Discrete Structures, Algorithms, and Applications

Sascha Kurz

Juli 2008

Habilitationsschrift

Discrete Structures, Algorithms, and Applications

Zur Erlangung der Lehrbefugnis im Fachgebiet Mathematik

der Fakultät Mathematik, Physik und Informatik
der Universität Bayreuth

vorgelegt von
Sascha Kurz

Bayreuth, Juli 2008

Contents

Chapter 1

Introduction

Continuous mathematics is a very powerful theory. Consider for instance the (rather easy) problem of maximizing the function $f_1 : [0,10] \to \mathbb{R}$, $x \mapsto 10x - x^2$. Since f_1 is continuous and differentiable we know a simple calculation pattern to determine a value $\tilde{x}$ that maximizes f_1 from high school: choose $\tilde{x}$ from $S = \{x \in [0,10] \mid f'(x) = 0\} \cup \{0,10\}$. In our case we have $S = \{0,5,10\}$ and $f(0) = 0$, $f(5) = 25$, $f(10) = 0$. Thus $\tilde{x} = 5$ maximizes the function f_1.

The beauty of this is: We have started with an infinite, even uncountable set $[0,10]$ and ask for an element $\tilde{x} \in [0,10]$ such that $f_1(\tilde{x}) \geqslant f_1(x)$ for all $x \in [0,10]$. And all we have to do is perform a simple algebraic calculation and merely check three elements of the infinite set $[0,10]$ to locate $\tilde{x}$.

Unfortunately in many applications the sets, which are searched for optimal values, are not continuous but discrete. This means that the objects can assume only distinct, separated values. By slightly changing our problem from the beginning we obtain a discrete problem. So let us consider the function $\hat{f}_1 : \{z \in \mathbb{Z} \mid 0 \leqslant z \leqslant 10\} \to \mathbb{Z}$, $x \mapsto 10x - x^2$. Now we have to localize an optimal solution in the discrete set $\{z \in \mathbb{Z} \mid 0 \leqslant z \leqslant 10\} = \{0,1,2,3,4,5,6,7,8,9,10\}$. Since this is a finite set in principle we are able to simply compare all $\hat{f}_1$-values

x	0	1	2	3	4	5	6	7	8	9	10
$\hat{f}_1(x)$	0	9	16	21	24	25	24	21	16	9	0

and determine $\tilde{x} = 5$ to be the optimal value. This can even be done in elementary school (if we use an appropriate formulation and notation of the problem). So, conceptionally, this approach is quite easy.

The big disadvantage of enumerating all possible values for x is that there can be a huge number of values. The number of 11 possible values for x in our example is not that big, but if we consider $\hat{f}_2 : \left\{z \in \mathbb{Z} \mid -10^6 \leqslant z \leqslant 10^6\right\} \to \mathbb{Z}$, $x \mapsto 10x - x^2$ the situation is changed drastically. Checking all 2000001 possible values for x manually would take a long time and is rather tedious.

In order to apply methods from continuous mathematics we may relax our problem and consider the function $f_2 : \left[-10^6, 10^6\right] \to \mathbb{R}$, $x \mapsto 10x - x^2$ instead. Let $\tilde{x}$ be an optimal solution maximizing f_2. Since $\left\{z \in \mathbb{Z} \mid -10^6 \leqslant z \leqslant 10^6\right\} \subset \left[-10^6, 10^6\right]$ we have $f_2(\tilde{x}) \geqslant f_2(x)$ for all $x \in \left\{z \in \mathbb{Z} \mid -10^6 \leqslant z \leqslant 10^6\right\}$. So in general we obtain an upper bound by relaxing a maximization problem. In our case we have $\tilde{x} = 5 \in \mathbb{Z}$ so that the optimal solution for f_2 is also optimal for the discrete function $\hat{f}_2$.

