|^2 \leq NE_R^2 + 2S_1. \tag{4.4}$$

We notice that this inequality has an underlying arithmetic meaning. In fact we have $S = S_0 + S_1$ and we observe that S counts the solutions to the equation

$$x_1^4 + \cdots + x_4^4 + y = {x'}_1^4 + \cdots + {x'}_4^4 + y' \tag{4.5}$$

subject to

$$0 < y, y' \leq Y, \quad \tfrac{1}{2}P_i < x_i, x_i' \leq P_i \quad (1 \leq i \leq 4), \tag{4.6}$$

by orthogonality.

4.2. **A change of variables.** The equation (4.5) can be rewritten in the following form

$$(x_1+h)^4 - {x_1}^4 = ({x_2}^4 - {x_2'}^4) + ({x_3}^4 - {x_3'}^4) + ({x_4}^4 - {x_4'}^4) + (y-y'), \tag{4.7}$$

where $h := x_1' - x_1$. We now focus only on those solutions, subject to (4.6), for which $h > 0$. By orthogonality, their number T is computed by the integral

$$T = \int_{\mathbb{R}/\mathbb{Z}} H_1 \, |f_2 f_3 f_4 g|^2 \, d\alpha, \tag{4.8}$$

where $H_1 = H(\alpha, P_1, 32P_1^{-3}P_2^4)$ is an exponential sum associated to the difference polynomial $\Delta(x,h) := (x+h)^4 - x^4$:

$$H(\alpha, X, Z) = \sum_{\substack{1\leq h\leq Z \\ \frac{1}{2}X<x\leq X-h}} e(\alpha[(x+h)^4 - x^4]). \tag{4.9}$$

Indeed every such solution satisfies

$$h = x_1' - x_1 \leq ({x_1'}^4 - x_1^4) x_1^{-3} \leq 4P_2^4 (\tfrac{1}{2}P_1)^{-3}$$

because of eq. (4.7) and the inequalities $P_3^4, P_4^4, Y \leq P_2^4$. The number S can be estimated by decomposing it naturally as $S = 2T + (S-2T)$. The term $S-2T$ accounts for the solutions of eq. (4.7) for which $h = 0$, i.e. it corresponds to an equation in fewer variables, since x_1 can be eliminated. The term $2T$ instead is

computed via the integral (4.8). This is easier to estimate than the integral in eq. (4.3), because its integrand is an exponential sum with fewer terms. Indeed H_1 only has $O(P_1^{-2}P_2^4) = O(P^{5/4})$ summands, which is noticeably less than the $O(P^2)$ terms of $|f_1|^2$. In particular, we record that the trivial estimate

$$H_1(\alpha) \ll P_1^{-2}P_2^4 \tag{4.10}$$

holds uniformly for all $\alpha \in \mathbb{R}/\mathbb{Z}$.

4.3. **A mollified version of $|S - 2T|$ near the origin.** Given the output (4.4) of Bessel's inequality, we actually need to estimate the term S_1, which is a portion of the integral $S = S_0 + S_1$ corresponding to the α that are bounded away from the origin. The idea is to decompose T somewhat analogously as $T_0 + T_1$ and then estimate S_1 as

$$S_1 \le |S_0 - 2T_0| + |2T_1| + |S - 2T|\,. \tag{4.11}$$

Since near the origin we have the estimates $g = Y + O(1)$ and $f_j = \nu_j + O(1)$, it is natural to compare the difference $S_0 - 2T_0$ with its mollified version $V - 2W$, where

$$V := Y^2 \int_{\mathbb{R}/\mathbb{Z}} |f_1\nu_2\nu_3\nu_4|^2\, d\alpha,$$
$$W := Y^2 \int_{\mathbb{R}/\mathbb{Z}} H_1\, |\nu_2\nu_3\nu_4|^2\, d\alpha.$$

Notice that we did not replace f_1 with its mollified version because we don't want to interfere with the change of variable that relates $|f_1|^2$ to H_1. In the following proposition we estimate the difference $V - 2W$ by looking at the underlying weighted diophantine equation.

Proposition 4.1.

$$V - 2W \ll Y^2 P_1 P_2^{-2} P_3^2 P_4^2 = Y^2.P^{\frac{7242}{4096}} \tag{4.12}$$

Proof. By orthogonality we have that

$$V = Y^2 \sum_{n\in\mathbb{Z}} r(n)\rho(n)$$

where $r(n) = r(n, P_1)$ is as in (4.17) and

$$\rho(n) := \sum_{\substack{\frac{1}{16}P_j^4 < z_j, z_j' \le P_j^4 \\ z_2+z_3+z_4-z_2'-z_3'-z_4'=n}} \tfrac{1}{4^6}(z_2 z_2' z_3 z_3' z_4 z_4')^{-3/4}.$$

Similarly, we have

$$W = Y^2 \sum_{n=1}^{\infty} r'(n)\rho(n)$$

where $r'(n)$ is as in (4.26). We notice immediately that

$$\rho(n) = 0 \qquad \text{for } |n| > 3P_2^4. \tag{4.13}$$

On the other hand we have

$$r(n) = 2r'(|n|) \qquad \text{for } 0 < |n| \le 4P_2^4 \tag{4.14}$$

because for $\frac{1}{2}P_1 < x, x' \le P_1$ the inequality $\left|x'^4 - x^4\right| \le 4P_2^4$ implies

$$|x' - x| \le (x'^4 - x^4)\min\{x, x'\}^{-3} \le 32P_1^{-3}P_2^4.$$

In other words by (4.13) and (4.14) we have

$$V - 2W = Y^2 r(0)\rho(0).$$

Since $r(0) = \frac{1}{2}P_1 + O(1)$ and

$$\rho(0) \ll P_2^4 P_3^8 P_4^8 (P_2^8 P_3^8 P_4^8)^{-3/4}$$

the proposition is proved. □

4.4. **Some useful estimates.** Before we proceed to study the difference between $|V - 2W|$ and "$|S_0 - 2T_0|$" (where T_0 has yet to be defined rigorously) we need to collect a few nontrivial estimates on integrals that involve $|\nu_j|^2$, $|f_j|^2$ and H_1. The first is similar to the one in Lemma 3.6.

Lemma 4.2.

$$\int_{\mathbb{R}/\mathbb{Z}} |\nu_j|^2 \, d\alpha \ll P_j^{-2}. \tag{4.15}$$

Proof. We estimate ν_j as in (3.15), so that the inequality follows from an elementary computation. □

Lemma 4.3. For every A, B, X we have

$$\int_A^{A+B} |f(\alpha, X)|^2 \, d\alpha \ll BX + X^{-2}\log X. \tag{4.16}$$

Proof. The integral (4.16) is estimated as in [3, eq.(17)] as follows. First, $|f(\alpha, X)|^2 = \sum_{n\in\mathbb{Z}} r(n, X)e(\alpha n)$ where

$$r(n, X) := \#\left\{(x, x') \;\middle|\; \begin{array}{c} x'^4 - x^4 = n \\ \frac{1}{2}X < x, x' \le X \end{array}\right\}. \tag{4.17}$$

Therefore

$$\int_A^{A+B} |f(\alpha, X)|^2 \, d\alpha = \sum_{n\in\mathbb{Z}} r(n, X) \int_A^{A+B} e(\alpha n) d\alpha. \tag{4.18}$$

If $n \neq 0$ the change of variable $\beta = \alpha n$ gives

$$\int_A^{A+B} e(\alpha n) d\alpha = \frac{1}{n}\int_{nA}^{nA+nB} e(\beta) d\beta \le \frac{2}{|n|},$$

hence

$$\int_A^{A+B} |f(\alpha, X)|^2 \, d\alpha = Br(0, X) + O\left(\sum_{n\neq 0} \frac{r(n, X)}{|n|}\right). \tag{4.19}$$

From the definition (4.17) we see that

$$\begin{aligned} r(0, X) &\ll X, \\ r(-n, X) &= r(n, X) && \text{for all } n, \\ r(n, X) &= 0 && \text{for } 0 < |n| \le \tfrac{1}{2}X^3 \text{ or } |n| > \tfrac{15}{16}X^4. \end{aligned}$$

Moreover we have that

$$\sum_{C<n\le C+\frac{1}{2}X^3} r(n,X)\le X$$

for every real C, because for every $x\in(X/2,X]$ there is at most one $x'\in(X/2,X]$ with $(C+x^4)<x'^4\le(C+x^4)+\frac{1}{2}X^3$. As a consequence, we have

$$\sum_{C<n\le C+D} r(n,X)\le 2DX^{-2}+O(X) \tag{4.20}$$

for all C,D,X. Therefore

$$\sum_{n\neq 0}\frac{r(n,X)}{|n|}\le 2\sum_{k=-1}^{\lfloor\log_2 X\rfloor}\frac{1}{2^kX^3}\sum_{2^kX^3<n\le 2^{k+1}X^3} r(n,X)\ll X^{-2}\log X \tag{4.21}$$

and (4.16) follows. □

Corollary 4.4. For all $1\le j\le 3$ we have

$$\int_{\mathfrak{B}_1^{(j)}}|f_j|^2\,\|\alpha\|^{-2}d\alpha\ll P_j^4\log P_j. \tag{4.22}$$

Proof. We divide the interval $\mathfrak{B}_1^{(j)}$, defined under (3.7), dyadically as follows

$$\mathfrak{B}_1^{(j)}\subseteq\bigcup_{k=-3}^{\lfloor 3\log_2 P_j\rfloor}\{\alpha\in\mathbb{R}/\mathbb{Z}:\ 2^kP_j^{-3}<\|\alpha\|\le 2^{k+1}P_j^{-3}\} \tag{4.23}$$

into pairs of intervals of length at most $2^kP_j^{-3}$. Hence by (4.16) we have

$$\int_{\mathfrak{B}_1^{(j)}}|f_j|^2\,\|\alpha\|^{-2}d\alpha\ll\sum_{k=-3}^{\lfloor 3\log_2 P_j\rfloor}P_j^{-2}(2^k+\log P_j)(2^{-2k}P_j^6)$$

that gives (4.22). □

Lemma 4.5. Let $\mathfrak{B}_1^{(2)}:=\{\alpha\in\mathbb{R}/\mathbb{Z}:\ \|\alpha\|>\frac{1}{8}P_2^{-3}\}$ as per (3.7), then

$$\int_{\mathfrak{B}_1^{(2)}}H_1\|\alpha\|^{-4}d\alpha\ll P_1^{-2}P_2^{12}\log P_1. \tag{4.24}$$

Proof. We proceed as in the proof of (4.16). First, we notice that for every A,B

$$\int_A^{A+B}H_1d\alpha\ll P_1^{-2}\log P_1. \tag{4.25}$$

Indeed, $H_1(\alpha)=\sum_{n=1}^{\infty}r'(n)e(\alpha n)$ where

$$r'(n):=\#\left\{(h,x)\,\middle|\,\begin{array}{c}(x+h)^4-x^4=n\\ 1\le h\le 32P_1^{-3}P_2^4\\ \frac{1}{2}P_1<x,x+h\le P_1\end{array}\right\}. \tag{4.26}$$

Therefore

$$\int_A^{A+B}H_1d\alpha=\sum_{n=1}^{\infty}r'(n)\int_A^{A+B}e(\alpha n)d\alpha\ll\sum_{n=1}^{\infty}\frac{r'(n)}{n} \tag{4.27}$$

as in (4.18)-(4.19). It is clear from (4.26) that

$$r'(n)=0\qquad\text{for } n\le\tfrac{1}{2}P_1^3 \text{ or } n>\tfrac{15}{16}P_1^4$$

and arguing as for (4.20) we get

$$\sum_{A<n\leq A+B} r'(n) \leq 2BP_1^{-2} + O(P_1). \tag{4.28}$$

Then (4.25) follows from (4.27) and (4.28) as in (4.21). Now we divide $\mathfrak{B}_1^{(2)}$ dyadically as in (4.23) and we obtain

$$\int_{\mathfrak{B}_1^{(2)}} H_1 \|\alpha\|^{-4} d\alpha \ll \sum_{k=-3}^{\lfloor 3\log_2 P_2 \rfloor} P_1^{-2} \log P_1 \cdot 2^{-4k} P_2^{12}.$$

The estimate (4.24) follows. □

4.5. **From S to T, through V and W.** For every measurable set $\mathfrak{B} \subseteq \mathbb{R}/\mathbb{Z}$, we recall the definition of the integral $S(\mathbf{P}, \mathfrak{B})$ and we define $T(\mathbf{P}, \mathfrak{B})$ as follows:

$$S(\mathbf{P}, \mathfrak{B}) := \int_{\mathfrak{B}} |f_1 f_2 f_3 f_4 g|^2 \, d\alpha,$$
$$T(\mathbf{P}, \mathfrak{B}) := \int_{\mathfrak{B}} H_1 \, |f_2 f_3 f_4 g|^2 \, d\alpha.$$

We also recall that S, S_0, S_1 denote $S(\mathbf{P}, \mathfrak{B})$ respectively for $\mathfrak{B} = \mathbb{R}/\mathbb{Z}, \mathfrak{B}_0^{(1)}, \mathfrak{B}_1^{(1)}$. We define $T = T_0 + T_1$ analogously, but for the new partition $\mathbb{R}/\mathbb{Z} = \mathfrak{B}_0^{(2)} \sqcup \mathfrak{B}_1^{(2)}$, where, as in (3.7):

$$\mathfrak{B}_0^{(2)} = \{\alpha : \|\alpha\| \leq \tfrac{1}{8} P_2^{-3}\} \qquad \mathfrak{B}_1^{(2)} = \{\alpha : \|\alpha\| > \tfrac{1}{8} P_2^{-3}\}.$$

In view of (4.11), the goal of this section is to prove that

$$|S_0 - 2T_0| \ll Y^2 P^{\frac{7242}{4096}} \approx Y^2 P^{1.768}.$$

Notice that $\mathfrak{B}_0^{(1)} \subseteq \mathfrak{B}_0^{(2)}$ and that the approximations $g \approx Y$ and $f_j \approx \nu_j$ for $2 \leq j \leq 4$ are valid on $\mathfrak{B}_0^{(2)}$, because $Y^{-2} \geq \frac{1}{8} P_2^{-3}$ by (2.1). We introduce the following integrals

$$V(\mathbf{P}, \mathfrak{B}) := Y^2 \int_{\mathfrak{B}} |f_1 \nu_2 \nu_3 \nu_4|^2 \, d\alpha,$$
$$W(\mathbf{P}, \mathfrak{B}) := Y^2 \int_{\mathfrak{B}} H_1 \, |\nu_2 \nu_3 \nu_4|^2 \, d\alpha,$$

then we define $V = V_0 + V_1$ (resp. $W = W_0 + W_1$) using $V(\mathbf{P}, \mathfrak{B})$ (resp. $W(\mathbf{P}, \mathfrak{B})$) and the partition $\mathbb{R}/\mathbb{Z} = \mathfrak{B}_0^{(1)} \sqcup \mathfrak{B}_1^{(1)}$ (resp. $\mathbb{R}/\mathbb{Z} = \mathfrak{B}_0^{(2)} \sqcup \mathfrak{B}_1^{(2)}$). Then we have $|S_0 - 2T_0| \leq E_S$, where

$$E_S := |S_0 - V_0| + |V_1| + |V - 2W| + |2W_1| + |2W_0 - 2T_0|. \tag{4.29}$$

We now dive into estimating the above five terms.

Proposition 4.6.

$$V_1 \ll_\epsilon P^\epsilon Y^2 P_1^4 P_2^{-6} P_3^2 P_4^2 = Y^2 P^{\frac{6218}{4096}+\epsilon}, \tag{4.30}$$

$$W_1 \ll_\epsilon P^\epsilon Y^2 P_1^{-2} P_2^6 P_3^{-6} P_4^2 = Y^2 P^{-\frac{54}{4096}+\epsilon}. \tag{4.31}$$

Proof. By (3.4) applied to ν_2 and (3.3) applied to ν_3, ν_4 we have

$$V_1 \ll Y^2 P_2^{-6} P_3^2 P_4^2 \int_{\mathfrak{B}_1^{(1)}} |f_1|^2 \|\alpha\|^{-2} d\alpha.$$

which gives (4.30) by (4.22). By (3.4) applied to ν_2, ν_3 and (3.3) for ν_4 we have

$$W_1 \ll Y^2 P_2^{-6} P_3^{-6} P_4^2 \int_{\mathfrak{B}_1^{(2)}} |H_1| \|\alpha\|^{-4} d\alpha.$$

The estimate (4.31) follows by (4.24). □

Proposition 4.7.

$$S_0 - V_0 \ll Y^2 P_1^{-2} P_2^2 P_3^2 P_4 = Y^2 P^{\frac{6069}{4096}}, \tag{4.32}$$

$$T_0 - W_0 \ll_\epsilon P^\epsilon Y^2 P_1^{-2} P_2^2 P_3^2 P_4 = Y^2 P^{\frac{6069}{4096}+\epsilon}. \tag{4.33}$$

Proof. Analogously to the computation in Proposition 3.5, by (3.6) and (3.2) we have

$$S_0 - V_0 \ll \int_{\mathfrak{B}_0^{(1)}} \left(Y \left| f_1^2 \mu_2^2 \mu_3^2 \mu_4^2 \right| + Y^2 \left| f_1^2 \right| \left(\left| \mu_2 \mu_3^2 \mu_4^2 \right| + \left| \mu_2^2 \mu_3 \mu_4^2 \right| + \left| \mu_2^2 \mu_3^2 \mu_4 \right| \right) \right) d\alpha,$$

where $\mu_j := \max\{|\nu_j|, 1\}$. We use the trivial estimate (3.3) for μ_2, μ_3, μ_4 and we use that $f_1 \ll \mu_1$ on $\mathfrak{B}_0^{(1)}$, by (3.6), to get

$$S_0 - V_0 \ll Y^2 P_2^2 P_3^2 P_4^2 \left(\tfrac{1}{Y} + \tfrac{1}{P_2} + \tfrac{1}{P_3} + \tfrac{1}{P_4} \right) \int_{\mathfrak{B}_0^{(1)}} |\mu_1|^2 d\alpha.$$

The integral to the right is $\ll P_1^{-2}$ by (4.15) and the fact that $\int_{\mathfrak{B}_0^{(1)}} 1 d\alpha \ll P_1^{-3}$. Since moreover $Y \geq P_4$, (4.32) follows.
Similarly, since $P_j^3 \leq P_2^3$ for all $j \geq 2$ and since $Y^2 \leq 8P_2^3$, we have by (3.6) and (3.2)

$$T_0 - W_0 \ll \int_{\mathfrak{B}_0^{(2)}} \left(Y \left| H_1 \mu_2^2 \mu_3^2 \mu_4^2 \right| + Y^2 |H_1| \left(\left| \mu_2 \mu_3^2 \mu_4^2 \right| + \left| \mu_2^2 \mu_3 \mu_4^2 \right| + \left| \mu_2^2 \mu_3^2 \mu_4 \right| \right) \right) d\alpha.$$

We apply (3.3) to μ_3, μ_4 and (4.10) to H_1 to get

$$T_0 - W_0 \ll Y^2 P_1^{-2} P_2^4 P_3^2 P_4^2 \left[\left(\tfrac{1}{Y} + \tfrac{1}{P_3} + \tfrac{1}{P_4} \right) \int_{\mathfrak{B}_0^{(2)}} |\mu_2|^2 d\alpha + \int_{\mathfrak{B}_0^{(2)}} |\mu_2| d\alpha \right].$$

The first integral is $\ll P_2^{-2}$ by (4.15) while the second integral is $\ll_\epsilon P^\epsilon P_2^{-3}$ by (3.14). The expression inside the square brackets is therefore $\ll P_2^{-2} P_4^{-1}$, hence we get (4.33). □

Finally, $V - 2W$ was estimated in eq. (4.12) and it turns out to be the main term in the right-hand side of (4.29). We conclude that

$$|S_0 - 2T_0| \leq E_S \ll Y^2 P^{\frac{7242}{4096}}. \tag{4.34}$$

5. Final estimates via the circle method

In this section we complete the proof of our main quantitative result, with a full application of the circle method and an induction on the number of variables in the underlying diophantine equation.

5.1. **Induction on the number of variables.** At this point, we still need to estimate the terms $|2T_1|$ and $|S-2T|$ in (4.11). We already commented briefly on the fact that $S-2T$ counts the number of solutions to the equation (4.5), subject to (4.6), together with $x_1' = x_1$. In particular if by $S^{(j)}$ we denote the number of solutions to the equation

$$x_j^4 + \cdots + x_4^4 + y = x_j'^4 + \cdots + x_j'^4 + y' \tag{5.1}$$

subject to

$$0 < y, y' \le Y, \quad \tfrac{1}{2}P_i < x_i, x_i' \le P_i \quad (j \le i \le 4),$$

we have $S - 2T \asymp P_1 S^{(2)}$. Now, eq. (5.1) has at least the "diagonal" solutions given by $y = y'$ and $x_i = x_i'$ for $j \le i \le 4$, hence

$$S^{(j)} \gg Y \prod_{i=j}^{4} P_j.$$

In particular, $S - 2T \gg P_1P_2P_3P_4Y$ and we cannot hope for a better estimate of this term. In the remainder of the section we will prove, by backward induction on j, that in fact

$$S^{(j)} \ll_\epsilon P^\epsilon Y \prod_{i=j}^{4} P_j \tag{5.2}$$

for $2 \le j \le 4$ and then we will show that

$$|2T_1| + |S-2T| \ll_\epsilon P^\epsilon P_1P_2P_3P_4Y = YN^{1-\gamma_0+\epsilon/4}, \tag{5.3}$$

where $\gamma_0 = 4059/16384$. Since by (4.4) we have

$$\sum_{\frac{1}{2}N < n \le N} \left|R(n) - \bar{R}(n)\right|^2 \le NE_R^2 + 2\left|S_0 - 2T_0\right| + 4\left|4T_1\right| + 2\left|S-2T\right|,$$

we finally get Theorem 1.2 by using (3.11), (4.34) and (5.3). The base step of induction is the following estimate of $S^{(4)}$.

Proposition 5.1.

$$S^{(4)} \ll P_4Y.$$

Proof. The number $S^{(4)}$ counts the solutions to the equation

$$x^4 + y = x'^4 + y' \tag{5.4}$$

subject to $\frac{1}{2}P_4 < x, x' \le P_4$ and $1 \le y, y' \le Y$. For every such solution, say with $x \le x'$, we have that

$$x'^4 - x^4 \le Y \le \tfrac{1}{2}P_4^3$$

and so $x'^4 - x^4 < (x+1)^4 - x^4$. This implies that (5.4) has only the diagonal solutions $x = x'$ and $y = y'$, therefore

$$S^{(4)} = (\tfrac{1}{2}P_4 + O(1))(Y + O(1)).$$

□

5.2. **Major arcs, central arc and minor arcs.** The equation (5.1), for $j \le 3$ is transformed via the substitution $x_j' = x_j + h$, like we did in section 4.2. To the resulting equation

$$(x_j+h)^4 - x_j^4 = (x_{j+1}^4 - (x_{j+1}')^4) + \cdots + (x_4^4 - (x_4')^4) + (y - y') \tag{5.5}$$

additionally constrained by $h > 0$, we attach the integrals

$$T^{(j)}(\mathbf{P}, \mathfrak{B}) := \int_{\mathfrak{B}} H_j \left| g \prod_{i=j+1}^{4} f_i \right|^2 d\alpha, \tag{5.6}$$

where $\mathfrak{B} \subseteq \mathbb{R}/\mathbb{Z}$ is a measurable set and where

$$H_j := H(\alpha, P_j, 32P_j^{-3}P_{j+1}^4)$$

is given by (4.9). The solutions to (5.5) corresponding to $h = 0$ are counted by

$$S^{(j)} - 2T^{(j)} = (\tfrac{1}{2}P_j + O(1))S^{(j+1)}. \tag{5.7}$$

We are going to estimate the integrals (5.6) with the circle method.
For every $1 \le j \le 3$ and every pair of coprime integers q, a with $q \ge 1$ we form

$$\mathfrak{M}^{(j)}(q,a) := \{\alpha \in \mathbb{R}/\mathbb{Z} : \|\alpha - a/q\| \le q^{-1}P_jP_{j+1}^{-4}\}, \tag{5.8}$$

and we define the j-th set of *major arcs* by

$$\mathfrak{M}^{(j)} := \bigcup_{q=2}^{P_j} \bigcup_{a \in (\mathbb{Z}/q\mathbb{Z})^*} \mathfrak{M}^{(j)}(q,a).$$

Notice that the intervals in the definition of $\mathfrak{M}^{(j)}$ are disjoint because for every two rational numbers a/q, A/Q with denominators $q \le Q \le P_j$ we have

$$\left|\frac{A}{Q} - \frac{a}{q}\right| \ge \frac{1}{qP_j} \ge \frac{1}{q}P_jP_{j+1}^{-4} + \frac{1}{Q}P_jP_{j+1}^{-4}$$

by (1.1). Notice that in the definition of $\mathfrak{M}^{(j)}$ we excluded the major arc centered at zero. For $j \in \{2,3\}$ we denote the j-th *central arc* by $\mathfrak{N}^{(j)} := \mathfrak{M}^{(j)}(1,0)$ and we define the j-th set of *minor arcs* $\mathfrak{m}^{(j)}$ so that $\mathbb{R}/\mathbb{Z} = \mathfrak{N}^{(j)} \sqcup \mathfrak{M}^{(j)} \sqcup \mathfrak{m}^{(j)}$ is a partition. For $j = 1$ we define the central arc by

$$\mathfrak{N}^{(1)} := \{\alpha : \tfrac{1}{8}P_2^{-3} < |\alpha| \le P_1P_2^{-4}\} = \mathfrak{M}^{(1)}(1,0) \cap \mathfrak{B}_1^{(2)} \tag{5.9}$$

and consider the partition $\mathfrak{B}_1^{(2)} = \mathfrak{N}^{(1)} \sqcup \mathfrak{M}^{(1)} \sqcup \mathfrak{m}^{(1)}$. For every $1 \le j \le 3$ we let $T_{\mathfrak{N}}^{(j)}, T_{\mathfrak{M}}^{(j)}, T_{\mathfrak{m}}^{(j)}$ denote $T^{(j)}(\mathbf{P}, \mathfrak{B})$ respectively for $\mathfrak{B} = \mathfrak{N}^{(j)}, \mathfrak{M}^{(j)}, \mathfrak{m}^{(j)}$. Finally, we define $T^{(1)} := T_1$ and $T^{(j)} := T^{(j)}(\mathbf{P}, \mathbb{R}/\mathbb{Z})$ for $j \in \{2,3\}$, so that

$$T^{(j)} = T_{\mathfrak{M}}^{(j)} + T_{\mathfrak{N}}^{(j)} + T_{\mathfrak{m}}^{(j)} \qquad (1 \le j \le 3). \tag{5.10}$$

5.3. **Estimates for H_j and the minor arc contribution.** It turns out that the minor arc component $T_{\mathfrak{m}}^{(j)}$ is the dominant term in $T^{(j)}$ for all $1 \le j \le 3$. Nevertheless, we are going to estimate it crudely for each $1 \le j \le 3$, as follows:

$$\left|T_{\mathfrak{m}}^{(j)}\right| \le \left(\sup_{\alpha \in \mathfrak{m}^{(j)}} |H_j|\right) \int_{\mathfrak{m}^{(j)}} |f_{j+1} \cdots f_4 g|^2 \, d\alpha \le \left(\sup_{\alpha \in \mathfrak{m}^{(j)}} |H_j|\right) S^{(j+1)}. \tag{5.11}$$

Thus we now need to bound from above the absolute value of the exponential sum H_j. Such estimate is proved as in [16, Lemma 1] using the Weyl differencing method:

Lemma 5.2. Let $H(\alpha, X, Z)$ be as in (4.9) with $Z \leq X$ and $|\alpha - a/q| \leq q^{-2}$ for some integers a, q. Then we have, for all $\epsilon > 0$:

$$H(\alpha, X, Z) \ll_\epsilon X^{1+\epsilon} Z (X^{-1} + q^{-1} + qX^{-3}Z^{-1})^{1/4},$$

where the implied constant depends only on ϵ.

Since H_j is a sum of terms with absolute value 1, it can be trivially estimated as

$$H_j(\alpha) \ll P_j^{-2} P_{j+1}^4 = P_j^{5/4}$$

for all $\alpha \in \mathbb{R}/\mathbb{Z}$ and for each $1 \leq j \leq 3$. From Lemma 5.2 can deduce better pointwise estimates for H_j in regions of interest to us.

Corollary 5.3. For all $1 \leq j \leq 3$ and all $\alpha \in \mathfrak{M}^{(j)}(q, a)$ with coprime $q, a \leq P_j$ we have

$$H_j(\alpha) \ll_\epsilon P^\epsilon P_j^{-2} P_{j+1}^4 \cdot q^{-1/4}. \tag{5.12}$$

Moreover, for each $1 \leq j \leq 3$ and all $\alpha \in \mathfrak{m}^{(j)}$ we have

$$H_j(\alpha) \ll_\epsilon P_j^\epsilon P_j^{-2} P_{j+1}^4 \cdot P_j^{-1/4} = P_j^{1+\epsilon}. \tag{5.13}$$

Proof. If $\alpha \in \mathfrak{M}^{(j)}(q, a)$ we apply Lemma 5.2 and we get (5.12) from $q \leq P_j$ and (1.1). Dirichlet's approximation theorem [17, Lemma 2.1] says that for every $\alpha \in \mathbb{R}$ and every $Q \geq 1$ there are integers a, q with $q \leq Q$ such that $|\alpha - a/q| \leq 1/(qQ)$. If $\alpha \in \mathfrak{m}^{(j)}$ we apply Dirichlet's theorem with $Q = P_j^{-1} P_{j+1}^4$. The corresponding fraction a/q satisfies $q > P_j$ by definition of $\mathfrak{m}^{(j)}$ and so Lemma 5.2 gives (5.13). □

Remark 5.4. By the same method, applying Dirichlet's theorem with $Q = P_2^3$, it is possible to prove that

$$H_1(\alpha) \ll_\epsilon P^\epsilon P_1^{-2} P_2^4 \cdot P_2^{-1/4} \tag{5.14}$$

for $\alpha \in \mathfrak{N}^{(1)}$. However, the trivial estimate $H_1(\alpha) \ll P_1^{-2} P_2^4$ will be sufficient for us in the treatment of the central arc $\mathfrak{N}^{(1)}$.

Focusing in particular on the minor arc estimate, for all $1 \leq j \leq 3$ we get

$$T_{\mathfrak{m}}^{(j)} \ll_\epsilon P^\epsilon P_j S^{(j+1)} \tag{5.15}$$

from (5.11) and (5.13). Combining (5.7), (5.10) and (5.15) we deduce that

$$S^{(j)} \ll_\epsilon P^\epsilon P_j S^{(j+1)} + T_{\mathfrak{M}}^{(j)} + T_{\mathfrak{N}}^{(j)} \qquad (2 \leq j \leq 3),$$
$$T_1 \ll_\epsilon P^\epsilon P_1 S^{(2)} + T_{\mathfrak{M}}^{(1)} + T_{\mathfrak{N}}^{(1)}.$$

This induction scheme, together with (5.7) for $j = 1$ and the base step (5.1), shows in particular that

$$|2T_1| + |S - 2T| \ll_\epsilon P^\epsilon (E_{\mathfrak{m}} + E_{\mathfrak{M}} + E_{\mathfrak{N}}),$$

where

$$E_{\mathfrak{M}} := \left|T_{\mathfrak{M}}^{(1)}\right| + P_1 \left|T_{\mathfrak{M}}^{(2)}\right| + P_1 P_2 \left|T_{\mathfrak{M}}^{(3)}\right|,$$
$$E_{\mathfrak{N}} := \left|T_{\mathfrak{N}}^{(1)}\right| + P_1 \left|T_{\mathfrak{N}}^{(2)}\right| + P_1 P_2 \left|T_{\mathfrak{N}}^{(3)}\right|,$$
$$E_{\mathfrak{m}} := Y P_1 P_2 P_3 P_4 = Y P^{\frac{12325}{4096}}.$$

Thus to prove the final estimate (5.3), as well as the intermediate claims (5.2), it is sufficient to prove that $E_{\mathfrak{M}}, E_{\mathfrak{N}} \ll E_{\mathfrak{m}}$.

5.4. **Treatment of the central arc.** Here we estimate the error terms coming from the central arcs of $T^{(1)}$, $T^{(2)}$ and $T^{(3)}$. In order to prove that $E_{\mathfrak{N}} \ll E_{\mathfrak{m}}$ it is enough to show, since $Y \le P^{\frac{4992}{4096}}$ by assumption, that $E_{\mathfrak{N}} \ll Y^2 P^{\frac{7333}{16384}}$.

Proposition 5.5.

$$T_{\mathfrak{N}}^{(1)} \ll_\epsilon P^\epsilon Y^2 P_1^{-2} P_2^8 P_3^{-6} P_4^2 = Y^2 P^{\frac{6602}{4096}+\epsilon}, \tag{5.16}$$

$$P_1 T_{\mathfrak{N}}^{(2)} \ll_\epsilon P^\epsilon Y^2 P_1 P_2^{-2} P_3^2 P_4^2 = Y^2 P^{\frac{7242}{4096}+\epsilon}, \tag{5.17}$$

$$P_1 P_2 T_{\mathfrak{N}}^{(3)} \ll_\epsilon P^\epsilon Y^2 P_1 P_2 P_3^{-1} P_4 = Y^2 P^{\frac{6917}{4096}+\epsilon}. \tag{5.18}$$

Proof. We have

$$T_{\mathfrak{N}}^{(1)} \le \left(\sup_{\alpha \in \mathfrak{N}^{(1)}} \left| H_1 f_4^2 g^2 \right| \right) \int_{\mathfrak{N}^{(1)}} |f_2 f_3|^2 \, d\alpha.$$

We also have $\mathfrak{N}^{(1)} \subseteq \mathfrak{B}_1^{(2)} \cap \mathfrak{B}_0^{(3)}$ (see (3.7) and (5.9)) since the inequalities

$$\tfrac{1}{8} P_2^{-3} < \|\alpha\| \le P_1 P_2^{-4} \le \tfrac{1}{8} P_3^{-3}$$

hold for every $\alpha \in \mathfrak{N}^{(1)}$. In particular f_3 is well approximated by ν_3 on $\mathfrak{N}^{(1)}$ and so $f_3(\alpha) \ll P_3^{-3} \|\alpha\|^{-1}$ by (3.6) and (3.4). Therefore

$$\int_{\mathfrak{N}^{(1)}} |f_2 f_3|^2 \, d\alpha \ll P_3^{-6} \int_{\mathfrak{B}_1^{(2)}} |f_2|^2 \, \|\alpha\|^{-2} d\alpha$$

which is $\ll_\epsilon P^\epsilon P_2^4 P_3^{-6}$ by (4.22). Hence (5.16) follows using the trivial estimates $H_1 \ll P_1^{-2} P_2^4$, $g \ll Y$ and $f_4 \ll P_4$. [(2)] We deal with $T_{\mathfrak{N}}^{(2)}$ similarly:

$$T_{\mathfrak{N}}^{(2)} \le \left(\sup_{\alpha \in \mathfrak{N}^{(2)}} \left| H_2 g^2 \right| \right) \int_{\mathfrak{M}^{(2)}(1,0)} |f_3 f_4|^2 \, d\alpha.$$

We estimate H_2 and g trivially as above. To estimate the integral instead, we observe that $\mathfrak{M}^{(2)}(1,0) \subseteq \mathfrak{B}_0^{(3)} \sqcup (\mathfrak{B}_0^{(4)} \setminus \mathfrak{B}_0^{(3)})$. On the interval $\mathfrak{B}_0^{(3)}$ we estimate f_4 trivially, while on $\mathfrak{B}_0^{(4)} \setminus \mathfrak{B}_0^{(3)}$ we proceed as in the previous case, so

$$T_{\mathfrak{N}}^{(2)} \ll Y^2 P_2^{-2} P_3^4 \left(P_4^2 \int_{\mathfrak{B}_0^{(3)}} |f_3|^2 \, d\alpha + P_4^{-6} \int_{\mathfrak{B}_1^{(3)}} |f_3|^2 \, \|\alpha\|^{-2} d\alpha \right). \tag{5.19}$$

Since on $\mathfrak{B}_0^{(3)}$ the approximation $f_3 = \nu_3 + O(1)$ holds, we have $|f_3|^2 = |\nu_3|^2 + O(P_3)$ and so the first integral in (5.19) is estimated as

$$\int_{\mathfrak{B}^{(3)}0} |f_3|^2 \, d\alpha \ll \int_{\mathbb{R}/\mathbb{Z}} |\nu_3|^2 \, d\alpha + P_3 \cdot \int_{\mathfrak{B}^{(3)}0} 1 d\alpha,$$

which is $\ll P_3^{-2}$ by (4.15). On the other hand the second integral of (5.19) is $\ll P_3^4 \log P_3$ by and (4.22), so (5.17) follows. Finally (5.18) follows simply from

$$T_{\mathfrak{N}}^{(3)} \le \left(\sup_{\alpha \in \mathfrak{N}^{(3)}} \left| H_3 g^2 \right| \right) \int_{\mathfrak{M}^{(3)}(1,0)} |f_4|^2 \, d\alpha,$$

[(2)] We could have saved $P_2^{-1/4}$ by using the more precise estimate (5.14), but this is not much actually.

estimating H_3 and g trivially and using (4.16) with $B = 2P_3P_4^{-4}$ to estimate the integral. □

5.5. **Treatment of the major arcs.** Here we estimate the error terms coming from the major arcs in $\mathfrak{M}^{(j)}$ (which exclude the central one). Since the Weyl sum g is small away from 0, we are able to estimate it nontrivially on $\mathfrak{M}^{(j)}$. For example we have the following proposition, that is obtained, *mutatis mutandis*, from [3, Lemma 2].

Proposition 5.6. For all $1 \leq j \leq 3$ we have, uniformly on $q > 1$:

$$\sum_{a \in (\mathbb{Z}/q\mathbb{Z})^*} \left(\sup_{\alpha \in \mathfrak{M}^{(j)}(q,a)} |g(\alpha, Y)|^2 \right) \ll qY. \tag{5.20}$$

This allows us to save one power of Y in the estimate for $E_{\mathfrak{M}}$. We will need also some estimates for the Weyl sums f_j. For this purpose the following result, taken from the book of Vaughan [17], is very useful.

Lemma 5.7. For every coprime q, a and every $\epsilon > 0$ we have

$$f(a/q + \beta, X) \ll q^{-1/4}\nu(\beta, X) + q^{1/2+\epsilon}(1 + X^4\|\beta\|)^{1/2} \quad \text{for all } \beta, \tag{5.21}$$

$$f(a/q + \beta, X) \ll q^{-1/4}\nu(\beta, X) + q^{1/2+\epsilon} \quad \text{if } \|\beta\| < \frac{1}{8qX^3}. \tag{5.22}$$

Proof. The estimates (5.21) and (5.22) follow from [17, Thm 4.1 and Thm 4.2]. □

In our case Lemma 5.7 is used to estimate the f_j in absolute value and in mean square over the major arcs, as in the following two corollaries.

Corollary 5.8. For all $1 \leq j \leq 3$, all coprime $q, a \leq P_j$ and all $\epsilon > 0$ we have

$$\int_{\mathfrak{M}^{(j)}(q,a)} |f_{j+1}|^2 \, d\alpha \ll_\epsilon P^\epsilon q^{-1} P_j^{1/2} P_{j+1}^{-2} \tag{5.23}$$

Proof. By (5.21) and (5.8) we have

$$\int_{\mathfrak{M}^{(j)}(q,a)} |f_{j+1}|^2 \, d\alpha \ll q^{-1/2} \int_{\mathbb{R}/\mathbb{Z}} |\nu_{j+1}|^2 \, d\alpha + q^\epsilon P_j \int_{\mathfrak{M}^{(j)}(q,a)} 1 \, d\alpha$$

and so (5.23) follows by (4.15), (1.1) and $q \leq P_j$. □

Corollary 5.9. For all $1 \leq i, j \leq 4$ with $j \geq i + 2$ and all coprime $q, a \leq P_i$ we have

$$\sup_{\alpha \in \mathfrak{M}^{(i)}(q,a)} |f_j(\alpha)| \ll_\epsilon P^\epsilon q^{-1/4} P_i^{3/4}. \tag{5.24}$$

Proof. For $\alpha \in \mathfrak{M}^{(i)}(q, a)$ we may estimate $|f_j(\alpha)|$ with (5.22) because the inequality

$$\frac{1}{q} P_i P_{i+1}^{-4} < \frac{1}{8q} P_j^{-3}$$

holds for P large enough. Then (5.24) follows from the trivial estimate $\nu(\beta, P_j) \ll P_j$ and the inequality $P_j \leq P_i^{3/4}$. □

We are now ready for the last computations. We recall that in order to have $E_{\mathfrak{M}} \ll E_{\mathfrak{m}}$ we need to show that $E_{\mathfrak{M}} \ll YP^{\frac{12325}{4096}}$.

Proposition 5.10.

$$T_{\mathfrak{M}}^{(1)} \ll_\epsilon P^\epsilon Y P_1^{5/4} P_2^2 = Y P^{\frac{11776}{4096}+\epsilon}, \tag{5.25}$$

$$P_1 T_{\mathfrak{M}}^{(2)} \ll_\epsilon P^\epsilon Y P_1 P_2^{1/4} P_3^2 = Y P^{\frac{10328}{4096}+\epsilon}, \tag{5.26}$$

$$P_1 P_2 T_{\mathfrak{M}}^{(3)} \ll_\epsilon P^\epsilon Y P_1 P_2 P_3^{-3/4} P_4^2 = Y P^{\frac{9790}{4096}+\epsilon}. \tag{5.27}$$

Proof. From the definitions we have

$$T_{\mathfrak{M}}^{(1)} \leq \sum_{\substack{2\leq q\leq P_1 \\ a\in(\mathbb{Z}/q\mathbb{Z})^*}} \sup_{\alpha\in\mathfrak{M}^{(1)}(q,a)} |g|^2 \cdot \left(\sup_{\alpha\in\mathfrak{M}^{(1)}(q,a)} \left|H_1 f_3^2 f_4^2\right| \right) \int_{\mathfrak{M}^{(1)}(q,a)} |f_2|^2 \, d\alpha.$$

We apply (5.20) to g, (5.12) to H_1 and (5.24) to estimate f_3, f_4. Together with (5.23) we get

$$T_{\mathfrak{M}}^{(1)} \ll_\epsilon P^\epsilon \sum_{q=2}^{\lfloor P_1 \rfloor} qY \cdot q^{-1/4} P_1^{-2} P_2^4 \cdot q^{-1/2} P_1^{3/2} \cdot q^{-1/2} P_1^{3/2} \cdot q^{-1} P_1^{1/2} P_2^{-2}$$

which gives (5.25). Similarly, to estimate $T_{\mathfrak{M}}^{(2)}$ we apply (5.20) to g, (5.12) to H_2, (5.24) to f_4 and (5.23) to f_3:

$$T_{\mathfrak{M}}^{(2)} \ll_\epsilon P^\epsilon \sum_{q=2}^{\lfloor P_2 \rfloor} qY \cdot q^{-1/4} P_2^{-2} P_3^4 \cdot q^{-1/2} P_2^{3/2} \cdot q^{-1} P_2^{1/2} P_3^{-2}$$

that gives (5.26). Finally, again by (5.20), (5.12) and (5.23) we have

$$T_{\mathfrak{M}}^{(3)} \ll_\epsilon P^\epsilon \sum_{q=2}^{\lfloor P_3 \rfloor} qY \cdot q^{-1/4} P_3^{-2} P_4^4 \cdot q^{-1} P_3^{1/2} P_4^{-2}$$

that gives (5.27). □

References

[1] J. Bourgain, C. Demeter, and L. Guth. Proof of the main conjecture in Vinogradov's mean value theorem for degrees higher than three. *Annals of Mathematics*, 184:633–682, 2016.

[2] J. Brüdern and T.D. Wooley. Additive representation in short intervals, II: sums of two like powers. *Math. Z.*, 286:179–196, 2017.

[3] S. Daniel. On gaps between numbers that are sums of three cubes. *Mathematika*, 44(1):1–13, 1997.

[4] H. Davenport. On Waring's problem for fourth powers. *Ann. Math.*, 40:731–747, 1939.

[5] J.M. Deshouillers, F. Hennecart, and B. Landreau. Sums of powers: an arithmetic refinement to the probabilistic model of Erdős and Rényi. *Acta Arithmetica*, 85(1):13–33, 1998.

[6] J.M. Deshouillers, F. Hennecart, and B. Landreau. On the density of sums of three cubes. In *Algorithmic number theory*, volume 4076 of *Lecture Notes in Comput. Sci.*, pages 141–155. Springer, Berlin, 2006.

[7] P. Erdős and A. Rényi. Additive properties of random sequences of positive integers. *Acta Arithmetica*, 6(1):83–110, 1960.

[8] L. Ghidelli. Arbitrarily long gaps between the values of positive-definite cubic and biquadratic diagonal forms. *Preprint accepted upon revisions by the Journal of the London Mathematical Society*, 2019.

[9] L. Ghidelli. Arithmethic properties of values of cubic and biquadratic theta functions. *Preprint*, 2019.

[10] G.H. Hardy and J.E. Littlewood. Some problems of "Partitio Numerorum" (VI): Further researches in Waring's problem. *Math. Z.*, 23:1–37, 1925.

[11] D.R. Heath-Brown. The circle method and diagonal cubic forms. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 356(1738):673–699, 1998.
[12] C. Hooley. On Waring's problem. *Acta Mathematica*, 157:49–97, 1966.
[13] C. Hooley. On some topics connected with Waring's problem. *J. reine angew. Math*, 369:110–153, 1986.
[14] C. Hooley. On Hypothesis K^* in Waring's problem. In *Sieve methods, exponential sums, and their applications in number theory, Cardiff, 1995. London Math. Soc. Lecture Series*, volume 237, pages 175–185. Cambridge University Press, 1997.
[15] L.B. Pierce. The Vinogradov mean value theorem after Wooley, and Bourgain, Demeter and Guth). *Séminaire Bourbaki, 69ième année*, pages 1134–1179, Juin 2017.
[16] R.C. Vaughan. On Waring's problem for smaller exponents. *Proc. Lond. Math. Soc.*, 52(3):445–463, 1986.
[17] R.C. Vaughan. *The Hardy-Littlewood method*. Number 2 in Cambridge tracts in mathematics. Cambridge University Press, 2 edition, 1997.
[18] R.C. Vaughan and T.D. Wooley. Waring's problem: a survey. *Number theory for the millennium 3*, pages 301–340, 2002.
[19] T.D. Wooley. Sums of three cubes. *Mathematica*, 47:53–61, 2000.
[20] T.D. Wooley. Nested efficient congruencing and relatives of Vinogradov's mean value theorem. *Proceedings of the London Mathematical Society*, 118(4):942–1016, 2019.

Chapter 8

Arithmetic properties of cubic and biquadratic theta series

ARITHMETIC PROPERTIES OF CUBIC AND BIQUADRATIC THETA SERIES

LUCA GHIDELLI

Abstract. A cubic (resp. biquadratic) theta series is a power series whose n-th coefficient is equal to 1 if n is a perfect cube (resp. fourth power) and zero otherwise. We improve on a result of Bradshaw by showing that such series is not a cubic (resp. biquadratic) algebraic number when evaluated at reciprocals of integers. The proof relies on a "nested gaps technique" for linear independence and on recent results by the author on Waring's problem for cubes and biquadrates.

Contents

1. Introduction

In this paper we consider numbers of the form

$$\theta_\ell(q) = \sum_{n=0}^{\infty} \frac{1}{q^{n^\ell}},$$

for $\ell \in \{3, 4\}$ and $q > 1$. These numbers can be thought as being values (at $z = 1/q$) of cubic/biquadratic generalizations of the well-known theta series $\sum_{n=0}^{\infty} z^{n^2}$. As usual for values of transcendental series, we expect that $\theta_\ell(q)$ is transcendental at algebraic inputs, possibly with some well-motivated exceptions. Our main result is the following.

Theorem 1.1. Let $\ell \in \{3, 4\}$, let $q \geq 2$ be an integer and suppose that $\theta_\ell(q)$ is algebraic. Then $\deg \theta_\ell(q) \geq \ell + 1$.

The proof of Theorem 1.1 is based on a variation of Bradshaw's technique of nested gaps for lacunary series. It also involves some delicate considerations about the natural numbers that can (or cannot) be represented as sums of three nonnegative cubes or as sums of four fourth powers. In Proposition 9.2 below we will quantify

Date: October 11, 2019.
2010 *Mathematics Subject Classification.* Primary: 11J17; Secondary: 11B05.

the conclusion of Theorem 1.1 by providing a measure of linear independence for the $(\ell+1)$-tuple $(1, \theta_\ell(q), \dots, \theta_\ell(q)^\ell)$.

1.1. **Notation.** We will denote the set of nonnegative integers by $\mathbb{N} := \{0, 1, \dots\}$ and by $\mathbb{N}_+ := \mathbb{N} - \{0\}$ the set of positive integers. The notation log will denote the natural logarithm and $\log_2$ will denote the logarithm in base 2.

2. Remarks on the method and comparison with the literature

To prove that a number is not algebraic, it is a common technique to seek for good rational approximations. Since $\theta_\ell(q)$ is defined as a series, it is natural to approximate it by its truncations. However their relatively slow rate of convergence implies only that $\theta_\ell(q)$ is irrational at integer inputs. The method of Bradshaw [4] improves on the above strategy when the series is "lacunary". It is based on the construction of "nested gaps" and on the following easy observation.

Remark 2.1. Let $S = \sum_{n\geq 0} s_n$ be a series for which a tail bound of the form $\left|\sum_{n\geq N} s_n\right| \leq f(N)$ is given. Suppose that for some $K, n_0 \in \mathbb{N}$ we have $s_{n_0+i} = 0$ for all $0 \leq i < K$: we say that the series S has a *gap* of length $\geq K$ at n_0. When we have such a gap, the bound for the tail at n_0 can be improved to $\left|\sum_{n\geq n_0} s_n\right| \leq f(n_0+K)$.

By applying this method to the (lacunary!) series representation of $\theta_\ell(q)^{\ell-1}$ Bradshaw was able to show [4, Theorem 2.0.1], for all integer ℓ, that $\theta_\ell(q)$ is not an algebraic number of degree $< \ell$. To extend the non-algebraicity of $\theta_\ell(q)$ up to degree $= \ell$ one faces technical difficulties related to Waring's problem (I thank Martin Rivard-Cooke for pointing this to me). More precisely, we need the existence of arbitrarily long sequences of consecutive integers none of which is a sum of ℓ nonnegative ℓ-th powers. This result was recently proved by the author [9, Thm. 1.1, 1.2, 8.8] for $\ell \in \{3, 4\}$ and is open for $\ell \geq 5$. The aim of this article is to check that this, together with the consideration of suitable "mild" gaps (see section 3), is enough for the proof of Theorem 1.1. As a side note, we would like to remark that our lower bound for the size of gaps between sums of fourth powers, although growing to infinity, it does so very slowly. Therefore, it came with some surprise that these estimates are in fact good enough to have arithmetic consequences on the biquadratic theta series.

In the literature variants of the above series have been considered. The irrationality and nonquadraticity of classical theta values $\theta_2(q)$ were studied by Duverney [7, 8]. Irrationality and irrationality measures of similar numbers have been considered in many works, such as [14, 5, 1]. Bézivin [3] proved the nonquadraticity of values of the more general Tschakaloff function $T_q(z) = \sum_{n=0}^{\infty} z^n q^{-n(n-1)/2}$. The results of Bézivin have been simplified by Bradshaw [4, Chapter 3] and extended by some authors [12]. Last but not least, a celebrated result of Nesterenko [13] implies that $\theta_2(q)$ is transcendental for all nonzero algebraic q satisfying $|q| > 1$ [2, Theorem 4]. His proof relies on an appropriate multiplicity estimate and it exploits the differential Ramanujan identities between the quasi-modular functions $E_2(q)$, $E_4(q)$ and $E_6(q)$.

3. A nested gaps principle for linear independence

In section 2 we mentioned that Bradshaw [4] took advantage of sufficiently large gaps for the series representation of $\theta_\ell(q)^{\ell-1}$, and that he applied a certain "nested gaps" argument to prove his results. We are going to reproduce a variation of his technique by considering the series representation of $\theta_\ell(q)^\ell$, for $\ell \in \{3, 4\}$, and by considering only those "gaps" that are followed by coefficients with controlled size. We call these gaps "mild" in Definition 3.2 below. Although we are ultimately

interested in (non)-algebraicity properties of $\theta_\ell(q)$, a careful inspection reveals that Bradshaw's method is more naturally seen as a lemma for *linear independence* of lacunary series. We think it is worthwile to recast Bradshaw's technique in this setting. However, we will not try to enunciate a criterion valid in maximal generality, in order not to obfuscate the underlying idea. We need a few definitions.

Definition 3.1. We define a $\frac{1}{2}$-*function* to be a powerseries $f(z) = \sum_{n\in\mathbb{N}} a_n z^n$ with integer coefficients that is absolutely convergent for all $|z| \leq 1/2$.

In particular, a $\frac{1}{2}$-function can be evaluated at reciprocals of integers $q \geq 2$.

Definition 3.2. Let $f(z) = \sum_{n\in\mathbb{N}} a_n z^n$ be a $\frac{1}{2}$-function, let $K \in \mathbb{N}_+$ and $E > 0$. We say that an index $n \in \mathbb{N}$ is a *mild gap point* for $f(z)$, with gap-length $\geq K$ and K-tail-norm $\leq E$, if $a_{n+k} = 0$ for all $0 \leq k < K$ and

$$\sum_{i=0}^{\infty} |a_{n+K+i}|\, 2^{-i} \leq E.$$

We denote by $\mathrm{MildGap}(f(z); K, E)$ the set of such mild gap points for f.

The next theorem is the promised criterion, abstracted from Bradshaw's method, for $\mathbb{Q}$-linear independence of the values $f(1/q)$, $g(1/q)$ of two lacunary $\frac{1}{2}$-functions at the reciprocal of an integer. It essentially states that the linear independence necessarily occurs when *pairs of (large enough) mild gaps of f can be found inside one (larger) gap of g*. As Damien Roy pointed out to me, the proof also yields a measure of linear independence between $f(1/q)$ and $g(1/q)$. We explore this quantitative refinement in section 9.

Theorem 3.3 (Nested Gaps Principle)**.** Let $q \geq 2$ be an integer and let $f(z) = \sum_{n\in\mathbb{N}} a_n z^n$ and $g(z) = \sum_{n\in\mathbb{N}} b_n z^n$ be $\frac{1}{2}$-functions. Suppose that for every $H > 0$ there are positive integers $K_1 \leq K_2 < K' \in \mathbb{N}_+$, indices $n' \leq n_1 < n_2 \in \mathbb{N}$ and real numbers $E, E' > 0$ such that:

(i) $n_1 + K_1 < n_2$ and $n_2 + K_2 \leq n' + K'$;
(ii) $n_1, n_2 \in \mathrm{MildGap}(f(z); K_1, E)$ and $n' \in \mathrm{MildGap}(g(z); K', E')$;
(iii) $\sum_{n=n_1}^{n_2-1} a_n q^{-n} \neq 0$;
(iv) $q^{K_1} > HE$ and $q^{K_2} > HE'$.

Then either $g(1/q) = 0$ or $f(1/q)$ and $g(1/q)$ are linearly independent over $\mathbb{Q}$.

Proof. Suppose the contrary. Then there exist integers α, β such that $\alpha \neq 0$ and

$$0 = \alpha f(1/q) + \beta g(1/q) = \sum_{n\in\mathbb{N}} \frac{R(n)}{q^n}, \tag{3.1}$$

where $R(n) := \alpha a_n + \beta b_n$. Let $H = \max\{|\alpha|, |\beta|\}$, then choose K_1, K_2, K', E, E' and n_1, n_2, n' as above. Now pick $i \in \{1, 2\}$ arbitrarily. By hypothesis (ii) and since $q \geq 2$ we have

$$\left| \sum_{n=n_i}^{\infty} \frac{R(n)}{q^n} \right| \leq |\alpha| \, \frac{E}{q^{n_i+K_1}} + |\beta| \, \frac{E'}{q^{n'+K'}}. \tag{3.2}$$

From the estimates (iv), eq. (3.1) and $n_i + K_2 \leq n' + K'$, we deduce that

$$\left| \sum_{n=0}^{n_i-1} \frac{R(n)}{q^n} \right| < \frac{2}{q^{n_i}}. \tag{3.3}$$

However, the left-hand side of eq. (3.3) is a rational number with denominator at most q^{n_i-1} and so it must be equal to zero. Having concluded this for both n_1 and

n_2, we deduce that

$$0 = \sum_{n=n_1}^{n_2-1} \frac{R(n)}{q^n} = \alpha \sum_{n=n_1}^{n_2-1} a_n q^{-n},$$

against hypothesis (iii). □

4. Simple tail bounds

In this section we present a pair of lemmas to estimate the "tail-norms" of a $\frac{1}{2}$-function when suitable bounds are known for its coefficients (see Definition 3.2).

Lemma 4.1. Let $(a_n)_{n\in\mathbb{N}}$ be a sequence of numbers with $|a_n| \le c(n+1)$ for all $n \in \mathbb{N}$ and some $c > 0$. Then for every $n_0 \in \mathbb{N}_+$ we have

$$\sum_{i=0}^{\infty} |a_{n_0+i}|\, 2^{-i} \le 8cn_0.$$

Proof. The positive function $\psi(x) = x2^{-x}$ satisfies $\psi(x) \ge \psi(2)$ for all $x \in [1,2]$ and it is monotone decreasing for $x > 1/\log 2 = 1.44269\ldots$, hence

$$\sum_{i=0}^{\infty} |a_{n_0+i}|\, 2^{-i} \le 2^{n_0+1} \sum_{n\ge n_0} \frac{c(n+1)}{2^{n+1}} \le c2^{n_0+1} \int_{n_0}^{\infty} \frac{t}{2^t} dt. \tag{4.1}$$

By partial integration we obtain

$$\int_{n_0}^{\infty} t2^{-t} dt = \frac{2^{-n_0}}{\log 2}\left(n_0 + \frac{1}{\log 2}\right) \le \left(\frac{1}{\log 2} + \frac{1}{(\log 2)^2}\right)\frac{n_0}{2^{n_0}} \le 4n_0 2^{-n_0}.$$

Together with eq. (4.1), this gives the lemma. □

Lemma 4.2. Let $(a_n)_{n\in\mathbb{N}}$ be as in Lemma 4.1 for some $c > 0$, and let $\kappa, n_0 \in \mathbb{N}_+$ with $n_0 + \kappa \le N$ and $\kappa \ge \log_2 N$ for some N. Suppose that for all $0 \le i < \kappa$ and some $E \ge 8c$ we have $|a_{n_0+i}| \le (3/2)^i E$. Then

$$\sum_{i=0}^{\infty} |a_{n_0+i}|\, 2^{-i} \le 5E.$$

Proof. From $|a_{n_0+i}| \le (3/2)^i E$ we get

$$\sum_{i=0}^{\kappa-1} |a_{n_0+i}|\, 2^{-i} \le \sum_{i=0}^{\infty} \left(\frac{3}{4}\right)^i \cdot E = 4E.$$

On the other hand, by Lemma 4.1 and the various inequalities relating the constants, we have

$$\sum_{i=\kappa}^{\infty} |a_{n_0+i}|\, 2^{-i} \le \frac{1}{2^\kappa} 8c(n_0+\kappa) \le \frac{1}{N} 8cN \le E.$$

□

5. Linear independence of powers of θ_ℓ

Fix $\ell \in \{3, 4\}$. For all $s \in \{1, \ldots, \ell\}$ and $n \in \mathbb{N}$ we set

$$r_{\ell,s}(n) = \#\{(n_1, \ldots, n_s) \in \mathbb{N}^s : \; n_1^\ell + \cdots + n_s^\ell = n\}$$

so that for all $q > 1$

$$\theta_\ell(q)^s = \sum_{n=0}^{\infty} \frac{r_{\ell,s}(n)}{q^n}.$$

We observe that $\theta_\ell(q)^s$ is the value at $1/q$ of the $\frac{1}{2}$-function

$$f_{\ell,s}(z) := \sum_{n=0}^{\infty} r_{\ell,s}(n) z^n$$

for all ℓ, s. Therefore we may apply Theorem 3.3 to prove the following criterion.

Proposition 5.1. Let $q \geq 2$ be an integer. Suppose that for every $J > 0$ there are $E, N > 0$, integers $K_1 \leq K_2 \in \mathbb{N}_+$ and $n_1, n_2 \in \mathrm{MildGap}(f_{\ell,\ell}; K_1, E)$ such that:

(i) $n_1 + K_1 < n_2$ and $n_2 + K_2 \leq N$;
(ii) $r_{\ell,\ell-1}(n) = 0$ for all $n_1 \leq n < n_2 + K_2$;
(iii) there exists $n_3 \in [n_1, n_2)$ with $r_{\ell,\ell}(n_3) > 0$;
(iv) $q^{K_1} > JE$ and $q^{K_2} > JN$.

Then either $\theta_\ell(q)$ is transcendental or it is algebraic with degree at least $\ell + 1$.

Proof. Suppose that $\theta_\ell(q)$ is algebraic of degree at most ℓ. Then there exist integers $\alpha_0, \ldots, \alpha_\ell$ with $\alpha_\ell \neq 0$ such that

$$\alpha_0 + \alpha_1 \theta_\ell(q) + \cdots + \alpha_\ell \theta_\ell(q)^\ell = 0. \tag{5.1}$$

We define $f(z) := f_{\ell,\ell}(z)$ and

$$g(z) := \alpha_0 + \alpha_1 f_{\ell,1}(z) + \cdots + \alpha_{\ell-1} f_{\ell,\ell-1}(z).$$

We notice that for all $s \leq \ell$ and all $n \in \mathbb{N}$ we have the (loose) estimate

$$0 \leq r_{\ell,s}(n) \leq (\sqrt[\ell]{n} + 1)^s \leq 2^\ell (n+1). \tag{5.2}$$

In particular for all $n \in \mathbb{N}$ the n-th coefficient of $g(z)$ has absolute value $\leq c(n+1)$ where $c = \ell \cdot 2^\ell \cdot \max\{|\alpha_i| : i < \ell\}$. We also notice that for $n_1 \leq n < n_2 + K_2$ the condition (ii) implies that $r_{\ell,s}(n) = 0$ for all $s < \ell$, i.e. that the n-th coefficient of $g(z)$ vanishes. By Lemma 4.1, this means that $n_1 \in \mathrm{MildGap}(g; K', E')$, where $K' = n_2 - n_1 + K_2$ and $E' = 8cN$. Moreover (iii) is equivalent to $\sum_{n=n_1}^{n_2-1} r_{\ell,\ell}(n) q^{-n} \neq 0$ because $r_{\ell,\ell}$ is nonnegative. Thus, for any $H > 0$, the hypotheses of the current proposition for $J = 8cH$ imply those of Theorem 3.3 with $n' = n_1$ and $E' = 8cN$. By eq. (5.1) the numbers $f(1/q),\ g(1/q)$ are linearly dependent. But since $f(1/q) > 0$ and $\alpha_\ell \neq 0$ we also have $g(1/q) \neq 0$, so we arrive at a contradiction. □

6. Sums of powers modulo M and existence of mild gaps

By the previous proposition, Theorem 1.1 is reduced to the problem of finding suitable mild gaps of $f_{\ell,\ell}$. In this section we present a proposition that provides "many" mild gaps of a prescribed type. This result is proved via an elementary technique known as the Maier matrix method [11]. We require the following definition: for every $m \in \mathbb{Z}$ and $M \in \mathbb{N}_+$ let

$$r_{\ell,\ell}(m, M) := \{(x_1, \ldots, x_\ell) \in (\mathbb{Z}/M\mathbb{Z})^\ell :\ x_1^\ell + \cdots + x_\ell^\ell \equiv m \bmod M\}.$$

Proposition 6.1. Let $K, M, m \in \mathbb{N}$ with $m + K < M$. Now let $\epsilon_0, \ldots, \epsilon_K > 0$ such that $r_{\ell,\ell}(m+k, M) \leq \epsilon_k M^{\ell-1}$ for all $0 \leq k \leq K$ and let $E_0, \ldots, E_K \in \mathbb{N}$ such that $\alpha < 1$, where

$$\alpha := \frac{\epsilon_0}{E_0 + 1} + \ldots + \frac{\epsilon_K}{E_K + 1}.$$

Then for each $N > 0$ with $N \geq M^\ell$ we have

$$\#\left\{ n \in [0, N-K) \;\middle|\; \begin{array}{c} n \equiv m \pmod M \\ r_{\ell,\ell}(n+k) \leq E_k \text{ for all } 0 \leq k \leq K \end{array} \right\} \geq \frac{1-\alpha}{2^\ell} \frac{N}{M}.$$

Proof. Let $L \in \mathbb{N}$ such that $L^\ell M^\ell \leq N < (L+1)^\ell M^\ell$ and let $I = L^\ell M^{\ell-1}$. It is not difficult [9, Prop. 8.4] to show that for all $0 \leq k \leq K$ we have

$$\sum_{i=0}^{I-1} r_{\ell,\ell}(m+k+iM) \leq L^\ell r_{\ell,\ell}(m+k, M).$$

From this we deduce that

$$\#\{i \in [0, I) : \ r_{\ell,\ell}(m+k+iM) > E_k\} \leq \frac{\epsilon_k I}{E_k + 1}.$$

Therefore the $0 \leq i < I$ such that $r_{\ell,\ell}(m+k+iM) \leq E_k$ for all $0 \leq k \leq K$ are at least

$$(1-\alpha)L^\ell M^{\ell-1} = (1-\alpha)\left(\frac{L}{L+1}\right)^\ell \frac{(L+1)^\ell M^\ell}{M} \geq \frac{1-\alpha}{2^\ell}\frac{N}{M}.$$

The proposition follows because for each such i we have $m + iM < N - K$. □

7. Key results from Waring's problem

In order to find mild gap points with gap-length K_1 using Proposition 6.1 it is crucial that we make $r_{\ell,\ell}(m+k, M)$ as small as possible for $k < K_1$ and that we can estimate it from above for larger values of k.

Lemma 7.1. Let $\ell \in \{3, 4\}$ and define the following auxiliary functions of T

$$\kappa_3(T) := \frac{\sqrt{T}}{(\log T)^2} \qquad \kappa_4(T) := \frac{\log\log T}{\log\log\log T}$$
$$\Xi_3(T) := \log\log T \qquad \Xi_4(T) := 1.$$

For each large enough T there are natural numbers M, m, K_1, with $\max\{2m, 4K_1\} < M$ and M even, and positive constants C_0, C_1, C_2, C_3 such that:

(i) $C_0 T \leq \log M \leq C_1 T$;
(ii) $K_1 \geq C_2 \cdot \kappa_\ell(T)$;
(iii) $r_{\ell,\ell}(m+k, M) \leq \frac{1}{2K_1} \cdot M^{\ell-1}$ for all $0 \leq k < K_1$;
(iv) $r_{\ell,\ell}(m', M) \leq e^{C_3 \Xi_\ell(T)} \cdot M^{\ell-1}$ for all $m' \in \mathbb{Z}$.

Proof. We are going to follow the arguments of [9, Sec. 8] applied to the diagonal form $F = x_1^\ell + \cdots + x_\ell^\ell$. In [9, Prop. 8.1] we construct a set $\mathcal{P}_F$ with positive density in the set of all primes $p \equiv 1 \pmod{\ell}$. By the Prime Number Theorem the product $M_T := \prod_p p$ of the primes $p \in \mathcal{P}_F \cap [1, T]$ satisfies $T \ll \log M_T \ll T$. In [9, Sec. 8.4] we prove that there exist two natural numbers $m, K_1 < M_T$ that fulfill condition (iii) provided that $T \geq \tau_\ell(\gamma_\ell, K_1)$, where [9, Def. 8.6]

$$\tau_3(\gamma, K) := \gamma K^2 (\log K)^4$$
$$\tau_4(\gamma, K) := \exp(\exp(\gamma K \log K))$$

and $\gamma_3, \gamma_4 > 0$ are some absolute constants. If T is large enough, we may take K_1 so that $C_2 \kappa_\ell(T) \leq K_1 < \frac{1}{2} M_T$ for some small enough C_2. By [9, Prop. 8.2] with $\mathcal{P}_1 = \emptyset$ and $\mathcal{P}_2 = \mathcal{P}_F \cap [1, T]$ we have that for all $m' \in \mathbb{Z}$ the inequality $r_{\ell,\ell}(m', M_T) \leq \xi M_T^{\ell-1}$ holds with $\xi > 0$ given by

$$\log \xi = (\ell-1)^\ell \sum_{p \in \mathcal{P}_F \cap [1,T]} p^{-(\ell-1)/2}.$$

The sum to the right is again estimated via the Prime Number Theorem (see [9, Lemma 5.4]): if $\ell = 3$ this sum is $\ll \log\log T$; if $\ell = 4$ it is bounded. In both cases we get the estimate $r_{\ell,\ell}(m', M) \leq e^{C_3 \Xi_\ell(T)} \cdot M^{\ell-1}$ for all $m' \in \mathbb{Z}$ and some C_3.

Finally, we define $M := 2M_T$. All the statements in the lemma now follow because for every $m' \in \mathbb{Z}$ we have

$$r_{\ell,\ell}(m', M) = r_{\ell,\ell}(m', 2) r_{\ell,\ell}(m', M_T) = 2^{\ell-1} r_{\ell,\ell}(m', M_T).$$

□

As we will see, the above lemma together with Proposition 6.1 implies that the series attached to $\theta_\ell(q)^\ell$ has gaps of arbitrarily large size. On the other hand, we need to produce two *distinct* such gaps inside a single gap attached to $\theta_\ell(q)^{\ell-1}$. The typical gap (in $[1, N]$) between sums of $\ell - 1$ perfect ℓ-th powers is of size $\approx N^{1/\ell}$. Therefore we need to show that most gaps between sums of ℓ perfect ℓ-th powers have size $\leq N^\gamma$ for some $\gamma < 1/\ell$. Such a result is easy to establish for $\ell = 3$ with the following greedy argument.

Lemma 7.2. For every $b \in \mathbb{N}$ there is $n \in (b - 25b^{8/27}, b]$ with $r_{3,3}(n) > 0$.

Proof. First notice that for every $B \in \mathbb{N}$ there is $x_1 \in \mathbb{N}$ such that $x_1^3 \leq B < (x_1 + 1)^3$. Such x_1 satisfies $B - x_1^3 \leq 6B^{2/3}$. Iterating this procedure, we find in turn $x_1, x_2, x_3 \in \mathbb{N}$ such that $0 \leq (B - x_1^3) - x_2^3 \leq 6(6B^{2/3})^{2/3}$ and

$$0 \leq B - x_1^3 - x_2^3 - x_3^3 \leq 6^{1+2/3+4/9} B^{8/27} < 25B^{8/27}.$$

The lemma follows by choosing $B = b$ and $n = x_1^3 + x_2^3 + x_3^3$. □

As mentioned above, the crucial point is that $8/27 < 1/3$. The greedy argument above, for $\ell = 4$, only gives x_1, x_2, x_3, x_4 such that

$$B - x_1^4 - x_2^4 - x_3^4 - x_4^4 = O(B^{(3/4)^4})$$

and $(3/4)^4 = \frac{5184}{16384} > 1/4$. One way to overcome this problem is to prove the existence of suitable $x_1, \dots, x_4$ via the so-called "circle method with diminishing ranges", which might be thought as a (nontrivial) improved version of the greedy argument. Since the proof is technical, we perform the required computation in a separate paper [10]. In that article, we extend to sums of four powers a result of Daniel for sums of three cubes [6] and in particular we are able to show the following [10, Corollary 1.2].

Lemma 7.3. For almost every $a \in \mathbb{N}$ there is $n \in (a - a^{\frac{4059}{16384}+\varepsilon}, a]$ with $r_{4,4}(n) > 0$, where $\varepsilon > 0$ is arbitrary.

By "almost every a" in the above lemma we mean that for every $\varepsilon > 0$ and all $\delta \in (0, 1)$ there is some $N_{\varepsilon,\delta} \in \mathbb{N}$ such that, for all $N \geq N_{\varepsilon,\delta}$ we have that the set

$$\mathcal{A}_N := \{a \in [1, N] : \ r_{4,4}(n) = 0 \text{ for all } n \in (a - a^{\frac{4059}{16384}+\varepsilon}, a]\} \tag{7.1}$$

has cardinality $\#\mathcal{A}_N \leq \delta N$.

8. Proof of Theorem 1.1

Fix $\ell \in \{3, 4\}$, an integer $q \geq 2$ and an arbitrary $J > 0$. Choose $\sigma_3 \in (3, \frac{27}{8})$ and $\sigma_4 \in (4, \frac{16384}{4059})$, then take $T = T(q, J, \sigma_\ell)$ large enough for the following arguments to be valid.

8.1. **Choice of parameters.** Given T, we choose M, m, K_1, C_i as in Lemma 7.1, then we set $N = M^{\sigma_\ell}$ and $K_2 = \frac{1}{2}M > 2K_1$. We also define $\xi_3 = (\log T)^{C_3}$ and $\xi_4 = \max\{C_3, 32/3\}$, and finally $E = 60\xi_\ell$. It is clear that the inequalities $q^{K_1} > JE$ and $q^{K_2} > JN$ hold if T is large enough. In other words, condition (iv) of Proposition 5.1 is fulfilled.

8.2. **A set of mild gap points.** We apply Proposition 6.1 with $K = K_2$ and:

(1) $\epsilon_k = \frac{1}{2K_1}$ and $E_k = 0$ for $0 \le k < K_1$;

(2) $\epsilon_{K_1+k} = \xi_\ell$ and $E_{K_1+k} = 12\xi_\ell(3/2)^k$ for $0 \le k \le K_2 - K_1$.

In addition to $m + K_2 < \frac{1}{2}M + \frac{1}{2}M = M$ and $M^\ell < M^{\sigma_\ell} = N$, we have

$$\alpha := \frac{\epsilon_0}{E_0+1} + \cdots + \frac{\epsilon_K}{E_K+1} < K_1\frac{1}{2K_1} + \sum_{k=0}^{\infty}\frac{\xi_\ell}{12\xi_\ell(3/2)^k} = \frac{3}{4}.$$

So Proposition 6.1 provides a set

$$\mathcal{B} = \{b_1 < b_2 < \ldots\} \subseteq [0, N-K_2) \cap (m + \mathbb{Z}M)$$

with cardinality $\#\mathcal{B} \ge N/(2^{\ell+2}M)$ such that $r_{\ell,\ell}(b_i+k) \le E_k$ for all $0 \le k \le K_2$. In particular, by condition (1) above we have that all elements of $\mathcal{B}$ are mild gap points for $f_{\ell,\ell}$ with gap-length $\ge K_1$. We recall from eq. (5.2) that $r_{\ell,\ell}(n) \le 2^\ell(n+1)$ for all $n \in \mathbb{N}$. Moreover we observe that

$$12\xi_\ell \ge 8 \cdot 2^\ell \quad \text{and} \quad \kappa := K_2 - K_1 \ge K_1 \ge \log_2 N$$

if T is large enough. Therefore, by Lemma 4.2 and condition (2), every $b_i \in \mathcal{B}$ has K_1-tail-norm $\le 5 \cdot 12\xi_\ell \le E$. In other words, we have $\mathcal{B} \subseteq \mathrm{MildGap}(f_{\ell,\ell}(z); K_1, E)$.

8.3. **"Nested" pairs of mild gaps.** We now seek to apply Proposition 5.1 to a pair of consecutive points $n_1 = b_i$, $n_2 = b_{i+1}$ from $\mathcal{B}$. We already argued that condition (iv) is satisfied by our choice of parameters. Condition (i) is fulfilled as well: $n_1 + K_1 < n_2$ because $b_i \equiv b_{i+1} \equiv m \pmod M$ and $K_1 < M$; while $n_2 + K_2 < N$ because $b_{i+1} \le \max\mathcal{B} < N - K_2$. In order to fulfill condition (ii) we need to exclude any b_i from the set

$$\mathcal{B}^{\mathrm{bad}} := \{b_i \in \mathcal{B} : \ \exists n \in [b_i, b_{i+1} + K_2] \text{ with } r_{\ell,\ell-1}(n) \ge 1\}.$$

Since $b_{i+1} + K_2 < b_{i+1} + M \le b_{i+2}$ for all $i \le \#\mathcal{B} - 2$, it is clear that

$$\#\mathcal{B}^{\mathrm{bad}} \le 2\sum_{n=0}^{N} r_{\ell,\ell-1}(n),$$

which in turn is $\le 2(\sqrt[\ell]{N}+1)^{\ell-1} \le 2^\ell N^{1-1/\ell}$. On the other hand, $\#\mathcal{B} \ge 2^{-\ell-2}N^{1-1/\sigma_\ell}$, so $\#\mathcal{B}^{\mathrm{bad}} < (\#\mathcal{B})/2$ if T (and so N) is sufficiently large. In particular, the complementary set $\mathcal{B}^{\mathrm{good}} := \mathcal{B} \setminus \mathcal{B}^{\mathrm{bad}}$ has cardinality at least $N/(2^{\ell+3}M)$. For every pair $(n_1, n_2) = (b_i, b_{i+1})$ with $b_i \in \mathcal{B}^{\mathrm{good}}$, condition (ii) of Proposition 5.1 is fulfilled.

8.4. **"Separated" pair of mild gaps.** If $\ell = 3$ then every pair $(n_1, n_2) = (b_i, b_{i+1})$ with $b_i \in \mathcal{B}^{\mathrm{good}}$ satisfies condition (iii) of Proposition 5.1. Indeed, recall that n_1 and n_2 are congruent (to m) modulo M, so $n_2 - n_1 \ge M$. By our choice of σ_3 we have

$$25n_2^{8/27} \le 25N^{8/27} < N^{1/\sigma_3} = M$$

for every T large enough, so the claim follows from Lemma 7.2. If $\ell = 4$ we define $\varepsilon = \frac{1}{2}(\sigma_4^{-1} - \frac{4059}{16384})$ and we consider the intervals of the form $I(a) := (a - a^{\frac{4059}{16384}+\varepsilon}, a]$, where a is an element of the set $\mathcal{A} \subseteq [1, N]$ given by

$$\mathcal{A} := \mathbb{N} \cap \bigcup_{b_i \in \mathcal{B}^{\mathrm{good}}} [b_i + \tfrac{1}{2}M, b_i + M).$$

We observe that $\frac{1}{2}M > N^{\frac{4059}{16384}+\varepsilon}$ for every T large enough, so each $I(a)$ with $a \in \mathcal{A}$ is contained in an interval (b_i, b_{i+1}), for some $b_i \in \mathcal{B}^{\mathrm{good}}$. Suppose that no pair $(n_1, n_2) = (b_i, b_{i+1})$ with $b_i \in \mathcal{B}^{\mathrm{good}}$ satisfies condition (iii) of Proposition 5.1. Then

for every $a \in \mathcal{A}$ and every $n \in I(a)$ we have $r_{4,4}(n) = 0$: in other words, $\mathcal{A} \subseteq \mathcal{A}_N$, where $\mathcal{A}_N$ is as in eq. (7.1). However,

$$\#\mathcal{A} = \tfrac{1}{2}M \cdot (\#\mathcal{B}^{\mathrm{good}}) \geq 2^{-\ell-4}N$$

and this contradicts Lemma 7.3, if T is large enough.

8.5. **Conclusion.** For every $J > 0$ we proved the existence of E, N, K_1, K_2 and n_1, n_2 that meet all requirements of Proposition 5.1. Theorem 1.1 follows.

9. Measure of linear independence

We present a quantitative version of the Nested Gaps Principle.

Proposition 9.1. Let $f(z)$, $g(z)$ and $q \geq 2$ be as in Theorem 3.3. Suppose there are positive integers $K_1 \leq K_2 < K' \in \mathbb{N}_+$, indices $n' \leq n_1 < n_2 \in \mathbb{N}$ and real numbers $E, E' > 0$ meeting all conditions (i)-(iv) of Theorem 3.3 for some $H > 0$. If α and β are integers with $\alpha \neq 0$ and $|\alpha| + |\beta| \leq H$ then

$$|\alpha f(1/q) + \beta g(1/q)| \geq q^{-n_2}.$$

Proof. We let $R(n) := \alpha a_n + \beta b_n$ and for $i \in \{1, 2\}$ we write

$$S_i = \sum_{n=0}^{n_i-1} \frac{R(n)}{q^n}.$$

Since $\alpha \neq 0$ we have that $S_2 - S_1 \neq 0$ by conditions (ii) and (iii). Thus, there exists $i_0 \in \{1, 2\}$ such that $S_{i_0} \neq 0$. Since S_{i_0} is a rational number with denominator $q^{-n_{i_0}+1}$, we have $S_{i_0} \geq q^{-n_{i_0}+1}$. On the other hand, as in the proof of Theorem 3.3 we get

$$\left| \sum_{n=n_{i_0}}^{\infty} \frac{R(n)}{q^n} \right| \leq \frac{|\alpha|\, E}{q^{n_{i_0}+K_2}} + \frac{|\beta|\, E'}{q^{n_{i_0}+K_1}} \leq q^{-n_{i_0}}. \tag{9.1}$$

Therefore

$$\alpha f(1/q) + \beta g(1/q) = S_{i_0} + \sum_{n=n_{i_0}}^{\infty} \frac{R(n)}{q^n} \geq \frac{q-1}{q^{n_{i_0}}} \geq q^{-n_2}.$$

□

From the above quantitative result we get the following measure of linear independence for the first powers of $\theta_\ell(q)$.

Proposition 9.2. Let $\ell \in \{3, 4\}$ and $\Theta := (1, \theta_\ell(q), \ldots, \theta_\ell(q)^\ell) \in \mathbb{R}^{\ell+1}$, where $q \geq 2$ is an integer. Let $P(\mathbf{T}) = \sum_{j=0}^{\ell} \alpha_j T_j$ be a nonzero linear form with integer coefficients satisfying $|\alpha_j| \leq q^A$ for some $A > 1.1$ and $\alpha_\ell \neq 0$. If $\ell = 3$ we have

$$|P(\Theta)| > \exp(-\log q \exp(c_3 A^2 (\log A)^2))$$

for some $c_3 > 0$, while if $\ell = 4$ we have

$$|P(\Theta)| > \exp(-\log q \exp\exp\exp(c_4 A \log A))$$

for some $c_4 > 0$.

Proof. We wish to apply Proposition 9.1 to the pair of $\frac{1}{2}$-functions

$$f(z) := f_{\ell,\ell}(z) \qquad\qquad g(z) := \sum_{j=0}^{\ell-1} \alpha_j f_{\ell,j}(z). \tag{9.2}$$

We let $c = \ell 2^\ell$ and $J = 8c(q^A + 1)$. Then we set

$$T := c_3' A^2 (\log A)^4 \qquad \text{(if } \ell = 3\text{)}$$
$$T := c_3' \exp\exp(c_4' A \log A) \qquad \text{(if } \ell = 4\text{)}$$

for some $c_\ell' > 0$ large enough and we choose K_1, K_2, N, E as in section 8.1. The above formula for T is chosen so that the inequality $q^{K_1} > JE$ holds if $c_\ell' > 0$ is larger than some absolute constant. Notice that if c_ℓ' is large enough we also have the inequality $q^{K_2} > JN$. Moreover, all the arguments of sections 8.2 to 8.4 are valid for every T larger than some T_0 independent of q and J. In particular, if c_ℓ' is large enough, there are some n_1, n_2 such that all the itemized conditions of Proposition 5.1 are fulfilled with this choice of J, K_1, K_2, N, E. As in the proof of Proposition 5.1 we then see that the hypotheses of Proposition 9.1 are fulfilled, with $n' = n_1$, $E' = 8cN$, $H = J/(8c)$ and $f(z), g(z)$ as in eq. (9.2). Since $n_2 < N$ and $\log N = O(T)$, we get from Proposition 9.1 the required estimate for $P(\Theta) = \alpha_\ell f(1/q) + g(1/q)$, for some $c_\ell > 0$. □

Notice that the hypothesis $\alpha_\ell \neq 0$ on $P(\mathbf{T})$ is not restrictive. In fact, if $\alpha_{\ell+1-h} = \cdots = \alpha_\ell = 0$ for some $h \geq 1$, we have that $P(\Theta) = \theta_\ell(q)^{-h} P'(\Theta)$ where $P'(\mathbf{T}) = \sum_{j=h}^{\ell} \alpha_{j-h} T_j$. We notice that $\theta_\ell(q) \leq \theta_\ell(2) \leq 2$ and so $|P(\Theta)| \geq 2^{-\ell} |P'(\Theta)|$. Therefore the estimates of Proposition 9.2 still hold if we replace c_ℓ by some larger absolute constant. However, we remark that in this situation one could apply Proposition 9.1 to the pair of $\frac{1}{2}$-functions

$$f(z) := f_{\ell,\ell-h}(z) \qquad\qquad g(z) := \sum_{j=0}^{\ell-h-1} \alpha_j f_{\ell,j}(z).$$

and obtain a measure of linear independendence that is single-exponential in "A" (as opposed to Proposition 9.2, where the estimate is doubly or quadruply exponential).

Acknowledgements

I would like to thank my supervisor Damien Roy for his encouragement and for his many comments on this work, especially the suggestion of computing a quantitative measure of linear independence. I am grateful to Martin Rivard-Cooke for introducing me to the problem of noncubicity of cubic theta values and for mentioning the need of new results in Waring's problem for cubes. This work was supported in part by a full International Scholarship from the Faculty of Graduate and Postdoctoral Studies of the University of Ottawa and by NSERC.

References

[1] M. Amou and M. Katsurada. Irrationality results for values of generalized Tschakaloff series. II. *J. Number Theory*, 104(1):132–155, 2004.

[2] D. Bertrand. Theta functions and transcendence. *Ramanujan J.*, 1(4):339–350, 1997.

[3] J. P. Bézivin. Sur les propriétés arithmétiques d'une fonction entière. *Math. Nachr.*, 190:31–42, 1998.

[4] R. Bradshaw. Arithmetic properties of values of lacunary series. Master's thesis, University of Ottawa, 2013.

[5] P. Bundschuh and I. Shiokawa. A measure for the linear independence of certain numbers. *Results Math.*, 7(2):130–144, 1984.

[6] S. Daniel. On gaps between numbers that are sums of three cubes. *Mathematika*, 44(1):1–13, 1997.

[7] D. Duverney. Propriétés arithmétiques d'une série liée aux fonctions thêta. *Acta arithmetica*, 64:175–188, 1993.

[8] D. Duverney. Sommes de deux carrés et irrationalité de valeurs de fonctions têta. *C. R. Acad. Sci. Paris Sér. I Math.*, 320:1041–1044, 1995.

[9] L. Ghidelli. Arbitrarily long gaps between the values of positive-definite cubic and biquadratic diagonal forms. *Preprint accepted upon revisions by the Journal of the London Mathematical Society*, 2019.
[10] L. Ghidelli. On gaps between sums of four fourth powers. *Preprint*, 2019.
[11] A. Granville. Unexpected irregularities in the distribution of prime numbers. *Proceedings of the International Congress of Mathematicians*, 1:388–399, 1995.
[12] C. Krattenthaler, I. Rochev, K. Väänänen, and W. Zudilin. On the non-quadraticity of values of the q-exponential function and related q-series. *Acta Arith.*, 136(3):243–269, 2009.
[13] Y. Nesterenko. Modular functions and transcendence problems. *C. R. Acad. Sci. Paris Sér. I Math.*, 322(10):909–914, 1996.
[14] T. Stihl. Arithmetische Eigenschaften spezieller Heinescher Reihen. *Math. Ann.*, 268(1):21–41, 1984.

Part III

Other results in Commutative Algebra and Combinatorics

Chapter 9

Multigraded Koszul complexes, filter-regular sequences and lower bounds for the multiplicity of the resultant

MULTIGRADED KOSZUL COMPLEXES, FILTER-REGULAR SEQUENCES AND LOWER BOUNDS FOR THE MULTIPLICITY OF THE RESULTANT

LUCA GHIDELLI

Abstract. The Rémond resultant attached to a multiprojective variety and a sequence of multihomogeneous polynomials is a polynomial form in the coefficients of the polynomials, which vanishes if and only if the polynomials have a common zero on the variety. We demonstrate that this resultant can be computed as a Cayley determinant of a multigraded Koszul complex, proving a key stabilization property with the aid of local Hilbert functions and the notion of filter-regular sequences. Then we prove that the Rémond resultant vanishes, under suitable hypotheses, with order at least equal to the number of common zeros of the polynomials. More generally, we estimate the multiplicity of resultants of multihomogeneous polynomials along prime ideals of the coefficient ring, thus considering for example the order of p-adic vanishing. Finally, we exhibit a corollary of this multiplicity estimate in the context of interpolation on commutative algebraic groups, with applications to Transcendental Number Theory.

Introduction

The theory of resultants is an old branch of Mathematics which provides important tools, both computational and theoretical, in many other fields. One of the most classical versions of a resultant, named after Macaulay [Mac02], is defined for a sequence $\underline{f} = (f_0, \dots, f_r)$ of $r+1$ homogeneous polynomials in $r+1$ variables $\mathbf{x} = (x_0, \dots, x_r)$ over a field k. The Macaulay resultant of $\underline{f}$ is an irreducible polynomial of the unknown coefficients of $\underline{f}$ uniquely determined, up to a multiplication by a constant, by the following property: it vanishes if and only if the polynomials in $\underline{f}$ admit a nontrivial common zero over an algebraic closure of k. It turns out that such implicit characterization gives rise to a mathematical object that can be computed explicitly [MS10, EM99, CLO13] and that satisfies several remarkable properties [CLO06, Stu98, Jou91, Jou95]. In this paper we discuss the following statement, together with its generalizations and applications.

Theorem 0.1. Suppose that the polynomials $\underline{f}$ have exactly N common roots, counting multiplicities. Then the Macaulay resultant, considered as a polynomial function, vanishes with multiplicity at least N when specialized at the coefficients of $\underline{f}$.

This is a useful property, that can be interpreted either as a multiplicity estimate for the resultant, or as an upper bound for the number of solutions to the nonlinear system given by $f_i = 0$ for $i = 0, \dots, r$. This theorem can be shown with a variety of methods when $r = 1$: for

Date: December 9, 2019.

2010 *Mathematics Subject Classification.* Primary: 13P15, 13D02, 13H15, 14L99, 16W70; Secondary: 11J81, 13C15, 14C17, 16W50.

instance one may use a formula of Poisson [Poi02] that expressess the resultant of $\underline{f} = (f_0, f_1)$, up to a nonzero multiplicative constant, as the symmetric polynomial

$$\prod_{i=1}^{d_0}\prod_{j=1}^{d_1}(\alpha_{0,i} - \alpha_{1,j})$$

in the roots of f_0 and f_1. A version of Theorem 0.1 was proved by Roy [Roy13, Theorem 5.2] for all $r \geq 1$, under the hypothesis that $k = \mathbb{C}$ and all the polynomials have equal degree. If the polynomials $\underline{f}$ have integer coefficients, there is an interesting arithmetic analogue of Theorem 0.1 given by Chardin [Cha93]. In this setting one often normalizes the irreducible polynomial that defines the Macaulay resultant to have integer coefficients, so the resultant of $\underline{f}$ is an integer R. Chardin proves, under suitable hypotheses, that if p is a prime number and the polynomials $\underline{f}$ have N common zeros modulo p, then $p^N \,|R$. In fact we observe in Remark 3.4 that with this point of view it is possible to prove a statement that is stronger than Chardin's. In [SS96] Scheja and Storch treat Theorem 0.1 and its arithmetic analogue as expressions of the same phenomenon, by working on polynomial algebras over integrally closed Noetherian domains and using a sufficiently general notion of "vanishing". For a somewhat unsimilar study of the multiplicity of the different, which can be thought as a geometric analog of the resultant of a gradient system of equations, we refer to Aluffi and Cukierman [AC93]. The goal of this paper is to prove a multiplicity estimate for multigraded Chow forms, known also as Rémond resultants. The main purpose is to deduce a corollary with potential applications in Algebraic Independence and Transcendental Number Theory.
The Macaulay resultant is only one of several notions in the rich theory of resultants. Many of the approaches to this theory are algebraic and express the resultant by means of determinantal formulas, we refer to [Dem84, Jou91] without the intent of completeness. However most generalizations come from geometric interpretations of the concept of resultants. In Algebraic Geometry one has the notion of Chow forms attached to arbitrary projective subvarieties [Phi01]. For comparison, the Macaulay resultant is a Chow form for the projective variety $\mathbb{P}^r_k$. Chow forms are important Intersection-Theoretic invariants and are applied in the theory of Heights [Phi91]. Moreover, the Chow forms of toric varieties [CLS11, Ful16] are often called sparse resultants and are important for computational reasons [Stu94, EM99]. Further generalizations with a geometric flavour, such as mixed resultants, can be found in the monograph of Gelfand-Kapranov-Zelevinski [GKZ94]. Multigraded Chow forms, or Rémond resultants, are attached to sequences of multihomogeneous polynomials $\underline{f}$ and multiprojective varieties/schemes $V \subseteq \mathbb{P}^{n_1}_k \times \cdots \times \mathbb{P}^{n_q}_k$. The Rémond resultant of $(\underline{f}, V)$ is a (not necessarily irreducible [DKS13, Example 1.31]) polynomial of the unknown coefficients of $\underline{f}$ with the following property: it vanishes if and only if the polynomials in $\underline{f}$ admit a nontrivial common zero in V over an algebraic closure of k. This is a notion of resultants which encapsulates most of the above definitions [DKS13, Remark 1.39].
In order to prove the aforementioned multiplicity estimate, we show that the resultants of Rémond can be computed as Cayley determinants of suitable multigraded Koszul complexes. This addresses a gap in the literature and it has other consequences. For example, it implies that the multiprojective resultants satisfy several classical formulas, such as the one that expresses the resultant as a gcd of the maximal minors of the Sylvester map [GKZ94, Theorem 34, Appendix A].

A notion of resultant for multihomogeneous polynomials is important in applications, especially when the set of variables $\mathbf{x} = (x_0, \dots, x_r)$ decomposes naturally in independent subcollections $\mathbf{x} = \mathbf{x}^{(1)} \cup \ldots \cup \mathbf{x}^{(q)}$. This is often the case, for example, in Transcendental and Algebraic Independence Theory, for which we refer to the book [NP01]. In this theory one typically starts with a tuple of numbers $\underline{\xi}$ that one wishes to prove algebraically independent (or $\mathbb{Q}$-linearly independent, or else), and then takes suitable combinations of $\underline{\xi}$ to fabricate a set of points $\Sigma \subseteq G(\mathbb{C})$ of an algebraic group $G = G_1 \times \cdots \times G_q$. One then assumes that the numbers $\underline{\xi}$ are not algebraically independent and constructs so-called auxiliary functions f_i that vanish with high order at all points of Σ, in hope to find a contradiction. For a detailed account on this method, we refer to the book of Waldschmidt [Wal00]. If r is the dimension of G and the auxiliary functions $\underline{f} = (f_0, \dots, f_r)$ are polynomials, one may consider their resultant. Since the polynomials f_i vanish simultaneously on Σ with high multiplicity, it follows from a suitable version of Theorem 0.1 that their resultant vanishes with high multiplicity as well: we explore this matter in more detail in Section 4. This construction might be seen as a way to package the information of several auxiliary polynomials $\underline{f}$ into a single "larger" auxiliary polynomial, namely their resultant. For examples of how the information on the multiplicity of the resultants is used to derive Diophantine results, in the context of interpolation on the commutative algebraic group $\mathbb{G}_a \times \mathbb{G}_m$, we refer to [Roy13, Ghi15, NR16].

Plan of the paper and methodology

The paper is subdivided as follows. In Section 1 we review the basic definitions and results we need from multigraded Commutative Algebra and multiprojective Geometry. In Section 2 we introduce the multigraded Koszul complex, we define the resultant as the determinant of sufficiently high multidegree slices of this complex, and we show that this definition coincides with that of Rémond. In Section 3 we prove our main multiplicity estimates and finally in Section 4 we present an application to the interpolation theory on commutative algebraic groups.

As we already remarked, the resultant is an algebraic invariant with a geometric interpretation. The most classical versions of the theory of resultants are formulated for polynomial algebras $A[\mathbf{x}]$, whereas some abstract geometric resultants are attached to $\mathcal{O}_X$-vector bundles, and their twists, over r-dimensional schemes [GKZ94]. One may find a middle ground by considering the "M-resultant" attached to an $A[\mathbf{x}]$-module. This unorthodox approach is the one that we adopt in this paper, see Remark 2.12 for a discussion on the hypotheses on $A[\mathbf{x}]$ and M. One reason for this choice is the fact that the module "M" that lurks under the definition of a (multigraded) Chow form does not necessarily have the structure of a polynomial algebra, if the underlying scheme is not a toric variety. Nevertheless, the multihomogeneous components of this multigraded module are free (cfr. Section 1.4), and this hypothesis turns out to be sufficient to guarantee the validity of our constructions. Therefore it is natural and not more difficult to allow M to be an essentially arbitrary module with this property, instead of restricting it only to the modules that arise in the construction of Chow forms. See Remark 2.11 for the recovery of the classical theory and the theory of Rémond, and see Remark 2.13 for a comparison with the geometric generalizations.

The algebraic theory of resultants is intimately related with the notion of regular sequences, and this reflects their fundamental intersection-theoretic nature. In this paper we use instead the more general notion of filter-regular elements: these are like regular elements that

"disregard" the irrelevant associated primes of the module (Proposition 1.2). Filter-regular elements and sequences are thus natural objects in multigraded Commutative Algebra and in fact it turns out that the theory of regular sequences alone is ill-suited for developing the theory of multigraded resultants/Chow forms. The geometric reason is that the "multiaffine cone" of a multiprojective variety may have singularities at the "multi-vertex" that are not Cohen-Macaulay: this may prevent the very existence of regular sequences with right length, see Remark 3.11.

Let us briefly discuss the definition of the M-resultant attached to an $A[\mathbf{x}]$-module M and a filter-regular sequence $\underline{f}$. One approach, adopted e.g. by Rémond [Rém01], is to define it as the annihilant form (or content, see Definition 2.6) of any multihomogeneous component of module $M/(\underline{f})M$ with sufficiently high multidegree. Notice that the $M/(\underline{f})M$ is the cokernel of the multigraded linear map

$$\begin{aligned} \partial_1 : M \times \cdots \times M &\to M \\ (m_0, \ldots, m_r) &\mapsto f_0 m_0 + \cdots + f_r m_r, \end{aligned} \tag{0.1}$$

known as the Sylvester map. In particular the divisor of the resultant detects the primes $\mathfrak{p} \subseteq A$ for which every multigraded slice ∂_1^ν with sufficiently high multidegree ν fails to be locally surjective at $\mathfrak{p}$. The Sylvester map can be completed to the left to form the multigraded Koszul complex $\mathbf{K}_\bullet = \mathbf{K}_\bullet(\underline{f}, M)$. Another approach to the construction of the resultant is to define it as the Cayley determinant of a sufficiently high multidegree slice of $\mathbf{K}_\bullet$. In particular the resultant detects when localizations of $\mathbf{K}_\bullet$ fail to be exact.

The results of the paper are organized as follows. The basic properties of filter-regular elements are derived all throughout Section 1, and in Proposition 2.2 we verify that the Koszul complex $\mathbf{K}_\bullet(\underline{f}, M)$ is acyclic if $\underline{f}$ is a filter-regular sequence. In Proposition 2.7 we prove that the divisor $\mathrm{div}_A((M/(\underline{f})M)_\nu$ stabilizes for ν large enough. The idea for proving this key stabilization property is that the multiplicity of $(M/(\underline{f})M)_\nu$ at some prime should be seen as a local Hilbert function, as ν varies. Our approach is therefore different than the one of Rémond [Rém01, Theorem 3.3], that uses elimination theory, and than the usual cohomological approach, see Remark 2.10. In Proposition 2.8 and Theorem 2.14 we prove that the two definitions of the multigraded M-resultant, respectively via the annihilant and via the determinant of the Koszul complex, coincide. In Theorem 3.3 we prove the main "$\mathfrak{p}$-adic" multiplicity estimate for M-resultants, and in Theorem 3.8 we deduce a multiplicity estimate for the Rémond resultant, in a form more suitable for geometrical applications. In Section 4 we introduce the theory of interpolation on a commutative algebraic group embedded in multiprojective space, we describe the primary decomposition of the so-called interpolation ideal and we discuss its relation with the surjectivity of the evaluation map. Finally, in Theorem 4.8 we state our main corollary, which is a lower bound on the multiplicity of the Chow form of the group at a sequence of interpolation polynomials.

Acknowledgements

I would like to gratefully acknowledge my supervisor Prof.Damien Roy for his wholehearted support, his careful reading of this paper and his valuable advices. I also thank an anonymous referee for bibliographical suggestions. This work was supported in part by the full International Scholarship of the University of Ottawa and the FGPS, in part by the International Doctoral Scholarship 712230205087, and in part by NSERC.

Contents

1. Preliminaries on multigraded commutative algebra

1.1. **Multigraded rings and modules.** Let $\mathbb{N}_+$ denote the set of positive integers and let $q \in \mathbb{N}_+$ be given. We say that a ring R is *multigraded* (or $\mathbb{N}^q$-graded if q may not be clear from the context) if it admits a decomposition $R = \bigoplus_{\mathbf{d}\in\mathbb{N}^q} R_{\mathbf{d}}$ such that $R_{\mathbf{a}}R_{\mathbf{b}} \subseteq R_{\mathbf{a}+\mathbf{b}}$ for every $\mathbf{a}, \mathbf{b} \in \mathbb{N}^q$. An R-module M is multigraded if it decomposes as $M = \bigoplus_{\mathbf{d}\in\mathbb{N}^q} M_{\mathbf{d}}$ and $R_{\mathbf{a}}M_{\mathbf{b}} \subseteq M_{\mathbf{a}+\mathbf{b}}$ for every $\mathbf{a}, \mathbf{b} \in \mathbb{N}^q$. Every element of $\mathbb{N}^q$ is called a *multidegree*. For every $p = 1, \dots, q$ we denote by $\mathbf{e}_p$ the multidegree corresponding to the p-th canonical basis vector of $\mathbb{N}^q$, i.e. such that $\mathbf{e}_{p,j} = \delta_{p,j}$, where δ is the Kronecker symbol. We also let $\mathbf{0}$ and $\mathbf{1}$ to be the elements of $\mathbb{N}^q$ with all the components equal to 0 and 1 respectively. We introduce on $\mathbb{N}^q$ the componentwise partial order $\leq$, such that $\mathbf{d}^{(1)} \leq \mathbf{d}^{(2)}$ if and only if $\mathbf{d}_i^{(1)} \leq \mathbf{d}_i^{(2)}$ for every $i = 1, \dots, q$. Then we state that a property holds *for* $\mathbf{d} \in \mathbb{N}^q$ *large enough* if there exists $\mathbf{d}^{(0)} \in \mathbb{N}^q$ such that the property holds for every $\mathbf{d} \in \mathbb{N}^q$ satisfying $\mathbf{d} \geq \mathbf{d}^{(0)}$. For each $\mathbf{d} \in \mathbb{N}^{\mathbf{d}}$ we say that $M_{\mathbf{d}}$ is a multihomogeneous component of M of multidegree $\mathbf{d}$ and we call every element of $M_{\mathbf{d}}$ a multihomogeneous element of M of multidegree $\mathbf{d}$. A multigraded R-module M is *eventually zero* if $M_{\mathbf{d}} = \{0\}$ for $\mathbf{d}$ large enough.

Remark 1.1. In the case $q = 1$, the notions of multigraded rings and modules coincide with the more common notions of graded rings and modules. The reader interested only

in the graded case can read all this article by replacing everywhere the words multigraded, multihomogeneous and multiprojective with graded, homogeneous and projective respectively.

1.2. **Multihomogeneous submodules and relevant ideals.** We say that an R-submodule $N \subseteq M$ is multihomogeneous if it is generated by multihomogeneous elements or, equivalently, if $N = \bigoplus_{\mathbf{d}\in\mathbb{N}^q} N \cap M_{\mathbf{d}}$. Given a multihomogeneous submodule N of M we have induced R-module structures on N and M/N, respectively with $N_{\mathbf{d}} = N \cap M_{\mathbf{d}}$ and $(M/N)_{\mathbf{d}} = M_{\mathbf{d}}/N_{\mathbf{d}}$. An ideal $I \subseteq R$ is multihomogeneous if it is a multihomogeneous submodule of R. We say that a multihomogeneous submodule $N \subseteq M$ is *irrelevant* if $N_{\mathbf{d}} = M_{\mathbf{d}}$ for $\mathbf{d}$ large enough or, equivalently, if M/N is eventually zero. A multihomogeneous submodule $N \subseteq M$ is *relevant* if it is not irrelevant. A multihomogeneous ideal $I \subseteq R$ is relevant (irrelevant) if I is a relevant (irrelevant) submodule of R.
If $\mathcal{F} \subseteq R$ is a family of multihomogeneous elements of R, we denote by $(\mathcal{F})$ the multihomogeneous ideal generated by them. If $N \subseteq M$ is a multihomogeneous submodule of a multigraded R-module M, if $\mathcal{N} \subseteq M$ is any family of multihomogeneous elements of M and if $\mathcal{F}$ is any family of multihomogeneous elements of R, then the colon submodule (module quotient) $(N :_M \mathcal{F}) := \{m \in M : \ fm \in N \ \forall f \in \mathcal{F}\}$ is a multihomogeneous submodule of M and the colon ideal (ideal quotient) $(N :_R \mathcal{N}) := \{r \in R : \ r\eta \in N \ \forall \eta \in \mathcal{N}\}$ is a multihomogeneous ideal of R. In particular $\mathrm{Ann}_R(M) := (0 :_R M)$ is a multihomogeneous ideal.
Given an R-module M we denote by $\mathrm{Ass}_R(M)$ the set of associated primes of M in R. If M is a multigraded R-module and $\mathfrak{p} \in \mathrm{Ass}_R(M)$, then $\mathfrak{p}$ is a multihomogeneous prime ideal of R and is equal to $(0 :_M m)$ for some multihomogeneous element $m \in M$ [Rém01, Lemme 2.5]. If R is Noetherian and M is a finitely generated R-module, then $\mathrm{Ass}_R(M)$ is a finite set.

1.3. **Filter-regular sequences and f-depth.** Given a multigraded ring R, a multigraded R-module M and a multihomogeneous element $f \in R$, we say that f is *filter-regular* for M if the colon submodule $(0 :_M f)$ is eventually zero or, equivalently, if the multiplication by f induces injective maps $M_\nu \xrightarrow{\cdot f} M_{\nu+\mathbf{d}}$ for ν large enough, where $\mathbf{d}$ is the multidegree of f. In particular, if M is eventually zero then any multihomogeneous $f \in R$ is filter-regular for M. A collection $\underline{f} = (f_0, \dots, f_r)$ of multihomogeneous elements of R is a *filter-regular sequence* for M if f_i is filter-regular for the module $M/(f_0, \dots, f_{i-1})M$ for $i = 0, \dots, r$. For every multihomogeneous ideal J of R and every multigraded module M we define $\text{f-depth}(J, M) \in \mathbb{N} \cup \{\infty\}$ to be the supremum of all the $r \in \mathbb{N}$ such that there exists a filter regular sequence $\underline{f} = (f_0, \dots, f_{r-1})$ for M with $f_i \in J$ for $i = 0, \dots, r-1$. The following fact is easy to prove. See for example [VM13, Proposition 2.5].

Proposition 1.2. Let R be a Noetherian multigraded ring, M a finitely generated multigraded R-module and $f \in R$ a multihomogeneous element. Then f is filter-regular for M if and only if is not contained in any *relevant* associated prime of M in R.

The above proposition is useful to prove the existence of filter-regular elements, expecially when coupled with the following multihomogeneous version of the Prime Avoidance lemma.

Lemma 1.3. Let R be a Noetherian multigraded ring, let $\mathfrak{p}_1, \dots, \mathfrak{p}_s$ be relevant multihomogenous primes of R and let I be a multihomogenous ideal of R with $I \not\subset \mathfrak{p}_i$ for $i = 1, \dots, s$. Then for every $\nu \in \mathbb{N}^q$ large enough there exists $f \in I$ multihomogeneous of multidegree ν such that $f \notin \mathfrak{p}_i$ for $i = 1, \dots, s$.

Proof. We may assume there are no inclusions among the $\mathfrak{p}_i$. For $i = 1, \ldots, s$ let $J_i := I \prod_{j \neq i} \mathfrak{p}_j$. It is well known that if a prime ideal contains the product of some ideals, then it must contain one of them. Moreover, the product of multihomogenoeus ideals is multihomogeneous. Therefore, there is $x_i \in J_i$ multihomogeneous of multidegree, say, $\mathbf{d}^{(i)}$, such that $x_i \notin \mathfrak{p}_i$. Since $\mathfrak{p}_i$ is relevant and R is Noetherian, we have that for all $\nu \in \mathbb{N}^q$ large enough there is $y_i \in R_{\nu - \mathbf{d}^{(i)}}$ with $y_i \notin \mathfrak{p}_i$. Then $f = \sum_{i=1}^s x_i y_i$ has the required property. □

Remark 1.4. Filter-regular sequences are related to superficial sequences and (mixed) multiplicity systems, and are widely used in the study of Rees algebras and Hilbert functions of local rings. See for example [TV10], [RV10], [VT15] or [KR94].

1.4. **Multigraded polynomial rings and componentwise free modules.** Given an integer $q \in \mathbb{N}_+$ as in Section 1.1 and positive natural numbers $n_1, \ldots, n_q \in \mathbb{N}_+$, we introduce the set of variables $\mathbf{x} = (x_{p,i})_{p=1,\ldots,q,\, i=0,\ldots,n_p}$ and for every $p = 1, \ldots, q$ we denote by $\mathbf{x}_p$ the subcollection $\mathbf{x}_p = (x_{p,0}, \ldots, x_{p,n_p})$. If A is any ring, we denote by $A[\mathbf{x}]$ the polynomial ring with coefficients in A and variables in $\mathbf{x}$. We consider on $A[\mathbf{x}]$ the unique $\mathbb{N}^q$-graded ring structure such that every nonzero constant $a \in A$ has multidegree $\mathbf{0}$ and that $x_{p,i}$ is multihomogeneous of multidegree $\mathbf{e}_p$, for every $p = 1, \ldots, q$ and $i = 0, \ldots, n_p$. We define a *componentwise free $A[\mathbf{x}]$-module* to be a finitely generated multigraded $A[\mathbf{x}]$-module M whose multihomogeneous components $M_{\mathbf{d}}$ are free A-modules of finite ranks.

1.5. **The Hilbert polynomial and the relevant dimension.** Given a ring A, we denote by Mod_A the category of finitely generated A-modules. An *additive integer-valued function* on Mod_A is a mapping $\lambda : Mod_A \to \mathbb{Z}$ satisfying $\lambda(M) = \lambda(M') + \lambda(M'')$ for every short exact sequence $0 \to M' \to M \to M'' \to 0$ in Mod_A. If $\mathbb{F}$ is a field, R is an Artinian ring and A is an integral domain with field of fractions $\mathbb{F}$, then the dimension $\dim_{\mathbb{F}}(-)$, the length $\ell(-)$ and the *generic rank* $\operatorname{rank}_A(-) = \dim_{\mathbb{F}}(- \otimes_A \mathbb{F})$ are additive integer valued functions respectively on $Mod_{\mathbb{F}}$, Mod_R and Mod_A.
If A is a Noetherian ring and M is a finitely generated multigraded $A[\mathbf{x}]$-module, then every multihomogeneous component $M_{\mathbf{d}}$ is a finitely generated A-module. If λ is an additive integer-valued function on Mod_A, we introduce the *Hilbert function* $h_{M,\lambda} : \mathbb{N}^q \to \mathbb{Z}$ given by $h_{M,\lambda}(\mathbf{d}) = \lambda(M_{\mathbf{d}})$.

Proposition 1.5. Let A be a field, an Artinian ring or a Noetherian integral domain, and let $\lambda(-)$ be $\dim_A(-)$, $\ell(-)$ or $\operatorname{rank}_A(-)$ respectively, as above. Then for every finitely generated $A[\mathbf{x}]$-module M there is a unique polynomial $P_{M,\lambda}$ in q variables and with coefficients in $\mathbb{Q}$, called the *Hilbert polynomial*, such that $h_{M,\lambda}(\mathbf{d}) = P_{M,\lambda}(\mathbf{d})$ for every sufficiently large $\mathbf{d} \in \mathbb{N}^q$.

Proof. The case of an Artinian ring includes the case of a field, which in turn implies the case of an integral domain. The standard reference is [Van29], although it actually covers only the bigraded case over a field. For a modern and more complete treatment of the field case see [Rém01, Theorem 2.10] or [MS05, Lemma 2.8]. For the Artinian ring case see [TV10, Theorem 2.6] or [HHRT97]. □

In case A is a Noetherian integral domain with fraction field $\mathbb{F}$ we also use the notation $H_M := P_{M,\operatorname{rank}_A} = P_{M \otimes_{A[\mathbf{x}]} \mathbb{F}[\mathbf{x}], \dim_{\mathbb{F}}}$ for brevity. In case $H_M \equiv 0$ (i.e. if and only if $M \otimes_{A[\mathbf{x}]} \mathbb{F}[\mathbf{x}]$ is eventually zero) we set $\text{dim-r}_A(M) := -1$. Otherwise, we denote the total degree of H_M by $\text{dim-r}_A(M)$ and we call it the *relevant dimension* of M. If $\text{dim-r}_A(M) = 0$ or $\text{dim-r}_A(M) = -1$

the Hilbert polynomial H_M is a constant nonnegative integer, and we define the *relevant degree* $\deg\text{-r}_A(M) \in \mathbb{N}$ to be this integer.

1.6. **Multiprojective subschemes and multisaturation.** Given a field k and $n \in \mathbb{N}_+$ we denote by $\mathbb{P}^n_k = \operatorname{Proj}(k[X_0, \dots, X_n])$ the projective space of dimension n over k. Given an integer $q \in \mathbb{N}_+$ as above and a collection $\mathbf{n} = (n_1, \dots, n_q) \in \mathbb{N}_+^q$ of positive natural numbers, we define $\mathbb{P}^{\mathbf{n}}_k := \mathbb{P}^{n_1}_k \times \cdots \times \mathbb{P}^{n_q}_k$ and we call it a *multiprojective space*. It is a reduced irreducible scheme over $\operatorname{Spec} k$ of dimension $|\mathbf{n}| := n_1 + \ldots + n_q$. Following [Rém01, Section 2.5], we see that its underlying (Zariski) topological space is naturally set-theoretically in bijection with the set of relevant multihomogeneous prime ideals of $k[\mathbf{x}]$. In fact, to every closed subscheme Z of $\mathbb{P}^{\mathbf{n}}_k$, which we call a *multiprojective subscheme*, is attached a multihomogeneous ideal $I \subseteq k[\mathbf{x}]$, called the *ideal of definition* of Z and denoted by $\mathcal{I}(Z)$. Conversely, every multihomogeneous ideal $I \subseteq k[\mathbf{x}]$ defines a multiprojective subscheme $\mathcal{Z}(I)$ such that $\mathcal{Z}(\mathcal{I}(Z)) = Z$ for every multiprojective subscheme Z of $\mathbb{P}^{\mathbf{n}}_k$. For every multihomogeneous ideal $I \subseteq k[\mathbf{x}]$ we define its *multisaturation* by $\overline{I} := \mathcal{I}(\mathcal{Z}(I))$, so that the ideals in the image of $Z \mapsto \mathcal{I}(Z)$ are those satisfying $I = \overline{I}$.

Proposition 1.6. The following are equivalent definitions for the multisaturation of I.

(i) $\overline{I} = \{f \in k[\mathbf{x}] : \exists \mathbf{d}_f \in \mathbb{N}^q \ fk[\mathbf{x}]_{\mathbf{d}_f} \subseteq I\}$.
(ii) $\overline{I}$ is maximal among all multihomogeneous ideals J such that $J_{\mathbf{d}} = I_{\mathbf{d}}$ for $\mathbf{d}$ large enough.
(iii) $\overline{I}$ is the intersection of the primary ideals of $k[\mathbf{x}]$ appearing in a minimal primary decomposition of I and corresponding to relevant primes.

Proof. (i) is proved in [Rém01, Proposition 2.17]. For (ii), the inclusion $I \subseteq \overline{I}$ is clear. $\overline{I}$ is generated by finitely many multihomogeneous elements $f_1, \dots, f_r$ and by (i) there are $\mathbf{d}_{f_1}, \dots, \mathbf{d}_{f_r} \in \mathbb{N}^q$ such that $f_i k[\mathbf{x}]_{\mathbf{d}_{f_i}} \subseteq I$. If $\mathbf{d}_1$ is an upperbound for the multidegrees of $f_1, \dots, f_r$ and if $\mathbf{d}_2$ is an upperbound for $\mathbf{d}_{f_1}, \dots, \mathbf{d}_{f_r}$, then for every $\mathbf{d} \geq \mathbf{d}_1 + \mathbf{d}_2$ we have $\overline{I}_{\mathbf{d}} = I_{\mathbf{d}}$. Moreover, if J is a multihomogeneous ideal of $k[\mathbf{x}]$ such that $J_{\mathbf{d}} = I_{\mathbf{d}}$ for $\mathbf{d}$ large enough, then $J \subseteq \overline{I}$ by (i). Finally, (iii) is a consequence of [Rém01, Lemme 2.4][(1)]. □

1.7. **More on the relevant dimension.** Given a Noetherian integral domain A with fraction field $\mathbb{F}$ and a finitely generated $A[\mathbf{x}]$-module M we defined the relevant dimension $\dim\text{-r}_A(M)$ in terms of the total degree of the Hilbert polynomial $P_{M,\operatorname{rank}_A}$. We denote by $\dim Z$ the dimension of a multiprojective subscheme Z of $\mathbb{P}^{\mathbf{n}}_{\mathbb{F}}$, and for a prime ideal $\mathfrak{p} \subseteq A[\mathbf{x}]$ we let $\dim(M_{\mathfrak{p}})$ be the Krull dimension of the module $M_{\mathfrak{p}}$, defined in terms of chains of prime ideals of $A[\mathbf{x}]_{\mathfrak{p}}$ containing the annihilator of $M_{\mathfrak{p}}$ [BH98, Appendix]. Then we define

$$e(M) := \max\{|\mathbf{n}| - \operatorname{ht}(\mathfrak{p}) : \ \mathfrak{p} \subseteq \mathbb{F}[\mathbf{x}] \text{ relevant prime}, \operatorname{Ann}_{\mathbb{F}[\mathbf{x}]}(M \otimes_A \mathbb{F}) \subseteq \mathfrak{p}\},$$

where $\operatorname{ht}(\mathfrak{p})$ denotes the height of $\mathfrak{p}$.

Proposition 1.7. Let A be a Noetherian integral domain with fraction field $\mathbb{F}$ and M a finitely generated $A[\mathbf{x}]$-module. Then

$$\dim\text{-r}_A(M) = \dim\text{-r}_{\mathbb{F}}(M \otimes_A \mathbb{F}) = \dim \mathcal{Z}(\operatorname{Ann}_{\mathbb{F}[\mathbf{x}]}(M \otimes_A \mathbb{F})) = e(M) = \max_{\mathfrak{p}} \dim(M_{\mathfrak{p}}),$$

where in the rightmost formula $\mathfrak{p}$ ranges through the relevant multihomogeneous primes of $A[\mathbf{x}]$ such that $\mathfrak{p} \cap A = (0)$.

[(1)] By (i), our $\overline{I}$ coincides with the characteristic ideal $\mathfrak{U}_{\emptyset}(I)$ of Rémond.

Proof. The first equality is clear, the second and the third are essentially proved in [Rém01, Theorem 2.10, Section 2.5], the last follows from the fact that the primes of $\mathbb{F}[\mathbf{x}]$ are in bijection with the primes of $A[\mathbf{x}]$ such that $\mathfrak{p} \cap A = (0)$. □

The next lemma shows that the operation of quotienting by a filter-regular sequence has the effect of decreasing the total degree of the Hilbert polynomial, by an amount at least equal to the lenght of the sequence.

Lemma 1.8. Let R be an Artinian ring and M a finitely generated multigraded $R[\mathbf{x}]$-module. Let $f \in R[\mathbf{x}]$ be a filter-regular element of multihomogeneous degree $\mathbf{d}$ for M and λ be the length function on Mod_R. If $P_{M,\lambda}$ is not the zero polynomial, then the total degree of $P_{M/fM,\lambda}$ is at least one less the total degree of $P_{M,\lambda}$. If $\mathbf{d} \geq \mathbf{1}$ then this inequality is indeed an equality.

Proof. Since $f \in R[\mathbf{x}]$ is filter-regular for M we have a short exact sequence $0 \to M_\nu \to M_{\nu+\mathbf{d}} \to (M/fM)_{\nu+\mathbf{d}} \to 0$ for ν large enough. From the additivity of λ we get $P_{M/fM,\lambda}(\nu + \mathbf{d}) = P_{M,\lambda}(\nu+\mathbf{d}) - P_{M,\lambda}(\nu)$ for ν large enough, which implies the first statement by inspection. The second part is similar, and uses the fact that the coefficients of a Hilbert polynomial corresponding to monomials of highest total degree are nonnegative [TV10, Theorem 2.6]. □

Corollary 1.9. Let A be a Noetherian integral domain, let M be a finitely generated $A[\mathbf{x}]$-module and let J be a multihomogeneous ideal of $A[\mathbf{x}]$. Then

$$\text{dim-r}_A(M/JM) \leq \max\{-1,\, \text{dim-r}_A(M) - \text{f-depth}(J, M)\}\,.$$

Remark 1.10. To see that the hypothesis $\mathbf{d} \geq \mathbf{1}$ in Lemma 1.8 is necessary, take $q = 2$, $n_1 = 2$, $n_2 = 1$, $M = A[\mathbf{x}]/(x_{2,1})$ and $f = x_{2,0}$, for which $\text{dim-r}_A(M) = 2$ and $\text{dim-r}_A(M/fM) = -1$. However, it is often possible to weaken this condition: see [Rém01, Theorem 2.10, (3)].

2. Koszul complexes and resultants

2.1. **Multigraded Koszul complexes.** Given a commutative ring R, an R-module M and a sequence $\underline{f} = (f_0, \ldots, f_r)$ of elements of R, the Koszul complex $\mathbf{K}_\bullet(\underline{f}, M)$ is a finite complex of R-modules given by

$$\mathbf{K}_\bullet(\underline{f}, M) \;:=\; 0 \to (\bigwedge^{r+1} L) \otimes M \xrightarrow{\partial_{r+1}} \ldots \xrightarrow{\partial_2} (\bigwedge^{1} L) \otimes M \xrightarrow{\partial_1} M \to 0,$$

where L is the free R-module R^{r+1} equipped with a basis $(e_0, \ldots, e_r)$, the tensor products are taken over R, and the differentials ∂_p are defined by

$$\partial_p(e_{i_1} \wedge \cdots \wedge e_{i_p} \otimes m) = \sum_{s=1}^{p} (-1)^{s+1} f_{i_s} e_{i_1} \wedge \cdots \wedge \widehat{e_{i_s}} \wedge \cdots \wedge e_{i_p} \otimes m.$$

The homology modules $\text{H}_p(\mathbf{K}_\bullet(\underline{f}, M))$ are denoted by $\text{H}_p(\underline{f}, M)$ for short and their direct sum $\text{H}_\bullet(\underline{f}, M)$ is called the *Koszul homology* of the sequence $\underline{f}$ with coefficients in M. For the 0-th and $(r+1)$-th homology modules we have the natural isomorphisms $\text{H}_0(\underline{f}, M) \cong M/(\underline{f})M$, $\text{H}_{r+1}(\underline{f}, M) \cong (0 :_M (\underline{f}))$. Moreover, the annihilator $\text{Ann}_R(\text{H}_\bullet(\underline{f}, M))$ contains both $\text{Ann}_R(M)$ and the ideal $(\underline{f})$. We refer to Section 1.6 of [BH98] for more on the general theory of Koszul complexes.

Suppose now that R is multigraded as in Section 1.1, M is a multigraded R-module, $\mathbb{d} = (\mathbf{d}^{(0)}, \ldots, \mathbf{d}^{(r)})$ is a collection of nonzero multidegrees and $\underline{f} = (f_0, \ldots, f_r)$ is a sequence

of multihomogeneous elements of R with multidegrees prescribed by $\mathtt{d}$. Then we can introduce on the R-modules $\mathbf{K}_p(\underline{f}, M) = (\bigwedge^p L) \otimes_R M$ the natural $\mathbb{N}^q$-grading for which $\deg_{\mathbb{N}^q}(e_{i_1} \wedge \cdots \wedge e_{i_p} \otimes m) = \mathbf{d}^{(i_1)} + \cdots + \mathbf{d}^{(i_p)} + \deg_{\mathbb{N}^q}(m)$, for m multihomogeneous. This is also done in [VT15, Section 3] and is similar to the homogeneous case [BH98, Remark 1.6.15] [Cha93]. We notice that the differentials preserve this grading, so that the homology modules inherit a multigraded structure. We then write $\mathbf{K}^\nu_\bullet(\underline{f}, M)$ and $\mathrm{H}^\nu_\bullet(\underline{f}, M)$ for the component of multidegree ν respectively of the Koszul complex and of the Koszul homology. If we denote the restricted differentials by

$$\partial^\nu_p : \mathbf{K}^\nu_p(\underline{f}, M) \to \mathbf{K}^\nu_{p-1}(\underline{f}, M)$$

then for every $\nu \in \mathbb{N}^q$ we see that $\mathbf{K}^\nu_\bullet(\underline{f}, M)$ is a complex of $R_{\mathbf{0}}$-modules with differentials ∂^ν_p and homology $\mathrm{H}^\nu_\bullet(\underline{f}, M)$.
The next proposition is an adaptation to filter-regular sequences of a classical result that relates the existence of regular sequences to the vanishing of higher Koszul homology. We give a proof along the lines of [Nor68, Section 8.5, Theorem 6], that uses the following definition.

Definition 2.1. Let R be a Noetherian multigraded ring, $\underline{f} = (f_0, \ldots, f_{s-1})$ a sequence of s multihomogeneous elements of R, and M a finitely generated multigraded R-module. If there is at least one integer $\lambda \in \{1, \ldots, s\}$ such that $\mathrm{H}_\lambda(\underline{f}, M)$ is not eventually zero, we define $\lambda(\underline{f}, M)$ to be the largest such integer. Otherwise, we set $\lambda(\underline{f}, M) := -\infty$.

Proposition 2.2. Let R, M and $\underline{f}$ be as in Definition 2.1, and let $J = (\underline{f})$. We have:

(i) if $\beta \in J$ is filter-regular for M, then $\lambda(\underline{f}, M/\beta M) = \lambda(\underline{f}, M) + 1$, where we let $-\infty + 1 := -\infty$;
(ii) f-depth$(J, M) = s - \lambda(\underline{f}, M)$, with f-depth$(J, M)$ as in Section 1.3 and $s - (-\infty) := \infty$;
(iii) if $\underline{f}$ is a filter-regular sequence for M, then $\mathbf{K}^\nu_\bullet(\underline{f}, M)$ is acyclic (i.e. its p-th homology modules vanish for $p \geq 1$) for ν large enough.

Proof. Let $\beta \in J$ be filter-regular for M. By definition, β is a multihomogeneous element of R. Let $\mathbf{d} \in \mathbb{N}^q$ be its multidegree. The R-module $M/\beta M$ is finitely generated and multigraded, so $\lambda(\underline{f}, M/\beta M)$ is defined. For every $\nu \in \mathbb{N}^q$ large enough we have an exact sequence

$$0 \to M_{\nu - \mathbf{d}} \xrightarrow{\beta} M_\nu \to (M/\beta M)_\nu \to 0,$$

where the first map is induced by the multiplication by β in M. The collection of these maps induce a long exact sequence in Koszul homology that at the level of multihomogeneous components takes the form

$$\to \mathrm{H}_\mu(\underline{f}, M)_{\nu - \mathbf{d}} \xrightarrow{\beta} \mathrm{H}_\mu(\underline{f}, M)_\nu \to \mathrm{H}_\mu(\underline{f}, M/\beta M)_\nu \to \mathrm{H}_{\mu-1}(\underline{f}, M)_{\nu - \mathbf{d}} \to$$

Since $\mathrm{H}_\mu(\underline{f}, M)$ is annihilated by all elements of J, the above exact sequence simplifies to

$$0 \to \mathrm{H}_\mu(\underline{f}, M)_\nu \to \mathrm{H}_\mu(\underline{f}, M/\beta M)_\nu \to \mathrm{H}_{\mu-1}(\underline{f}, M)_{\nu - \mathbf{d}} \to 0$$

For $\mu > \lambda(\underline{f}, M) + 1$ both $\mathrm{H}_\mu(\underline{f}, M)$ and $\mathrm{H}_{\mu-1}(\underline{f}, M)$ are eventually zero modules and hence we obtain $\mathrm{H}_\mu(\underline{f}, M/\beta M)_\nu = 0$ as well for sufficiently large ν. In particular, if $\lambda(\underline{f}, M) = -\infty$ we have $\lambda(\underline{f}, M/\beta M) = -\infty$ as well. On the other hand if $\lambda(\underline{f}, M) \geq 0$ and $\mu = \lambda(\underline{f}, M) + 1$, we obtain an isomorphism

$$\mathrm{H}_\mu(\underline{f}, M/\beta M)_\nu \cong \mathrm{H}_{\mu-1}(\underline{f}, M)_{\nu - \mathbf{d}}$$

for sufficiently large $\nu \in \mathbb{N}^q$, which shows that $\mathrm{H}_\mu(\underline{f}, M/\beta M)$ is not eventually zero. Therefore, we have $\lambda(\underline{f}, M/\beta M) = \lambda(\underline{f}, M) + 1$, which is (i).
To prove (ii), first suppose that f-depth$(J, M) = 0$. This means that no element of J is filter-regular for M. In this case all multihomogeneous elements of J are contained in the union of the relevant associated primes of M by Proposition 1.2, and so by Lemma 1.3 all of J is contained in one of them, say $\mathfrak{p}$. Write $\mathfrak{p} = (0 :_R m)$ for a multihomogeneous element $m \in M$. Since $\mathfrak{p}$ is a relevant prime, Rm is a module which is not eventually zero and is contained into the colon module $(0 :_M \mathfrak{p})$, which in turn is contained in $(0 :_M J)$. This proves that $\mathrm{H}_s(\underline{f}, M) = (0 :_M J)$ is not eventually zero and so $\lambda(\underline{f}, M) = s$.
Now assume that f-depth$(J, M) > 0$ and f-depth$(J, M) \neq \infty$. Then by definition there exists $\beta \in J$ that is filter-regular for M and f-depth$(J, M/\beta M) =$ f-depth$(J, M) - 1$. Then we have, by induction on f-depth(J, M) and (i) above, that

$$\text{f-depth}(J, M) = \text{f-depth}(J, M/\beta M) + 1 = s + 1 - \lambda(\underline{f}, M/\beta M) = s - \lambda(\underline{f}, M).$$

On the other hand, if f-depth$(J, M) = \infty$, we can find a filter-regular sequence $\underline{\beta} = (\beta_1, \ldots, \beta_n)$ for M, with n arbitrarily large. Let $N = M/(\underline{\beta})M$. Then by repeatedly using (i) we get $\lambda(\underline{f}, N) = \lambda(\underline{f}, M) + n$. However, we clearly have $\lambda(\underline{f}, N) \leq s$, so we get a contradiction if $\lambda(\underline{f}, M) \geq 0$ and $n > s$.
Finally, suppose that $\underline{f}$ is a filter-regular sequence for M. Then f-depth$(J, M) \geq s$ and so, by (ii) above, we get $\lambda(\underline{f}, M) \leq 0$. This exactly means that $\mathbf{K}^\nu_\bullet(\underline{f}, M)$ is acyclic for ν large enough. □

Remark 2.3. Another approach to prove Proposition 2.2 is to to use the fact that a multihomogeneous element $f \in R$ is filter-regular for M if and only if it is regular for $M_{\geq \mathbf{d}} = \bigoplus_{\mathbf{d}' \geq \mathbf{d}} M_{\mathbf{d}'}$ for some $\mathbf{d} \in \mathbb{N}^q$.

Remark 2.4. Proposition 2.2 also shows that all maximal filter-regular sequences for M in J have the same number of elements.

2.2. **Contents and divisors of torsion modules.** Given a Noetherian integral domain A and a finitely generated A-module M, we say that M is a *torsion module* if $\mathrm{Ann}_A(M) \neq 0$. If $\mathfrak{p}$ is a prime ideal of A, then the localization $M_\mathfrak{p}$ is not the zero module if and only if $\mathrm{Ann}_A(M) \subseteq \mathfrak{p}$. In particular, choosing $\mathfrak{p} = \{0\}$, we see that M is a torsion A-module if and only if $M \otimes_A \mathbb{F} = 0$, where $\mathbb{F}$ is the field of fractions of A. Moreover, if $\mathfrak{p}$ is a prime ideal of height 1 and M is a torsion A-module, then $M_\mathfrak{p}$ is a torsion $A_\mathfrak{p}$-module and thus it has finite length $\ell(M_\mathfrak{p})$. This length is nonzero if and only if $\mathfrak{p}$ is a minimal associated prime of M in A.

Definition 2.5. If A is a Noetherian integral domain, we denote by $\mathrm{Div}(A)$ the free abelian group generated by the primes $\mathfrak{p}$ of A of height 1. If M is a torsion A-module we define $\mathrm{div}_A(M) \in \mathrm{Div}(A)$ by

$$\mathrm{div}_A(M) := \sum_\mathfrak{p} \ell(M_\mathfrak{p})[\mathfrak{p}],$$

where the sum ranges over all primes $\mathfrak{p}$ of A of height 1. If M is not torsion, we define $\mathrm{div}(M) = 0$.

We refer to [Bou72][Chap. 7, par.4] for the theory of divisors of torsion modules. In case A is an UFD ring, every prime of height 1 is principal, generated by an irreducible (prime) element $\pi \in A$, well defined up to multiplication by a unit $u \in A^\times$.

Definition 2.6. If A is an UFD ring, M is a torsion A-module and $\mathrm{irr}(A)$ is a choice of representatives for the irreducible elements of A, we define the *content* $\chi_A(M) \in A$ of M by the formula

$$\chi_A(M) := \prod_{\pi \in \mathrm{irr}(A)} \pi^{\ell(M_{(\pi)})}.$$

In elimination theory, the content of a torsion module is sometimes called *annihilant form* [DD00, Definition 1.22]. This notion is also related to the MacRae invariants and the zeroth Fitting ideals.
The following is a technical result that we will use in the next paragraph to be able to define the resultant. Together with Corollary 2.16 below, it generalizes [Rém01, Theorem 3.3], but our proof is considerably different, since we cannot make use of multihomogeneous elimination theory here. Instead, we make the key observation that the multiplicities appearing in the divisors under consideration can be computed as local Hilbert functions. Then to prove that they are eventually constant, it suffices to show that the corresponding Hilbert polynomials have degree zero.

Proposition 2.7. Let A be a Noetherian integral domain and M a finitely generated multigraded $A[\mathbf{x}]$-module that is projective as an A-module. Let also $\text{dim-r}_A(M) = r$ and $\underline{f} = (f_0, \dots, f_r)$ be a filter-regular sequence for M in $A[\mathbf{x}]$. Then there is $\nu_0 \in \mathbb{N}^q$ such that $\mathrm{div}_A((M/(\underline{f})M)_\nu) = \mathrm{div}_A((M/(\underline{f})M)_{\nu_0}) \neq 0$ for every $\nu \geq \nu_0$.

Proof. Let $N = M/(\underline{f})M$ and let $\mathbb{F}$ be the fraction field of A. Since $\underline{f}$ is a filter-regular sequence for M of length $\text{dim-r}_A(M) + 1$ we see that $\text{dim-r}_A(N) = -1$ by Corollary 1.9. This implies that $(N_\nu) \otimes_A \mathbb{F} = 0$ for $\nu \in \mathbb{N}^q$ large enough, which is equivalent to say that N_ν is a torsion A-module, or that $\mathrm{Ann}_A(N_\nu) \neq 0$.
We now show that the ideal $\mathrm{Ann}_A(N_\nu)$ is constant for ν large enough. Indeed M is generated as an $A[\mathbf{x}]$-module by finitely many elements with multidegrees bounded above by some $\nu_1 \in \mathbb{N}^q$. For $\nu_1 \leq \nu \leq \nu'$ we have $\mathrm{Ann}_A(N_\nu) \subseteq \mathrm{Ann}_A(N_{\nu'})$ and so we conclude by the noetherianity of A. The discussion preceding Definition 2.5 shows that a prime $\mathfrak{p}$ of height 1 appears in $\mathrm{div}_A(N_\nu)$ if and only if $\mathfrak{p} \supseteq \mathrm{Ann}_A(N_\nu)$. Since the latter is constant for ν large enough, we deduce that also the prime ideals appearing in $\mathrm{div}_A(N_\nu)$ form a fixed finite set for ν large enough.
Let $\mathfrak{p}$ be such a prime and let $(-)_\mathfrak{p}$ denote the localization at that prime. We will show that the number $\ell((N_\nu)_\mathfrak{p})$ is fixed for ν large enough. Let π be any nonzero element of $(\mathrm{Ann}_A(N_\nu))_\mathfrak{p} = \mathrm{Ann}_{A_\mathfrak{p}}((N_\nu)_\mathfrak{p}) \subseteq \mathfrak{p}A_\mathfrak{p}$ and let $L_\mathfrak{p} := M_\mathfrak{p}/(\pi)M_\mathfrak{p}$. Since the sequence $\underline{f}$ is filter-regular for M, we deduce by Proposition 2.2 (iii) that $\mathrm{H}^\nu_i(\underline{f}, M) = 0$ for $i = 1, \dots, r+1$ and ν large enough. Since $A_\mathfrak{p}$ is a flat A-module, if we apply the localization functor $(-)_\mathfrak{p} = (-) \times_A A_\mathfrak{p}$ to an exact sequence of A-modules (i.e. with trivial homology) we get an exact sequence of A-modules or, alternatively, of $A_\mathfrak{p}$-modules. Therefore $\mathrm{H}^\nu_i(\underline{f}, M \otimes_A A_\mathfrak{p}) = 0$ for $i = 1, \dots, r+1$ and ν large enough, where we still denote by $\underline{f}$ the induced sequence of elements in $A_\mathfrak{p}[\mathbf{x}]$. We have that $M_\mathfrak{p}$ is a finitely generated $A_\mathfrak{p}[\mathbf{x}]$-module and $A_\mathfrak{p}$ is a Noetherian integral domain with $\mathbb{F}$ as its fraction field. Therefore $\text{dim-r}_{A_\mathfrak{p}}(M_\mathfrak{p}) = \text{dim-r}_\mathbb{F}(M \otimes_A \mathbb{F}) = \text{dim-r}_A(M) = r$ by Proposition 1.7. Moreover, each multihomogeneous component of M, being a direct summand of M, is a (finitely generated) projective A-module. Therefore every multihomogeneous component of $M_\mathfrak{p}$ is a free $A_\mathfrak{p}$-module of finite rank. In other words, $M_\mathfrak{p}$ is a componentwise free $A_\mathfrak{p}[\mathbf{x}]$-module. Since π is nonzero in $A_\mathfrak{p}$, we have a short exact sequence $0 \to M_\mathfrak{p} \xrightarrow{\alpha} M_\mathfrak{p} \xrightarrow{\beta} L_\mathfrak{p} \to 0$, where

α is induced by the multiplication by π and β is the canonical projection. This short exact sequence induces a long exact sequence (of $A_{\mathfrak{p}}[\mathbf{x}]$-modules) in Koszul homology, which at the level of multihomogeneous components reads as

$$\cdots \to \mathrm{H}_i^{\nu}(\underline{f}, M_{\mathfrak{p}}) \to \mathrm{H}_i^{\nu}(\underline{f}, L_{\mathfrak{p}}) \to \mathrm{H}_{i-1}^{\nu}(\underline{f}, M_{\mathfrak{p}}) \to \cdots,$$

from which we deduce that $\mathrm{H}_i^{\nu}(\underline{f}, L_{\mathfrak{p}}) = 0$ for $i = 2, \ldots, r+1$ and ν large enough. In other words, by Definition 2.1, we have $\lambda(\underline{f}, L_{\mathfrak{p}}) \leq 1$. Therefore, by Proposition 2.2 (ii), we have f-depth$(\underline{f}, L_{\mathfrak{p}}) \geq r$, which means there exists a sequence $\underline{g} = (g_0, \ldots, g_{r-1})$ of multihomogeneous elements of $A_{\mathfrak{p}}[\mathbf{x}]$ contained in the ideal $(\underline{f})$ which is filter-regular for the multigraded module $L_{\mathfrak{p}}$. Let now $k = A_{\mathfrak{p}}/(\pi)A_{\mathfrak{p}}$, which is an Artinian ring, because $\mathfrak{p}A_{\mathfrak{p}}$ is a prime of height 1 in the integral domain $A_{\mathfrak{p}}$ and π is a nonzero element of $\mathfrak{p}A_{\mathfrak{p}}$. Let $p : A_{\mathfrak{p}} \to k$ be the natural projection and let λ be the length function on Mod_k as in Section 1.5. We notice that $L_{\mathfrak{p}}$ has a $k[\mathbf{x}]$-module structure that induces the $A_{\mathfrak{p}}[\mathbf{x}]$-module structure. In particular, the sequence $p(\underline{g})$ is still a filter-regular sequence of length r for $L_{\mathfrak{p}}$. Moreover it's easy to see that $L_{\mathfrak{p}}$ is a componentwise free $k[\mathbf{x}]$-module and that it satisfies the equality $\lambda((L_{\mathfrak{p}})_{\nu}) = \lambda(k) \cdot \mathrm{rank}_{A_{\mathfrak{p}}}((M_{\mathfrak{p}})_{\nu})$, from which we deduce that the Hilbert polynomial $P_{L_{\mathfrak{p}},\lambda}$ has degree r. Then, a repeated use of Lemma 1.8 shows that $P_{L_{\mathfrak{p}}/(p(\underline{g}))L_{\mathfrak{p}},\lambda}$ has degree at most zero, and so is eventually constant. Since $M_{\mathfrak{p}}/(\pi, \underline{f})M_{\mathfrak{p}}$ is a quotient of $L_{\mathfrak{p}}/(p(\underline{g}))L_{\mathfrak{p}}$, we have that the Hilbert function $\mathbf{d} \mapsto \lambda((M_{\mathfrak{p}}/(\pi, \underline{f})M_{\mathfrak{p}})_{\mathbf{d}})$ is constant as well, for $\mathbf{d}$ large enough. By our choice of π, for ν large enough the A-module $(M_{\mathfrak{p}}/(\pi, \underline{f})M_{\mathfrak{p}})_{\nu}$ is nothing but $(N_{\mathfrak{p}})_{\nu}$, and its lenght is the same whether we consider it as a k-module or as an $A_{\mathfrak{p}}$-module. Therefore we deduce that $\ell((N_{\nu})_{\mathfrak{p}})$ is constant for ν large enough. □

2.3. **Cayley determinants and resultants.** Let A be a Noetherian integral domain with fraction field $\mathbb{F}$ and let $\mathbf{C}_{\bullet}$ be a finite complex of A-modules

$$0 \to C_s \xrightarrow{d_s} \ldots \xrightarrow{d_1} C_0 \to 0.$$

We say that $\mathbf{C}_{\bullet}$ is *generically exact* if the complex $\mathbf{C}_{\bullet} \otimes_A \mathbb{F}$ is an exact sequence of $\mathbb{F}$-vector spaces or, equivalently, if all the homology modules of $\mathbf{C}_{\bullet}$ are torsion A-modules. If $\mathbf{C}_{\bullet}$ is a finite generically exact complex of free A-modules of finite rank and $\{\underline{b}_i\}_{0 \leq i \leq s}$ is a system of A-bases for the modules C_i, we can find a partition $\underline{b}_i = \underline{b}_i' \cup \underline{b}_i''$, with $\underline{b}_0'' = \underline{b}_s' = \emptyset$, inducing a decomposition $C_i = C_i' \oplus C_i''$, such that the matrix representations of the differentials d_i take the form $\left(\begin{smallmatrix} a_i & \phi_i \\ b_i & c_i \end{smallmatrix}\right)$, where the ϕ_i are square matrices with nonzero determinant. Then the *Cayley determinant* of the complex $\mathbf{C}_{\bullet}$ with respect to the above choices of A-bases and partitions is the element of $\mathbb{F}^{\times}$ given by $\prod_{i=1}^{s} \det(\phi_i)^{(-1)^{i+1}}$. It can be shown that another choice of A-bases and partitions changes this value by multiplication with an invertible element of A. Therefore, we can define unambiguously an element $\det_A(\mathbf{C}_{\bullet}) \in \mathbb{F}^{\times}/A^{\times}$, which we still call the determinant of $\mathbf{C}_{\bullet}$. For more on the Cayley determinant, see [GKZ94, Appendix A].

Proposition 2.8. Let A be a Noetherian UFD ring, M a componentwise free $A[\mathbf{x}]$-module with dim-r$_A(M) = r$ and $\underline{f} = (f_0, \ldots, f_r)$ a filter-regular sequence for M. Then $\mathbf{K}_{\bullet}^{\nu}(\underline{f}, M)$ is generically exact for ν large enough and

$$\det_A(\mathbf{K}_{\bullet}^{\nu}(\underline{f}, M)) = \chi_A((M/(\underline{f})M)_{\nu}) \pmod{A^{\times}}$$

for $\nu \in \mathbb{N}^q$ large enough.

Proof. Let $N = M/(\underline{f})M$. Since $\underline{f}$ is a filter-regular sequence for M of length $\dim\text{-r}_A(M)+1$, we see that $\dim\text{-r}_A(N) = -1$ by Corollary 1.9, so $N_\nu \otimes \mathbb{F} = 0$ and $N_\nu = \mathrm{H}_0^\nu(\underline{f}, M)$ is torsion, if ν is large enough. Moreover, by Proposition 2.2 (iii) we see that $\mathrm{H}_p^\nu(\underline{f}, M) = 0$ for all ν large enough and all $p \geq 1$. Therefore $\mathbf{K}_\bullet^\nu(\underline{f}, M)$ is generically exact for ν large enough, and so we can consider $\det_A(\mathbf{K}_\bullet^\nu(\underline{f}, M))$. Let D_ν any element of $\mathbb{F}$ representing it, and denote by $\mathrm{ord}_\pi : \mathbb{F} \to \mathbb{Z} \cup \{\infty\}$ the valuation associated to any prime element $\pi \in A$. The thesis then amounts to proving that, for ν large enough, $\mathrm{ord}_\pi(D_\nu) = \mathrm{ord}_\pi(\chi_A(N_\nu))$ for every prime element π of A. However, the right-hand side equals $\ell((N_\nu)_{(\pi)})$ by definition, whereas the left-hand side equals $\sum_i (-1)^i \ell(\mathrm{H}_i^\nu(\underline{f}, M)_{(\pi)})$ by [GKZ94, Theorem 30, Appendix A, p.493] (cfr. also [Cha93, Proposition 2]). Since $N_\nu = \mathrm{H}_0^\nu(\underline{f}, M)$ and $\mathrm{H}_p^\nu(\underline{f}, M) = 0$ for all ν large enough and all $p \geq 1$, the thesis follows. □

We now remark that Proposition 2.7 and Proposition 2.8 together imply that there exists $\nu_0 \in \mathbb{N}^q$ such that $\det_A(\mathbf{K}_\bullet^\nu(\underline{f}, M)) = \det_A(\mathbf{K}_\bullet^{\nu_0}(\underline{f}, M))$ for every $\nu \geq \nu_0$. In other words, $\det_A(\mathbf{K}_\bullet^\nu(\underline{f}, M)$ stabilizes at a well-defined nonzero element of $A/A^\times \subseteq \mathbb{F}/A^\times$, for ν large enough.

Definition 2.9. Let A be a Noetherian UFD ring, M a componentwise free $A[\mathbf{x}]$-module with $\dim\text{-r}_A(M) = r$ and $\underline{f} = (f_0, \ldots, f_r)$ a filter-regular sequence for M. Then we define the *M-resultant* $\mathrm{RES}_A(\underline{f}, M) \in A/A^\times$ of $\underline{f}$ with respect to M by

$$\mathrm{RES}_A(\underline{f}, M) := \det_A(\mathbf{K}_\bullet^\nu(\underline{f}, M))$$

for $\nu \in \mathbb{N}^q$ large enough.

Remark 2.10. The usual way of proving the stabilization of $\det_A(\mathbf{K}_\bullet^\nu(\underline{f}, M))$ is via the vanishing of certain cohomology modules [GKZ94, Jou95]. In a sense, our approach of relating it to $\chi_A((M/(\underline{f})M)_\nu)$ and interpreting it as a collection of local Hilbert functions is more direct. However, it should be noted that the stabilization of Hilbert functions to Hilbert polynomials is related to cohmological results such as the vanishing theorem of Serre [Har77]. In the case of the Macaulay resultant, or more generally when M is a polynomial algebra, we may take $\nu \geq \nu_0$ in Definition 2.9 for some explicit ν_0 [SS96, Theorem 2.2]. In general the value of ν_0 depends on the Castelnuovo-Mumford regularity of M [Cas93, MB66, Cha07], see also [MS04, BC17] for multigraded Castelnuovo-Mumford regularity.

Remark 2.11. The theory of [Cha93] is recovered with the module $M = A[\mathbf{x}]$, while the theory of [Rém01] corresponds to the elimination ring $A = k[\mathbf{u}]$, and the module $M = (k[\mathbf{x}]/I) \otimes k[\mathbf{u}]$, where k is a field and $I \subseteq k[\mathbf{x}]$ is a multihomogeneous ideal of $k[\mathbf{x}]$ (see Section 2.4).

Remark 2.12. For the sake of simplicity in this paper we usually assume that M is a componentwise free $A[\mathbf{x}]$-module and that A is a Noetherian UFD ring. This is enough for our purposes, because of Remark 2.11. However our constructions, conveniently adapted, can be performed under weaker hypotheses, for example if M is just projective over A and A is any integrally closed Noetherian integral domain. See for example [GKZ94, Appendix A] for a general definition of the Cayley determinant. Of course our presentation extends to the case in which M is a multigraded module over some multigraded ring that is *standard graded* (terminology of [TV10]), i.e. that is generated over $R_{\mathbf{0}}$ by elements with minimal multidegree.

We have not attempted to cover the case of polynomial algebras $A[\mathbf{x}]$ whose variables have arbitrary weight/multidegree. For the reader interested in this case, we refer to [SS96, SS01].

Remark 2.13. As we mentioned in the introduction, the theory of resultants formulated for modules allude to a generalization to vector bundles over schemes. This point of view is adopted for the mixed resultants in [GKZ94, Chapter 3, Sec. 3], but some comments are in order. Indeed, while the classical resultants and the mixed resultants are always irreducible, in the theory of Rémond and of this paper, they might not be [DKS13, Example 1.31]. The reason is that in multiprojective setting the relevant line bundles come from projection on factors and thus they are not very ample. This forces one to allow multiplicities, in oder to have a well-behaved theory, including, for example, an analogue of [GKZ94, Theorem 3.10].

2.4. **Rémond's definition of the resultant.** Let k be a field, $k[\mathbf{x}]$ a multigraded polynomial ring as in Section 1.4, I a multihomogeneous ideal of $k[\mathbf{x}]$ and $M = k[\mathbf{x}]/I$. For every multidegree $\mathbf{d} \in \mathbb{N}^q$ we denote by $\mathfrak{M}_{\mathbf{d}}$ the collection of monomials of multidegree $\mathbf{d}$ in the variables $\mathbf{x}$. Let $r = \text{dim-r}_k(M)$ and let $\mathbb{d} = (\mathbf{d}^{(0)}, \dots, \mathbf{d}^{(r)})$ be a collection of nonzero multidegrees. For $i = 0, \dots, r$ and $\mathfrak{m} \in \mathfrak{M}_{\mathbf{d}^{(i)}}$ we introduce a variable $u_{\mathfrak{m}}^{(i)}$. The collection of variables $\mathbf{u} = (u_{\mathfrak{m}}^{(i)} : \ 0 \leq i \leq r, \ \mathfrak{m} \in \mathfrak{M}_{\mathbf{d}^{(i)}})$ is called the collection of *generic coefficients*. For $i = 0, \dots, r$ we also consider the subcollection $\mathbf{u}^{(i)} = (u_{\mathfrak{m}}^{(i)} : \ \mathfrak{m} \in \mathfrak{M}_{\mathbf{d}^{(i)}})$ and the *generic polynomial* of multidegree $\mathbf{d}^{(i)}$ defined by

$$U_i := \sum_{\mathfrak{m} \in \mathfrak{M}_{\mathbf{d}^{(i)}}} u_{\mathfrak{m}}^{(i)} \mathfrak{m},$$

which is a multihomogeneous element of multidegree $\mathbf{d}^{(i)}$ in the polynomial ring $k[\mathbf{u}^{(i)}][\mathbf{x}]$. Let $M[\mathbf{u}] := M \otimes_k k[\mathbf{u}]$, $\underline{\mathrm{U}} = (U_0, \dots, U_r)$ and $\mathcal{M}(I) := M[\mathbf{u}]/(\underline{\mathrm{U}})M[\mathbf{u}]$. We observe that $k[\mathbf{u}]$ is an UFD ring and that $M[\mathbf{u}]$ is a componentwise free $k[\mathbf{u}][\mathbf{x}]$-module such that $\text{dim-r}_{k[\mathbf{u}]}(M[\mathbf{u}]) = \text{dim-r}_k(M) = r$. In [Rém01] it is proved, using multihomogeneous elimination, that $\mathcal{M}(I)$ is a torsion $k[\mathbf{u}]$-module and that $\chi_{k[\mathbf{u}]}(\mathcal{M}(I)_\nu)$ is equal, for $\nu \in \mathbb{N}^q$ large enough, to a fixed element $\text{rés}_{\mathbb{d}}(I) \in k[\mathbf{u}]$, called the *resultant form* of index $\mathbb{d}$ attached to I. The aim of this paragraph is to prove the following.

Theorem 2.14. With the notation above, $\text{rés}_{\mathbb{d}}(I) = \text{RES}_{k[\mathbf{u}]}(\underline{\mathrm{U}}, M[\mathbf{u}]) \pmod{k^\times}$.

To prove this, we adapt to our situation a classical result about the generic polynomials [Jou80, pp. 6-8], saying that $\underline{\mathrm{U}}$ is a filter-regular sequence for the $\mathbb{N}^q$-graded $k[\mathbf{u}][\mathbf{x}]$-module $M[\mathbf{u}]$. By means of Proposition 2.8 this will imply that $\text{rés}_{\mathbb{d}}(I)$ coincides with the M-resultant (see Definition 2.9), up to elements of $k[\mathbf{u}]^\times = k^\times$, and so Theorem 2.14.

Lemma 2.15. Let R be any commutative ring, M an R-module and S a finite set. Let $(r_i)_{i \in S}$ be a set of elements $r_i \in R$ and $\mathbf{v} = (v_i)_{i \in S}$ a collection of independent variables, both indexed by S. Denote $M[\mathbf{v}] := M \otimes_R R[\mathbf{v}]$, let J be the ideal of R generated by $(r_i)_{i \in S}$ and let $V := \sum_{i \in I} r_i v_i \in R[\mathbf{v}]$. Then $(0 :_{M[\mathbf{v}]} V) \subseteq (0 :_M J^\infty)[\mathbf{v}]$, where $(0 :_M J^\infty) := \bigcup_{n \in \mathbb{N}} (0 :_M J^n)$.

Proof. We consider the elements of $M[\mathbf{v}]$ as polynomials in the variables $\mathbf{v}$ and with coefficients in M. More precisely, we consider the $\mathbb{N}^{|S|}$-grading on $R[\mathbf{v}]$ (and thus on $M[\mathbf{v}]$) induced by requiring that all elements of R have degree $\mathbf{0}$ and that for all $i \in S$ the element v_i has degree $\mathbf{e}_i$, where the $\mathbf{e}_i$ are the canonical basis elements of $\mathbb{N}^{|S|}$ and $\mathbf{0}$ is the trivial element. Let

now $m \in (0 :_{M[\mathbf{v}]} V)$, so that $mV = 0$ in $M[\mathbf{v}]$. We write $m = \sum_{\alpha \in \mathbb{N}^{|S|}} m_\alpha \mathbf{v}^\alpha$ and we will eventually prove that $m_\alpha \in (0 :_M J^\infty)$ for every α.
Fix $i \in S$. Let LEX_i be a monomial lexicographic order on $\mathbb{N}^{|S|}$ so that $\mathbf{e}_i > \mathbf{e}_j \ \forall j \in S$, $j \neq i$. We now prove that $\forall \alpha \in \mathbb{N}^{|S|} \ \exists n \in \mathbb{N}$ such that $m_\alpha r_i^n \in M$. By contradiction, let α be a counterexample to this claim, maximal with respect to LEX_i. Comparing terms of multidegree $\alpha + \mathbf{e}_i$ in the equality $mV = 0$, we see that

$$m_\alpha r_i + \sum_{j \neq i, \ \alpha_j \neq 0} m_{\alpha + \mathbf{e}_i - \mathbf{e}_j} r_j = 0. \tag{2.1}$$

We notice that all the $\alpha + \mathbf{e}_i - \mathbf{e}_j$ appearing in this formula, if any, are bigger than α with respect to LEX_i. Therefore by assumption the corresponding $m_{\alpha+\mathbf{e}_i-\mathbf{e}_j}$ vanish when multiplied by certain power of r_i. Thus, if we multiply both sides of the equation (2.1) by a suitable power of r_i we get a contradiction. Let now $N \in \mathbb{N}$ be big enough, so that $\forall \alpha \ \forall i \in S$ we have $m_\alpha r_i^N = 0$. Then we can deduce that for every α we have $m_\alpha \in (0 :_M J^{N \cdot |S|})$. □

Corollary 2.16. With the notation above we have that $\underline{U} = (U_0, \dots, U_r)$ is a filter-regular sequence for $M[\mathbf{u}]$ in $k[\mathbf{u}][\mathbf{x}]$.

Proof. For $i = 0, \dots, r$ let $\tilde{\mathbf{u}}_i := \mathbf{u} \setminus \mathbf{u}^{(i)}$, let $\widetilde{M}_i := M[\tilde{\mathbf{u}}_i]/(U_0, \dots, U_{i-1})M[\tilde{\mathbf{u}}_i]$ and let $M_i := M[\mathbf{u}]/(U_0, \dots, U_{i-1})M[\mathbf{u}]$, so that $M_i = \widetilde{M}_i[\mathbf{u}^{(i)}]$. Lemma 2.15 with $\mathbf{v} = \mathbf{u}^{(i)}$, $J = (\mathfrak{M}_{\mathbf{d}^{(i)}})$ and $V = U_i$ gives that for every $m \in (0 :_{M_i} U_i)$ there is $N(m) \in \mathbb{N}$ such that $m\mathfrak{M}_{N(m)\mathbf{d}^{(i)}} = 0$ in M_i, and so that $mk[\mathbf{u}][\mathbf{x}]_\nu = 0$ for all $\nu \geq N(m)\mathbf{d}^{(i)}$. Since $k[\mathbf{u}][\mathbf{x}]$ is Noetherian, M_i is Noetherian as well, and so $(0 :_{M_i} U_i)$ is generated over $k[\mathbf{u}][\mathbf{x}]$ by finitely many multihomogeneous elements $m_1, \dots, m_\ell$, respectively with multidegrees $\nu_1, \dots, \nu_\ell$. Then, $(0 :_{M_i} U_i)_\nu = \sum_{j=1}^{\ell} m_j k[\mathbf{u}][\mathbf{x}]_{\nu - \nu_j} = 0$ for every $\nu \in \mathbb{N}^q$ such that $\nu \geq \nu_j + N(m_j)\mathbf{d}^{(i)}$, $\forall j = 0, \dots, \ell$. This means that U_i is filter-regular for the module M_i. □

Remark 2.17. Despite the lost of irreducibility, it is comforting to acknowledge that the theory of Rémond resultants retains some of the essential features of the theory of resultants, such as the computability via Cayley determinants. As we have seen, this is because the Cayley determinant, thanks to [GKZ94, Theorem 30, Appendix A, p.493], detects the multiplicities in the divisor of a complex.

3. Lower bounds for the multipicity of the resultant

3.1. **The order function induced by a prime ideal.** Let A be a Noetherian integral domain, let $\mathfrak{p}$ be a nonzero prime ideal of A, and let $\mathfrak{m}_\mathfrak{p} := \mathfrak{p}A_\mathfrak{p}$ be the maximal ideal of the localization $A_\mathfrak{p}$ of A at $\mathfrak{p}$. For every $n \in \mathbb{N}$ the *n-th symbolic power* of $\mathfrak{p}$ is $\mathfrak{p}^{(n)} := \mathfrak{m}_\mathfrak{p}^n \cap A$. The following proposition (see [ZS58, Vol.1, Ch. IV, Sec. 12] gives alternative definitions for symbolic powers.

Proposition 3.1. We have $\mathfrak{p}^{(n)} = \{a \in A : \ \exists b \in A - \mathfrak{p} \text{ with } ab \in \mathfrak{p}^n\}$. Moreover, $\mathfrak{p}^{(n)}$ is the smallest $\mathfrak{p}$-primary ideal of A that contains $\mathfrak{p}^n$. In particular if $\mathfrak{p}$ is maximal then $\mathfrak{p}^n = \mathfrak{p}^{(n)}$.

As a consequence of Krull's intersection theorem we have $\bigcap_{n=0}^{\infty} \mathfrak{m}_\mathfrak{p}^n = \{0\}$, and so we can consider the *order function* $\mathrm{ord}_\mathfrak{p} : A_\mathfrak{p} \to \mathbb{N} \cup \{+\infty\}$ associated to the filtration $\{\mathfrak{m}_\mathfrak{p}^n\}_{\mathbf{n} \in \mathbb{N}}$, given by $\mathrm{ord}_\mathfrak{p}(0) = +\infty$ and $\mathrm{ord}_\mathfrak{p}(a) = n$ if $a \in \mathfrak{m}_\mathfrak{p}^n - \mathfrak{m}_\mathfrak{p}^{n+1}$ [Bou72, Ch. III, Sec. 2.2]. The order function $\mathrm{ord}_\mathfrak{p}$ satisfies $\mathrm{ord}_\mathfrak{p}(a+b) \geq \min\{\mathrm{ord}_\mathfrak{p}(a), \mathrm{ord}_\mathfrak{p}(b)\}$ and $\mathrm{ord}_\mathfrak{p}(ab) \geq \mathrm{ord}_\mathfrak{p}(a) + \mathrm{ord}_\mathfrak{p}(b)$

for all $a, b \in A_{\mathfrak{p}}$. Moreover, it satisfies a *weak homomorphism property*: if $a, b \in A_{\mathfrak{p}}$ and $\mathrm{ord}_{\mathfrak{p}}(b) = 0$, then $\mathrm{ord}_{\mathfrak{p}}(ab) = \mathrm{ord}_{\mathfrak{p}}(a)$. The restriction of $\mathrm{ord}_{\mathfrak{p}}$ to A is the order function with respect to the filtration $\{\mathfrak{p}^{(n)}\}_{n\in\mathbb{N}}$. Moreover, if $\overline{a} \in A/A^{\times}$ we define $\mathrm{ord}_{\mathfrak{p}}(\overline{a})$ to be the order of any element of A representing $\overline{a}$. This is a good definition because $\mathrm{ord}_{\mathfrak{p}}(u) = 0$ for every $u \in A^{\times}$.

Remark 3.2. Geometrically speaking, an element $a \in A_{\mathfrak{p}}$ is a rational function over $\operatorname{Spec} A$, regular in a neighbourhood of $\mathfrak{p}$. Then $\mathrm{ord}_{\mathfrak{p}}(a)$ is interpreted as the multiplicity of vanishing of a at $\mathfrak{p}$. See also the Zariski-Nagata Theorem [Eis95, Chapter 3.9] about this interpretation.

3.2. **The multiplicity of the resultant along a prime ideal.** Let A be a Noetherian UFD ring with fraction field $\mathbb{F}$, let M be a componentwise free $A[\mathbf{x}]$-module as in Section 1.4, with $\text{dim-r}_A(M) = r$ and let $\underline{f} = (f_0, \dots, f_r)$ be a filter-regular sequence in $A[\mathbf{x}]$ for M. Let also $\mathfrak{p}$ be a prime ideal of A, $k_{\mathfrak{p}} = A_{\mathfrak{p}}/\mathfrak{p}A_{\mathfrak{p}}$ the residue field, $\overline{M} = M \otimes_A k_{\mathfrak{p}}$, $\pi : A[\mathbf{x}] \to k_{\mathfrak{p}}[\mathbf{x}]$ the natural projection, $(\pi(\underline{f}))$ the ideal of $k_{\mathfrak{p}}[\mathbf{x}]$ generated by $\pi(f_0), \dots, \pi(f_r)$ and $N := \overline{M}/(\pi(\underline{f}))\overline{M}$.

Theorem 3.3. With the above notation, consider the resultant $\textsc{res}_A(\underline{f}, M) \in A/A^{\times}$ as in Definition 2.9 and suppose that $\text{f-depth}((\pi(\underline{f})), \overline{M}) = r$. Then

$$\mathrm{ord}_{\mathfrak{p}}(\textsc{res}_A(\underline{f}, M)) \geq \text{deg-r}_{k_{\mathfrak{p}}}(N).$$

Remark 3.4. If $M = \mathbb{Z}[\mathbf{x}]$ and p is a prime number, $\ell := \text{deg-r}_{\mathbb{F}_p}(N)$ counts with multiplicity the number of common zeros modulo p of the polynmials $\underline{f}$. Then from Theorem 3.3 we recover the result of Chardin mentioned in the introduction: the resultant $R(\underline{f}) := \textsc{res}_A(\underline{f}, M)$ is an integer divisible by p^{ℓ}. Now we observe that we can say more if instead we use the module $M = \mathbb{Z}[\mathbf{u}, \mathbf{x}]$, the sequence of generic polynomials $\underline{\mathrm{U}}$ and the U-resultant $R(\underline{\mathrm{U}}) \in \mathbb{Z}[\mathbf{u}]/\{\pm 1\}$. Let $\mathfrak{p} \subseteq \mathbb{Z}[\mathbf{u}]$ be the kernel of the morphism that maps the generic coefficients $u_{\mathbf{m}}^{(i)}$ to the corresponding coefficients of f_i, composed with the reduction modulo p. Then again $N \cong \mathbb{F}_p[\mathbf{x}]/(\underline{f})$ has relevant degree equal to ℓ and so we get

$$\mathrm{ord}_{\mathfrak{p}}(R(\underline{\mathrm{U}})) \geq \ell. \tag{3.1}$$

If we expand in Taylor series the polynomial $R(\underline{\mathrm{U}})$, at the point corresponding to the coefficients of $\underline{f}$, we rediscover that $p^{\ell} \,|R(\underline{f})$, but we also prove more: all partial derivatives $\partial_{u_{\mathbf{m}}^{(i)}} R(\underline{f})$ are divisible by $p^{\ell-1}$ and more generally all iterated derivatives $\partial_{\mathbf{u}}^{\alpha} R(\underline{f})$ of order $|\alpha| < \ell$ are divisible by $p^{\ell-|\alpha|}$.

Remark 3.5. We recall from Section 1.3 that $\text{f-depth}((\pi(\underline{f})), \overline{M})$ is the maximal length of a filter-regular sequence for $\overline{M}$ made of elements of $(\pi(\underline{f}))$, and we recall from Section 1.5 that the relevant dimension $\text{dim-r}_A(M)$ is the total degree of the Hilbert polynomial H_M. By Corollary 1.9 $\text{f-depth}((\pi(\underline{f})), \overline{M})$ can be seen as a *codimension* of N with respect to $\overline{M}$. Since $\text{dim-r}_A(M) = r$ we have $\text{dim-r}_{k_{\mathfrak{p}}}(\overline{M}) = r$ and this, together with $\text{f-depth}((\pi(\underline{f})), \overline{M}) = r$, implies $\text{dim-r}_{k_{\mathfrak{p}}}(N) \leq 0$. Therefore H_N is constant and $\text{deg-r}_{k_{\mathfrak{p}}}(N) = H_N$ is defined.

The proof of Theorem 3.3 relies on the computation of the resultant via Cayley determinants and Koszul complexes, as done in Section 2.3, on an adaptation of techniques already used by Chardin in [Cha93], and on the following easy lemma.

Lemma 3.6. Let D be an $s \times s$ square matrix with entries in $A_{\mathfrak{p}}$, let $\overline{D}$ be the matrix with entries in $k_{\mathfrak{p}}$ obtained from D by reduction modulo $\mathfrak{p}A_{\mathfrak{p}}$ and let $\mathrm{corank}(\overline{D})$ denote the

codimension of the image of the $k_\mathfrak{p}$-linear map represented by $\overline{D}$. Then $\mathrm{ord}_\mathfrak{p}(\det(D)) \geq \mathrm{corank}(\overline{D})$.

Proof. Since $k_\mathfrak{p}$ is a field, we can find two invertible $s \times s$ matrices A, B with coefficients in $k_\mathfrak{p}$ such that $A\overline{D}B$ is a block matrix $\left(\begin{smallmatrix} I & 0 \\ 0 & 0 \end{smallmatrix}\right)$ with the first block being square of size $s - \mathrm{corank}(\overline{D})$. We lift arbitrarily A and B to matrices $\widetilde{A}$ and $\widetilde{B}$ with entries in $A_\mathfrak{p}$ and we notice that $\mathrm{ord}_\mathfrak{p}(\det(\widetilde{A})) = \mathrm{ord}_\mathfrak{p}(\det(\widetilde{B})) = 0$. Then all the entries of the last $\mathrm{corank}(\overline{D})$ columns (or rows) of the matrix $\widetilde{A}D\widetilde{B}$ belong to $\mathfrak{p}A_\mathfrak{p}$, and thus we obtain from Laplace's expansion that $\mathrm{ord}_\mathfrak{p}(\det(\widetilde{A}D\widetilde{B})) \geq \mathrm{corank}(\overline{D})$. We conclude what we wanted using the multiplicativity of the determinant and the weak homomorphism property of the order function $\mathrm{ord}_\mathfrak{p}$ (see Section 3.1). □

Proof of Theorem 3.3. If $\dim_{k_\mathfrak{p}}(N) = -1$, then $\text{deg-r}_{k_\mathfrak{p}}(N) = 0$ and the thesis is trivial. Therefore, we suppose N is not eventually zero. Since all the homogeneous components of M are free A-modules of finite rank, for every $\nu \in \mathbb{N}^q$ the complex $\mathbf{K}^\nu_\bullet := \mathbf{K}^\nu_\bullet(\underline{f}, M)$ is a finite complex of free A-modules of finite rank. We can therefore choose a system $\{\underline{b}^{(\nu)}_p\}_{0 \leq p \leq r+1}$ of A-bases for the modules $\mathbf{K}^\nu_p(\underline{f}, M)$. When we change scalars from A to $k_\mathfrak{p}$ we can consider the induced $k_\mathfrak{p}$-bases, which we still call $\underline{b}^{(\nu)}_p$, for the $k_\mathfrak{p}$-vector spaces $\overline{\mathbf{K}}^\nu_p := \mathbf{K}^\nu_p(\pi(\underline{f}), \overline{M}) \cong \mathbf{K}^\nu_p \otimes_A k_\mathfrak{p}$. Since $\underline{f}$ is filter-regular for M, we have by Proposition 2.8 that for ν large enough the complex $\mathbf{K}^\nu_\bullet$ is generically exact and $\det_A(\mathbf{K}^\nu_\bullet) = \mathrm{RES}_A(\underline{f}, M) \pmod{A^\times}$. In addition to this, $\mathrm{H}_0(\overline{\mathbf{K}}^\nu_\bullet) = N_\nu$ for every $\nu \in \mathbb{N}^q$ and so $\dim_{k_\mathfrak{p}}(\mathrm{H}_0(\overline{\mathbf{K}}^\nu_\bullet)) = \text{deg-r}_{k_\mathfrak{p}}(N)$ for ν large enough. Moreover, since N is not eventually zero and $\text{f-depth}((\pi(\underline{f})), \overline{M}) = r$, Proposition 2.2 (ii) implies that the homology modules $\mathrm{H}_p(\overline{\mathbf{K}}^\nu_\bullet)$ vanish for $p \geq (r+1) - r + 1 = 2$ and ν large enough. Let $\nu \in \mathbb{N}^q$ such that all the above requirements hold for $\nu' \geq \nu$ and denote by $\overline{\partial}^\nu_p$ the differentials of $\overline{\mathbf{K}}^\nu_\bullet$, induced by the differentials ∂^ν_p of $\mathbf{K}^\nu_\bullet$. By the vanishing of the higher homology, we can find by elementary linear algebra (see for example [Cha93]) a partition of the bases $\underline{b}^{(\nu)}_p = \underline{b}^{(\nu)}_{p,1} \cup \underline{b}^{(\nu)}_{p,2}$ for $p = 1, \ldots, r+1$, with $\underline{b}^{(\nu)}_{r+1,1} = \emptyset$, inducing decompositions of $\overline{\mathbf{K}}^\nu_p$ and $\mathbf{K}^\nu_p$, such that for $p = 2, \ldots, r+1$ the matrix representations of the differentials $\overline{\partial}^\nu_p$ (resp. ∂^ν_p) take the form $\begin{pmatrix} \overline{a}_p & \overline{\phi}_p \\ \overline{b}_p & \overline{c}_p \end{pmatrix}$ (resp. $\left(\begin{smallmatrix} a_p & \phi_p \\ b_p & c_p \end{smallmatrix}\right)$), where $\overline{\phi}_p$ (resp. ϕ_p) is a square matrix with entries in $k_\mathfrak{p}$ (resp. A) and nonzero determinant (resp. determinant in $A - \mathfrak{p}$) for $p = 2, \ldots, r+1$. For $p = 0$ we consider the trivial partition $\underline{b}^{(\nu)}_0 = \underline{b}^{(\nu)}_0 \cup \emptyset$, that induces a block matrix representation of ∂^ν_1 of the form $\left(\begin{smallmatrix} a_1 & \phi_1 \end{smallmatrix}\right)$. From the fact that the complex $\mathbf{K}^\nu_\bullet$ is generically exact we deduce that also the matrix ϕ_1 must be square. Then, by definition, the Cayley determinant of $\mathbf{K}^\nu_\bullet$ with respect to the above choices of bases and partitions is given by

$$\det_A(\mathbf{K}^\nu_\bullet) = \prod_{i=1}^{r+1} \det(\phi_i)^{(-1)^{i+1}}.$$

By the above construction we have $\mathrm{ord}_\mathfrak{p}(\phi_i) = 0$ for $i = 2, \ldots, r+1$ and from Lemma 3.6 we have $\mathrm{ord}_\mathfrak{p}(\phi_1) \geq \dim_{k_\mathfrak{p}}(\mathrm{H}_0(\overline{\mathbf{K}}^\nu_\bullet))$. By the above choice of ν and the weak homomorphism property of the order function $\mathrm{ord}_\mathfrak{p}$ we deduce

$$\mathrm{ord}_\mathfrak{p}(\det_A(\mathbf{K}^\nu_\bullet)) \geq \text{deg-r}_{k_\mathfrak{p}}(N).$$

Since $\mathrm{RES}_A(\underline{f}, M) = \det_A(\mathbf{K}^\nu_\bullet) \pmod{A^\times}$ for ν large enough, we conclude what we wanted. □

3.3. **The order of vanishing at a sequence of polynomials.** In this paragraph we focus specifically on the Rémond resultant attached to a multihomogeneous ideal as in Section 2.4 and therefore we work in a multiprojective setting as in Section 1.6. Let $k[\mathbf{x}]$ be a multigraded polynomial ring with k an infinite field, let $I \subseteq k[\mathbf{x}]$ be a multihomogeneous ideal with $\dim \mathcal{Z}(I) = r$. Let $\mathbb{d} = (\mathbf{d}^{(0)}, \dots, \mathbf{d}^{(r)})$ be a collection of nonzero multidegrees, and let $k[\mathbf{u}]$ and $\mathrm{rés}_{\mathbb{d}}(I) \in k[\mathbf{u}]$ be as in Section 2.4.
For every $(r+1)$-tuple of polynomials $\underline{f} = (f_0, \dots, f_r)$ of $k[\mathbf{x}]$ with multidegrees prescribed by $\mathbb{d}$ there exists, by the universal property of polynomial rings, a unique k-algebra map $\mathrm{eval}_{\underline{f}} : k[\mathbf{u}] \to k$ that maps the generic coefficients $u^{(i)}_{\mathfrak{m}}$ to the corresponding coefficients of f_i and restricts to the identity on k. This also means that every $R \in k[\mathbf{u}]$ induces a map

$$R(\cdot) : \ k[\mathbf{x}]_{\mathbf{d}^{(0)}} \times \cdots \times k[\mathbf{x}]_{\mathbf{d}^{(r)}} \longrightarrow k$$

given by $R(\underline{f}) := \mathrm{eval}_{\underline{f}}(R)$. We observe that the kernel of the map $\mathrm{eval}_{\underline{f}}$ is a maximal ideal. Then let $\mathrm{ord}_{\underline{f}} : k[\mathbf{u}] \to \mathbb{N} \cup \{+\infty\}$ be the order function corresponding to it as in Section 3.1.

Remark 3.7. One can show that for every $R \in k[\mathbf{u}]$ the value $\mathrm{ord}_{\underline{f}}(R)$ is the largest power of t dividing $T = R(\underline{f} + t\underline{U}) \in k[\mathbf{u}][t]$, where $\underline{U} = (U_0, \dots, U_r)$ is the sequence of generic polynomials as in Section 2.4, and T is defined as above by means of the universal property of polynomial rings.

Theorem 3.8. Let J be a multihomogeneous ideal of $k[\mathbf{x}]$ such that $I \subseteq J$ and $\dim \mathcal{Z}(J) = 0$. Suppose also that, for every $i = 0, \dots, r-1$, we have $\dim \mathcal{Z}(J_{\mathbf{d}^{(i)}}) = 0$ and that, for every relevant $\mathfrak{p} \in \mathrm{Ass}_{k[\mathbf{x}]}(k[\mathbf{x}]/J_{\mathbf{d}^{(i)}}k[\mathbf{x}])$, the local ring (module) $(k[\mathbf{x}]/I)_{\mathfrak{p}}$ is Cohen-Macaulay of (Krull) dimension r. Then the resultant form $\mathrm{rés}_{\mathbb{d}}(I)$ vanishes to order at least $\deg(J)$ at each $(r+1)$-tuple $\underline{f} = (f_0, \dots, f_r) \in J_{\mathbf{d}^{(0)}} \times \cdots \times J_{\mathbf{d}^{(r)}}$, i.e. $\mathrm{ord}_{\underline{f}}(\mathrm{rés}_{\mathbb{d}}(I)) \geq \deg(J)$.

Remark 3.9. Geometrically speaking, we require that $\mathcal{Z}(J_{\mathbf{d}^{(i)}})$ is supported on a finite set of points, located on components of $\mathcal{Z}(I)$ with maximal dimension, and that $\mathcal{Z}(I)$ has mild singularities at these points.

Proof of Theorem 3.8. We adapt an idea from [Roy13, Theorem 5.2] and consider the affine space $\mathbb{A}_{\mathbb{d}}$ over $\mathrm{Spec}\, k$ corresponding to the finite dimensional k-vector space $k[\mathbf{x}]_{\mathbf{d}^{(0)}} \times \cdots \times k[\mathbf{x}]_{\mathbf{d}^{(r)}}$. Then $\mathcal{V} = J_{\mathbf{d}^{(0)}} \times \cdots \times J_{\mathbf{d}^{(r)}}$ is a k-vector subspace of $\mathbb{A}_{\mathbb{d}}$ and so it is an algebraic subset of it, irreducible and closed in the Zariski topology. We pospone the proof of the following fact.

Lemma 3.10. Under the hypotheses of Theorem 3.8 there exists a Zariski dense subset $\mathcal{U}$ of $\mathcal{V}$ such that for every $\underline{f} = (f_0, \dots, f_r) \in \mathcal{U}$ the subsequence $(f_0, \dots, f_{r-1})$ is filter-regular for $M = k[\mathbf{x}]/I$.

Given this fact, we apply Theorem 3.3 to get $\mathrm{ord}_{\underline{f}}(\mathrm{rés}_{\mathbb{d}}(I)) \geq \text{deg-r}_k(k[\mathbf{x}]/(I, \underline{f}))$ for every $\underline{f} \in \mathcal{U}$. From $(I, \underline{f}) \subseteq J$ and $\deg(J) := \text{deg-r}_k(k[\mathbf{x}]/J)$ we deduce in particular that $\mathrm{ord}_{\underline{f}}(\mathrm{rés}_{\mathbb{d}}(I)) \geq \deg(J)$ for every $\underline{f} \in \mathcal{U}$. To conclude it then suffices to see that the set $\{\underline{f} \in \mathbb{A}_{\mathbb{d}} : \mathrm{ord}_{\underline{f}}(\mathrm{rés}_{\mathbb{d}}(I)) \geq \deg(J)\}$ is Zariski closed. This is true because this is the common zero locus of a collection of polynomial functions $\{\mathcal{D}\,\mathrm{rés}_{\mathbb{d}}(I)\}_{\mathcal{D}} \subseteq k[\mathbf{u}]$, where $\mathcal{D}$ ranges

through the differential operators on $k[\mathbf{u}]$ which are partial derivatives of order at most $\deg(J)$. □

Proof of Lemma 3.10. Let $\mathcal{A} = \bigcup_{i=0}^{r-1} \mathrm{Ass}_{k[\mathbf{x}]}(k[\mathbf{x}]/J_{\mathbf{d}^{(i)}}k[\mathbf{x}])$. We will prove that for every $i = 0, \dots, r$ there exists a Zariski dense subset $\mathcal{U}_i \subseteq \mathcal{V}$ with the following properties:

(i) for every $\underline{f} = (f_0, \dots, f_r) \in \mathcal{U}_i$ the subsequence $\underline{f}_{(i)} := (f_0, \dots, f_{i-1})$ is filter-regular for $M = k[\mathbf{x}]/I$;
(ii) for every relevant prime $\mathfrak{p} \in \mathcal{A}$ the module $M_{\underline{f},i} := M/(\underline{f}_{(i)})M$ is locally Cohen-Macaulay at $\mathfrak{p}$;
(iii) for every relevant prime $\mathfrak{p} \in \mathcal{A}$ and every $\mathfrak{q} \in \mathrm{Ass}_{k[\mathbf{x}]}(M_{\underline{f},i})$ with $\mathfrak{q} \subseteq \mathfrak{p}$ we have $\dim \mathcal{Z}(\mathfrak{q}) = r - i$.

The degenerate case $i = 0$ is provided by the hypothesis and $\mathcal{U}_0 = \mathcal{V}$. Let $\mathcal{U}_i$ satisfy the requirements for some $i \leq r-1$, let $\underline{f} \in \mathcal{U}_i$ and consider the following finite collection of k-subspaces of $J_{\mathbf{d}^{(i)}}$:

$$\mathcal{S}_{\underline{f}} = \{\mathfrak{q} \cap J_{\mathbf{d}^{(i)}} : \mathfrak{q} \in \mathrm{Ass}_{k[\mathbf{x}]}(M_{\underline{f},i}),\ \dim \mathcal{Z}(\mathfrak{q}) \geq 0\}.$$

Since $\dim \mathcal{Z}(J_{\mathbf{d}^{(i)}}) = 0$ we see that if $\mathfrak{q}$ is any multihomogeneous prime of $k[\mathbf{x}]$ containing $J_{\mathbf{d}^{(i)}}$, then either $\mathfrak{q}$ is irrelevant ($\dim \mathcal{Z}(\mathfrak{q}) = -1$) or $\dim \mathcal{Z}(\mathfrak{q}) = 0$ and $\mathfrak{q} \in \mathcal{A}$ because in particular $\mathfrak{q}$ is minimal over $J_{\mathbf{d}^{(i)}}$. In either case, also by condition (iii) above, no such $\mathfrak{q}$ appears in the definition of $\mathcal{S}_{\underline{f}}$. Therefore $\mathcal{S}_{\underline{f}}$ is a finite collection of proper k-subspaces of $J_{\mathbf{d}^{(i)}}$. Since k is an infinite field, their union $S_{\underline{f}} := \cup \mathcal{S}_{\underline{f}}$ is a proper Zariski-closed subset of $J_{\mathbf{d}^{(i)}}$. We now define

$$\mathcal{U}_{i+1} := \bigcup_{\underline{f} \in \mathcal{U}_i} \{(f_0, \dots, f_{i-1})\} \times (J_{\mathbf{d}^{(i)}} - S_{\underline{f}}) \times \{(f_{i+1}, \dots, f_r)\}.$$

For $\underline{f} \in \mathcal{U}_i$ the closure of $J_{\mathbf{d}^{(i)}} - S_{\underline{f}}$ is all of $J_{\mathbf{d}^{(i)}}$, so in particular it contains f_i. Then $\mathcal{U}_{i+1}$ is dense in $\mathcal{V}$, because its closure contains $\mathcal{U}_i$. For every $\underline{f} = (f_0, \dots, f_r) \in \mathcal{U}_{i+1}$ the element f_i is filter-regular for $M_{\underline{f},i}$ by Proposition 1.2 and for every relevant prime $\mathfrak{p} \in \mathcal{A}$ it is a regular element for the localization $(M_{\underline{f},i})_{\mathfrak{p}}$, which is Cohen-Macaulay. Therefore $(M_{\underline{f},i+1})_{\mathfrak{p}}$ is Cohen-Macaulay as well. Moreover, by unmixedness, all associated primes $\mathfrak{q}'$ of $M_{\underline{f},i+1}$ containing $\mathfrak{p}$ are minimal ones. Since they are in particular minimal primes for the ideals $(\mathfrak{q}, f_i)$, where $\mathfrak{q}$ is an associated prime of $M_{\underline{f},i}$ containing $\mathfrak{p}$, every such $\mathfrak{q}'$ satisfies $\dim \mathcal{Z}(\mathfrak{q}') = r - i - 1$. We can then continue by induction and we conclude what we wanted when $i = r$. □

Remark 3.11. In Theorem 3.3 we used the notion of f-depth, defined in terms of filter-regular sequences, instead of the more common notion of depth, involving regular sequences. Indeed, the former is more natural (many of our statements are true 'for $\mathbf{d}$ large enough') and more general (a regular sequence is also filter-regular). Moreover, it was essential in order to prove Lemma 3.10 (and so Theorem 3.8), imposing only mild conditions on the multiprojective subvariety $\mathcal{Z}(I)$. Namely, we assumed it to be locally Cohen-Macaulay (e.g. smooth is enough) at a finite number of points.
In fact, to have the analogous statement with regular sequences, one needs $\mathcal{Z}(I)$ to be arithmetically Cohen Macaulay (ACM), which means that the whole coordinate ring $k[\mathbf{x}]/I$

is Cohen-Macaulay (thus also at the irrelevant primes). This is a strong global condition, but it is satisfied, for example, in the case $\mathcal{Z}(I) = \mathbb{P}^{\mathbf{n}}_k$ studied in [Roy13, Theorem 5.2].
We give an example, taken from [VT01], of a family of non-ACM varieties. Let $q = 1$, $n_1 = 2m - 1$ and $I = (x_{2k} : 0 \leq k < m) \cap (x_{2k+1} : 0 \leq k < m)$. Then $\mathcal{Z}(I)$ corresponds to an $(m-1)$-dimensional projective variety but the $k[\mathbf{x}]$-module $k[\mathbf{x}]/I$ has only depth $= 1$. Indeed, it's not possible to extend the regular sequence $\{x_0 + x_1\}$, since after factoring it out, x_0 annihilates all the monomials. For an example of a non-CM integral domain see [Hai10].

4. Polynomials vanishing at prescribed directions

4.1. **Preliminaries on commutative algebraic groups.** Let $G_1, \ldots, G_q$ be connected commutative algebraic groups defined over $\mathbb{C}$. We recall that they are smooth quasi-projective varieties by the structure theorem of Chevalley and Barsotti and their set of complex points $G_1(\mathbb{C}), \ldots, G_q(\mathbb{C})$ have a structure of complex Lie groups. Let $\overline{G}_1, \ldots, \overline{G}_q$ be suitable projective compactifications of them, embedded in projective spaces by $\theta_i : \overline{G}_i \hookrightarrow \mathbb{P}^{n_i}_{\mathbb{C}}$ for $i = 1, \ldots, q$. We then put $G = G_1 \times \ldots \times G_q$, $\overline{G} = \overline{G}_1 \times \cdots \times \overline{G}_q$, $\mathbb{P}^{\mathbf{n}}_{\mathbb{C}} = \mathbb{P}^{n_1}_{\mathbb{C}} \times \cdots \times \mathbb{P}^{n_q}_{\mathbb{C}}$ and $\theta = \theta_1 \times \cdots \times \theta_q : \overline{G} \hookrightarrow \mathbb{P}^{\mathbf{n}}_{\mathbb{C}}$. Thus, we consider G as a Zariski open subscheme of a multiprojective reduced closed subscheme $\overline{G}$ of the multiprojective space $\mathbb{P}^{\mathbf{n}}_{\mathbb{C}}$. For $i = 1, \ldots, q$ we consider in $\mathbb{P}^{n_i}_{\mathbb{C}}$ a set of projective coordinates $\mathbf{x}_i = (x_{i,0}, \ldots, x_{i,n_i})$ and the affine coordinate chart U_i defined by $\{x_{i,0} \neq 0\}$. We consider in $\mathbb{P}^{\mathbf{n}}_{\mathbb{C}}$ the set of multiprojective coordinates $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_q)$, the affine chart $U = U_1 \times \cdots \times U_q$, and the multigraded coordinate ring $\mathbb{C}[\mathbf{x}]$. We denote by $\mathfrak{G} \subseteq \mathbb{C}[\mathbf{x}]$ the multihomogeneous ideal of definition of $\overline{G}$, which is a prime ideal because $\overline{G}$ is irreducible, being the closure of a connected algebraic group. We also let $\pi_i : \mathbb{P}^{\mathbf{n}}_{\mathbb{C}} \to \mathbb{P}^{n_i}_{\mathbb{C}}$ and use the same symbol to indicate the projections $\overline{G} \to \overline{G}_i$ and $G \to G_i$.
Let $T_e G(\mathbb{C}) = T_{e_1} G_1(\mathbb{C}) \times \cdots \times T_{e_q} G_q(\mathbb{C})$ be the tangent space at the identity, identified with the Lie algebra $\mathfrak{g} = \mathfrak{g}_1 \times \cdots \times \mathfrak{g}_q$ of invariant derivations on $G(\mathbb{C})$. This Lie algebra is commutative since the Lie group $G(\mathbb{C})$ is commutative. Let $\Delta = \{\partial_1, \ldots, \partial_d\} \subseteq \mathfrak{g}$ be a set of linearly independent invariant derivations and let $\Sigma = \{\gamma_1, \ldots, \gamma_\ell\} \subseteq G(\mathbb{C})$ be a finite set of complex points of G. We assume that $\Sigma \subset U(\mathbb{C})$ (see [MW81, p. 492] for how to reduce the general case to this one). For every $\sigma \in \mathbb{N}^d$ we define the differential operator $\partial^\sigma = \partial_1^{\sigma_1} \ldots \partial_d^{\sigma_d}$ of order $|\sigma| = \sigma_1 + \cdots + \sigma_d$ and for every $\underline{m} \in \mathbb{Z}^\ell$ we define the point $\underline{m}\gamma = m_1\gamma_1 + \cdots + m_\ell\gamma_\ell$. Since we assumed that Σ is contained in the affine chart $U = \{x_{1,0} \neq 0\} \cap \ldots \cap \{x_{q,0} \neq 0\}$ we can give the following definition, as is done in [Fis05] for the homogeneous case.

Definition 4.1. Given Σ, Δ as above and a positive integer T, we define for every multidegree $\mathbf{d}$ the evaluation operator

$$\begin{array}{rccc} \mathrm{ev}_{\Sigma,T,\mathbf{d}} : & \mathbb{C}[\mathbf{x}]_{\mathbf{d}} & \longrightarrow & \mathbb{C}^{|\Sigma|\binom{T-1+d}{d}} \\ & P & \mapsto & \left(\partial^\sigma \left(\frac{P}{x_{1,0}^{d_0} \cdots x_{q,0}^{d_q}}\right)(\gamma) : \ |\sigma| < T, \gamma \in \Sigma\right) \end{array}$$

Remark 4.2. One can slightly generalize the datum of Δ, Σ, T introducing the concept of a *ponderated set*, as in [Phi96] or [Gal14]. Moreover one can enlarge this setting to quasi-projective varieties with an action of G [Nak95] or even to non-commutative algebraic groups [Hui15], under suitable hypothesis on the projective embedding.

4.2. **The interpolation ideal.** Throughout this paragraph we keep the setting and the notations for $G, \theta, \Sigma, \Delta, T$ introduced in Section 4.1. We define in this multiprojective setting the main ideal $I^{\Sigma,T}$ of the theory of interpolation on commutative algebraic groups, which is the ideal generated by the multihomogeneous polynomials vanishing in Σ with order T in the directions prescribed by Δ. We then describe the multiprojective subscheme it defines and its relation with the surjectivity of the map $\mathrm{ev}_{\Sigma,T,\mathbf{d}}$ introduced in Definition 4.1.

Definition 4.3. For every multidegree $\mathbf{d} \in \mathbb{N}^q$ we let

$$I^{\Sigma,T}_{\mathbf{d}} := \ker(\mathrm{ev}_{\Sigma,T,\mathbf{d}})$$

and then we define $I^{\Sigma,T} := \bigoplus_{\mathbf{d}\in\mathbb{N}^q} I^{\Sigma,T}_{\mathbf{d}}$.

We observe that $I^{\Sigma,T}$ is a multihomogeneous ideal of $\mathbb{C}[\mathbf{x}]$ which contains $\mathfrak{G}$. The following result is a 'trivial' form of an interpolation lemma. In general the objective of an interpolation lemma is to achieve better estimates for the multidegree $\mathbf{d}_{ev}$. Here we essentially reproduce Lemma 4.2 of [Fis05] in multihomogeneous setting.

Proposition 4.4. Let $\mathbf{d}_{ev} \in \mathbb{N}^q$ have all its coordinates equal to $T\,|\Sigma|$. Then for every $\mathbf{d} \geq \mathbf{d}_{ev}$ the map $\mathrm{ev}_{\Sigma,T,\mathbf{d}}$ is surjective.

Proof. Let $\mathbf{1} := (1, \ldots, 1) \in \mathbb{N}^q$ as in Section 1.1, and $\mathbf{d} \geq \mathbf{d}_{ev} = T\,|\Sigma|\,\mathbf{1}$. For every $\nu \in \mathbb{N}^q$ let $z(\nu) \in k[\mathbf{x}]_\nu$ be given by the formula

$$z(\nu) := \prod_{p=1}^{q} x_{p,0}^{\nu_p}.$$

Let $\gamma, \delta \in \Sigma$ be distinct and let $i \in \{1, \ldots, d\}$. We now exhibit the existence of polynomials $L_{\gamma,\delta}, M_{\gamma,i} \in k[\mathbf{x}]_{\mathbf{1}}$ such that:

(i) $L_{\gamma,\delta}$ vanishes at δ and not at γ;
(ii) $M_{\gamma,i}$ vanishes at γ and $\partial_j(M_{\gamma,i}/z(\mathbf{1}))(\gamma) = \delta_{i,j}$ for all $j = \{1, \ldots, d\}$,

where $\delta_{i,j}$ is Kronecker's symbol. Then, given $\gamma \in \Sigma$ and $\sigma \in \mathbb{N}^d$ with $|\sigma| < T$, we construct a polynomial $P_{\gamma,\sigma} \in k[\mathbf{x}]_{\mathbf{d}}$ such that:

(i) $\partial^\sigma(P_{\gamma,\sigma}/z(\mathbf{d}))(\gamma) \neq 0$;
(ii) $\partial^\tau(P_{\gamma,\sigma}/z(\mathbf{d}))(\gamma) = 0$ for every $\tau \leq \sigma$ with $\tau \neq \sigma$;
(iii) $\partial^\tau(P_{\gamma,\sigma}/z(\mathbf{d}))(\delta) = 0$ for every $\delta \in \Sigma - \{\gamma\}$ and every $\tau \in \mathbb{N}^d$ with $|\tau| < T$.

It is clear that these polynomials will witness the surjectivity of $\mathrm{ev}_{\Sigma,T,\mathbf{d}}$.
Since $\gamma \neq \delta$ there are i, j, p with $1 \leq p \leq q$ and $0 \leq i < j \leq n_p$ such that the linear form $\delta_{p,i}x_{p,j} - \delta_{p,j}x_{p,i}$ vanishes at δ and not at γ. We thus define

$$L_{\gamma,\delta} := (\delta_{p,i}x_{p,j} - \delta_{p,j}x_{p,i}) \prod_{p'\neq p} x_{p',0}.$$

Since the derivations $\partial_1, \ldots, \partial_d$ are linearly independent, the following matrix, with d rows and $|\mathbf{n}| = n_1 + \ldots + n_q$ columns,

$$\left[\partial_j\left(\frac{x_{p,k}}{x_{p,0}}\right)(\gamma)\right]_{\substack{j:\,1\leq j\leq d\\(p,k):1\leq p\leq q, 1\leq k\leq n_p}} = \left[\partial_j\left(\frac{x_{p,k}\prod_{p'\neq p} x_{p',0}}{z(\mathbf{1})}\right)(\gamma)\right]_{\substack{j:\,1\leq j\leq d\\(p,k):1\leq p\leq q, 1\leq k\leq n_p}}$$

has rank d. Therefore for every $i \in \{1, \dots, d\}$ there is $\widetilde{M}_{\gamma,i} \in k[\mathbf{x}]_{\mathbf{1}}$ such that $\partial_j(\widetilde{M}_{\gamma,i}/z(\mathbf{1}))(\gamma) = \delta_{i,j}$ for all $j \in \{1, \dots, d\}$. Then we define $M_{\gamma,i}$ by adding to $\widetilde{M}_{\gamma,i}$ a suitable multiple of $z(\mathbf{1})$ so that $M_{\gamma,i}(\gamma) = 0$. Finally, we define, for $\gamma \in \Sigma$ and $\sigma \in \mathbb{N}^d$ with $|\sigma| < T$:

$$P_{\gamma,\sigma} = z(\mathbf{d} - (|\sigma| + (|\Sigma| - 1)T)\mathbf{1}) \prod_{i=1}^{d} M_{\gamma,i}^{\sigma_i} \prod_{\delta \in \Sigma \setminus \{\gamma\}} L_{\gamma,\delta}^{T}.$$

□

The following proposition employs a qualitative modification of a long division algorithm from [Roy13].

Proposition 4.5. Let $\mathbf{d} \in \mathbb{N}^q$ such that $\mathrm{ev}_{\Sigma,T,\mathbf{d}}$ is surjective and let $\mathbf{d}' \in \mathbb{N}^q$ such that $\mathbf{d}' \geq \mathbf{d} + \mathbf{1}$. Then $(I_{\mathbf{d}'}^{\Sigma,T}) = I^{\Sigma,T} \cap \mathbb{C}[\mathbf{x}]_{\geq \mathbf{d}'}$.

Proof. Let $\mathbf{d}'' \geq \mathbf{d}'$. We need to show that $I_{\mathbf{d}''}^{\Sigma,T} = \mathbb{C}[\mathbf{x}]_{\mathbf{d}''-\mathbf{d}'} I_{\mathbf{d}'}^{\Sigma,T}$. We denote by $(\mathbf{e}_p)_{1 \leq p \leq q}$ the canonical basis of $\mathbb{N}^q$ as in Section 1.1. We will prove the assertion assuming $\mathbf{d}'' = \mathbf{d}' + \mathbf{e}_p$ for some p. The general case then follows by induction because $\mathbb{C}[\mathbf{x}]_{\mathbf{a}}\mathbb{C}[\mathbf{x}]_{\mathbf{b}} = \mathbb{C}[\mathbf{x}]_{\mathbf{a}+\mathbf{b}}$ for every $\mathbf{a}, \mathbf{b} \in \mathbb{N}^q$. Let Q be any element of $I_{\mathbf{d}''}^{\Sigma,T}$. We can write $Q = \sum_{i=0}^{n_p} P_i x_{p,i}$ for some $P_i \in \mathbb{C}[\mathbf{x}]_{\mathbf{d}'}$. Since $\mathrm{ev}_{\Sigma,T,\mathbf{d}}$ is surjective, for every $i = 1, \dots, n_p$ we can find $R_i \in \mathbb{C}[\mathbf{x}]_{\mathbf{d}}$ such that $\mathrm{ev}_{\Sigma,T,\mathbf{d}}(R_i) = \mathrm{ev}_{\Sigma,T,\mathbf{d}'}(P_i)$. Then we write

$$\begin{aligned} Q &= \sum_{i=0}^{n_p} P_i x_{p,i} - \sum_{i=1}^{n_p} R_i x_{p,0} x_{p,i} + \sum_{i=0}^{n_p} R_i x_{p,0} x_{p,i} \\ &= \sum_{i=1}^{n_p} x_{p,i}(P_i - x_{p,0} R_i) + x_{p,0}(P_0 + \sum_{i=1}^{n_p} R_i x_{p,i}). \end{aligned}$$

We notice that $P_i - x_{p,0} R_i \in I_{\mathbf{d}'}^{\Sigma,T}$ by construction and $Q \in I^{\Sigma,T}$. Therefore also $P_0 + \sum_{i=1}^{n_p} R_i x_{p,i}$ is in $I_{\mathbf{d}'}^{\Sigma,T}$ and this concludes the proof. □

Proposition 4.6. The subscheme $\mathcal{Z}(I^{\Sigma,T})$ is zero-dimensional and $\deg(I^{\Sigma,T}) = |\Sigma| \binom{T-1+d}{d}$.

Proof. By Definition 4.3 and Proposition 4.4 we have, for $\mathbf{d}$ sufficiently large, the following exact sequence of $\mathbb{C}$-vector spaces:

$$0 \to I_{\mathbf{d}}^{\Sigma,T} \hookrightarrow \mathbb{C}[\mathbf{x}]_{\mathbf{d}} \xrightarrow{\mathrm{ev}_{\Sigma,T,\mathbf{d}}} \mathbb{C}^{|\Sigma|\binom{T-1+d}{d}} \to 0$$

which immediately implies that the value of the Hilbert function $\dim_{\mathbb{C}} \left(\mathbb{C}[\mathbf{x}]/I^{\Sigma,T}\right)_{\mathbf{d}}$ is constantly equal to $|\Sigma|\binom{T-1+d}{d}$ for every $\mathbf{d}$ sufficiently big. The degree of the attached Hilbert polynomial is therefore zero, and its only nonzero term is $|\Sigma|\binom{T-1+d}{d}$. □

Proposition 4.7. For every $\gamma \in \Sigma$ the ideal $I^{\{\gamma\},1}$ is prime and $I^{\{\gamma\},T}$ is $I^{\{\gamma\},1}$-primary. The minimal primary decomposition of $I^{\Sigma,T}$ is $I^{\Sigma,T} = \bigcap_{\gamma \in \Sigma} I^{\{\gamma\},T}$.

Proof. $I^{\{\gamma\},1}$ is generated by the multihomogeneous polynomials vanishing at the point γ or, in other words, is the ideal of definition for the reduced irreducible multiprojective scheme corresponding to that point. Then $I^{\{\gamma\},1}$ is a prime ideal. From Leibnitz rule we get $(I^{\{\gamma\},1})^T \subseteq I^{\{\gamma\},T} \subseteq I^{\{\gamma\},1}$ and so the radical of $I^{\{\gamma\},T}$ is $I^{\{\gamma\},1}$. This implies that $I^{\{\gamma\},1}$ is the only minimal prime over $I^{\{\gamma\},T}$. Moreover, $I^{\{\gamma\},T}$ is multisaturated, because if $f \in \mathbb{C}[\mathbf{x}]$

is multihomogeneous and $f\mathbb{C}[\mathbf{x}]_{\mathbf{d}} \subseteq I^{\{\gamma\},T}$ for some $\mathbf{d} \in \mathbb{N}^q$, then in particular $fz \in I^{\{\gamma\},T}$ for $z = \prod_{p=1}^{q} x_{p,0}^{d_p}$ and so $f \in I^{\{\gamma\},T}$. By multisaturation and Proposition 1.6, since moreover $\mathcal{Z}(I^{\{\gamma\},T})$ is zero-dimensional by Proposition 4.6, $I^{\{\gamma\},T}$ cannot have embedded associated primes. Therefore its minimal primary decomposition consists of only one primary ideal, necessarily equal to $I^{\{\gamma\},T}$ itself. Finally, the equality $I^{\Sigma,T} = \bigcap_{\gamma\in\Sigma} I^{\{\gamma\},T}$ is clear, and since the ideals appearing in this formula are primary ideals corresponding to distinct prime ideals without mutual inclusions, this gives an irredundant primary decomposition for $I^{\Sigma,T}$. □

4.3. **The main corollary.** For this paragraph we keep the notations of Section 4.1 and we denote by n_G the dimension of G. The following is the corollary we aimed for.

Theorem 4.8. Let $\mathfrak{d} = (\mathbf{d}^{(0)}, \dots, \mathbf{d}^{(n_G)})$ be a collection of multidegrees such that $\mathrm{ev}_{\Sigma,T,\mathbf{d}^{(i)}}$ is surjective for all $i = 0, \dots, n_G - 1$. Then the resultant $\mathrm{rés}_{\mathfrak{d}}(\mathfrak{G})$ of index $\mathfrak{d}$ attached to the prime ideal $\mathfrak{G}$ vanishes with multiplicity at least $|\Sigma|\binom{T-1+d}{d}$ on every (n_G+1)-uple of polynomials in $I^{\Sigma,T}_{\mathbf{d}^{(0)}} \times \cdots \times I^{\Sigma,T}_{\mathbf{d}^{(n_G)}}$.

Proof. By Proposition 4.5 we have $(I^{\Sigma,T}_{\mathbf{d}^{(i)}}) = I^{\Sigma,T} \cap \mathbb{C}[\mathbf{x}]_{\geq \mathbf{d}^{(i)}}$ for every $i = 0, \dots, n_G - 1$. Therefore for the same values of i we have that $\mathcal{Z}(I^{\Sigma,T}_{\mathbf{d}^{(i)}}) = \mathcal{Z}(I^{\Sigma,T})$ and, by Proposition 1.6, that the ideals $I^{\Sigma,T}_{\mathbf{d}^{(i)}}$ and $I^{\Sigma,T}$ have the same relevant associated ideals. By Proposition 4.7 these primes correspond to reduced irreducible multiprojective subschemes supported on the points of Σ. Since $\Sigma \subseteq G(\mathbb{C})$ and G is an algebraic group we see that $\mathcal{Z}(\mathfrak{G})$ is smooth at every such point and is therefore locally a complete intersection. $\mathbb{C}[\mathbf{x}]$ being Cohen-Macaulay at every localization, we deduce that for every relevant $\mathfrak{p} \in \mathrm{Ass}_{\mathbb{C}[\mathbf{x}]}(\mathbb{C}[\mathbf{x}]/I^{\Sigma,T})$ the local ring $(\mathbb{C}[\mathbf{x}]/\mathfrak{G})_{\mathfrak{p}}$ is Cohen-Macaulay as well. The thesis is then a corollary of Theorem 3.8 and Proposition 4.6. □

Remark 4.9. The hypothesis of Theorem 4.8 are satisfied if the multidegrees $\mathbf{d}^{(i)}$ are large enough, thanks to the trivial estimate given in Proposition 4.4. In practice, one may want to apply the theorem in an optimal situation and therefore may seek for sharper conditions that imply the surjectivity of the maps $\mathrm{ev}_{\Sigma,T,\mathbf{d}^{(i)}}$. This is exactly the objective of an interpolation lemma, for which we refer the reader, for example, to [Fis03], [Fis05] or [FN14].

References

[AC93] P. Aluffi and F. Cukierman. Multiplicities of discriminants. *Manuscripta mathematica*, 78(3):245–258, 1993.

[BC17] N. Botbol and M. Chardin. Castelnuovo mumford regularity with respect to multigraded ideals. *Journal of Algebra*, 474:361–392, 2017.

[BH98] W. Bruns and J. Herzog. *Cohen-Macaulay rings*. Cambridge studies in advanced mathematics, 39. Cambridge University Press, revised edition, 1998.

[Bou72] N. Bourbaki. *Commutative algebra: Chapters 1-9*. Paris, Hermann, 1972.

[Cas93] G. Castelnuovo. Sui multipli di una serie lineare di gruppi di punti appartenente ad una curva algebrica. *Rendiconti del Circolo Matematico di Palermo (1884-1940)*, 7:89–110, 1893.

[Cha93] M. Chardin. The Resultant via a Koszul Complex. *Computational Algebraic Geometry*, 109:29–39, 1993.

[Cha07] M. Chardin. Some results and questions on Castelnuovo-Mumford regularity. *Lecture Notes in Pure and Applied Mathematics*, 254:1, 2007.

[CLO06] D. A. Cox, J. B. Little, and D. O'shea. *Using algebraic geometry*, volume 185. Springer Science & Business Media, 2006.

[CLO13] D. A. Cox, J. B. Little, and D. O'Shea. *Ideals, varieties, and algorithms: an introduction to computational algebraic geometry and commutative algebra*. Springer Science & Business Media, 2013.

[CLS11] D. A. Cox, J. B. Little, and H. K. Schenck. *Toric varieties*, volume 124 of *Graduate Studies in Mathematics*. Providence, RI: American Mathematical Society, 2011.

[DD00] C. D'Andrea and A. Dickenstein. Explicit formulas for the multivariate resultant. *Journal of Pure and Applied Algebra*, 164(1-2):59–86, 2000.

[Dem84] M. Demazure. Une définition constructive du résultant. *Notes informelles de calcul formel 2, prépublication du Centre de Mathématiques de l'École Polytechnique*, 1984.

[DKS13] C. D'Andrea, T. Krick, and M. Sombra. Heights of varieties in multiprojective spaces and arithmetic Nullstellensätze. *Annales scientifiques de l'École Normale Supérieure*, 46(4):549–627, 2013.

[Eis95] D. Eisenbud. *Commutative algebra*, volume 150 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1995. With a view toward algebraic geometry.

[EM99] I. Z. Emiris and B. Mourrain. Matrices in elimination theory. *Journal of Symbolic Computation*, 28(1-2):3–44, 1999.

[Fis03] S. Fischler. *Contributions à l'étude diophantienne des polylogarithmes et des groupes algébriques*. PhD thesis, Université Pierre et Marie Curie (Paris VI), 2003.

[Fis05] S. Fischler. Interpolation on algebraic groups. *Compositio Mathematica*, 141(4):907–925, 2005.

[FN14] S. Fischler and M. Nakamaye. Seshadri constants and interpolation on commutative algebraic groups. *Annales de l'institut Fourier*, 64(3):1269–1289, 2014.

[Ful16] W. Fulton. *Introduction to toric varieties.*, volume 131. Princeton University Press, 2016.

[Gal14] A. Galateau. Un théorème de zéros dans les groupes algébriques commutatifs. *Publications mathématiques de Besançon*, 1:35–44, 2014.

[Ghi15] L. Ghidelli. Heights of multiprojective cycles and small value estimates in dimension two. Master's thesis, University of Pisa, 2015.

[GKZ94] I. M. Gelfand, M. Kapranov, and A. Zelevinski. *Discriminants, resultants and multidimensional determinants*. Springer Science & Business Media, 1994.

[Hai10] D. Hailong. Two questions about cohen-macaulay rings. MathOverflow, 2010. URL:http://mathoverflow.net/q/21488 (version: 2010-04-15).

[Har77] R. Hartshorne. *Graduate texts in mathematics*, volume 52. Springer, 1977.

[HHRT97] M. Herrmann, E. Hyry, J. Ribbe, and Z. Tang. Reduction Numbers and Multiplicities of Multigraded Structures. *Journal of Algebra*, 197(2):311–341, 1997.

[Hui15] M. Huicochea. On the multiplicity estimates. *arXiv preprint arXiv:1511.09145*, 2015.

[Jou80] J.-P. Jouanolou. Idéaux résultants. *Advances in Mathematics*, 37(3):212–238, 1980.

[Jou91] J.-P. Jouanolou. Le formalisme du résultant. *Advances in Mathematics*, 90(2):117–263, 1991.

[Jou95] J.-P. Jouanolou. Aspects invariants de l'élimination. *Advances in Mathematics*, 114(1):1 – 174, 1995.

[KR94] D. Kirby and D. Rees. Multiplicities in graded rings. I. The general theory. In *Commutative algebra: syzygies, multiplicities, and birational algebra*, volume 159, pages 218–277. Contemporary mathematics, 1994.

[Mac02] F. S. Macaulay. Some formulae in elimination. *Proceedings of the London Mathematical Society*, 1(1):3–27, 1902.

[MB66] D. Mumford and G. M. Bergman. *Lectures on Curves on an Algebraic Surface*. Number 59 in Annals of Mathematics Studies. Princeton University Press, 1966.

[MS04] D. Maclagan and G. G. Smith. Multigraded castelnuovo-mumford regularity. *J. Reine Angew. Math.*, 571:179–212, 2004.

[MS05] D. Maclagan and G. Smith. Uniform bounds on multigraded regularity. *Journal of Algebraic Geometry*, 14(1):137–164, 2005.

[MS10] A. Y. Morozov and S. R. Shakirov. New and old results in resultant theory. *Theoretical and Mathematical Physics*, 163(2):587–617, 2010.

[MW81] D. W. Masser and G. Wüstholz. Zero estimates on group varieties I. *Inventiones Mathematicae*, 64(3):489–516, 1981.

[Nak95] M. Nakamaye. Multiplicity estimates and the product theorem. *Bulletin de la Société Mathématique de France*, 123:155–188, 1995.

[Nor68] D. G. Northcott. *Lessons on rings, modules and multiplicities*. Cambridge University Press, Collier-Macmillan Ltd., 1968.

[NP01] Y. Nesterenko and P. Philippon. *Introduction to Algebraic Independence Theory*. Number 1752 in Lecture Notes in Mathematics. Springer Science & Business Media, 2001.

[NR16] N. A. V. Nguyen and D. Roy. A small value estimate in dimension two involving translations by rational points. *International Journal of Number Theory*, 12(05):1273–1293, 2016.

[Phi91] P. Philippon. Sur des hauteurs alternatives. I. *Mathematische Annalen*, 289(2):255–284, 1991.

[Phi96] P. Philippon. Nouveaux Lemmes de Zéros dans les Groupes Algébriques Commutatifs. *Rocky Mountain Journal of Mathematics*, 26(3):1069–1088, 1996.

[Phi01] P. Philippon. Diophantine Geometry. In Y. Nesterenko and P. Philippon, editors, *Introduction to Algebraic Independence Theory*, chapter 6, pages 83–94. Springer-Verlag, 2001.

[Poi02] S.-D. Poisson. Memoire sur l'elimination dans les equations algebriques. *Journal de l'ecole polytechnique*, 4(11):199, 1802.

[Rém01] G. Rémond. Élimination multihomogène. In Y. Nesterenko and P. Philippon, editors, *Introduction to Algebraic Independence Theory*, chapter 5, pages 53–82. Springer-Verlag, 2001.

[Roy13] D. Roy. A Small Value Estimate for $\mathbb{G}_a \times \mathbb{G}_m$. *Mathematika*, 59(02):333–363, 2013.

[RV10] M. E. Rossi and G. Valla. *Hilbert functions of filtered modules*, volume 9. Springer Science & Business Media, 2010.

[SS96] G. Scheja and U. Storch. The divisor of the resultant. *Beiträge zur Algebra und Geometrie*, 37(1):149–159, 1996.

[SS01] G. Scheja and U. Storch. *Regular sequences and resultants*. AK Peters/CRC Press, 2001.

[Stu94] B. Sturmfels. On the newton polytope of the resultant. *Journal of Algebraic Combinatorics*, 3(2):207–236, 1994.

[Stu98] B. Sturmfels. Introduction to resultants. In *AMS Proceedings of Symposia in Applied Mathematics*, volume 53, pages 25–39, 1998.

[TV10] N. V. Trung and J. K. Verma. Hilbert functions of multigraded algebras, mixed multiplicities of ideals and their applications. *Journal of Commutative Algebra*, 2(4):515–565, 2010.

[Van29] B. L. Van der Waerden. On Hilbert's function, series of composition of ideals and a generalization of the theorem of Bézout. *Proc. Roy. Acad. Amsterdam*, 31:749–770, 1929.

[VM13] D. Q. Viet and N. T. Manh. Mixed multiplicities of multigraded modules. In *Forum Mathematicum*, volume 25, pages 337–361, 2013.

[VT01] A. Van Tuyl. *Sets of Points in Multi-Projective Spaces and their Hilbert Function*. PhD thesis, Queen's University, Kingston, Canada, 2001.

[VT15] D. Q. Viet and T. T. H. Thanh. The Euler-Poincaré characteristic and mixed multiplicities. *Kyushu Journal of Mathematics*, 69(2):393–411, 2015.

[Wal00] M. Waldschmidt. *Diophantine approximation on linear algebraic groups: transcendence properties of the exponential function in several variables*, volume 326. Springer Science & Business Media, 2000.

[ZS58] O. Zariski and P. Samuel. *Commutative algebra, Volume I*. The University Series in Higher Mathematics. D. Van Nostrand Company, Inc., Princeton, New Jersey, 1958. With the cooperation of I. S. Cohen.

–

Chapter 10

On the largest planar graphs with everywhere positive combinatorial curvature

On the largest planar graphs with everywhere positive combinatorial curvature

Luca Ghidelli[1]

Department of Mathematics and Statistics, University of Ottawa, Canada

Abstract

A planar PCC graph is a simple connected planar graph with everywhere positive combinatorial curvature which is not a prism or an antiprism and with all vertices of degree at least 3. We prove that every planar PCC graph has at most 208 vertices, thus answering completely a question raised by DeVos and Mohar. The proof is based on a refined discharging technique and on an accurate low-scale combinatorical description of such graphs. We also prove that all faces in a planar PCC graph have at most 41 sides, and this result is sharp as well.

Keywords: planar graph, combinatorial curvature, positive curvature; discharging, linear optimization, local-global
2010 MSC: Primary: 05C10, 05C30; Secondary: 90C05, 57M15, 05B45

Introduction

Let $\mathcal{S}$ be a surface (connected 2-dimensional manifold) and let G be a graph 2-cell embedded in $\mathcal{S}$, without loops or multiple edges. Then there is on $\mathcal{S}$ an induced structure of polyhedral surface, that is an abstract metric space made of regular polygons with some of the vertices and edges identified. For every vertex v of G we consider the sum $\theta(v)$ of the angles incident in v, and we define its *combinatorial curvature* by $K(v) = 1 - \frac{\theta(v)}{2\pi}$. See formula (1.1) for an equivalent definition. The interested reader is referred to [1, 2, 3, 4] and the introduction of [5] for historical notes and comparison with other notions of curvature on graphs.
If all the vertices of G have strictly positive combinatorial curvature and have degree at least 3, then G is necessarily finite, and $\mathcal{S}$ is either the sphere or the projective plane [6, 7]. There are four infinite families of such graphs: the prisms, the antiprisms and their projective analogues. All other graphs with the above properties will be called *PCC graphs*, and there is only a finite number of them. In this work we mainly focus on the planar case, i.e. with G embedded in the sphere, because every projective PCC graph can be lifted to a planar one.

DeVos and Mohar [6] proved that all planar PCC graphs have at most 3444 vertices, and they asked for a sharp bound. The first conjectured answer was 120, corresponding to the great rhombic icosidodecahedron, but then the lower bound was improved to 138 in [8] and to 208 in [9, 10].

On the other hand, as it was already observed by DeVos and Mohar, much more effort is required to ameliorate the upper bound on the number of vertices. The paper [11] lowers it to 579, but unfortunately it contains a mistake [10, Sec. 8]. Oldridge [10] lowered the bound to 244, conditionally on a result that we prove in section 5, and an unconditional upper bound of 380 vertices was recently given by Oh [12].

The purpose of this article is to provide a complete solution to the problem by showing that the bound 208 is optimal. In section 1 we set some preliminary notation and lemmas, and in section 2 we outline our strategy. The full proof occupies everything from section 2 to section 7. Our result settles also the analogous problem in the projective setting, namely that all projective PCC graphs have at most 104 vertices.

In section 5 we prove a useful result of independent interest in the classification of PCC graphs: every face of a planar PCC graph can have at most 41 edges. In the literature about PCC graphs the faces with at least 42 edges are called *big faces* [6] or *monster faces* [10]. These faces appear very often as annoying special cases that require ad-hoc arguments to be dealt with. Zhang [11] was able to prove that a big face in a PCC graph has at most 290 vertices, while Oh [12] showed that a PCC graph has at most one big face, with no more than 190 vertices. Our result show that these faces do not, in fact, exist.

In section 8 we present some examples which show that our results are sharp. First, we exhibit the known examples of planar PCC graphs with exactly 208 vertices. Then we show a systematic way to construct, for any given $N \in \{3, \dots, 41\}$, a PCC graph G_N containing a face with size N. This construction shows that our result on big faces is sharp, it disproves a conjecture made in [8] and solves a problem raised by Oldridge [10].

Another important theme in this paper is the notion of $\heartsuit$-triangles, see Definition 1.4. We discover that the $\heartsuit$-triangles in a very large PCC graph tend to organize in cyclical structures, which we call *chains*. We take advantage of this phenomenon in section 7.2 to prove that there are no PCC graphs with exactly 209 vertices, via a simple argument. We also use chains of $\heartsuit$-triangles to find one of the graphs with 208 vertices and to construct the graphs G_N.

There is an active area of research that explores, as in the present paper, structural theorems on polyhedral graphs with curvature bounds. The interested reader is referred to [13, 14, 15, 16, 17] for further research on planar graphs with nonnegative curvature and to [18, 19] for graphs on spherical and hyperbolic polyhedral surfaces.

The main technique that is used in this field, as well as in the present paper, is called *discharging*. The discharging method is a flexible technique in structural graph theory that is used to reduce a "global" statement to a number of "local" verifications. It was introduced more than a century ago [20] and it has been used as an essential tool in the proof of celebrated results such as the Four Color

Map Theorem. We refer to [21, Sec 3.1], to [22, 23] and to the first section of [24] for more on this technique.
To apply the discharging method, one is required to define suitable *discharging rules*. In the present paper, this is done in section 3. The choice of weights in the discharging rules is essentially the result of a linear optimization problem. Therefore, in theory, a discharging argument may be performed in an automated way by a computer program, see Oldridge's thesis [10]. The author believes that the ideas of the present paper, together with the methods of Oldridge, will enable further results in the classification of PCC graphs.

Remark 0.1. In the arxiv version of the present paper [21] we perform the long case-analysis in great detail and we support the text with several tables. In order not to obfuscate the arguments with tedious verifications, here we articulate the proofs with a more succinct and readable style. Only occasionally, we omit the proofs of lemmas that can be proved via a straightforward diagram-chasing. Namely, we do it for Lemmas 1.2, 4.6, 5.2 to 5.5, 7.1, 7.4 and 7.5 and for the inequality $n_{TS}(v) \leq 3$ in rule (R_{TS}). The reader interested in carefully verifying some portions of the paper, including the lemmas above, is invited to consult also the arxiv version.

Contents

1. Notation for graph-theoretic objects and multisets

1.1. Basic notation

Let G be a finite simple connected planar graph and let $\mathcal{V}$, $\mathcal{E}$, $\mathcal{F}$ be respectively the set of vertices, edges and faces (including the face containing the point at infinity). Given $x \in \mathcal{V}$, $y \in \mathcal{E}$ and $z \in \mathcal{F}$ we define $E^{(v)}(x)$ to be the set of edges meeting at x, $F^{(v)}(x)$ the multiset of faces touching x, $V^{(e)}(y)$ the set (pair) of endpoints of y, $F^{(e)}(y)$ the multiset (of cardinality 2) of faces touching y by either side, $V^{(f)}(z)$ the multiset of vertices in the boundary of z, and $E^{(f)}(z)$

the multiset of edges in the boundary of z. The *degree* $\deg(v)$ of a vertex $v \in \mathcal{V}$ is the cardinality of $E^{(v)}(v)$ (or of $F^{(v)}(v)$). The *size* $|\sigma|$ of a face $\sigma \in \mathcal{F}$ is the cardinality of $E^{(f)}(\sigma)$ (or of $V^{(f)}(\sigma)$); for example a *triangle* is a face $\sigma \in \mathcal{F}$ of size 3. The *face vector* $f(v)$ of $v \in \mathcal{V}$ and the *side vector* $s(e)$ of $e \in \mathcal{E}$ are the multisets of the sizes of the elements respectively of $F^{(v)}(v)$ and $F^{(e)}(e)$.

1.2. PCC graphs and admissible vertices

We say that a multiset $f = (n_1, \dots, n_d)$ of integers $n_i \geq 3$ is *admissible* if $d \geq 3$ and $K(f) > 0$, where

$$K(n_1, \dots, n_d) := 1 - \frac{d}{2} + \sum_{i=1}^{d} \frac{1}{n_i}. \tag{1.1}$$

Given $v \in \mathcal{V}$, we say that $K(v) := K(f(v))$ is the *combinatorial curvature* of v. It is straightforward to make a list of all the admissible multisets, and this is done e.g. in [6, Table 1]. For the reader's convenience we copy this list in table 1.1.

$f(v)$	where	$f(v)$	where
$(3, a, b)$	$3 \leq a \leq 6,\ a \leq b$	$(3, 4, 4, a)$	$4 \leq a \leq 5$
$(3, 7, a)$	$7 \leq a \leq 41$	$(3, 3, 3, 3, a)$	$3 \leq a \leq 5$
$(3, 8, a)$	$8 \leq a \leq 23$	$(4, 4, a)$	$4 \leq a$
$(3, 9, a)$	$9 \leq a \leq 17$	$(4, 5, a)$	$5 \leq a \leq 19$
$(3, 10, a)$	$10 \leq a \leq 14$	$(4, 6, a)$	$6 \leq a \leq 11$
$(3, 11, a)$	$11 \leq a \leq 13$	$(4, 7, a)$	$7 \leq a \leq 9$
$(3, 3, 3, a)$	$3 \leq a$	$(5, 5, a)$	$5 \leq a \leq 9$
$(3, 3, 4, a)$	$4 \leq a \leq 11$	$(5, 6, 6)$	
$(3, 3, 5, a)$	$5 \leq a \leq 7$	$(5, 6, 7)$	

Table 1.1: Table of admissible face vectors.

Definition 1.1. A *(planar) PCC graph* is a finite simple planar graph G such that $f(v)$ is admissible for all $v \in \mathcal{V}$, and such that G is not a prism or an antiprism.

In other words, all vertices $v \in \mathcal{V}$ of a PCC graph G have positive combinatorial curvature, and satisfy $\deg(v) \geq 3$. We recall that *prism* (of order N) is a planar graph with exactly $2N$ vertices, two faces of size N and N faces of size 4, such that $f(v) = \{4, 4, N\}$ for every vertex v, while an *antiprism* (of order N) is a planar graph with exactly $2N$ vertices, two faces of size N and $2N$ faces of size 3, such that $f(v) = \{3, 3, 3, N\}$ for every vertex v. In the rest of the article G will denote a planar PCC graph, unless we explicitly state otherwise.

1.3. Notation for multisets

Given a (multi)set S we denote its cardinality by $\#S$, multiplicities taken into account. When we list the elements of a multiset, multiple elements occur more than once in the list, according to their multiplicity. However, we employ three different types of brackets to contain such a list.

- We use curly brackets $\{\cdots\}$ if we don't specify any order on the elements.
- We use round brackets $(\cdots)$ to list, weakly increasingly according to the partial order induced by the usual linear order of $\mathbb{N}$, the elements of $f(v)$ and $s(e)$, or also $F^{(v)}(v)$ and $F^{(e)}(e)$ for every $v \in \mathcal{V}$ and $e \in \mathcal{E}$.
- We use angle brackets $\langle\cdots\rangle$ to list, counterclockwise according to the *cyclic order* induced by an orientation of the plane, the elements of $E^{(v)}(v)$, $F^{(v)}(v)$, $f(v)$, $V^{(f)}(\sigma)$ and $E^{(f)}(\sigma)$, for every $v \in \mathcal{V}$ and $\sigma \in \mathcal{F}$.

We write $\mathcal{A} = \{a_1, \ldots, a_n\} \subseteq \mathcal{B}$ if all elements a_i appear in $\mathcal{B}$ with multiplicity greater than or equal to the multiplicity of a_i in $\mathcal{A}$. If $\mathcal{A}$ and $\mathcal{B}$ are linearly ordered, we write $(a_1, \ldots, a_n) \subseteq \mathcal{B}$ to emphasize that the order in $\mathcal{A}$ equals the one induced by $\mathcal{B}$. If instead $\mathcal{B}$ has a cyclic order, we write $\langle a_1, \ldots, a_n\rangle \subseteq \mathcal{B}$ to say that $a_1, \ldots, a_n$ appear in $\mathcal{B}$ as *consecutive* elements. If $\sigma \in \mathcal{F}$, then a priori $V^{(f)}(\sigma)$ and $E^{(f)}(\sigma)$ are just multisets, see fig. 1.1. However, if G is a PCC graph we have the following lemma, which simplifies our exposition.

Lemma 1.2 ([21, Lemma 2.2])**.** *Let G be a PCC graph and let $\sigma \in \mathcal{F}$ such that $|\sigma| \leq 6$ or $|\sigma| \geq 12$. Then $V^{(f)}(\sigma)$ and $E^{(f)}(\sigma)$ are actually sets.*

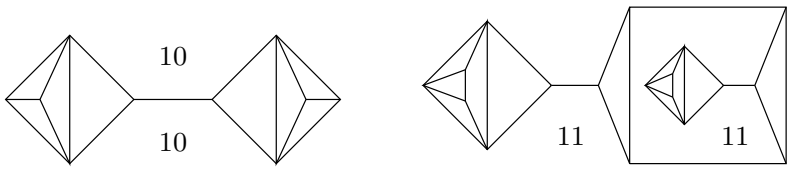
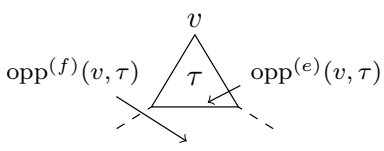
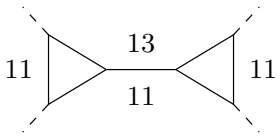


Figure 1.1: PCC graphs with faces having multiple edges on their boundary.

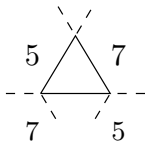
1.4. Other definitions



The opposite face/edge.



A blue edge.



A ♡-triangle.

Figure 1.2: Illustrations for three definitions.

The following notion will be useful in several places:

Definition 1.3. Given a triangle $\tau \in \mathcal{F}$ and a vertex $v \in V^{(f)}(\tau)$ we denote by $\mathrm{opp}^{(e)}(v,\tau)$ and $\mathrm{opp}^{(f)}(v,\tau)$ the "opposite edge" and the "opposite face", i.e. $\mathrm{opp}^{(e)}(v,\tau) = w_1w_2$ and $F^{(e)}(w_1w_2) = \{\tau, \mathrm{opp}^{(f)}(v,\tau)\}$ if $V^{(f)}(\tau) = \{v, w_1, w_2\}$.

In section 8 we will mention two examples of large PCC graphs, with 208 vertices each. The one in fig. 8.1 makes use of faces with size 3, 5 and 7 arranged in chains of $\heartsuit$-triangles, a notion which will play an important role in section 7.2.

Definition 1.4. We say that $\tau \in \mathcal{F}$ with $|\tau| = 3$ is a *$\heartsuit$-triangle* if

$$\forall\ v \in V^{(f)}(\tau) \quad f(v) = (3,3,5,7).$$

The example due to Nicholson and Sneddon instead is built with faces of size 3, 11 and 13. In order to study similar configurations in section 3.1, section 4.3 and section 4.4, we need the following notion of *blue*-edges, α-vertices and β-vertices.

Definition 1.5. Let $v_1, v_2 \in \mathcal{V}$ with $s(v_1v_2) = (11,13)$ and $f(v_1) = f(v_2) = (3,11,13)$, and for $i = 1,2$ let $\tau_i \in F^{(v)}(v_i)$ with $|\tau_i| = 3$. If both $\left|\mathrm{opp}^{(f)}(v_1,\tau_1)\right| = 11$ and $\left|\mathrm{opp}^{(f)}(v_2,\tau_2)\right| = 11$ we say that e is a *blue*-edge and we say that its endpoints v_1, v_2 are β-vertices. Otherwise we say that v_1 and v_2 are α-vertices.

2. Statement of results and strategy

The goal of this article is to prove the following.

Theorem 2.1. *Let G be a planar PCC graph. Then $\#\mathcal{V} \leq 208$.*

As we remarked in the introduction, and as we will show in section 8, there are examples of PCC graphs with 208 vertices, so Theorem 2.1 is sharp. A *projective* PCC graph is a finite simple graph G' that is 2-cell embedded in the projective plane $\mathbb{P}^2$ and such that its pull-back G, through the 2-fold covering of $\mathbb{P}^2$ by the sphere, is a planar PCC graph. The number of vertices of the pull-back G is twice the number of the vertices of G', therefore Theorem 2.1 has the following consequence.

Corollary 2.2. *A projective PCC graph has at most 104 vertices.*

It is easy to see that the large planar PCC graphs discussed in section 8 descend to projective PCC graphs with 104 vertices [9, 10], hence Corollary 2.2 is sharp as well.
In this section we overview our proof of Theorem 2.1. We begin with the following observation: by the Euler-Poincaré formula and a double-counting argument we have

$$\sum_{v\in\mathcal{V}} K(v) = \sum_{v\in\mathcal{V}} 1 - \sum_{e\in\mathcal{E}} \sum_{v\in V^{(e)}(e)} \frac{1}{2} + \sum_{\sigma\in\mathcal{F}} \sum_{v\in V^{(f)}(\sigma)} \frac{1}{|\sigma|} = \#\mathcal{V} - \#\mathcal{E} + \#\mathcal{F} = 2. \tag{2.1}$$

In particular we have $\#\mathcal{V} < 209$ if and only if the the average value of $c_v := K(v) - \frac{2}{209}$ for $v \in \mathcal{V}$ is strictly positive. It is difficult to estimate efficiently

from below the "curvature-contribution" c_v because some vertices of a PCC graph may have very small curvature (e.g. $K(3,7,41) = \frac{1}{1722}$). Our strategy is to *discharge* the function c_v on $v \in \mathcal{V}$ to a function on another set $\widetilde{\mathcal{F}}$, with the same (weighted) average, and then to estimate the new function pointwise. More precisely, we introduce two auxiliary indeterminate objects $\diamondsuit$, $\spadesuit$, which we call *auxiliary faces*, and we consider the set of *discharge faces* $\widetilde{\mathcal{F}}$ given by

$$\widetilde{\mathcal{F}} := \{\diamondsuit, \spadesuit\} \cup \{\sigma \in \mathcal{F} : \ |\sigma| \notin \{3,4,6,8,9,10,12\}\}.$$

In section 3 we will construct a map $\phi : \ \mathcal{V} \times \widetilde{\mathcal{F}} \to \mathbb{Q}_{\geq 0}$, called *pairing*, such that

$$\sum_{\sigma \in \widetilde{\mathcal{F}}} \phi(v,\sigma) = 1 \qquad \forall\, v \in \mathcal{V}. \tag{2.2}$$

For all $\sigma \in \widetilde{\mathcal{F}}$ we define the weight $\phi(\sigma)$ and the discharged contribution $c(\sigma)$ by

$$\phi(\sigma) := \sum_{v \in \mathcal{V}} \phi(v,\sigma) \quad \text{ and } \quad c(\sigma) := \sum_{v \in \mathcal{V}} c_v\, \phi(v,\sigma).$$

The following quantities will also be useful in estimating $c(\sigma)$, for $\sigma \in \widetilde{\mathcal{F}}$:

$$c_+(\sigma) := \sum_{\substack{v \in \mathcal{V} \\ c_v \geq 0}} c_v\, \phi(v,\sigma) \quad \text{ and } \quad c_-(\sigma) := \sum_{\substack{v \in \mathcal{V} \\ c_v < 0}} c_v\, \phi(v,\sigma).$$

The crucial observation is contained in the following lemma.

Lemma 2.3. *We have $\#\mathcal{V} \leq 208$ if and only if $c(G) := \sum_{\sigma \in \widetilde{\mathcal{F}}} c(\sigma) > 0$.*

Proof. By formulas (2.1) and (2.2) we have

$$\sum_{\sigma \in \widetilde{\mathcal{F}}} c(\sigma) = \sum_{\sigma \in \widetilde{\mathcal{F}}} \sum_{v \in \mathcal{V}} c_v \phi(v,\sigma) = \sum_{v \in \mathcal{V}} c_v = \sum_{v \in \mathcal{V}} \left(K(v) - \frac{2}{209} \right) = \frac{2(209 - \#\mathcal{V})}{209}, \tag{2.3}$$

which is strictly positive if and only if $\#\mathcal{V}$ is strictly smaller than 209. □

It is clear that $c(\sigma) = 0$ whenever $\phi(\sigma) = 0$, so we may restrict our attention to $\widetilde{\mathcal{F}}_{\phi \neq 0} := \{\sigma \in \widetilde{\mathcal{F}} : \phi(\sigma) \neq 0\}$. The pairing $\phi : \ \mathcal{V} \times \widetilde{\mathcal{F}} \to \mathbb{Q}_{\geq 0}$ defined in section 3 is carefully designed so that for every $\sigma \in \widetilde{\mathcal{F}}_{\phi \neq 0}$ the number $c(\sigma)$ is positive or, if negative, very small in absolute value. We will prove the following proposition thorough the case analysis performed in section 4 and section 6.

Proposition 2.4. *Let G be a PCC graph and $\sigma \in \widetilde{\mathcal{F}}_{\phi \neq 0}$ with $\sigma \neq \diamondsuit$. Then $c(\sigma) > 0$. Moreover $c(\diamondsuit) > -0.01$, and so $c(G) > c(\sigma) - 0.01$ for all $\sigma \in \widetilde{\mathcal{F}} \setminus \{\diamondsuit\}$.*

As we show in section 6.2, Proposition 2.4 is enough to conclude that $\#\mathcal{V} \leq 210$. Moreover, it implies that $\#\mathcal{V} = 210$ if and only if for all $v \in \mathcal{V}$ we have $f(v) \in \{(3,3,5,6),(5,6,7)\}$, but we show in section 7.1 that this is impossible. In section 7.2 we discover that $\heartsuit$–triangles contained in very large PCC graphs tend to organize in a cyclical pattern. Using this phenomenon we are able to show that there cannot be PCC graphs with 209 vertices, thus proving Theorem 2.1.

3. Description of the discharging weights

We will define the discharging pairing ϕ as the pointwise sum of two functions $\phi_1, \phi_2 : \mathcal{V} \times \widetilde{\mathcal{F}} \to \mathbb{Q}_{\geq 0}$. In order to better describe this construction, we divide the vertices into seven categories, which we call *types*. For some heuristics that motivate our complicated definition of ϕ, we refer to [21, Sec 4.7].

3.1. Seven types of vertices

Definition 3.1. We say that $v \in \mathcal{V}$ is a $\diamondsuit$-vertex if and only if $f(v) = (5, 6, 7)$ or $f(v) = (3, 3, 5, 7)$.

Definition 3.2. We say that $v \in \mathcal{V}$ is a $\spadesuit$-vertex if and only if $f(v)$ is one of the following multisets: $(3, 3, a)$ with $5 \leq a \leq 10$, $(3, 5, a)$ with $5 \leq a \leq 10$, $(3, 6, a)$ with $6 \leq a \leq 10,\ a \neq 7$, $(3, a, 12)$ with $5 \leq a \leq 10$, $(3, a, 13)$ with $5 \leq a \leq 10$, $(3, a, 19)$ with $a \in \{3, 6, 7, 8\}$, $(4, 4, a)$ with $6 \leq a \leq 41,\ a \notin \{7, 11, 13, 19\}$, $(4, 6, a)$ with $a \in \{6, 8, 9, 10\}$, $(5, 5, a)$ with $5 \leq a \leq 9$, $(5, 6, 6)$, $(3, 3, 3, 19)$, $(3, 3, 4, 8)$, $(3, 3, 4, 9)$, $(3, 3, 4, 10)$, $(3, 3, 5, 5)$, $(3, 3, 5, 6)$.

Definition 3.3. We say that $v \in \mathcal{V}$ is a big vertex if and only if $N \in f(v)$ for some $N \geq 42$.

We remark that actually a PCC cannot contain any big vertex, as we will prove in Theorem 5.1.

Definition 3.4. We say that $v \in \mathcal{V}$ is a regular vertex if and only if $f(v)$ is one of the following multisets. $(3, 3, a) : 13 \leq a \leq 41,\ a \neq 19$, $(3, 5, a) : a \in \{40, 41\}$, $(3, 6, a) : 14 \leq a \leq 41,\ a \neq 19$, $(3, 7, a) : 14 \leq a \leq 41,\ a \neq 19$, $\{3, a, 11\} : 6 \leq a \leq 12,\ a \neq 11$, $(3, 8, a) : 14 \leq a \leq 22,\ a \neq 19$, $(3, 9, a) : 8 \leq a \leq 10,\ 14 \leq a \leq 17$, $(3, 10, 14)$, $(4, 4, a) : a \in \{5, 7, 11, 13, 19\}$, $(4, 5, a) : 8 \leq a \leq 18,\ a \neq 11$, $(4, 6, 7)$, $(4, 6, 11)$, $(4, 7, 8)$, $(4, 7, 9)$, $(3, 3, 3, a) : 13 \leq a \leq 41,\ a \neq 19$, $(3, 3, 4, 11)$. For every regular vertex v we choose an integer n_v appearing in $f(v)$ with multiplicity one. If $f(v) = (3, 11, 12)$ we set $n_v = 11$, if $f(v) = (4, 5, a)$ for some a we set $n_v = 5$ and if $f(v) \in \{(4, 7, 8), (4, 7, 9)\}$ we set $n_v = 7$. In all other cases we set $n_v = \max f(v)$.

Definition 3.5. We say that $v \in \mathcal{V}$ is a semi-regular vertex if and only if $f(v)$ is listed in table 3.1. For every semi-regular vertex v we choose a string div_v (displayed in the table) of the form $\frac{1}{2}[n_v] + \frac{1}{2}[\spadesuit]$, $\frac{1}{2}[n_v] + \frac{1}{2}[n_v]'$ or $r_v[m_v] + (1 - r_v)[n_v]$ for some $m_v, n_v \in f(v)$ and $r_v \in [0, 1]$.

$f(v)$	where	div_v
$(3,5,11)$		$\frac{1}{2}[11]+\frac{1}{2}[\spadesuit]$
$(3,5,a)$	$14 \le a \le 19$	$\frac{1}{2}[5]+\frac{1}{2}[a]$
$(3,5,a)$	$20 \le a \le 39$	$\frac{1}{2}[a]+\frac{1}{2}[\spadesuit]$
$(3,11,11)$		$\frac{1}{2}[11]+\frac{1}{2}[11]'$
$(3,11,13)$	v is α-vertex	$\frac{1}{7}[11]+\frac{6}{7}[13]$
$(3,11,13)$	v is β-vertex	$\frac{3}{7}[11]+\frac{4}{7}[13]$
$(4,5,5)$		$\frac{1}{2}[5]+\frac{1}{2}[5]'$
$(4,5,a)$	$a \in \{7,11\}$	$\frac{1}{2}[5]+\frac{1}{2}[a]$
$(4,5,19)$		$\frac{3}{4}[5]+\frac{1}{4}[19]$
$(4,7,7)$		$\frac{1}{2}[7]+\frac{1}{2}[7]'$

Table 3.1: Mnemonics for semi-regular vertices.

Definition 3.6. We say that $v \in \mathcal{V}$ is a TS-vertex if and only if for all $\kappa \in f(v)$ is a triangle of a square.

Definition 3.7. We say that $v \in \mathcal{V}$ is a potentially-special vertex if and only if v is not a $\diamondsuit$-vertex, a $\spadesuit$-vertex, a big vertex, a regular vertex, a semi-regular vertex nor a TS-vertex. If v is a potentially-special vertex and $f(v)$ is $(3,3,11)$, $(3,7,8)$, $(3,7,9)$, $(4,5,6)$, $(3,3,3,11)$, $(3,3,4,5)$, or $(3,3,4,7)$, we respectively set $n_v = 11,7,7,5,11,5,7$ and $r_v = \frac{1}{2},\frac{1}{2},\frac{1}{2},\frac{1}{2},\frac{1}{3},\frac{1}{2},\frac{1}{4}$. If $(3,4,a)$ with $5 \le a \le 41$ and $a \notin \{6,8,9,10,12\}$ we set $n_v = a$ and $r_v = \frac{1}{2}$. In all other cases we set $r_v = 0$.

3.2. The pairing: part 1

We now define the "regular part" of the pairing, namely the function ϕ_1.

Definition 3.8. Let $\phi_1 : \mathcal{V} \times \widetilde{\mathcal{F}} \to \mathbb{Q}_{\ge 0}$ be the only function that satisfies:

(i) $\phi_1(v,\diamondsuit) = 1$ if v is a $\diamondsuit$-vertex;

(ii) $\phi_1(v,\spadesuit) = 1$ if v is a $\spadesuit$-vertex or a big vertex;

(iii) $\phi_1(v,\sigma) = 1$ if v is a regular vertex, $\sigma \in F^{(v)}(v)$ and $|\sigma| = n_v$;

(iv) $\phi_1(v,\sigma) = \frac{1}{2}$ and $\phi_1(v,\spadesuit) = \frac{1}{2}$ if v is a semi-regular vertex, $\mathrm{div}_v = \frac{1}{2}[n_v]+\frac{1}{2}[\spadesuit]$, $\sigma \in F^{(v)}(v)$ and $|\sigma| = n_v$;

(v) $\phi_1(v,\sigma) = \frac{m}{2}$ if v is a semi-regular vertex, $\mathrm{div}_v = \frac{1}{2}[n_v]+\frac{1}{2}[n_v]'$, $\sigma \in F^{(v)}(v)$ with multiplicity m, and $|\sigma_v| = n_v$;

(vi) $\phi_1(v,\sigma) = r_v$ and $\phi_1(v,\sigma') = 1-r_v$ if v is a semi-regular vertex, $\mathrm{div}_v = r_v[m_v]+(1-r_v)[n_v]$, $\sigma,\sigma' \in F^{(v)}(v)$, $|\sigma| = m_v$ and $|\sigma'| = n_v$;

(vii) $\phi_1(v,\sigma) = r_v$ if v is a potentially-special vertex, $r_v \neq 0$, $\sigma \in F^{(v)}(v)$ and $|\sigma| = n_v$;

(viii) $\phi_1(v,\sigma) = 1$ if $f(v) = (3,4,4,5)$, $\sigma \in F^{(v)}(v)$, $|\sigma| = 5$, and v is *not special*, as explained in the rule $(R_{(3,4,4,5)})$ below;

(ix) $\phi_1(v,\sigma) = 0$ in all other cases.

3.3. The pairing: part 2

The most delicate part of this paper is the definition of ϕ_2.

Definition 3.9. The "special part" of the pairing is the function $\phi_2 : \mathcal{V} \times \widetilde{\mathcal{F}} \to \mathbb{Q}_{\geq 0}$ that is given by the rules (R_{TS}), $(R_{(3,3,a)})$, $(R_{(3,4,a)})$, $(R_{(3,a,b)})$, $(R_{(3,3,3,a)})$, $(R_{(4,5,6)})$, $(R_{(3,3,4,a)})$, $(R_{(3,4,4,5)})$ and $(R_{(3,3,3,3,5)})$ when $v \in \mathcal{V}$ is a TS-vertex or a potentially-special vertex, and is zero otherwise (see the paragraphs below and fig. 3.1).

A vertex $v \in \mathcal{V}$ will be called *special* if there exists $\sigma \in \widetilde{\mathcal{F}} \setminus \{\diamondsuit, \spadesuit\}$ with $\phi_2(v,\sigma) \neq 0$. We will also say that v *is special to* σ in this case. In the following paragraphs we will describe the necessary and sufficient conditions for a vertex to be special. In order to simplify the exposition, we will indicate the values of $\phi_1(v,\sigma)$ and $\phi_2(v,\sigma)$, for $(v,\sigma) \in \mathcal{V} \times \widetilde{\mathcal{F}}$, only when they are nonzero.

(R_{TS}): Special rules for TS vertices

If $v \in \mathcal{V}$ is a TS-vertex, then each face $\kappa \in F^{(v)}(v)$ is a square or a triangle. We denote by $\mathcal{F}_{TS}(v)$ the set of faces $\sigma \in \mathcal{F}$ with $|\sigma| \in \{11, 40, 41\}$ that share at least one edge with some $\kappa \in F^{(v)}(v)$. Then let $n_{TS}(v) := \#\mathcal{F}_{TS}(v)$ and notice [21, Lemma 5.1] that $0 \leq n_{TS}(v) \leq 3$. We set $\phi_2(v,\spadesuit) = 1 - \frac{n_{TS}(v)}{3}$, and for every $\sigma \in \mathcal{F}_{TS}(v)$ we set $\phi_2(v,\sigma) = \frac{1}{3}$. In particular v is *special* if and only if $n_{TS} \geq 1$.

$(R_{(3,3,a)})$: Special rules for (3,3,a)

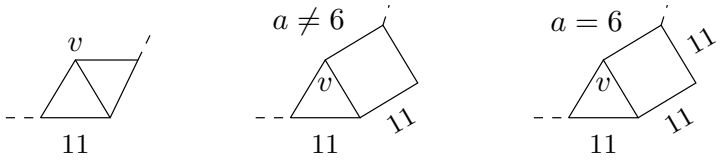
Let $v \in \mathcal{V}$ with $f(v) = (3,3,a)$, where $a \in \{11, 12\}$, then recall from Definition 3.7 $r_v = \frac{1}{2}$ if $a = 11$ and $r_v = 0$ if $a = 12$. We write $F^{(v)}(v) = (\tau_1, \tau_2, \sigma)$ and let $\sigma_i = \mathrm{opp}^{(f)}(v, \tau_i)$ for $i \in \{1,2\}$. If both $|\sigma_1|, |\sigma_2| \neq 11$, then we set $\phi_2(v,\spadesuit) = 1 - r_v$. Otherwise, if $|\sigma_j| = 11$ for some $j \in \{1,2\}$, we set $\phi_2(v,\sigma_j) = 1 - r_v$.

$(R_{(3,4,a)})$: Special rules for (3,4,a)

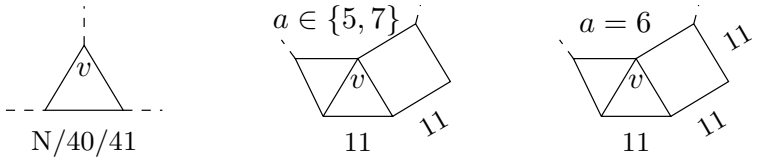
Let $v \in \mathcal{V}$ with $f(v) = (3,4,a)$, where $5 \leq a \leq 41$. Write $F^{(v)}(v) = (\tau, \kappa, \sigma)$ and $V^{(f)}(\kappa) = \langle v, v_1, v_2, v_3 \rangle$ with $s(vv_1) = (3,4)$. Then consider $\sigma_1, \sigma_2, \sigma_3 \in \mathcal{F}$ such that $\sigma_1 = \mathrm{opp}^{(f)}(v,\tau)$, $F^{(e)}(v_1v_2) = \{\kappa, \sigma_2\}$ and $F^{(e)}(v_2v_3) = \{\kappa, \sigma_3\}$. If $a \neq 6$ and $|\sigma_1| = 11$, then we let $\phi_2(v,\sigma_1) = 1 - r_v$, so v is special to σ_1. Otherwise, if $|\sigma_1| \neq 11$ we let $\phi_2(v,\spadesuit) = 1 - r_v$. If $a = 6$ then we consider the set

$$\mathcal{A}_v := \{\sigma' \in \mathcal{F} : \ |\sigma'| = 11 \text{ and } \sigma' = \sigma_j \text{ for some } j \in \{1,2,3\}\},$$

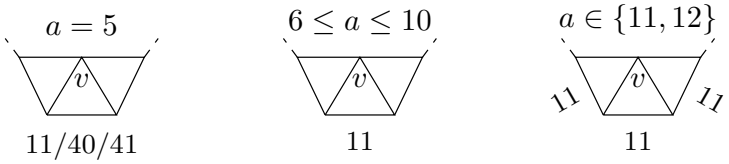
and let $a_v := \#\mathcal{A}_v$. Then we let $\phi_2(v,\spadesuit) = 1 - \frac{a_v}{2}$ and $\phi_2(v,\phi') = \frac{1}{2}$ for every $\sigma' \in \mathcal{A}_v$. Notice that necessarily $a_v \leq 2$.



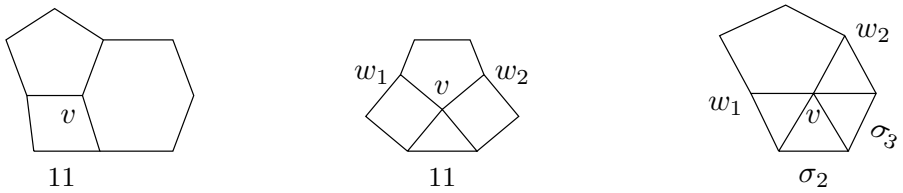
Case $f(v) = (3, 3, a)$. Case $f(v) = (3, 4, a)$.



Case $f(v) = (3, a, b)$. Case $f(v) = (3, 3, 4, a)$.



Case $f(v) = (3, 3, 3, a)$



Case $f(v) = (4, 5, 6)$. Case $f(v) = (3, 4, 4, 5)$. Case $f(v) = (3, 3, 3, 3, 5)$.

Figure 3.1: Illustration of the special rules in the definition of ϕ.

$\big(R_{(3,a,b)}\big)$: *Special rules for (3,a,b)*

Let $v \in \mathcal{V}$ with $f(v) = (3, a, b)$, where $6 \leq a \leq 10$ and $7 \leq b \leq 10$. Write $F^{(v)}(v) = (\tau, \sigma_1, \sigma_2)$ and let $\sigma = \mathrm{opp}^{(f)}(v, \tau)$. If $(a, b) = (6, 7)$ and $|\sigma| \in \{40, 41\}$ then v is special to σ with $\phi_2(v, \sigma) = 1$. If $(a, b) \neq (6, 7)$, $14 \leq |\sigma| \leq 41$ and $|\sigma| \neq 19$ then v is special to σ, with $\phi_2(v, \sigma) = 1 - r_v$. Notice that necessarily if $b = 8$, then $|\sigma| \leq 23$, if $b = 9$, then $|\sigma| \leq 17$, and if $b = 10$, then $|\sigma| \leq 14$.

$\big(R_{(4,5,6)}\big)$: *Special rules for (4,5,6)*

Let $v \in \mathcal{V}$ with $f(v) = (4, 5, 6)$. Write $F^{(v)}(v) = (\kappa, \sigma, \sigma')$, so that $|\kappa| = 4$ and $|\sigma| = 5$, and let $V^{(f)}(\kappa) = \langle v, v_1, v_2, v_3 \rangle$ with $s(vv_1) = (4, 6)$. Then, consider $\sigma_1 \in \mathcal{F}$ such that $F^{(e)}(v_1 v_2) = \{\kappa, \sigma_1\}$. If $|\sigma_1| = 11$, then v is special to σ_1 with $\phi_2(v, \sigma_1) = \frac{1}{2}$. Otherwise, it is not special and $\phi_2(v, \sigma_1) = \frac{1}{2}$.

$\big(R_{(3,3,3,a)}\big)$: *Special rules for (3,3,3,a)*

Let $v \in \mathcal{V}$ with $f(v) = (3,3,3,a)$, where $5 \leq a \leq 12$. Let $F^{(v)}(v) = \langle \tau_1, \tau_2, \tau_3, \sigma \rangle$ with $|\sigma| = a$, and let $\sigma_i = \mathrm{opp}^{(f)}(v, \tau_i)$ for $i \in \{1,2,3\}$. First we consider the cases $5 \leq a \leq 10$. If $a = 5$ and $|\sigma_2| \in \{11, 40, 41\}$, or $6 \leq a \leq 10$ and $|\sigma_2| = 11$, then v is special to σ_2 with $\phi_2(v, \sigma_2) = 1$. Otherwise $\phi_2(v, \spadesuit) = 1$ and v is not special. If $a \in \{11, 12\}$ instead we consider the set

$$\mathcal{A}_v := \{\sigma' \in \mathcal{F} : \ |\sigma'| = 11 \text{ and } \sigma' = \sigma_j \text{ for some } j \in \{1,2,3\}\},$$

and let $a_v := \#\mathcal{A}_v$. If $a = 11$ we let $r = r_v = \frac{1}{3}$, while if $a = 12$ we let $r = \frac{1}{2}$ and recall that $r_v = 0$. Then we let $\phi_2(v, \spadesuit) = 1 - r_v - a_v \cdot r$ and for every $\sigma' \in \mathcal{A}_v$ we set $\phi_2(v, \phi') = r$.

$\big(R_{(3,3,4,a)}\big)$: *Special rules for (3,3,4,a)*

This set of rules is similar to $(R_{(3,4,a)})$. Let $v \in \mathcal{V}$ with $f(v) = \langle 3,3,4,a \rangle$, where $a \in \{5,6,7\}$. Write $F^{(v)}(v) = \langle \tau_1, \tau_2, \kappa, \sigma \rangle$ with $|\kappa| = 4$ and $|\sigma| = a$, and let $V^{(f)}(\kappa) = \langle v, v_1, v_2, v_3 \rangle$ with $s(vv_1) = (3,4)$. Then consider $\sigma_1, \sigma_2, \sigma_3 \in \mathcal{F}$ such that $\sigma_1 = \mathrm{opp}^{(f)}(v, \tau_2)$, $F^{(e)}(v_1 v_2) = \{\kappa, \sigma_2\}$ and $F^{(e)}(v_2 v_3) = \{\kappa, \sigma_3\}$. If $a \in \{5,7\}$ and $|\sigma_1| = 11$, then we let $\phi_2(v, \sigma_1) = 1 - r_v$. Otherwise, if $|\sigma_1| \neq 11$ we let $\phi_2(v, \spadesuit) = 1 - r_v$. If $a = 6$ then we consider the set

$$\mathcal{A}_v := \{\sigma' \in \mathcal{F} : \ |\sigma'| = 11 \text{ and } \sigma' = \sigma_j \text{ for some } j \in \{1,2,3\}\},$$

and let $a_v := \#\mathcal{A}_v$. Then we let $\phi_2(v, \spadesuit) = 1 - \frac{a_v}{2}$ and $\phi_2(v, \phi') = \frac{1}{2}$ for every $\sigma' \in \mathcal{A}_v$. Notice that necessarily $a_v \leq 2$.

$\big(R_{(3,4,4,5)}\big)$: *Special rules for (3,4,4,5)*

Let $v \in \mathcal{V}$ with $f(v) = \langle 4,3,4,5 \rangle$. Write $= F^{(v)}(v) = (\tau, \kappa_1, \kappa_2, \sigma)$, consider $w_1, w_2 \in \mathcal{V}$ such that $s(vw_1) = s(vw_2) = (4,5)$, and let $\sigma_1 = \mathrm{opp}^{(f)}(v, \tau)$. If $|\sigma_1| = 11$ and $f(w_1), f(w_2) \notin \{(4,5,a) : \ 14 \leq a \leq 19\}$ then we set $\phi_1(v, \sigma) = 0$ and $\phi_2(v, \sigma_1) = 1$, so v is special to σ_1. If otherwise $f(v) = \langle 3,4,4,5 \rangle$, or if $f(v) = \langle 4,3,4,5 \rangle$ but the above condition doesn't hold, we set $\phi_1(v, \sigma) = 1$ and $\phi_2(v, \sigma_1) = 1$, so v is not special.

$\big(R_{(3,3,3,3,5)}\big)$: *Special rules for (3,3,3,3,5)*

Let $v \in \mathcal{V}$ with $f(v) = (3,3,3,3,5)$. Let $F^{(v)}(v) = \langle \tau_1, \tau_2, \tau_3, \tau_4, \sigma \rangle$ with $|\sigma| = 5$, and let $\mathcal{A} := \{(3,4,5), (3,3,4,5), (3,4,4,5)\}$. For $i \in \{1,4\}$ consider $w_i \in \mathcal{V}$ such that $F^{(e)}(vw_i) = \{\tau_i, \sigma\}$, and for $j \in \{2,3\}$ let $\sigma_j = \mathrm{opp}^{(f)}(v, \tau_j)$.
If $f(w_1) \in \mathcal{A}$ and $|\sigma_2| = 11$, then v is special to σ_2 with $\phi_2(v, \sigma_2) = 1$. Symmetrically, if $f(w_4) \in \mathcal{A}$ and $|\sigma_3| = 11$ we set $\phi_2(v, \sigma_3) = 1$. Finally, also if $|\sigma_j| \in \{40, 41\}$ for some $j \in \{2,3\}$, we set $\phi_2(v, \sigma_j) = 1$. If none of the above conditions hold, we set $\phi_2(v, \spadesuit) = 1$ and v is not special.

3.4. The pairing

Finally, we define the pairing used for the discharging method.

Definition 3.10. We define $\phi : \mathcal{V} \times \widetilde{\mathcal{F}} \to \mathbb{Q}_{\geq}0$ as the pointwise sum of the functions ϕ_1 and ϕ_2 from Definitions 3.8 and 3.9.

It is straightforward to check that ϕ satisfies the fundamental property eq. (2.2).

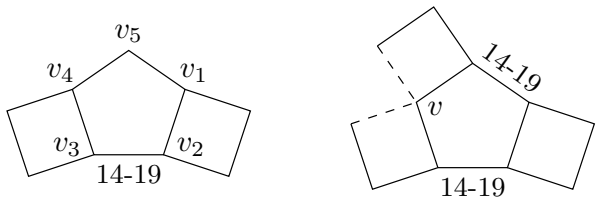
4. Case-by-case analysis of the discharge faces

4.1. Analysis of faces with 5 edges

Let $\sigma \in \widetilde{\mathcal{F}}$ with $|\sigma| = 5$ and $\phi(\sigma) \neq 0$. With our construction of the pairing, we have $\phi(v, \sigma) \neq 0$ if and only if $v \in V^{(f)}(\sigma)$, v is not a special vertex of type $(3, 4, 4, 5)$, and either $4 \in f(v)$ or $f(v) = (3, 5, a)$ with $14 \leq a \leq 19$. Let $A = \#\mathcal{A}$, $B = \#\mathcal{B}$ and $L = \#\mathcal{L}$, where

$$\begin{aligned}
\mathcal{A} &:= \{v \in V^{(f)}(\sigma) : \ f(v) = (4, 5, a), \ 14 \leq a \leq 19\}, \\
\mathcal{B} &:= \{v \in V^{(f)}(\sigma) : \ f(v) = (3, 5, a), \ 14 \leq a \leq 19\}, \\
\mathcal{L} &:= \{e \in E^{(f)}(\sigma) : \ s(e) = (5, a), \ 14 \leq a \leq 19\}.
\end{aligned}$$

Notice that $2L = A + B \leq 4$, since it is both even and less than 5.



Case $A = 2$, $B = 0$. Case $A = 4$.

Figure 4.1: Illustrations for $|\sigma| = 5$.

Proposition 4.1. *Let $\sigma \in \widetilde{\mathcal{F}}_{\phi \neq 0}$ with $|\sigma| = 5$. Then $c(\sigma) > 0.002$.*

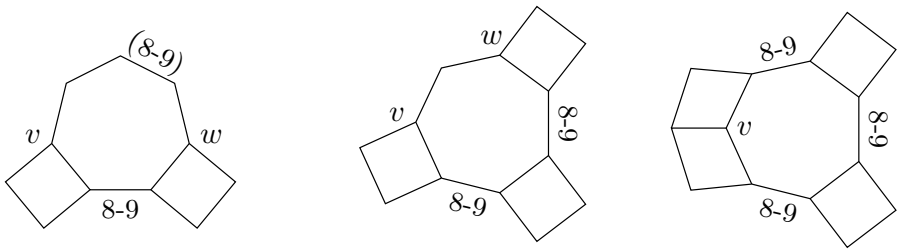
Proof. By inspection we have that $c_-(\sigma) \geq A \cdot \frac{3}{4} c(4, 5, 19)$. If $A = 0$ we have $c(\sigma) = c_+(\sigma) \geq \frac{1}{2} c(4, 5, 11) > 0.015$. If $B \geq 1$ instead, we have $\exists v \in \mathcal{B}$ and $A \leq 3$, hence $c(\sigma) \geq \frac{1}{2} c(3, 5, 19) + 3\frac{3}{4} c(4, 5, 19) > 0.022$. If $B = 0$ and $A = 2$ we let $V^{(f)}(\sigma) = \langle v_1, v_2, v_3, v_4, v_5 \rangle$ with $\mathcal{A} = \{v_2, v_3\}$. Then according to rule $(R_{(3,4,4,5)})$, neither v_1 nor v_4 can be a special vertex of type $(3, 4, 4, 5)$, since they are both consecutive in σ to a vertex in $\mathcal{A}$. Therefore we get $c(\sigma) \geq 2\frac{1}{2} c(4, 5, 11) + 2\frac{3}{4} c(4, 5, 19) > 0.019$. Finally, if $A = 4$ we let $\{v\} = V^{(f)}(\sigma) \setminus \mathcal{A}$, so either $f(v) = (4, 4, 5)$ or $f(v) = (3, 4, 4, 5)$ with v not special, by rule $(R_{(3,4,4,5)})$. In both cases, $c(\sigma) \geq 4\frac{3}{4} c(4, 5, 19) + c(3, 4, 4, 5) > 0.002$. □

4.2. Analysis of faces with 7 edges

Let $\sigma \in \widetilde{\mathcal{F}}$ with $|\sigma| = 7$ and $\phi(\sigma) \neq 0$. With our construction of the pairing, we have $\phi(v, \sigma) \neq 0$ if and only if $v \in V^{(f)}(\sigma)$, and either $4 \in f(v)$, $8 \in f(v)$, or $9 \in f(v)$. Let $A = \#\mathcal{A}$, $B = \#\mathcal{B}$ and $L = \#\mathcal{L}$, where

$$\begin{aligned}
\mathcal{A} &:= \{v \in V^{(f)}(\sigma) : \ f(v) = (4, 7, a), \ 8 \leq a \leq 9\}, \\
\mathcal{B} &:= \{v \in V^{(f)}(\sigma) : \ f(v) = (3, 7, a), \ 8 \leq a \leq 9\}, \\
\mathcal{L} &:= \{e \in E^{(f)}(\sigma) : \ s(e) = (7, a), \ 8 \leq a \leq 9\}.
\end{aligned}$$

Notice that $2L = A + B \leq 6$, since it is both even and less than 7.



Case $A \in \{2, 4\}$, $B = 0$. Case $A = 4$, $B = 0$. Case $A = 6$.

Figure 4.2: Illustrations for $|\sigma| = 7$.

Proposition 4.2. *Let $\sigma \in \widetilde{\mathcal{F}}_{\phi \neq 0}$ with $|\sigma| = 7$. Then $c(\sigma) > 0.0095$.*

Proof. First, notice that $c_-(\sigma) > A \cdot c(4, 7, 9)$. If $A = 0$ we have $c(\sigma) \geq \frac{1}{4} c(3, 3, 4, 7) > 0.012$ by checking all possible v with $\phi(v, \sigma) \neq 0$. If $B \geq 1$, instead, we have $\exists v \in \mathcal{B}$ and $A \leq 5$, so $c(\sigma) \geq \frac{1}{2} c(3, 7, 9) + 5c(4, 7, 9) > 0.0095$. If $B = 0$ and $A \in \{2, 4\}$ we argue as in "*Case $B = 0$ and $A = 2$*" of section 4.1 and we get $c(\sigma) \geq 2\frac{1}{4} c(3, 3, 4, 7) + 2c(4, 7, 9) > 0.012$. □

Finally, if $A = 6$ and $\{v\} = V^{(f)}(\sigma) \setminus \mathcal{A}$, we have $f(v) = (4, 4, 7)$, and so $c(\sigma) \geq c(4, 4, 7) + 6c(4, 7, 9) > 0.098$.

4.3. Analysis of faces with 11 edges

The faces with 11 edges in a PCC graph exhibit rich combinatorial complexity around them, and many of the vertices in their boundaries may have very small curvature. Therefore we are required to perform a more careful analysis than in other sections. In particular, we exploit much more heavily the machinery of special vertices. See also Definition 1.5 for the notion of blue edges, α-vertices and β-vertices.

Let $\sigma \in \widetilde{\mathcal{F}}$ with $|\sigma| = 11$ and $\phi(\sigma) \neq 0$. With our construction of the pairing, we have $\phi(v, \sigma) \neq 0$ if and only if $v \in V^{(f)}(\sigma)$ or v is a vertex special to σ. A vertex

can be special to σ as a consequence of all rules except $(R_{(3,a,b)})$. Let $A = \#\mathcal{A}$, $B = \#\mathcal{B}$, $C = \#\mathcal{C}$ and $D = \#\mathcal{D}$, where

$$\begin{aligned}
\mathcal{A} &:= \{v \in V^{(f)}(\sigma) : \ v \text{ is an } \alpha\text{-vertex}\},\\
\mathcal{B} &:= \{v \in V^{(f)}(\sigma) : \ v \text{ is a } \beta\text{-vertex}\},\\
\mathcal{C} &:= \{v \in V^{(f)}(\sigma) : \ f(v) = (3,11,11)\},\\
\mathcal{D} &:= \{v \in V^{(f)}(\sigma) : \ f(v) \in \{(3,11,12),(4,6,11),(3,3,4,11)\}\}.
\end{aligned}$$

and notice that A, B, C are even.

Lemma 4.3. *Suppose that $\exists v \in \mathcal{V}$ which is special to σ or that there is $v \in V^{(f)}(\sigma)$ that is not in $\mathcal{A} \cup \mathcal{B} \cup \mathcal{C} \cup \mathcal{D}$. Then $c(\sigma) > 0.001$.*

Proof. In the first case we have $c_+(\sigma) \geq c(3,4,4,5) > 0.023$ by checking all possible vertices special to σ. In the second case we have at least one vertex $v \in V^{(f)}(\sigma)$ with $f(v) \in \{(3,a,11) : \ a \leq 6\} \cup \{(4,4,11),(3,3,3,11)\}$ or at least two vertices $v_1, v_2 \in V^{(f)}(\sigma)$ with $f(v_i) \in \{(3,a,11),\ 7 \leq a \leq 10\} \cup \{(4,5,11)\}$, so $c_+(\sigma) \geq \frac{1}{3}c(3,3,3,11) > 0.027$. On the other hand, the negative courvature-contributions to σ come only from $v \in \mathcal{A} \cup \mathcal{B} \cup \mathcal{D}$ and we easily check that $c_-(\sigma) \geq 11 \cdot c(4,6,11) > -0.022$ In any case $c_+(\sigma) + c_-(\sigma) > 0.001$. □

We are now going to show that some configurations on the boundary of σ imply the existence of vertices special to σ. In fact, most of the special rules in section 3.3 were designed exactly to avoid these configurations. We let

$$\begin{aligned}
\mathcal{D}_1 &:= \{v \in V^{(f)}(\sigma) : \ f(v) = \langle 3,4,3,11\rangle\},\\
\mathcal{D}_2 &:= \{v \in V^{(f)}(\sigma) : \ f(v) = \langle 3,3,4,11\rangle \text{ (or } f(v) = \langle 4,3,3,11\rangle)\},\\
\mathcal{D}_6 &:= \{v \in V^{(f)}(\sigma) : \ f(v) = (4,6,11)\},\\
\mathcal{D}_{12} &:= \{v \in V^{(f)}(\sigma) : \ f(v) = (3,11,12)\}.
\end{aligned}$$

Lemma 4.4. *Let $v_1v_2 \in E^{(f)}(\sigma)$ so that either: (i) $v_1 \in \mathcal{D}_{12}$ and $v_2 \in \mathcal{D}_1$; (ii) $v_1 \in \mathcal{D}_{12}$ and $v_2 \in \mathcal{D}_2$; (iii) $v_1, v_2 \in \mathcal{D}_1$; (iv) $v_1 \in \mathcal{A}$ and $v_2 \in \mathcal{D}_1$; or (v) $v_1, v_2 \in \mathcal{D}_2$ and $s(v_1v_2) = (3,11)$. Then $\exists v \in \mathcal{V}$ which is special to σ.*

Proof. Notice that in every case $s(v_1v_2) = (3,11)$, so set $F^{(e)}(v_1v_2) = \{\tau, \sigma\}$ with $|\tau| = 3$ and let $w \in \mathcal{V}$ such that $V^{(e)}(\tau) = \{v_1, v_2, w\}$. In (i) we must have that $f(w) = (3,4,12)$, so w is special to σ by $(R_{(3,4,a)})$. In (ii) we have $f(w) \in \{(3,3,12),(3,3,3,12)\}$, so w is special to σ by either $(R_{(3,3,a)})$ or $(R_{(3,3,3,a)})$. In (iii) we have that either w is a TS-vertex or $f(w) = \langle 4,3,4,5\rangle$. In the first case w is special to σ by (R_{TS}). In the second case it's easy to see by diagram-chasing that w is not adjacent to vertices of type $(4,5,a)$ for $14 \leq a \leq 19$. Therefore w is special to σ by $(R_{(3,4,4,5)})$. In (iv) we have $f(w) = (3,4,13)$, so w is special to σ by $(R_{(3,4,a)})$. In the last case (v) we define w_1 as in fig. 4.3. If $f(w) = (3,3,3,a)$ with $a \geq 13$, then $f(w_1) = (3,4,a)$ and w_1 is special to σ by $(R_{(3,4,a)})$. Otherwise, w is special to σ by either (R_{TS}), $(R_{(3,3,3,a)})$ or $(R_{(3,3,3,3,5)})$. □

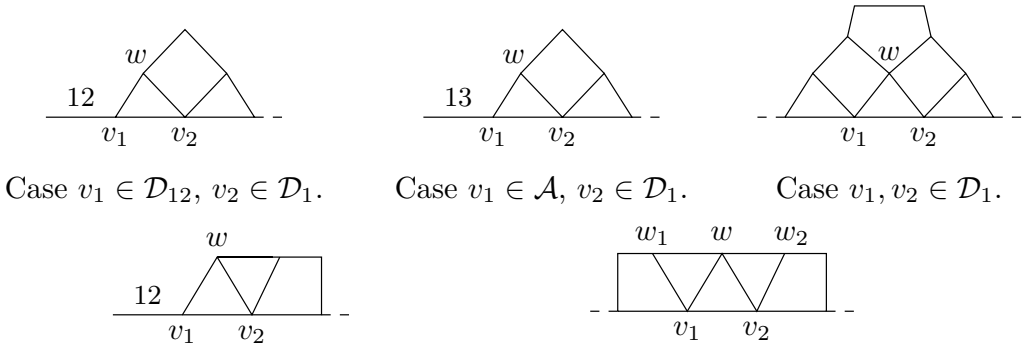


Figure 4.3: Illustrations for Lemma 4.4.

Lemma 4.5. *Let $v_1, v_2, v_3, v_4 \in V^{(f)}(\sigma)$ be consecutive vertices on σ and suppose that one of the following 6 cases holds: (i) $v_1, v_3 \in \mathcal{D}_2$ and $v_2 \in \mathcal{D}_1$; (ii) $v_1 \in \mathcal{D}_1$, $v_2 \in \mathcal{D}_2$ and $v_3 \in \mathcal{D}_6$; (iii) $v_1 \in \mathcal{D}_1$, $v_2, v_3 \in \mathcal{D}_2$ and $v_4 \in \mathcal{A}$; (iv) $v_1 \in \mathcal{C}$, $v_2 \in \mathcal{D}_2$ and $v_3 \in \mathcal{D}_6$; (v) $v_1 \in \mathcal{C}$, $v_2, v_3 \in \mathcal{D}_2$ and $v_4 \in \mathcal{A}$; (vi) $v_1 \in \mathcal{C}$, $v_2, v_3 \in \mathcal{D}_2$ and $v_4 \in \mathcal{D}_1$. Then $\exists v \in \mathcal{V}$ which is special to σ.*

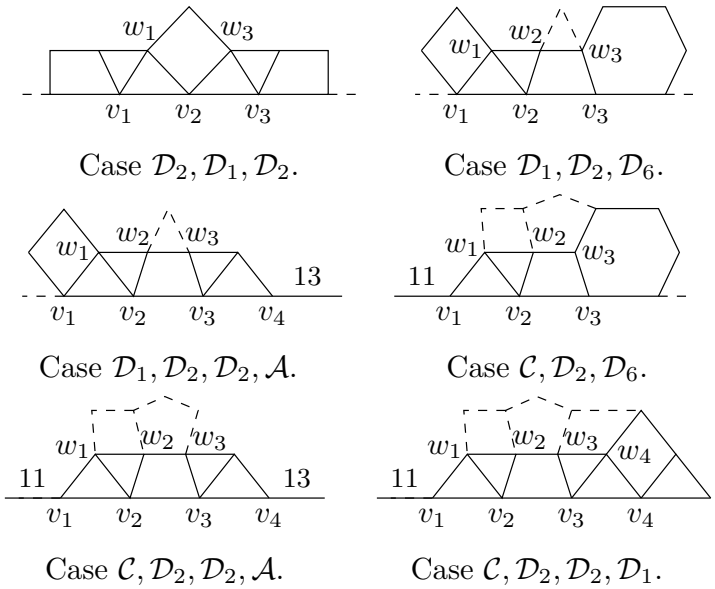


Figure 4.4: Illustrations for Lemma 4.5.

Proof. In case (i) we notice that $F^{(v)}(v_2) = \langle \tau_2, s_1, \tau_1, \sigma \rangle$ with $|\tau_1| = |\tau_2| = 3$ and $|s_1| = 4$, and let $V^{(f)}(s_1) = \langle w_1, v_2, w_3, w_2 \rangle$. It is impossible to have both $f(w_1) = (3, 3, 4, a)$ and $f(w_3) = (3, 3, 4, b)$ with $8 \leq a, b \leq 11$, so for some $i \in \{1, 3\}$ we have that w_i is special to σ by either (R_{TS}) or $(R_{(3,3,4,a)})$. In the remaining cases (ii)-(vi) let w_1, w_2, w_3 be as in fig. 4.4. If (ii) holds, then an

easy diagram-chasing reveals that w_1 is special to σ by (R_{TS}) or by $(R_{(3,3,4,a)})$, or w_2 is special to σ by $(R_{(3,4,a)})$, or w_3 is special to σ by either $(R_{(3,4,a)})$ or $(R_{(3,3,4,a)})$. In (iii) we have that w_1 is special by (R_{TS}) or $(R_{(3,3,4,a)})$, w_2 is special by $(R_{(3,4,a)})$, or w_3 is special to σ by (R_{TS}). In case (iv) we have that w_1 is special to σ by $(R_{(3,3,a)})$ or $(R_{(3,3,3,a)})$, or w_2 is special to σ by (R_{TS}), or w_3 is special to σ by $(R_{(4,5,6)})$. Similarly, if (v) holds, then either w_1 is special by $(R_{(3,3,a)})$ or $(R_{(3,3,3,a)})$, or w_2 is special by (R_{TS}), or w_3 is special to σ by $(R_{(3,4,a)})$ or $(R_{(3,3,4,a)})$. Finally, in case (vi) we have that w_1 is special by $(R_{(3,3,a)})$ or $(R_{(3,3,3,a)})$, or w_2 is special by (R_{TS}), or w_3 is special by $(R_{(3,4,a)})$ or $(R_{(3,3,4,a)})$, or w_4 is special to σ by (R_{TS}), where w_4 is as in fig. 4.4. □

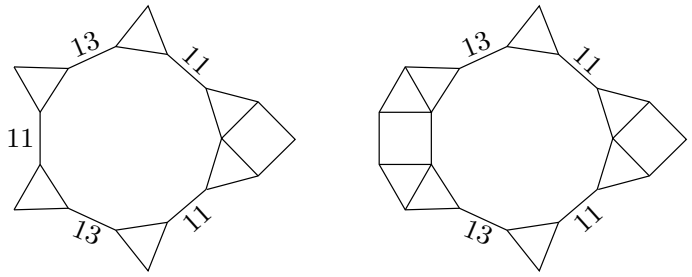
However, if we try to arrange the vertices in $\mathcal{A} \cup \mathcal{B} \cup \mathcal{C} \cup \mathcal{D}$ around the boundary of σ taking into account the constrained given by Lemma 4.4 and Lemma 4.5 to avoid the formation of special vertices, we get the following result.

Lemma 4.6 ([21, Lemmas 8.2, 8.4])**.** *Suppose that $V^{(f)}(\sigma) = \mathcal{A} \cup \mathcal{B} \cup \mathcal{C} \cup \mathcal{D}$ and that there are no special vertices to σ. Then there exist consecutive vertices v_1, v_2, v_3 on σ such that $v_2 \in \mathcal{D}_1$ and $v_1, v_3 \in \mathcal{C}$. Moreover we either have $C \geq 6$ or $A \geq 4$.*

By the above results, we can now conclude considering two final cases.

Proposition 4.7. *Let $\sigma \in \widetilde{\mathcal{F}}_{\phi \neq 0}$ with $|\sigma| = 11$. Then $c(\sigma) > 0.0003$.*

Proof. If one of the hypotheses of Lemma 4.3 is fulfilled, then $c(\sigma) > 0.001$. Otherwise, by Lemma 4.6 we have $C \geq 6$ or $A \geq 4$. If $C \geq 6$, then $c_+(\sigma) \geq 6\frac{1}{2}c(3, 11, 11) > 0.01674$ and $B \leq 4$. However $c_-(\sigma) > B\frac{3}{7}c(3, 11, 13) + (11 - B - C)c(3, 11, 12) > -0.01644$, and so $c(\sigma) > 0.0003$. If $A \geq 4$ instead, we easily see with the aid of Lemma 4.6 that necessarily $A = 4$, $B = 0$, $C = 4$, $D = 3$. As a consequence, $c_+(\sigma) = 4\frac{1}{2}c(3, 11, 11) > 0.01116$ and $c_-(\sigma) = 4\frac{1}{7}c(3, 11, 13) + 3c(3, 11, 12) > -0.01084$, hence $c(\sigma) > 0.00032$. □



Example with $C \geq 6$. Example with $A \geq 4$.

Figure 4.5: Examples of $\sigma \in \widetilde{\mathcal{F}}$ with $|\sigma| = 11$ and very small $c(\sigma)$.

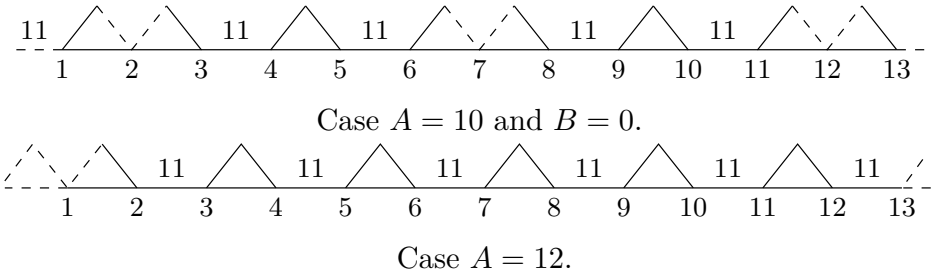


Figure 4.6: Illustrations for $|\sigma| = 13$.

4.4. Analysis of faces with 13 edges
Let $\sigma \in \widetilde{\mathcal{F}}$ with $|\sigma| = 13$ and $\phi(\sigma) \neq 0$. With our construction of the pairing, we have $\phi(v, \sigma) \neq 0$ if and only if $v \in V^{(f)}(\sigma)$, and $f(v)$ is one of the following 5 multisets: (3,3,13), (3,4,13), (3,11,13), (4,4,13), (3,3,3,13). As in the previous section, we refer to Definition 1.5 for the definition of α-vertices, β-vertices and blue edges. Let $A = \#\mathcal{A}$, $B = \#\mathcal{B}$, $C = \#\mathcal{C}$, $L = \#\mathcal{L}$ and $M = \#\mathcal{M}$, where

$$\begin{aligned}
\mathcal{A} &= \{v \in V^{(f)}(\sigma) : \ v \text{ is an } \alpha\text{-vertex}\},\\
\mathcal{B} &= \{v \in V^{(f)}(\sigma) : \ v \text{ is a } \beta\text{-vertex}\},\\
\mathcal{C} &= \{v \in V^{(f)}(\sigma) : \ f(v) \in \{(3,3,13),(3,4,13),(4,4,13),(3,3,3,13)\}\},\\
\mathcal{L} &= \{e \in E^{(f)}(\sigma) : \ s(e) = (11,13)\},\\
\mathcal{M} &= \{e \in E^{(f)}(\sigma) : \ e \text{ is a blue edge}\}.
\end{aligned}$$

We have $B = 2M$ and $2L = A + B \leq 12$, so A, B are even and $A + B \leq 12$. Moreover $C \geq 1$, because otherwise for every $v \in V^{(f)}(\sigma)$ we would have $f(v) = (a, b, 13)$ for some $a \in \{3, 4\}$ and $5 \leq b \leq 11$, which is not possible because 13 is odd.

Proposition 4.8. *Let $\sigma \in \widetilde{\mathcal{F}}_{\phi\neq 0}$ with $|\sigma| = 13$. Then $c(\sigma) > 0.00003$.*

Proof. First of all, observe that $c(\sigma) \geq (\frac{6}{7}A + \frac{4}{7}B) \cdot c(3,11,13) + C \cdot c(4,4,13)$. If $A+B \leq 8$ we get $c(\sigma) > 0.00967$. If $A+B = 10$ and $B \geq 2$ we get $c(\sigma) > 0.00005$. If $A = 10$ and $B = 0$ we necessarily have $C = 3$, so $c(\sigma) > 0.12995$. Finally, if $A + B = 12$ we necessarily have $A = 4$ and $B = 8$, so $c(\sigma) > 0.00003$. □

4.5. Analysis of faces with N edges, where $14 \leq N \leq 39$ and $N \neq 19$
Let $\sigma \in \widetilde{\mathcal{F}}$ with $\phi(\sigma) \neq 0$ and $|\sigma| = N$, where $14 \leq N \leq 39$ and $N \neq 19$. With our construction of the pairing, we have $\phi(v, \sigma) \neq 0$ if and only if $v \in \widehat{V}_3 \cup \widehat{V}_{sp}$, where

$$\begin{aligned}
\widehat{V}_3 &:= \{v \in V^{(f)}(\sigma) : \ 3 \in f(v)\},\\
\widehat{V}_{sp} &:= \{v \in \mathcal{V} : \ v \text{ is special to } \sigma\}.
\end{aligned}$$

Notice that for all $v \in \widehat{V}_{sp}$ we have $f(v) = (3, a, b)$ for some $a, b \in \{7, 8, 9, 10\}$. In order to prove $c(\sigma) > 0$ we are going to discharge the curvature-contribution from $\widehat{V}_3 \cup \widehat{V}_{sp}$ to $\widehat{E} := \{e \in E^{(f)}(\sigma) : \ s(e) = (3, N)\}$.

Definition 4.9. For all $v \in \widehat{V}_{sp}$ let $\hat{\tau}_v$ be the only triangle in $F^{(v)}(v)$ and let $\hat{e}_v := \mathrm{opp}^{(e)}(v, \hat{\tau}_v) \in \widehat{E}$. For every $v \in \widehat{V}_3$ we let $\widehat{E}_v := E^{(v)}(v) \cap \widehat{E}$.

Then we define the discharging function as follows.

Definition 4.10. Let $\hat{\phi} : (\widehat{V}_3 \cup \widehat{V}_{sp}) \times \widehat{E} \to \mathbb{Q}_{\geq 0}$ be the only function such that: (i) $\hat{\phi}(v, e) = (\#\widehat{E}_v)^{-1}\phi(v, \sigma)$ for all $v \in \widehat{V}_3$ and $e \in \widehat{E}_v$ (ii) $\hat{\phi}(v, \hat{e}_v) = \phi(v, \sigma)$ if $v \in \widehat{V}_{sp}$; (iii) $\hat{\phi}(v, e) = 0$ otherwise.

Notice that $\sum_{e \in \widehat{E}} \hat{\phi}(v, e) = \phi(v, \sigma)$ for all $v \in \widehat{V}_3 \cup \widehat{V}_{sp}$. Moreover we have

$$c(\sigma) = \sum_{e \in \widehat{E}} \hat{c}(e), \tag{4.1}$$

where $\hat{c}(e) := \sum_{v \in \widehat{V}} c_v \hat{\phi}(v, e)$ for all $e \in \widehat{E}$.

Proposition 4.11. *Let $\sigma \in \widetilde{\mathcal{F}}$ with $\phi(\sigma) \neq 0$ and $|\sigma| = N$, where $14 \leq N \leq 39$ and $N \neq 19$. Then $c(\sigma) > 0.0002$.*

Proof. It suffices to prove $\hat{c}(e) > 0.0002$ for every $e = v_1 v_2 \in \widehat{E}$. Let

$$\mathcal{A} := \{v \in V^{(f)}(\sigma) : \ f(v) = (3, a, N), \ a \in \{7, 8, 9, 10\}, .$$

If $v_1 \notin \mathcal{A}$ or $v_2 \notin \mathcal{A}$, we simply have $\hat{c}(e) = c_{v_1}\hat{\phi}(v_1, e) + c_{v_2}\hat{\phi}(v_2, e)$ and so $\hat{c}(e) \geq c(3, 8, 23) + \frac{1}{2}c(3, 3, 3, N) > 0.0002$. If $v_1, v_2 \in \mathcal{A}$ then $e = \hat{e}_v$ for some $v \in \widehat{V}_{sp}$. Therefore $\hat{c}(e) \geq 2 \cdot c(3, 8, 23) + c(3, 10, 10) > 0.007$. □

4.6. Analysis of faces with 19 edges

Let $\sigma \in \widetilde{\mathcal{F}}$ with $|\sigma| = 19$ and $\phi(\sigma) \neq 0$. We have $\phi(v, \sigma) \neq 0$ if and only if $v \in V^{(f)}(\sigma)$ and either $4 \in f(v)$ or $5 \in f(v)$. Let $A = \#\mathcal{A}$, with

$$\mathcal{A} = \{v \in V^{(f)}(\sigma) : \ f(v) = \{(4, 5, 19)\}.$$

Proposition 4.12. *Let $\sigma \in \widetilde{\mathcal{F}}_{\phi \neq 0}$ with $|\sigma| = 19$. Then $c(\sigma) > 0.0065$.*

Proof. If $A = 0$ we have $c(\sigma) = c_+(\sigma) \geq \frac{1}{2}c(3, 5, 19) > 0.038$. Otherwise we have $1 \leq A \leq 18$, and there exist $v \in \mathcal{A}$ and $w \in V^{(f)}(\sigma) \backslash \mathcal{A}$ that are consecutive on σ, hence $4 \in f(w)$ or $5 \in f(w)$. In this case $c_w \phi(w, \sigma) \geq \frac{1}{2}c(3, 5, 19)$, so $c(\sigma) \geq A\frac{1}{4}c(4, 5, 19) + \frac{1}{2}c(3, 5, 19) > 0.0065$. □

4.6.1. Analysis of faces with 40 or 41 edges

Let $\sigma \in \widetilde{\mathcal{F}}$ with $\phi(\sigma) \neq 0$ and $|\sigma| = N$, where $N \in \{40, 41\}$. With our construction of the pairing, we have $\phi(v, \sigma) \neq 0$ if and only if $v \in \widetilde{V}_3 \cup \widetilde{V}_{sp}$, where

$$\widetilde{V}_3 := \{v \in V^{(f)}(\sigma) : \ 3 \in f(v)\},$$
$$\widetilde{V}_{sp} := \{v \in \mathcal{V} : \ v \text{ is special to } \sigma\}.$$

Moreover we can write $\widetilde{V}_{sp} = \widetilde{V}_{sp}^{(1)} \cup \widetilde{V}_{sp}^{(2)}$, where

$$\widetilde{V}_{sp}^{(1)} := \{v \in \widetilde{V}_{sp} : \ f(v) \in \{(3,6,7),(3,7,7)\}\},$$
$$\widetilde{V}_{sp}^{(2)} := \{v \in \widetilde{V}_{sp} : \ f(v) \in \{(3,3,3,5),(3,3,3,3,5)\}\}.$$

To prove $c(\sigma) > 0$ we will discharge c_v from $\widetilde{V}_3 \cup \widetilde{V}_{sp}$ to $\widetilde{E} \cup \{\clubsuit\}$, where $\widetilde{E} := \{e \in E^{(f)}(\sigma) : \ s(e) = (3,N)\}$ and $\clubsuit$ is an auxiliary symbol.

Definition 4.13. For all $v \in \widetilde{V}_{sp}^{(1)}$ we let $\widetilde{\tau}_v$ be the only triangle in $F^{(v)}(v)$ and let $\widetilde{e}_v := \mathrm{opp}^{(e)}(v, \widetilde{\tau}_v) \in \widetilde{E}$. For every $v \in \widetilde{V}_3$ let $\widetilde{E}_v := E^{(v)}(v) \cap \widetilde{E}$.

We define the discharging function as follows.

Definition 4.14. Let $\widetilde{\phi} : (\widetilde{V}_3 \cup \widetilde{V}_{sp}) \times (\widetilde{E} \cup \{\clubsuit\}) \to \mathbb{Q}_{\geq 0}$ be the only function such that: (i) $\widetilde{\phi}(v,e) = (\#\widetilde{E}_v)^{-1}\phi(v,\sigma)$ for all $v \in \widetilde{V}_3$ and $e \in \widetilde{E}_v$; (ii) $\widetilde{\phi}(v,\widetilde{e}_v) = \phi(v,\sigma)$ for $v \in \widetilde{V}_{sp}^{(1)}$; (iii) $\widetilde{\phi}(v,\clubsuit) = \phi(v,\sigma)$ if $v \in \widetilde{V}_{sp}^{(2)}$; (iv) $\widetilde{\phi}(v,e) = 0$ otherwise.

Notice that $\sum_{e \in \widetilde{E} \cup \{\clubsuit\}} \widetilde{\phi}(v,e) = \phi(v,\sigma)$ for all $v \in \widetilde{V}_3 \cup \widetilde{V}_{sp}$. For every $e \in \widetilde{E} \cup \{\clubsuit\}$ we let $\widetilde{c}(e) := \sum_{v \in \widetilde{V}} c_v \widetilde{\phi}(v,e)$ and

$$\widetilde{c}_-(\sigma) := \sum_{e \in \widetilde{E}} \min\{0, \widetilde{c}(e)\} \quad \text{and} \quad \widetilde{c}_+(\sigma) := \sum_{e \in \widetilde{E}} \max\{0, \widetilde{c}(e)\},$$

so that $c(\sigma) = \widetilde{c}_-(\sigma) + \widetilde{c}_+(\sigma) + \widetilde{c}(\clubsuit)$. We begin with an estimate of $\widetilde{c}_-(\sigma)$.

Lemma 4.15. *We have $\widetilde{c}_-(\sigma) > -0.066$.*

Proof. We notice that the vertices $v \in \widetilde{V}_3 \cup \widetilde{V}_{sp}$ with $c_v < 0$ are those in

$$\mathcal{A} := \{v \in V^{(f)}(\sigma) : \ f(v) = (3,7,N)\}.$$

So, if $e = v_1v_2 \in \widetilde{E}$ and $v_1, v_2 \notin \mathcal{A}$ we have $\widetilde{c}(e) > 0$. If $v_1 \in \mathcal{A}$ and $v_2 \notin \mathcal{A}$ then $\widetilde{c}(e) \geq c(3,7,N) + \frac{1}{2}c(3,3,3,N) > -0.0016$. If instead $v_1, v_2 \in \mathcal{A}$, then $e = \widetilde{e}(v)$ for some $v \in \widetilde{V}_{sp}^{(1)}$, so $\widetilde{c}(e) = 2 \cdot c(3,7,N) + c(3,7,7) > 0$. A rough estimate is then $\widetilde{c}_-(\sigma) \geq 41 \cdot (-0.0016) > -0.066$. □

Our strategy from this point is based on the observation that the most efficient way to build a large PCC graph from a large face such as σ, is to have a pattern of three consecutive vertices $e_1, e_2, e_3 \in V^{(f)}(\sigma)$ with $s(e_1) = (7,N)$ and $s(e_2) = s(e_3) = (3,N)$ repeated along the boundary of σ (compare with fig. 8.1). In the following three lemmas we prove that it is "inefficient" to do otherwise. Then in Proposition 4.19 we will exploit the fact that $N \in \{40,41\}$ is not divisible by three, and so the inefficient configurations are unavoidable.

Lemma 4.16. *Suppose $\exists v \in V^{(f)}(\sigma)$ such that $f(v) = (3,3,N)$, or $4 \in f(v)$, or $5 \in f(v)$. Then $\widetilde{c}_+(\sigma) > 0.078$.*

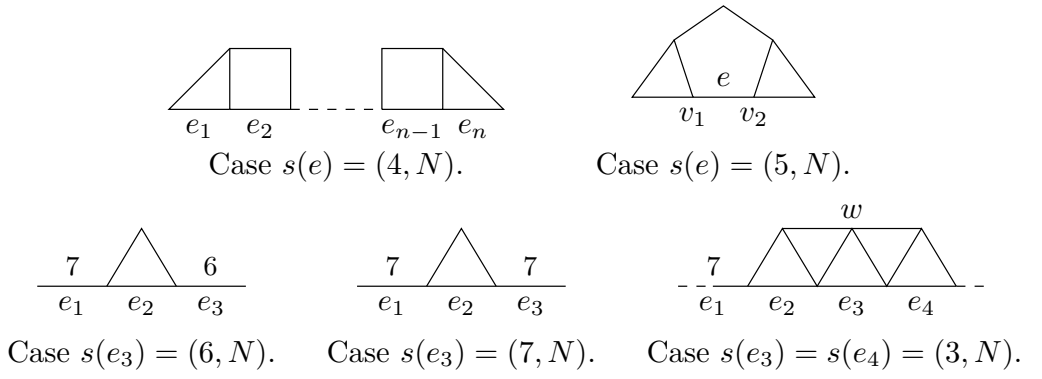


Figure 4.7: Some unavoidable configurations analyzed in Lemmas 4.16 to 4.18.

.

Proof. If $f(v) = (3, 3, N)$, take $e \in E^{(v)}(v)$. Then $\widetilde{c}_+(\sigma) \geq \widetilde{c}(e)$ and $\widetilde{c}(e) \geq \frac{1}{2}c(3, 3, N)+c(3, 7, N) > 0.0081$. If $4 \in f(v)$, then there exist $e_0, \ldots, e_n \in E^{(f)}(\sigma)$ consecutive on σ, with $2 \leq n \leq N$, such that $e_0, e_n \in \widetilde{E}$ and $s(e_1), \ldots, s(e_{n-1}) = (4, N)$. When $n < N$ we get $\widetilde{c}_+(\sigma) \geq \widetilde{c}(e_0)+\widetilde{c}(e_n) > 2\cdot(\frac{1}{2}c(3, 4, N)+c(3, 7, N)) > 0.080$. When $n = N$ instead we have $\widetilde{c}_+(\sigma) \geq \widetilde{c}(e_0) > 2 \cdot \frac{1}{2}c(3, 4, N) > 0.098$. Finally, if $5 \in f(v)$, then there exist $e_0, e_1, e_2 \in E^{(f)}(\sigma)$ consecutive on σ with $e_0, e_2 \in \widetilde{E}$ and $s(e_1) = (5, N)$. Then $\widetilde{c}_+(\sigma) \geq \widetilde{c}(e_0) + \widetilde{c}(e_2) > 2 \cdot (c(3, 5, N) + c(3, 7, N)) > 0.078$. □

Lemma 4.17. *Suppose $\exists e_1, e_2, e_3 \in E^{(f)}(\sigma)$ consecutive on σ such that $s(e_1) = (7, N)$ but $s(e_3) \neq (3, N)$. Then $\widetilde{c}_+(\sigma) > 0.078$.*

Proof. If $s(e_3) \in \{(4, N), (5, N)\}$ we have $\widetilde{c}_+(\sigma) > 0.078$ by Lemma 4.16. If $s(e_3) \in \{(6, N), (7, N)\}$ we have that $e_2 \in \widetilde{E}$ and $e_2 = \widetilde{e}_v$ for some $v \in \widetilde{V}_{sp}^{(1)}$. Then $\widetilde{c}_+(\sigma) \geq \widetilde{c}(e_2)$ and $\widetilde{c}(e_2) \geq c(4, 7, N) \times 2 + c(3, 7, 7) > 0.091$. □

Lemma 4.18. *Suppose there exist $e_1, e_2, e_3, e_4 \in E^{(f)}(\sigma)$ consecutive edges on σ, with $s(e_1) = (7, N)$ and $s(e_2) = s(e_3) = s(e_4) = (3, N)$. Then $\widetilde{c}(\clubsuit) > 0.023$.*

Proof. Let $w \in \mathcal{V}$ such that $e_3 = \mathrm{opp}^{(e)}(w, \tau)$ for some triangle $\tau \in \mathcal{F}$. We observe that $\{3, 3, 3\} \subseteq f(w)$, but $f(w) \neq (3, 3, 3, a)$ for $a \geq 6$. Therefore $w \in \widetilde{V}_{sp}^{(2)}$ by (R_{TS}), $(R_{(3,3,3,a)})$ or $(R_{(3,3,3,3,5)})$. So $\widetilde{c}(\clubsuit) \geq c(3, 3, 3, 3, 5) > 0.023$. □

Proposition 4.19. *Let $\sigma \in \widetilde{\mathcal{F}}_{\phi\neq 0}$ with $|\sigma| = N \in \{40, 41\}$. Then $c(\sigma) > 0.011$.*

Proof. First, by Lemma 4.15 we have $\widetilde{c}_-(\sigma) > 41 \cdot (-0.0016) > -0.066$. We may suppose that $f(v) \in \{(3, 6, N), (3, 7, N), (3, 3, 3, N)\}$ for all $v \in V^{(f)}(\sigma)$ because otherwise $\widetilde{c}(\sigma) > 0.012$ by Lemma 4.15. Let $L = \#\mathcal{L}$ with

$$\mathcal{L} := \{e \in E^{(f)}(\sigma) : \ s(e) = (7, N)\}.$$

If $L \leq 12$ we check directly that $c(\sigma) \geq 2L \cdot c(3, 7, N) + (N - 2L) \cdot c(3, 3, 3, 7) > 0.0352$. If $L \geq 14 > N/3$ then there exist $e_1, e_2, e_3 \in E^{(f)}(\sigma)$ consecutive on

σ such that $s(e_1) = s(e_3) = (7, N)$, so again $\widetilde{c}(\sigma) > 0.012$ by Lemma 4.17. Suppose now $L = 13$ and notice that $40 = 12 \cdot 3 + 4$ and $41 = 12 \cdot 3 + 5 = 11 \cdot 3 + 4 + 4$. If Lemma 4.17 doesn't apply, then we immediately see that there exist $e_1, e_2, e_3, e_4 \in E^{(f)}(\sigma)$ consecutive edges on σ, with $s(e_1) = (7, N)$ and $s(e_2) = s(e_3) = s(e_4) = (3, N)$. In this case we have $\widetilde{c}(\clubsuit) > 0.023$ by Lemma 4.18 and we deduce that $c(\sigma) > 26 \cdot c(3, 7, N) + (N-26) \cdot c(3, 3, 3, N) + \widetilde{c}(\clubsuit) > 0.011$. $\square$

5. There are no faces with more than 41 edges

In this section we prove a result of independent interest in the classification of PCC graphs, namely that a PCC graph cannot have a face of size greater or equal to 42.

Theorem 5.1. *Suppose G is a PCC graph. Then for all $\sigma \in \mathcal{F}$ we have $|\sigma| \leq 41$.*

Before we start the proof, we pause to record two useful lemmas. Both essentially say that two large faces in a PCC graph cannot be too close without merging, and are proved through a tedious but strightforward diagram chasing. The first lemma is a refinement of [5, Lemma 4.2] and it states that two adjacent vertices on G, both sited on the boundary of some "large" faces, must be consecutive boundary vertices on the same face.

Lemma 5.2 ([21, Lemma 2.4])**.** *Let G be a PCC graph, let $v_1, v_2 \in \mathcal{V}$ and let $\sigma_1, \sigma_2 \in \mathcal{F}$ with $\sigma_i \in F^{(v)}(v_i)$ and $|\sigma_i| \geq 20$. If $v_1 v_2 \in \mathcal{E}$ (i.e. v_1, v_2 are adjacent in G), then $\sigma_1 = \sigma_2$ and $v_1 v_2 \in E^{(f)}(\sigma)$ (i.e. v_1, v_2 are consecutive on σ).*

The second lemma states that a "small" face κ cannot share two or more edges with "large" faces.

Lemma 5.3 ([21, Lemma 2.5])**.** *Let $\sigma_1, \sigma_2, \kappa \in \mathcal{F}$ with $|\sigma_1|, |\sigma_2| \geq 20$ and $|\kappa| \leq 6$. Let also $e_1, e_2 \in E^{(f)}(\kappa)$ with $e_i \in E^{(f)}(\sigma_i)$. Then $\sigma_1 = \sigma_2$ and $e_1 = e_2$.*

We are now ready to prove the theorem.

Proof of Theorem 5.1. Let $\sigma \in \mathcal{F}$ with $|\sigma| = N \geq 42$. Our strategy to prove that σ cannot exist is to find a set $\mathcal{V}_{42}$ of vertices with $V^{(f)}(\sigma) \subseteq \mathcal{V}_{42} \subseteq \mathcal{V}$ whose sum of the curvatures exceeds 2, thus contradicting (2.1). This idea of looking at the neighborhood of a big face to control its size can be traced back to [6]. First we consider the vertices on the boundary of σ and we write $V^{(f)}(\sigma) = \bigcup_{k=1}^{6} \mathcal{A}_k$ where:

$$\begin{aligned}
\mathcal{A}_1 &:= \{v \in V^{(f)}(\sigma) : \ f(v) = (3, 3, 3, N)\}, \\
\mathcal{A}_2 &:= \{v \in V^{(f)}(\sigma) : \ f(v) = (4, 4, N)\}, \\
\mathcal{A}_k &:= \{v \in V^{(f)}(\sigma) : \ f(v) = (3, k, N)\}, \qquad \text{if } 3 \leq k \leq 6.
\end{aligned}$$

We notice that: $K(v) = \frac{1}{N} + \frac{1}{6}$ if $v \in \mathcal{A}_3$; $K(v) = \frac{1}{N} + \frac{1}{12}$ if $v \in \mathcal{A}_4$; $K(v) = \frac{1}{N} + \frac{1}{30}$ if $v \in \mathcal{A}_5$; and $K(v) = \frac{1}{N}$ if $v \in \mathcal{A}_1 \cup \mathcal{A}_2 \cup \mathcal{A}_6$. Next, we consider some vertices

at "distance one" from the boundary of σ, namely $v \in \mathcal{C}_1 \cup \mathcal{C}_2$, where

$$\begin{aligned}
\mathcal{T} &:= \{\tau \in \mathcal{F} :\ V^{(f)}(\tau) = \{v, v_1, v_2\} \text{ with } v_1, v_2 \in \mathcal{A}_1 \cup \mathcal{A}_5 \cup \mathcal{A}_6\};\\
\mathcal{C}_1 &:= \{v \in \mathcal{V} \setminus V^{(f)}(\sigma) :\ \exists \tau \in \mathcal{T} \text{ with } v \in V^{(f)}(\tau)\};\\
\mathcal{C}_2 &:= \{v \in \mathcal{V} \setminus V^{(f)}(\sigma) :\ \exists v' \in \mathcal{A}_2 \text{ with } vv' \in \mathcal{E}\}.
\end{aligned}$$

With the aid of Lemma 5.3 it is not difficult to show that $\mathcal{C}_1$ and $\mathcal{T}$ are in natural bijection (see also fig. 5.1).

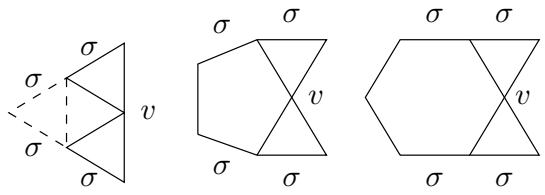


Figure 5.1: Some configurations that arise if $v \in \mathcal{C}_1$ belongs to two triangles in $\mathcal{T}$.

Lemma 5.4 ([21, Lemma 13.2])**.** *For every $v \in \mathcal{C}_1$ there is a unique $\tau_v \in \mathcal{T}$ with $v \in V^{(f)}(\tau_v)$. Conversely, for all $\tau \in \mathcal{T}$ there exists exactly one $v \in \mathcal{C}_1$ with $v \in V^{(f)}(\tau)$.*

Similarly, C_2 is in bijection with $\mathcal{A}_2$.

Lemma 5.5 ([21, Lemma 13.3])**.** *For every $v \in \mathcal{C}_2$ there is a unique $v' \in \mathcal{A}_2$ with $vv' \in \mathcal{E}$. Conversely, for every $v \in \mathcal{A}_2$ there is a unique $v' \in \mathcal{C}_2$ with $vv' \in \mathcal{E}$.*

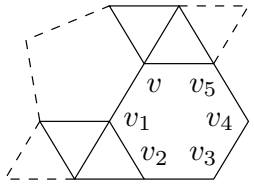
Finally, we single out the special subsets $\mathcal{C}_{15}, \mathcal{C}_{16} \subseteq \mathcal{C}_1$ and we consider the vertices $v \in \mathcal{D}$ at "distance at most two" from the boundary of σ, where:

$$\begin{aligned}
\mathcal{C}_{15} &:= \{v \in \mathcal{C}_1 :\ V^{(f)}(\tau_v) = \{v, v_1, v_2\} \text{ with } v_1 \in \mathcal{A}_1 \text{ and } v_2 \in \mathcal{A}_5\};\\
\mathcal{C}_{16} &:= \{v \in \mathcal{C}_1 :\ V^{(f)}(\tau_v) = \{v, v_1, v_2\} \text{ with } v_1 \in \mathcal{A}_1 \text{ and } v_2 \in \mathcal{A}_6\};\\
\mathcal{D} &:= \{v \in \mathcal{V} :\ \exists v' \in \mathcal{C}_{16} \text{ with } vv' \in \mathcal{E} \text{ and } s(vv') = (5, 6)\}.
\end{aligned}$$

Notice that $(4, 4) \subseteq f(v)$ for all $v \in \mathcal{C}_2$ and $(5, 6) \subseteq f(v)$ for all v in $\mathcal{D}$. We infer from this that $\mathcal{D}$ is disjoint from $\mathcal{C}_2$ and from $V^{(f)}(\sigma)$. We could prove that it is disjoint from $\mathcal{C}_1$ as well, but it is simpler, and sufficient for our purposes, to show only the following.

Lemma 5.6. *If $v \in \mathcal{C}_1 \cap \mathcal{D}$, then $v \in \mathcal{C}_{15}$.*

Proof. If $v \in \mathcal{C}_1 \cap \mathcal{D}$ we have $(3, 5, 6) \subseteq f(v)$, and so $f(v) \in \{(3, 5, 6), (3, 3, 5, 6)\}$. Let $\sigma_1, \sigma_2 \in F^{(v)}(v)$ with $|\sigma_1| = 6$ and $|\sigma_2| = 5$, and let $V^{(f)}(\sigma_1) = \langle v, v_1, \ldots, v_5 \rangle$ so that $E^{(f)}(vv_1) = \{\sigma_1, \sigma_2\}$. By the definition of $\mathcal{D}$, we have that $v_1 \in \mathcal{C}_{16}$. This easily implies that $v_2 \in \mathcal{A}_6$, and so necessarily $s(v_2v_3) = (6, N)$. Let now $\tau_v \in F^{(v)}(v)$ as in Lemma 5.4. We notice that we cannot have $v_5 \in V^{(f)}(\tau_v)$. Indeed, otherwise we would have $v_5 \in \mathcal{A}_6$ and so $s(v_4v_5) = (6, N)$. But this is in



Case $v \in \mathcal{C}_1 \cap \mathcal{D}$.

Figure 5.2: Illustration for Lemma 5.6.

contradiction with Lemma 5.3. Therefore we must have $F^{(v)}(v) = \langle \tau_v, \tau', \sigma_1, \sigma_2 \rangle$ for some $\tau' \in \mathcal{F}$ with $|\tau'| = 3$. Let $V^{(f)}(\tau_v) = \{v, w_1 w_2\}$ with $s(vw_1) = (3, 5)$. Then $\{\tau_v, \sigma_1\} \subseteq F^{(v)}(w_1)$ and $\{\tau_v, \tau'\} \subseteq F^{(v)}(w_2)$. Since $\tau_v \in \mathcal{T}$, we must have $w_1 \in \mathcal{A}_5$ and $w_2 \in \mathcal{A}_1$, so $v \in \mathcal{C}_{15}$. □

For every $v \in \mathcal{D}$ we associate the subset $\mathcal{S}_{\mathcal{D}}(v) \subseteq \mathcal{C}_1$ given by

$$\mathcal{S}_{\mathcal{D}}(v) := \{v' \in \mathcal{C}_{16} \text{ with } vv' \in \mathcal{E} \text{ and } s(vv') = (5, 6)\} \cup (\{v\} \cap \mathcal{C}_1)$$

and we notice in passing that $s_v := \#\mathcal{S}_{\mathcal{D}}(v) \leq 2$.
Now, we define $\mathcal{V}_{42} := V^{(f)}(\sigma) \cup \mathcal{C}_1 \cup \mathcal{C}_2 \cup \mathcal{D}$ and we claim that the sum of curvatures $\sum_{v \in \mathcal{V}_{42}} K(v)$ exceeds 2. To prove this, we discharge these curvatures to a new set $\mathcal{F}_{42}$ according to the following pairing.

Definition 5.7. Let $\mathcal{F}_{42} := \mathcal{A}_2 \cup \mathcal{B}$ where

$$\mathcal{B} := \{e \in E^{(f)}(\sigma) : \ s(e) = (3, N)\}.$$

and for every $v \in \mathcal{C}_1$ let $e_v := \text{opp}^{(e)}(v, \tau_v) \in \mathcal{B}$. There exists a unique function $\phi_{42} : \mathcal{V}_{42} \times \mathcal{F}_{42} \to \mathbb{Q}_{\geq 0}$ such that:

- $\phi_{42}(v, e_1) = \phi_{42}(v, e_2) = \frac{1}{2}$ for $v \in \mathcal{A}_1 \cup \mathcal{A}_3$ and $\mathcal{B} \cap E^{(v)}(v) = \{e_1, e_2\}$;
- $\phi_{42}(v, e) = 1$ for $v \in \mathcal{A}_4 \cup \mathcal{A}_5 \cup \mathcal{A}_6$ and $e \in \mathcal{B} \cap E^{(v)}(v)$;
- $\phi_{42}(v, v) = 1$ for $v \in \mathcal{A}_2$;
- $\phi_{42}(v, v') = 1$ for $v \in \mathcal{C}_2$ and $v' \in \mathcal{A}_2$ with $vv' \in \mathcal{E}$;
- $\phi_{42}(v, e_v) = 1$ for $v \in \mathcal{C}_1 \setminus \mathcal{D}$;
- $\phi_{42}(v, e_{v'}) = 1/s_v$ for $v \in \mathcal{D}$ and all $v' \in \mathcal{S}_{\mathcal{D}}(v)$;

and is zero otherwise.

Notice that for every $v \in \mathcal{V}_{42}$ we have $\sum_{x \in \mathcal{F}_{42}} \phi_{42}(v, x) = 1$. For every $x \in \mathcal{F}_{42}$ we define

$$c_x := \sum_{v \in \mathcal{V}_{42}} \phi_{42}(v, x) K(v);$$

$$\omega_x := \sum_{v \in V^{(f)}(\sigma)} \phi_{42}(v, x),$$

so that

$$\begin{aligned}\sum_{x\in\mathcal{F}_{42}} c_x &= \sum_{x\in\mathcal{F}_{42}}\sum_{v\in\mathcal{V}_{42}} \phi_{42}(v,x)K(v) = \sum_{v\in\mathcal{V}_{42}} K(v);\\ \sum_{x\in\mathcal{F}_{42}} \omega_x &= \sum_{x\in\mathcal{F}_{42}}\sum_{v\in V^{(f)}(\sigma)} \phi_{42}(x,v) = \sum_{v\in V^{(f)}(\sigma)} 1 = N.\end{aligned} \tag{5.1}$$

Notice also that for every $x \in \mathcal{F}_{42}$ all summands in the definition of c_x are nonnegative. Therefore any partial sum of them will give an estimate of c_x from below. We are going to show the following

Lemma 5.8. *For every $x \in \mathcal{F}_{42}$ we have $c_x > \omega_x\left(\frac{1}{N}+\frac{1}{42}\right)$.*

Proof. If $x \in \mathcal{A}_2$, then there is $v \in \mathcal{C}_2$ with $\phi_{42}(v,x) = 1$. We have that $\omega_x = \phi_{42}(x,x) = 1$ and it is easy to check that $K(v) \geq \frac{1}{30}$. Hence $c_x \geq \frac{1}{N}+\frac{1}{30} > \omega_x\left(\frac{1}{N}+\frac{1}{42}\right)$. If instead $x \in \mathcal{B}$, we have $x = v_1v_2$ for some $v_1, v_2 \in V^{(f)}(\sigma)\setminus\mathcal{A}_2$. The lemma is easy to prove if v_1 or v_2 is in $\mathcal{A}_3\cup\mathcal{A}_4$, therefore now we assume that $v_1, v_2 \in \mathcal{A}_1\cup\mathcal{A}_5\cup\mathcal{A}_6$ and so $v_1v_2 = e_v$ for some $v \in \mathcal{C}_1$. If both $v_1, v_2 \in \mathcal{A}_5\cup\mathcal{A}_6$ we have that $\omega_x = 2$ and $f(v) \in \{(3,5,5),(3,5,6),(3,6,6),(3,3,5,6)\}$ and so the inequality $c_x > \omega_x\left(\frac{1}{N}+\frac{1}{42}\right)$ is an easy check. Similarly, if both $v_1, v_2 \in \mathcal{A}_1$ we have $\omega_x = 1$ and $K(v) \geq \frac{1}{30}$, with equality if $f(v) = (3,3,3,3,5)$. Hence $c_x \geq \frac{1}{2}\cdot\frac{1}{N}+\frac{1}{2}\cdot\frac{1}{N}+\frac{1}{30} > \omega_x\left(\frac{1}{N}+\frac{1}{42}\right)$. Thus, without loss of generality, there are only two cases missing.

If $v_1 \in \mathcal{A}_1$ and $v_2 \in \mathcal{A}_5$, then $\omega_x = \frac{1}{2}+1 = \frac{3}{2}$ and $(3,3,5) \subseteq f(v)$. If $v \in \mathcal{D}$, then $f(v) = (3,3,5,6)$ and $\phi_{42}(v,x) = \frac{1}{2}$, so $\phi_{42}(v,x)K(v) = \frac{1}{2}\cdot\frac{1}{30} = \frac{1}{60}$. If otherwise $v \notin \mathcal{D}$, then $\phi_{42}(v,x) = 1$ and $K(v) \geq \frac{1}{210}$ with equality if $f(v) = (3,3,5,7)$. In any case $c_x \geq \frac{1}{2}\cdot\frac{1}{N}+\left(\frac{1}{N}+\frac{1}{30}\right)+\frac{1}{210} = \frac{3}{2}\left(\frac{1}{N}+\frac{2}{3}\frac{8}{210}\right) > \omega_x\left(\frac{1}{N}+\frac{1}{42}\right)$, because $\frac{1}{42} = \frac{15}{630} < \frac{16}{630}$.

If $v_1 \in \mathcal{A}_1$, $v_2 \in \mathcal{A}_6$, then $\omega_x = \frac{3}{2}$ and $(3,3,6) \subseteq f(v)$. If $f(v) \neq (3,3,5,6)$, then $K(v) \geq \frac{1}{12}$ and the lemma is easily obtained. Otherwise $K(v) = \frac{1}{30}$, and there is $w \in \mathcal{D}$ with $s(vw) = (5,6)$. If $w \in \mathcal{D}\setminus\mathcal{C}_1$ and $s_w = 1$, then $\phi_{42}(w,x) = 1$ and $K(w) \geq \frac{1}{210}$, with equality if $f(w) = (5,6,7)$. Else, we have $\phi_{42}(w,x) = \frac{1}{2}$ and $K(w) \geq \frac{1}{30}$, with equality if $f(w) \in \{(5,6,6),(3,3,5,6)\}$. In any case $\phi_{42}(w,x)K(w) \geq \frac{1}{210}$ and so $c_x \geq \frac{1}{2}\cdot\frac{1}{N}+\frac{1}{N}+\frac{1}{30}+\frac{1}{210} = \frac{3}{2}\left(\frac{1}{N}+\frac{2}{3}\frac{8}{210}\right) > \omega_x\left(\frac{1}{N}+\frac{1}{42}\right)$, because $\frac{1}{42} = \frac{15}{630} < \frac{16}{630}$.

□

By (2.1) we have $\sum_{v\in\mathcal{V}_{42}} K(v) \leq 2$. Then, by (5.1) and Lemma 5.8 we get

$$2 \geq \sum_{x\in\mathcal{F}_{42}} c_x > \sum_{x\in\mathcal{F}_{42}} \omega_x\left(\frac{1}{N}+\frac{1}{42}\right) = 1+\frac{N}{42},$$

which is a contradiction if $N \geq 42$. Thus Theorem 5.1 is proved. □

6. Analysis of the auxiliary faces and proving $\#\mathcal{V} \leq 210$

6.1. Analysis of the auxiliary face $\spadesuit$

Suppose that $\phi(\spadesuit) \neq 0$ and let $v \in \mathcal{V}$ with $\phi(v, \spadesuit) \neq 0$. We notice that v cannot be a regular vertex or a $\diamondsuit$-vertex. Moreover, v cannot be a big vertex by Theorem 5.1.

Proposition 6.1. *If $\phi(v, \spadesuit) \neq 0$, then $c(\spadesuit) \geq c_v \phi(v, \spadesuit) > 0.0006$.*

Proof. If v is semi-regular, then $\phi(v, \spadesuit) = \frac{1}{2}$ and $f(v) = (3, 5, a)$ for $a = 11$ or $20 \leq a \leq 39$. Therefore $c_v \phi(v, \spadesuit) \geq \frac{1}{2} \cdot c(3, 5, 39) > 0.024$. If v is a TS-vertex then $c_v \phi(v, \spadesuit) \geq \frac{1}{3} \cdot c(3, 4, 4, 4) > 0.024$. If v is a $\spadesuit$-vertex, then $\phi(v, \spadesuit) = 1$. Since c_v is smaller when the entries of $f(v)$ are larger, it suffices to check the cases $f(v) = (3, 8, 19)$, $(3, 10, 13)$, $(4, 4, 41)$, $(4, 6, 10)$, $(5, 5, 9)$, $(5, 6, 6)$, $(3, 3, 3, 19)$, $(3, 3, 4, 10)$ and $(3, 3, 5, 6)$. We obtain $c_v \geq c(3, 10, 13) > 0.0006$. Similarly, if v is potentially-special then $\phi(v, \spadesuit) \geq \frac{1}{3}$ and $c_v \geq 0.023$. □

For the arguments of section 7.2 we will need the above analysis performed with more accuracy under the additional assumption $5 \in f(v)$.

Lemma 6.2. *If there is $\phi(v, \spadesuit) \neq 0$ with $5 \in f(v)$ and $f(v) \neq (5, 5, 9)$, then $c(\spadesuit) > 0.015$.*

Proof. If v is semi-regular then $c_v \phi(v, \spadesuit) > 0.024$. If v is a $\spadesuit$-vertex and $f(v) \neq (5, 5, 9)$ then $\phi(v, \spadesuit) = 1$ and the case-analysis gives $c_v \geq c(5, 5, 8) > 0.015$. Finally, if v is potentially-special, then $f(v) \in \{(3, 3, 3, 5), (3, 3, 3, 3, 5)\}$ and $\phi(v, \spadesuit) = 1$, or $f(v) \in \{(3, 4, 5), (4, 5, 6), (3, 3, 4, 5)\}$ and $\phi(v, \spadesuit) = \frac{1}{2}$, hence $c_v \phi(v, \spadesuit) > 0.023$. □

6.2. Analysis of the auxiliary face $\diamondsuit$*, and proving* $\#\mathcal{V} \leq 210$

By definition of ϕ we have $\phi(v, \diamondsuit) \neq 0$ if and only if $v \in \mathcal{Z}$, where

$$\mathcal{Z} = \{v \in \mathcal{V} : \ f(v) \in \{(5, 6, 7), (3, 3, 5, 7)\}\}.$$

We let $Z = \#\mathcal{Z}$ and we notice that, for every $v \in \mathcal{Z}$, we have $K(v) = \frac{2}{210}$, $\phi(v, \diamondsuit) = 1$ and $c_v = \left(\frac{2}{210} - \frac{2}{209}\right) = -\frac{2}{210 \cdot 209}$.

Proposition 6.3. *If G is a PCC graph, then $Z \leq 210$ and $c(\diamondsuit) > -0.0096$.*

Proof. We already observed in equation (2.1) that $\sum_{v \in \mathcal{V}} K(v) = 2$. Therefore

$$Z \cdot \frac{2}{210} = \sum_{v \in \mathcal{Z}} K(v) \leq 2$$

which implies that $Z \leq 210$. Thus we also have

$$c(\diamondsuit) = -\frac{2Z}{210 \cdot 209} \geq -\frac{2}{209} > -0.0096. \tag{6.1}$$

□

We remark that the above proposition, together with Propositions 4.1, 4.2, 4.7, 4.8, 4.11, 4.12 and 4.19, concludes the proof of Proposition 2.4.
We also deduce the following important partial result towards our main theorem:

Corollary 6.4. *If G is a PCC graph, then $\#\mathcal{V} \leq 210$. Moreover, we have equality if and only if $Z = 210$.*

Proof. By Proposition 2.4, Proposition 6.3 and formula (2.3), we get

$$\frac{2(209 - \#\mathcal{V})}{209} \geq -\frac{2Z}{210 \cdot 209} \geq \frac{-2}{209},$$

from which the corollary follows. □

7. Conclusion via double-counting and ♡-triangles

7.1. Proving the upper bound $\#\mathcal{V} \leq 209$

The above Corollary 6.4 forces a rigid combinatorial description in case $\#\mathcal{V} = 210$. The following lemma, easy to prove, summarizes such description.

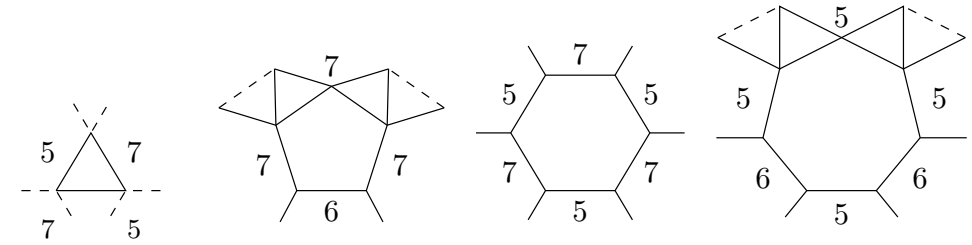


Figure 7.1: Illustrations for Lemma 7.1.

Lemma 7.1 ([21, Lemma 16.1])**.** *Suppose that $\#\mathcal{V} = 210$ and let $\sigma \in \mathcal{F}$. Then $|\sigma| \in \{3, 5, 6, 7\}$ and the faces around σ (i.e. meeting σ at the boundary) are arranged as depicted in fig. 7.1.*

However, we show that this situation is impossible via a double-counting argument.

Corollary 7.2. *There is no PCC graph with exactly 210 vertices.*

Proof. Let $A = \#\mathcal{A}$, $B = \#\mathcal{B}$, $C = \#\mathcal{C}$, where

$$\mathcal{A} = \{\sigma \in \mathcal{F} : |\sigma| = 5\},$$
$$\mathcal{B} = \{\sigma \in \mathcal{F} : |\sigma| = 6\},$$
$$\mathcal{C} = \{\sigma \in \mathcal{F} : |\sigma| = 7\}.$$

From Lemma 7.1 we derive the following system of equalities, by double-counting the edges $e \in \mathcal{E}$ with $s(e) = (5,6), (5,7)$, or $(6,7)$:

$$\begin{cases} A = 3B \\ 2A = 3C \\ 3B = 2C \end{cases}$$

which is evidently inconsistent. □

7.2. Graph surgery along chains of ♡-triangles and conclusion

In the following lemma we use the data that we have acquired so far to considerably reduce the combinatorial complexity surrounding pentagons in an hypothetical PCC graph with 209 vertices.

Lemma 7.3. *Suppose that $\#\mathcal{V} = 209$, let $\sigma \in \mathcal{F}$ with $|\sigma| = 5$ and let*

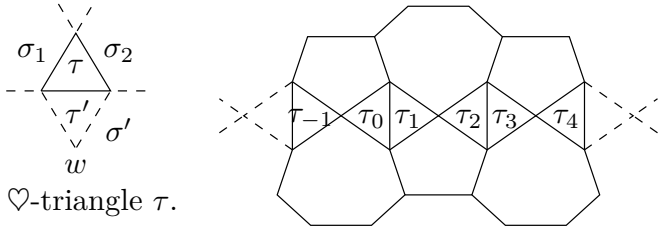
$$\mathcal{S} = \{(4,4,5), (3,4,4,5)\} \cup \{(4,5,a) : \ 14 \leq a \leq 19\}.$$

Then either all $v \in V^{(f)}(\sigma)$ satisfy $f(v) \in \mathcal{S}$, or all $v \in V^{(f)}(\sigma)$ are ♢-vertices.

Proof. Since $\#\mathcal{V} = 209$, we notice that $c(G) = 0$ by (2.3). If $\phi(\sigma) \neq 0$ then by the arguments of section 4.1 we see that either $c(\sigma) > 0.015$ or the *Case* $A = 4$ (of section 4.1) applies to σ. The first case is impossible since it gives $c(G) > 0.005$ by Proposition 2.4, while in the second case we get $f(v) \in \mathcal{S}$ for all $v \in V^{(f)}(\sigma)$. Now assume that $\phi(\sigma) = 0$ and let $w \in V^{(f)}(\sigma)$. If w is a ♠-vertex then $f(w) = (5,5,9)$, otherwise $c(G) > 0.005$ by Lemma 6.2 and Proposition 2.4. If w is not a ♠-vertex or a ♢-vertex, then it is special to some σ' with $|\sigma'| \in \{11, 40, 41\}$. However, we cannot have $|\sigma'| \in \{40, 41\}$, otherwise $c(\sigma') > 0.011$ by Proposition 4.19 and so $c(G) > 0.001$. Since $\phi(w, \sigma) = 0$, we only have three possibilities: $f(v) \in \{(3,3,3,5), (3,4,4,5), (3,3,3,3,5)\}$. The case $f(w) = (3,3,3,5)$ is impossible because it implies $c_+(\sigma') \geq c_w \phi_2(w, \sigma') > 0.190$ and so, since $c_-(\sigma') > -0.022$ (see the proof of Lemma 4.3), we get $c(\sigma') > 0.168$. This gives $c(G) > 0.158$ by Proposition 2.4. Also the case $f(w) = (3,3,3,3,5)$ is impossible, because then by $(R_{(3,3,3,3,5)})$ w is consecutive on σ to $w_1 \in V^{(f)}(\sigma)$ with $(3,4,5) \subseteq f(w_1)$. But then using $(R_{(3,4,4,5)})$ we get that w_1 is not special and so $\phi(\sigma) \neq 0$. Therefore $f(w) = (3,4,4,5)$. Then by $(R_{(3,4,4,5)})$ w is consecutive on σ to $w_1, w_2 \in V^{(f)}(\sigma)$ with $4 \in f(w_1), f(w_2)$. Since $\phi(\sigma) = 0$ we necessarily have that w_1 and w_2 are special vertices with $f(w_1) = f(w_2) = (3,4,4,5)$. By repeating the argument we deduce that all $v \in V^{(f)}(\sigma)$ satisfy $f(v) = (3,4,4,5)$. Summing up, we conclude that either for all $v \in V^{(f)}(\sigma)$ we have $f(v) \in \mathcal{S}$, or all $v \in V^{(f)}(\sigma)$ satisfy $f(v) \in \{(5,6,7), (3,3,5,7), (5,5,9)\}$. However, if $V^{(f)}(\sigma) = \langle v_1, \ldots, v_5 \rangle$ and $f(v_1) = (5,5,9)$, we necessarily get $f(v_i) = (5,5,9)$ for all $1 \leq i \leq 5$, but this is impossible since $\#V^{(f)}(\sigma)$ is odd. □

We recall from Definition 1.4 that a triangle $\tau \in \mathcal{F}$ is a ♡-triangle if all its vertices are ♢-vertices. It's easy to see, as in Lemma 7.1, that if τ is a ♡-triangle then there exist uniquely $v_\tau \in V^{(f)}(\tau)$ and $e_\tau \in E^{(f)}(\tau)$ with $f(v_\tau) = \langle 3,5,3,7 \rangle$ and $s(e_\tau) = (3,3)$. Conversely, using Lemma 7.3 we can easily prove the following.

Lemma 7.4 ([21, Lemma 17.2])**.** *Suppose that* $\#\mathcal{V} = 209$*, let* $\tau \in \mathcal{F}$ *with* $|\tau| = 3$ *and let* $v \in V^{(f)}(\sigma)$ *with* $f(v) = \langle 3, 5, 3, 7\rangle$*. Then* τ *is a* $\heartsuit$*-triangle. Moreover, also* $\mathrm{opp}^{(f)}(v, \tau)$ *is a* $\heartsuit$*-triangle.*



A $\heartsuit$-triangle τ.

Figure 7.2: An illustration for Lemma 17.2 and (a portion of) a chain of $\heartsuit$-triangles.

A remarkable consequence of Lemma 7.4 is that, given a $\heartsuit$-triangle $\tau \in \mathcal{F}$, there is an unique sequence of $\heartsuit$-triangles

$$\ldots, \tau_{-2}, \tau_{-1}, \tau = \tau_0, \tau_1, \tau_2, \ldots$$

such that for every $n \in \mathbb{Z}$ the triangles τ_{2n} and τ_{2n-1} meet at the vertex $v_{\tau_{2n-1}} = v_{\tau_{2n}}$ with $f(v_{\tau_{2n}}) = \langle 5, 3, 7, 3\rangle$, while τ_{2n} and τ_{2n+1} share the common edge $e_{\tau_{2n}} = e_{\tau_{2n+1}}$. Since the set $\mathcal{F}$ is finite, every such sequence of $\heartsuit$-triangles must become periodic. In other words, $\exists L \geq 1$ such that $\tau_L = \tau_0$. If L is the least positive integer with this property, we say that $\{\tau_k\}_{k\in\mathbb{Z}}$ is a *chain of* $\heartsuit$*-triangles of length* L. It is clear that the length of a chain of $\heartsuit$-triangles must be even, but we can be more precise if we look at the alternating sequence of 5-sided and 7-sided faces neighboring the chain.

Lemma 7.5 ([21, Lemma 17.3])**.** *The length of a chain of* $\heartsuit$*-triangles is a multiple of 4. A chain of* $\heartsuit$*-triangles of length* $4m$ *consists of exactly* $4m$ *triangles and involves exactly* $10m$ *pairwise distinct edges and* $6m$ *distinct vertices, organized in a 2-cell complex embedded in the sphere, homeomorphic to the union of* $2m$ *disks glued together at* $2m$ *boundary points in a circular structure.*

Using this notion of chain of $\heartsuit$-triangles we can perform a "surgery trick" to prove the following.

Lemma 7.6. *Suppose that there exists a PCC graph with 209 vertices. Then there also exists a PCC graph with at least 210 vertices.*

Proof. Let G be a PCC graph with 209 vertices. From Lemma 2.3 and Proposition 2.4 we see that the set of its $\diamondsuit$-vertices is non-empty. Moreover, from Lemma 7.3 we deduce that there is a face σ with $|\sigma| = 5$ for which all $v \in V^{(f)}(\sigma)$ are $\diamondsuit$-vertices. Since the number of edges of σ is odd, there must exist two edges $e_1, e_2 \in E^{(f)}(\sigma)$ consecutive on σ with $s(e_1), s(e_2) \neq (5, 7)$. Then the common endpoint of e_1, e_2 is a vertex v with $f(v) = \langle 3, 5, 3, 7\rangle$. By Lemma 7.4 this implies the existence of a $\heartsuit$-triangle, and the previous discussion shows the existence of a chain of $\heartsuit$-triangles. By Lemma 7.5 this chain contains $6m$ vertices for some $m \in \mathbb{N}$.

Let C be the support of this chain. By the Jordan-Schönflies theorem the complement of C consists of two connected open sets U_1, U_2 homeomorphic to the unit ball. For $i = 1, 2$ let n_i be the number of vertices of G contained in U_i, and suppose without loss of generality that $n_1 \geq n_2$. We have that $n_1 + n_2 + 6m = 209$ is odd, thus $n_1 \geq n_2 + 1$.
The embedding of G in the sphere induces a 2-cell complex structure on both $\overline{U_1}, \overline{U_2}$. We notice that their boundary structure (including the face vectors of all the vertices) is isomorphic. Therefore, we can perform a *graph surgery* and replace the 2-cell complex structure on $\overline{U_2}$ with an homeomorphic copy of the 2-cell complex structure on $\overline{U_1}$, without altering the geometric and combinatorial data in a neighborhood of C. This construction gives a new PCC graph G' with $n_1 + n_1 + 6m \geq 210$ vertices. □

However, we already proved that no PCC graph can exist with at least 210 vertices. Therefore by Lemma 7.6, Corollary 6.4, and Corollary 7.2 we deduce Theorem 2.1.

8. PCC graphs with 208 vertices and faces with given size

The first examples of PCC graphs with 208 vertices were found by Nicholson and Sneddon in [9]. These graphs are cleverly built using faces with 3,4,11 and 13 edges, arranged as in fig. 4.5. See also [17] for discussions on these graphs. Another example of a PCC graph with 208 vertices, containing faces with 3,5,7 and 39 edges, is displayed in fig. 8.1. This graph was discovered in 2011 by the author (private communication with Prof. Jamie Sneddon) and later independently re-discovered by Oldridge [10].
The PCC graph in fig. 8.1 is constructed modularly around a closed chain of ♡-triangles, by repeating 26 times a "♡-motif" that consists of 5 triangles, a 5-sided face and a 7-sided face. Oldridge observed that by allowing only $2N$ repetitions in this construction, we could exhibit PCC graphs containing a pair of faces with size $|\sigma| = 3N$, for each $1 \leq N \leq 13$, thus disproving a previous conjecture from [8, pag. 29] about the impossibility of faces with size $|\sigma| \geq 23$. In [10, Sec. 6.3] it is proposed the open problem of exhibiting a PCC graph containing a face with size $|\sigma| \geq 23$ not divisible by three. We now provide a simple solution to this problem. It suffices to repeat the ♡-motif as before, but around an *open* chain of ♡-triangles, as in fig. 8.2. In this way each ♡-motif contributes three edges to the "outer" face. Moreover, in order to end up with a PCC graph, it is necessary to add two "closing caps" at the extremities of the chain, in such a way that only admissible vertices are produced. This can be done without difficulty: in fig. 8.2 we use 4 triangles and two pentagons, so each cap contributes five edges to the outer face. By using one less triangle it is possible to construct a cap that contributes only four edges.
With this construction we are able to produce, for any $8 \leq N \leq 41$, a PCC graph G_N that contains a face σ with $|\sigma| = N$. Since PCC-graphs with a face with size $|\sigma| \in \{3, 4, 5, 6, 7\}$ are easy to find, we conclude that all sizes $3 \leq N \leq 41$ are admissible for faces in a PCC graph.

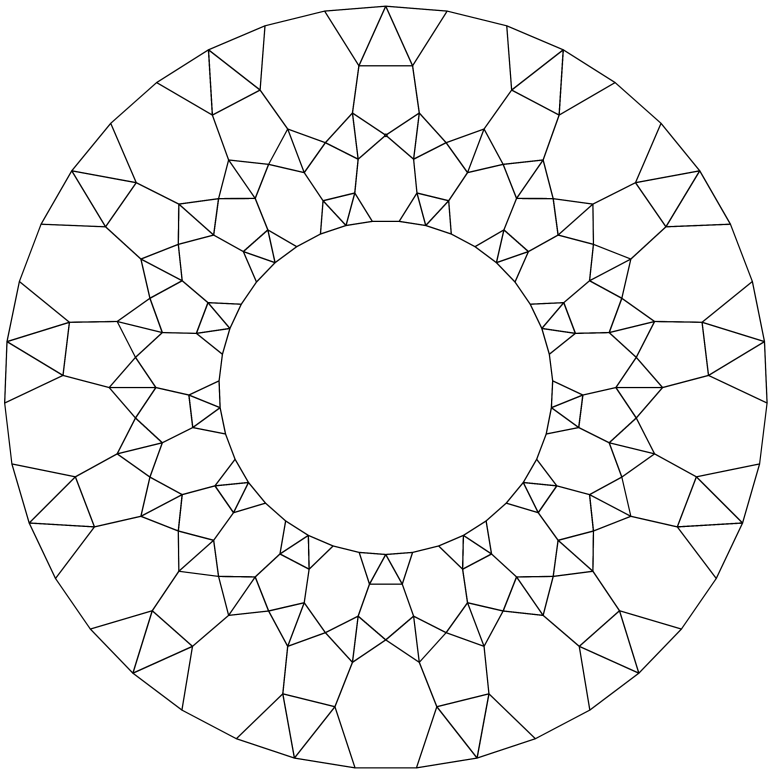


Figure 8.1: A PCC graph with 208 vertices and 3,5,7,39-sided faces.

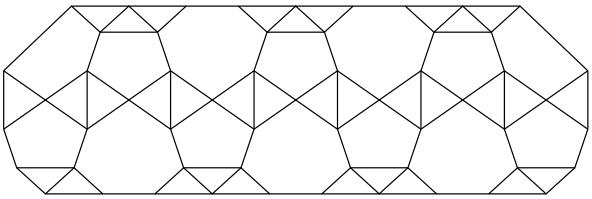


Figure 8.2: An example of a PCC graph containing a face σ with $|\sigma| = 25$.

Acknowledgements

This work was supported in part by the full scholarship (Corso Ordinario) granted by the Scuola Normale Superiore (Pisa, Italy), and in part by the full International scholarship awarded by the Faculty of Graduate and Postdoctoral Studies (Ottawa, Canada). The author is grateful to Bobo Hua for noticing that our main result on the size of planar PCC graphs resolves also the corresponding problem for projective PCC graphs.

References

[1] L. Najman, P. Romon (Eds.), Modern Approaches to Discrete Curvature, volume 2184 of *Lecture Notes in Mathematics*, Springer International Publishing, Cham, 2017.

[2] Y. Higuchi, Combinatorial curvature for planar graphs, Journal of Graph Theory 38 (2001) 220–229.

[3] B. Hua, Y. Lin, Curvature notions on graphs, Frontiers of Mathematics in China 11 (2016) 1275–1290.

[4] S. Kamtue, Combinatorial, Bakry-Émery, Ollivier's Ricci curvature notions and their motivation from Riemannian geometry (2018). arXiv:1803.08898.

[5] B. Chen, G. Chen, Gauss-Bonnet formula, finiteness condition, and characterizations of graphs embedded in surfaces, Graphs and Combinatorics 24 (2008) 159–183.

[6] M. DeVos, B. Mohar, An analogue of the Descartes-Euler formula for infinite graphs and Higuchi's conjecture, Transactions of the American Mathematical Society 369 (2007) 3287–3300.

[7] B. Chen, The Gauss-Bonnet formula of polytopal manifolds and the characterization of embedded graphs with nonnegative curvature, Proceedings of the American Mathematical Society 137 (2009) 1601–1611.

[8] T. Réti, E. Bitay, Z. Kosztolányi, On the polyhedral graphs with positive combinatorial curvature, Acta Polytechnica Hungarica 2 (2005) 19–37.

[9] R. Nicholson, J. Sneddon, New Graphs with thinly spread positive combinatorial curvature, New Zealand Journal of Mathematics 41 (2011) 39–43.

[10] P. R. Oldridge, Characterizing the polyhedral graphs with positive combinatorial curvature, Ph.D. thesis, University of Victoria, 2017.

[11] L. Zhang, A result on combinatorial curvature for embedded graphs on a surface, Discrete Mathematics 308 (2008) 6588–6595.

[12] B.-G. Oh, On the number of vertices of positively curved planar graphs, Discrete Mathematics 340 (2017) 1300–1310.

[13] P. W. Fowler, S. Nikolić, R. De Los Reyes, W. Myrvold, Distributed curvature and stability of fullerenes, Physical Chemistry Chemical Physics 17 (2015) 23257–23264.

[14] B. Hua, Y. Su, The first gap for total curvatures of planar graphs with nonnegative curvature (2017). arXiv:1709.05309.

[15] B. Hua, Y. Su, The set of vertices with positive curvature in a planar graph with nonnegative curvature, Advances in Mathematics 343 (2019) 789–820.

[16] B. Hua, Y. Su, Total curvature of planar graphs with nonnegative combinatorial curvature (2017). arXiv:1703.04119.

[17] M. L. Childs, Topological graph theory and graphs of positive combinatorial curvature, 2016.

[18] Y. Akama, B. Hua, Hyperbolic polyhedral surfaces with regular faces (2018). arXiv:1807.10762.

[19] Y. Akama, B. Hua, Y. Su, Areas of spherical polyhedral surfaces with regular faces (2018). arXiv:1804.11033.

[20] P. Wernicke, Über den kartographischen Vierfarbensatz, Mathematische Annalen 58 (1904) 413–426.

[21] L. Ghidelli, On the largest planar graphs with everywhere positive combinatorial curvature (Extended arxiv version) (2017). arXiv:1708.08502.

[22] O. Borodin, Colorings of plane graphs: A survey, Discrete Mathematics 313 (2013) 517–539.

[23] D. W. Cranston, D. B. West, An introduction to the discharging method via graph coloring, Discrete Mathematics 340 (2017) 766–793.

[24] R. Radoičić, G. Tóth, The discharging method in combinatorial geometry and the Pach-Sharir conjecture, in: J. Pach, R. Pollack (Eds.), Surveys on Discrete and Computational Geometry: Twenty Years Later, volume 453 of *Contemporary Mathematics*, American Mathematical Society, 2008, pp. 319–342.

Bibliography

[1] M. Aigner and G. M. Ziegler. *Proofs from the Book (6th ed.)*. Springer, 2018.

[2] A. Balog and T. D. Wooley. Sums of two squares in short intervals. *Canadian Journal of Mathematics*, 52(04):673–694, 2000.

[3] R. Bambah and S. Chowla. On numbers which can be expressed as a sum of two squares. In *Proc. Nat. Inst. Sci. India*, volume 13, pages 101–103, 1947.

[4] E. Bank, L. Bary-Soroker, and A. Fehm. Sums of two squares in short intervals in polynomial rings over finite fields. *American Journal of Mathematics*, 140(4):1113–1131, 2018.

[5] J. M. Basilla. On the solution of $x^2 + dy^2 = m$. *Proceedings of the Japan Academy, Series A, Mathematical Sciences*, 80(5):40–41, 2004.

[6] P. T. Bateman and H. G. Diamond. *Analytic number theory: an introductory course*, volume 1. World Scientific, Co. Pte. Ltd., Singapore, 2004.

[7] P. T. Bateman and M. E. Low. Prime numbers in arithmetic progressions with difference 24. *The American Mathematical Monthly*, 72(2):139–143, 1965.

[8] B. C. Berndt. *Ramanujan's notebooks, Part IV*. Springer Science & Business Media, 1994.

[9] B. C. Berndt, S. Kim, and A. Zaharescu. The Circle problem of Gauss and the divisor problem of Dirichlet — still unsolved. *The American Mathematical Monthly*, 125(2):99–114, 2018.

[10] B. C. Berndt and R. A. Rankin. *Ramanujan: Letters and commentary*, volume 9. American Mathematical Soc., 1995.

[11] B. C. Berndt, K. S. Williams, and R. J. Evans. *Gauss and Jacobi sums*. New York: Wiley, 1998.

[12] J. Binet. Note sur une question relative a la théorie des nombres. *Comptes Rendus Acad. Sci. Paris*, 12:248–250, 1841. Reprinted, Sphinx-Oedipe, 4, 1909, 29-30.

[13] E. Bombieri. Counting points on curves over finite fields. In *Séminaire Bourbaki vol. 1972/73, Exposé 430*, pages 234–241. Springer, 1974.

[14] E. Bombieri and H. Iwaniec. On the order of $\zeta(\frac{1}{2} + it)$. *Annali della Scuola Normale Superiore di Pisa - Classe di Scienze*, 13(3):449–472, 1986.

[15] J. Bourgain and N. Watt. Mean square of zeta function, circle problem and divisor problem revisited. *arXiv preprint arXiv:1709.04340*, 2017.

[16] R. Bradshaw. Arithmetic properties of values of lacunary series. Master's thesis, University of Ottawa, 2013.

[17] Brahmagupta. *Brāhmasphuṭasiddhānta*. Edited by Sudhākara Dvivedin (Benares, 1902) and by Ram Swarup Sharma (New Delhi, 1966), c. 628.

[18] J. Brillhart. Note on representing a prime as a sum of two squares. *Mathematics of Computation*, 26(120):1011–1013, 1972.

[19] L. Brünjes. *Forms of Fermat equations and their zeta functions*. World Scientific, 2004.

[20] R. D. Carmichael. *Diophantine analysis*. Number 16 in Incorporated. New York: John Wiley & sons, 1915.

[21] A. Choudhry. On equal sums of cubes. *The Rocky Mountain journal of mathematics*, pages 1251–1257, 1998.

[22] A. D. Christopher. A partition-theoretic proof of fermat's two squares theorem. *Discrete Mathematics*, 339(4):1410–1411, 2016.

[23] F. R. Chung. Several generalizations of Weil sums. *Journal of Number Theory*, 49(1):95–106, 1994.

[24] F. Clarke, W. Everitt, L. Littlejohn, and S. Vorster. H.J.S. Smith and the Fermat two squares theorem. *The American Mathematical Monthly*, 106(7):652–665, 1999.

[25] H. Cohen. *A course in computational algebraic number theory*, volume 138. Springer Science & Business Media, 2013.

[26] H. T. Colebrooke. *Algebra with arithmetic and mensuration: From the Sanscrit of Brahmegupta and Bháscara*. London, John Murray, 1817.

[27] G. Cornacchia. Su di un metodo per la risoluzione in numeri interi dell'equazione $\sum_{h=0}^{n} C_h x^{n-h} y^h = P$. *Giornale di Matematiche di Battaglini*, 46:33–90, 1908.

[28] H. Davenport. *Multiplicative number theory*, volume 74 of *Graduate Texts in Mathematics*. Springer-Verlag New York, 1980.

[29] J.-M. De Koninck and F. Luca. *Analytic number theory: Exploring the anatomy of integers*, volume 134 of *Graduate Studies in Mathematics*. American Mathematical Society, 2012.

[30] C. de la Vallée-Poussin. Recherches analytiques de la théorie des nombres premiers. *Annales de la Société scientifique de Bruxelles*, 20:183–256, 1896.

[31] A.-M. Décaillot. Géométrie des tissus. Mosaïques. Échiquiers. Mathématiques curieuses et utiles. *Revue d'histoire des mathématiques*, 8(2):145–206, 2002.

[32] R. Dedekind. *Sur la théorie des nombres entiers algébriques*. Bulletin des sciences mathématiques, 1877. Translated into English by John Stillwell, Cambridge University Press, 1996.

[33] R. Dedekind. *Über die Theorie der ganzen algebraischen Zahlen*. Supplement XI to Dirichlet, L.: Vorlesungen über Zahlentheorie, 1894.

[34] R. Descartes. Progymnasmata de solidorum elementis. In *Oeuvres de Descartes*, volume X, pages 265–276, 2008.

[35] J.-M. Deshouillers, F. Hennecart, and B. Landreau. Sums of powers: an arithmetic refinement to the probabilistic model of Erdős and Rényi. *Acta Arithmetica*, 85(1):13–33, 1998.

[36] J.-M. Deshouillers, F. Hennecart, and B. Landreau. On the density of sums of three cubes. In *International Algorithmic Number Theory Symposium*, pages 141–155. Springer, Berlin, 2006.

[37] M. DeVos and B. Mohar. An analogue of the Descartes-Euler formula for infinite graphs and Higuchi's conjecture. *Transactions of the American Mathematical Society*, 369(7):3287–3300, 2007.

[38] L. E. Dickson. *History of the theory of numbers: Diophantine Analysis*, volume 2. Courier Corporation, 2013.

[39] R. Dietmann and C. Elsholtz. Longer gaps between values of binary quadratic forms. *arXiv preprint arXiv:1810.03203*, 2018.

[40] A. Diophantus. Arithmetica, c. 200-300.

[41] P. G. L. Dirichlet. Beweis des Satzes, daß jede unbegrenzte arithmetische Progression, deren erstes Glied und Differenz ganze Zahlen ohne gemeinschaftlichen Factor sind, unendlich viele Primzahlen enthält. *Abhandlungen der Königlich Preussischen Akademie der Wissenschaften*, 45:81, 1837.

[42] C. Elsholtz. A combinatorial approach to sums of two squares and related problems. In *Additive Number Theory*, pages 115–140. Springer, 2010.

[43] P. Erdős. On a new method in elementary number theory which leads to an elementary proof of the prime number theorem. *Proceedings of the National Academy of Sciences of the United States of America*, 35(7):374, 1949.

[44] P. Erdős and A. Rényi. Additive properties of random sequences of positive integers. *Acta Arithmetica*, 6(1):83–110, 1960.

[45] L. Euler. Variae observationes circa series infinitas. *Commentarii Academiae Scientiarum Petropolitanae*, 9:160–188, 1737.

[46] L. Euler. Letter CXXV, 1749. Letter to Goldbach, Berlin, 12th April.

[47] L. Euler. De numerus qui sunt aggregata duorum quadratorum. *Novi commentarii academiae scientiarum Petropolitanae (1752/3)*, pages 3–40, 1758. Included in Opera Omnia: Series 1, Volume 2, pp. 295–327 Reprinted in Commentationes Arithmeticae 1, 1849, pp. 155-173 [E228b].

[48] L. Euler. Demonstratio theorematis Fermatiani omnem numerum primum formae $4n+1$ esse summam duorum quadratorum. *Novi commentarii academiae scientiarum Petropolitanae (1754/5)*, pages 3–13, 1760. Included in Opera Omnia: Series 1, Volume 2, pp. 328–337. Reprinted in Commentationes Arithmeticae 1, 1849, pp. 210-215 [E241a].

[49] L. Euler. Solutio generalis quorundam problematum Diophanteorum, quae vulgo nonnisi solutiones speciales admittere videntur. *Novi commentarii academiae scientiarum Petropolitanae (1756/7)*, pages 155–184, 1761. Included in Opera Omnia: Series 1, Volume 2, pp. 428–458. Reprinted in Commentationes Arithmeticae 1, 1849, pp. 193–209 [E255].

[50] L. Euler. *Leonhardi Euleri opera omnia*. Typis et in aedibus BG Teubneri, 1915.

[51] J. A. Ewell. A Simple Proof of Fermat's Two-Square Theorem. *The American Mathematical Monthly*, 90(9):635–637, 1983.

[52] L. P. Fibonacci. *Liber Quadratorum*. Ed. Orlando, FL:Academic Press, 1225. The Book of Squares, An annotated translation into modern English, (1928).

[53] B. Frénicle. Letter X, Brouncker to Wallis, Oct. 13, 1657. French transl. in Oeuvres de Fermat, 111, 419-420.

[54] J. B. Friedlander. Sifting short intervals II. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 92, No. 3, pages 381–384. Cambridge University Press, 1982.

[55] J. B. Friedlander and H. Iwaniec. *Opera de cribro*, volume 57. American Mathematical Soc., 2010.

[56] R. Gar-El and L. Vaserstein. On the diophantine equation $a^3 + b^3 + c^3 + d^3 = 0$. *Journal of Number Theory*, 94(2):219–223, 2002.

[57] C. F. Gauss. *Disquisitiones arithmeticae*. Lipsiae, in commissis apud Germ. Fleischer, Jux., 1801. Translated into English by Arthur A. Clarke, Yale University Press, vol. 157, 1966. Revised edition by William C. Waterhouse with the help of Cornelius Greither and A. W. Grootendorst, Springer, 1986.

[58] C. F. Gauss. De nexu inter multitudinem classium, in quas formae binariae secundi gradus distribuntur, earumque determinantem. In E. Schering, editor, *Werke*, volume 2, pages 269–291. Königlichen Gesellschaft der Wissenschaften, 1826.

[59] C. F. Gauss. Letter of Gauss to Enke (1849). *In: Werke, II*, pages 444–447, 1872.

[60] C. F. Gauss. Tafel der Frequenz der Primzahlen (1792-3). *In: Werke, II*, pages 436–442, 1872.

[61] L. Ghidelli. Arbitrarily long gaps between the values of positive-definite cubic and biquadratic forms. *Preprint*, —-:000, 2019.

[62] L. J. Goldstein. A history of the prime number theorem. *The American Mathematical Monthly*, 80(6):599–615, 1973.

[63] D. Goldston, S. Graham, J. Pintz, and C. Yıldırım. Small gaps between products of two primes. *Proceedings of the London Mathematical Society*, 98(3):741–774, 2009.

[64] J. Grace. The four square theorem. *Journal of the London Mathematical Society*, 1(1):3–8, 1927.

[65] A. Granville. On elementary proofs of the Prime Number Theorem for Arithmetic Progressions, without characters. In *Proceedings of the 1993 Amalfi Conference on Analytic Number Theory*, pages 157–194, 1993.

[66] A. Granville. Unexpected irregularities in the distribution of prime numbers. In *Proceedings of the International Congress of Mathematicians*, pages 388–399. Springer, Birkhäuser, Basel, 1995.

[67] B. J. Green and S. Lindqvist. Monochromatic Solutions to $x+y=z^2$. *Canadian Journal of Mathematics*, pages 1–27, 2018.

[68] J. Hadamard. Sur la distribution des zéros de la fonction $\zeta(s)$ et ses conséquences arithmétiques. *Bulletin de la Societé mathematique de France*, 24:199–220, 1896.

[69] G. Hardy. *Ramanujan: Twelve lectures on subjects suggested by his life and work, 3rd ed.* Chelsea, 1999.

[70] G. H. Hardy. On the expression of a number as the sum of two squares. *Quart. J. Math.*, 46:263–283, 1915.

[71] G. H. Hardy and J. E. Littlewood. Some problems of 'Partitio numerorum'(VI): Further researches in Waring's Problem. *Mathematische Zeitschrift*, 23(1):1–37, 1925.

[72] K. Hardy, J. B. Muskat, and K. S. Williams. A deterministic algorithm for solving $n = fu^2 + gv^2$ in coprime integers u and v. *Mathematics of Computation*, 55(191):327–343, 1990.

[73] G. Harman. Sums of Two Squares in Short Intervals. *Proceedings of the London Mathematical Society*, 3(2):225–241, 1991.

[74] D. Heath-Brown. Fermat's two squares theorem. *Invariant*, 11:3–5, 1984.

[75] C. Hermite. Note au sujet de l'article précédent. *J. Math. Pures Appl*, 13:15, 1848.

[76] C. Hooley. On the representations of a number as the sum of two cubes. *Mathematische Zeitschrift*, 82(3):259–266, 1963.

[77] C. Hooley. On the intervals between numbers that are sums of two squares. *Acta Math.*, 127:279–297, 1971.

[78] C. Hooley. On the intervals between numbers that are sums of two squares. III. *Journal für die reine und angewandte Mathematik*, 267(1):207–218, 1974.

[79] C. Hooley. On some topics connected with Waring's problem. *Journal für die reine und angewandte Mathematik*, 369(1):110–153, 1986.

[80] C. Hooley. On the intervals between numbers that are sums of two squares: IV. *Journal für die reine und angewandte Mathematik*, 452:79–110, 1994.

[81] C. Hooley. On Hypothesis K^* in Waring's problem. *Sieve methods, exponential sums, and their applications in number theory, Cardiff, 1995*, 1997.

[82] M. N. Huxley. On the difference between consecutive primes. *Inventiones mathematicae*, 15(2):164–170, 1971.

[83] M. N. Huxley. The integer points close to a curve II. *Analytic Number Theory: The Halberstam Festschrift 2*, 139:487–516, 1996.

[84] M. N. Huxley. Exponential sums and lattice points III. *Proceedings of the London Mathematical Society*, 87(3):591–609, 2003.

[85] K. Ireland and M. Rosen. *A classical introduction to modern number theory*, volume 84. Springer Science & Business Media, 2013.

[86] A. Ivić, E. Krätzel, M. Kühleitner, and W. G. Nowak. Lattice points in large regions and related arithmetic functions: Recent developments in a very classic topic. In W. Schwarz and J. Steuding, editors, *Proc. ELAZ-Conf. May 24–28, 2004*, pages 89—-128. Stuttgart: Franz Steiner Verlag, 2006.

[87] H. Iwaniec. The half dimensional sieve. *Acta Arithmetica*, 1(29):69–95, 1976.

[88] H. Iwaniec and E. Kowalski. *Analytic number theory*, volume 53. Providence RI: American Mathematical Society, 2004.

[89] H. Iwaniec and C. J. Mozzochi. On the divisor and circle problems. *Journal of Number theory*, 29(1):60–93, 1988.

[90] N. W. Johnson. Convex polyhedra with regular faces. *Canadian Journal of Mathematics*, 18:169–200, 1966.

[91] A. Kalmynin. Intervals between numbers that are sums of two squares. *arXiv preprint arXiv:1706.07380*, 2017.

[92] N. M. Katz. Crystalline cohomology, Dieudonné modules, and Jacobi sums. In *Automorphic forms, representation theory and arithmetic*, pages 165–246. Springer, berlin, Heidelberg, 1981.

[93] M. Kraitchik. *Le problème des reines*, volume 3, chapter XIII, pages 300–356. Stevens Frères, Bruxelles, 1930.

[94] J.-L. Lagrange. Recherches d'arithmétique. *Nouveaux mémoires de l'Académie royale des sciences et belles-lettres de Berlin*, 1775. Première Partie in Nouv. Mém. Acad. Berlin 1773/275; Seconde Partie ibid 1775/351. In Gallica-Math:Œuvres complètes, Lagrange, tome 3.

[95] E. G. H. Landau. Über die Einteilung der positiven ganzen Zahlen in vier Klassen nach der Mindeszahl der zu ihrer additiven Zusammensetzung erforderlichen Quadrate. *Arch. Math. Phys.*, 13:305–312, 1908.

[96] E. G. H. Landau. *Handbuch der Lehre von der Verteilung der Primzahlen, Band II.* Leipzig, Berlin: Teubner, 1909.

[97] E. G. H. Landau. Über die Gitterpunkte in einem Kreise. II. *Nachr. Ges. Wiss. Göttingen*, pages 161–71, 1915.

[98] L. C. Larson. A theorem about primes proved on a chessboard. *Mathematics Magazine*, 50(2):69–74, 1977.

[99] F. Le Lionnais. *Les nombres remarquables*, volume 1407. Hermann, 1983.

[100] A. M. Legendre. *Essai sur la Theorie de Nombres, 1st ed.* Paris, p. 19, 1798.

[101] A. M. Legendre. *Essai sur la Theorie de Nombres, 2nd ed.* Paris, p. 394, 1808.

[102] W. J. LeVeque. *Topics in Number Theory*, volume 2. Addison-Wesley Publishing Company, Inc, Reading, 1961.

[103] M. Livio. *The golden ratio: The story of phi, the world's most astonishing number.* New York: Broadway Books, 2002.

[104] É. Lucas. *Application de l'arithmétique à la construction de l'armure des satins réguliers.* Gustave Retaux, Libraire-Éditeur, 1867.

[105] K. Mahler. Note on hypothesis K of Hardy and Littlewood. *Journal of the London Mathematical Society*, 1(2):136–138, 1936.

[106] Y. I. Manin. *Cubic forms: algebra, geometry, arithmetic*, volume 4. Elsevier, 1986.

[107] K. Matomäki and J. Teräväinen. On the Möbius function in all short intervals. *arXiv preprint arXiv:1911.09076*, 2019.

[108] J. Maynard. Sums of two squares in short intervals. In *Analytic Number Theory*, pages 253–273. Springer, 2015.

[109] F. Mertens. Ein Beitrag zur analytischer Zahlentheorie. *J. Reine Angew*, 78:46–62, 1874.

[110] F. Mertens. Über Dirichlet's Beweis des Satzes, daß jede unbegrenzte arithmetische Progression, deren Differenz zu ihren Gliederen teilerfremd ist, unendlich viele Primzahlen enthält. *Sitzber. Wiener Akad.*, 106:254–286, 1897.

[111] H. L. Montgomery. *Ten lectures on the interface between analytic number theory and harmonic analysis*. Number 84 in Regional conference series in mathematics. American Mathematical Soc., 1994.

[112] F. Morain and J.-L. Nicolas. On Cornacchia's algorithm for solving the diophantine equation $u^2 + dv^2 = m$. *Projet*, 1000:a1, 1990.

[113] P. Moree and J. Cazaran. On a claim of Ramanujan in his first letter to Hardy. *Expositiones Mathematicae*, 17(4):289–311, 1999.

[114] C. J. Moreno. Sur le problème de Kummer. *Enseign. Math*, 20(2):45–51, 1974.

[115] M. R. Murty and V. K. Murty. *Non-vanishing of L-functions and applications*. Modern Birkhäuser Classics. Springer Science & Business Media, 2012.

[116] R. Murty, N. Thain, et al. Primes in certain arithmetic progressions. *Functiones et Approximatio Commentarii Mathematici*, 35:249–259, 2006.

[117] D. J. Newman. Simple analytic proof of the prime number theorem. *The American Mathematical Monthly*, 87(9):693–696, 1980.

[118] R. Nicholson and J. Sneddon. New Graphs with thinly spread positive combinatorial curvature. *New Zealand Journal of Mathematics*, 41:39–43, 2011.

[119] B.-G. Oh. On the number of vertices of positively curved planar graphs. *Discrete Mathematics*, 340(6):1300–1310, 2017.

[120] P. R. Oldridge. *Characterizing the polyhedral graphs with positive combinatorial curvature*. PhD thesis, University of Victoria, 2017.

[121] T. Piezas III. A collection of algebraic identities. `https://sites.google.com/site/tpiezas/Home`, 2010.

[122] V. A. Plaksin. The distribution of numbers representable as a sum of two squares. *Izvestiya Rossiiskoi Akademii Nauk. Seriya Matematicheskaya*, 51(4):860–877, 1987.

[123] V. A. Plaksin. Letter to the editor: correction to the paper "The distribution of numbers representable as a sum of two squares". *Izvestiya Rossiiskoi Akademii Nauk. Seriya Matematicheskaya*, 56(4):908–909, 1992.

[124] Plato. *The republic*. New York: Oxford University Press, 1994.

[125] G. Pólya. Uber die "doppelt-periodischen" Losupgen des n-Damen Problems, Mathematische Unterhal-tungen und Spiele. *Dr. W. Ahrens. Zweiter Band, BG Teubner, Leipzig*, pages 363–374, 1918.

[126] S. Ramanujan. *Notebooks*. Tata Institute of Fundamental Research, Bombay, 1957.

[127] T. Réti, E. Bitay, and Z. Kosztolányi. On the polyhedral graphs with positive combinatorial curvature. *Acta Polytechnica Hungarica*, 2(2):19–37, 2005.

[128] I. Richards. On the gaps between numbers which are sums of two squares. *Advances in Mathematics*, 46(1):1–2, 1982.

[129] R. Richardson. *On the number of integers expressible as the sum of two squares*. Msc thesis, Youngstown State University, 2009.

[130] D. S. Richeson. *Euler's gem: the polyhedron formula and the birth of topology*, volume 6. Springer, 2008.

[131] B. Riemann. Üeber die anzahl der primzahlen unter einer gegebenen grösse. *Monat. der Königl. Preuss. Akad. der Wissen. zu Berlin aus der Jahre 1859*, pages 671–680, 1860. also, *Gesammelte math. Werke und wissensch. Nachlass*, 2. Aufl. 1892, 145—155.

[132] D. Roy. An arithmetic criterion for the values of the exponential function. *Acta Arithmetica*, 97:183–194, 2001.

[133] D. Roy. A small value estimate for $\eth_a \times \eth_m$. *Mathematika*, 59(2):333–363, 2013.

[134] Schwering, K. Vereinfachte Lösungen des Eulerschen Aufgabe $x^3+y^3+z^3+\nu^3 = 0$. *Archiv Math. Phys.*, 2(3):280–284, 1902.

[135] A. Selberg. An elementary proof of Dirichlet's theorem about primes in an arithmetic progression. *Annals of Mathematics*, pages 297–304, 1949.

[136] A. Selberg. An elementary proof of the prime-number theorem. *Ann. Math*, 50(2):305–313, 1949.

[137] J.-P. Serre. *Lectures on $N_X(p)$*, volume 11. Chapman & Hall/CRC Research Notes in Mathematics. AK Peters/CRC Press, Boca Raton, FL, 2016.

[138] J.-A. Serret. Sur un théorème relatif aux nombres entiers. *Journal de mathématiques pures et appliquées*, pages 12–14, 1848.

[139] Seva (http://mathoverflow.net/users/9924/seva). Consecutive non-quadratic residues, 2014. Mathoverflow, `http://mathoverflow.net/q/161279` (version: 2014-03-28).

[140] H. N. Shapiro. On primes in arithmetic progressions (II). *Annals of Mathematics*, pages 231–243, 1950.

[141] P. Shiu. The gaps between sums of two squares. *The Mathematical Gazette*, 97(539):256–262, 2013.

[142] W. Sierpiński. O pewnem zagadnieniu z rachunku funkcyj asymptotycznych. *Prace matematyczno-fizyczne*, 1(17):77–118, 1906.

[143] W. Sierpiński. Sur un problème du calcul des fonctions asymptotiques. In S. Hartman, K. Kuratowski, E. Marczewski, A. Mostowski, Andrezj Schinzel, R. Sikorski, and M. Stark, editors, *Œuvres choisies*, volume 1. Académie Polonaise des Sciences, Institut mathématique, 1974.

[144] J. H. Silverman. Taxicabs and sums of two cubes. *The American mathematical monthly*, 100(4):331–340, 1993.

[145] H. Smith et al. De compositione numerorum primorum formae $4\lambda + 1$ ex duobus quadratis. *Journal für die reine und angewandte Mathematik*, 50:91, 1855.

[146] H. J. S. Smith. *The collected mathematical papers of Henry John Stephen Smith Vol. 2*. Clarendon Press, 1894.

[147] A. Spivak. Крылатые квадраты (winged squares). *In* Суммы квадратов*, lecture notes for the mathematical circle at Moscow State University, 15th lecture*, 2007.

[148] S. A. Stepanov. On the number of points of a hyperelliptic curve over a finite prime field. *Mathematics of the USSR-Izvestiya*, 3(5):1103, 1969.

[149] S. Stevin. *L'arithmétique de Simon Stevin de Bruges*. Bonaventure et Abraham Elzevier, 1625. Annotated by Albert Girard.

[150] P. Tannery and C. Henry, editors. *Œuvres de Fermat*. Gauthier-Villars et fils, 1891.

[151] T. Tao. Mertens' theorems. `https://terrytao.wordpress.com/2013/12/11/mertens-theorems/`.

[152] G. Tenenbaum. *Introduction to analytic and probabilistic number theory*, volume 163 of *Graduate studies in mathematics*. Providence, Rhode Island: American Mathematical Society, 3rd edition, 2015.

[153] J. Thunders. The circle problem. `http://www.math.niu.edu/~jthunder/Courses/2014Fall/680/sec1/circle.pdf`, Fall 2014. Online notes for the course Math 680, Analytic Number Theory, at the Northern Illinois University.

[154] V. Tikhomirov. Three paths to Mt.Fermat-Euler. *Quantum Magazine*, pages 5–7, May/June 1994.

[155] User 14052: t0rajir0u. Three approaches to sums of squares. *ArtOfProblemSolving community blog c1157: Annoying Precision*, Aug. 5, 2008.

[156] J. Uspensky and M. Heaslet. *Elementary number theory, 1st ed.* McGraw-Hill, 1939.

[157] A. van der Poorten. The Hermite-Serret Algorithm and 12^2+33^2. In *Cryptography and Computational Number Theory*, pages 129–136. Springer, 2001.

[158] R. C. Vaughan and T. D. Wooley. Waring's problem: a survey. *Number theory for the millennium*, 3:301–340, 2002.

[159] G. Voronoï. Sur une fonction transcendante et ses applications à la sommation de quelques séries. In *Annales scientifiques de l'École Normale Supérieure*, volume 21, pages 207–267, 1904.

[160] A. Weil. On the Riemann hypothesis in function-fields. *Proceedings of the National Academy of Sciences of the United States of America*, 27(7):345, 1941.

[161] A. Weil. On some exponential sums. *Proceedings of the National Academy of Sciences of the United States of America*, 34(5):204, 1948.

[162] A. Weil. Jacobi sums as "Grossencharaktere". *Transactions of the American Mathematical Society*, 73(3):487–495, 1952.

[163] D. G. Wells. *The Penguin Dictionary of Curious and Interesting Numbers*. London: Penguin, 1986.

[164] A. S. Werebrusov. Réponse 2179. Tables de solutions d'équations cubiques. *L'intermédiaire des Mathématiciens*, 9:164–165, 1902. & 11:96—97, 1904. & 11:289, 1904.

[165] K. S. Williams. Merten's theorem for arithmetic progressions. *Journal of Number Theory*, 6(5):353–359, 1974.

[166] T. Wooley. Sums of three cubes. *Mathematica*, 47:53–61, 2000.

[167] D. Zagier. A one-sentence proof that every prime $p \equiv 1 \pmod 4$ is a sum of two squares. *Amer. Math. Monthly*, 97(2):144, 1990.

[168] D. Zagier. Newman's short proof of the prime number theorem. *The American mathematical monthly*, 104(8):705–708, 1997.

[169] V. A. Zalgaller. Convex polyhedra with regular faces. In *Seminars in Mathematics, Steklov Math. Inst., Leningrad*, 1969.