That this coincidence does not occur in general can be seen by considering the discrete function $\hat{f}_3 : \{z \in \mathbb{Z} \mid 0 \leqslant z \leqslant 10\} \to \mathbb{Z}$, $x \mapsto 9x - x^2$ and its relaxation $f_3 : [0,10] \to \mathbb{R}$, $x \mapsto 9x - x^2$. The optimal solution for the continuous problem is given by $\tilde{x} = 4.5 \notin \mathbb{Z}$. Clearly there is yet another clever workaround that enables us to apply continuous optimization in this case. We may consider the two subproblems of maximizing $f_{3,1} : [0,4] \to \mathbb{R}$, $x \mapsto 9x - x^2$ and $f_{3,2} : [5,10] \to \mathbb{R}$, $x \mapsto 9x - x^2$. The optimal solution of the first subproblem is given by $\tilde{x}_1 = 4 \in \mathbb{Z}$ with $f_{3,1}(\tilde{x}_1) = 20$ and the optimal solution of the second subproblem is given by $\tilde{x}_2 = 5 \in \mathbb{Z}$ with $f_{3,2}(\tilde{x}_2) = 20$. Thus $\tilde{x}_1 = 4$ and $\tilde{x}_2 = 5$ are optimal solutions of our discrete problem of maximizing $\hat{f}_3$ since f_3 is concave.

The intention of the introductory example is to show that discrete problems may in some sense be harder to solve than continuous problems. Often the very powerful methods from mathematical analysis can not be applied directly. In some cases there are workarounds to deal with discrete problems using continuous techniques in several subproblems. More examples can be found, e. g. in the theory of integer linear programming.

1 Mathematics in the computer age

The invention of the computer continues to have a great impact on the progression of mathematics. Calculation patterns have to be reassessed from another point of view.

The continuous way of maximizing $\hat{f}_2$ is very clever and few steps have to be performed by a human. Checking 2000001 cases is really simpleminded, lengthy, tedious, and error-prone if it is performed by a human.

Since the invention of the computer or mechanical calculating machines we can pass some calculations to machines. The following simple and short computer program maximizes $\hat{f}_2$ in less than a second on a customary personal computer:

```
champion = 0
for x = −1000000 to 1000000 do
    if 9 · x − x^2 > 9 · champion − champion^2 then
```

$\mathtt{champion} = x$
print $\mathtt{champion}$

Clearly, in this case we can also perform the continuous way of maximizing $\hat{f}_2$ on a computer by using a computer algebra system. Nowadays, it is no problem for computer algebra systems to differentiate functions and to perform algebraic and analytic transformations. However such a system needs a lot of steps to perform an operation like "differentiate a given function f" compared to the steps needed in the above described simple enumeration scheme.

For practical applications it is not important how a problem is solved but how long it takes to solve it and how much it costs. Clearly these aspects are not new. But the metric changed from counting in human working days to something more intricate. Now we can buy ready-made devices to complete a certain task or we invent specialized machines. Some things can either be done by a computer or by a human. We can buy standard software or customized software. There are endless alternatives how to solve a practical problem. And in most cases everything is measured in costs.

Concerning the alternative between customized machines and standard computers we can state that the current trend in (super-) computing is to use one standard machine for allmost all types of problems due to cost reasons. For every application there may be different software that solves the problem on such a machine. So mathematicians, computer scientists, and other scientists are asked to develop algorithms to solve problems on given machines.

Today the most successful and reasonably priced calculating machines can only manipulate bits which mathematically means that only boolean functions $\{0,1\}^n \to \{0,1\}$ can be applied. So these are discrete machines which are simpleminded but very fast. These facts imply that one has to study discrete structures and algorithms.

2 Facing the discrete nature of some mathematical problem types

In the previous section we have argued that many real-world problems are nowadays solved using discrete machines. So, at least in some cases, these machines have to approximate continuous values by discrete values in order to solve continuous problems.

The more important reason to study discrete structures and corresponding algorithms is that, undeniably, some mathematical problems indeed have a discrete structure. Working with continuous mathematics on discrete problems means to approximate discrete problems by continuous problems. Due to the big potency of the methods from continuous mathematics this indirect route was and in some cases still is successful in applications.

During the last decades mathematicians have made great progress in the development of algorithms based on discrete structures. In combination with the present-day computers the more direct route of solving discrete problems using discrete techniques on discrete machines becomes more and more successful in many applications.

In this context we would like to mention a recent result in the theory of mixed-integer nonlinear programming [5]. Here the authors do not assume that the optimization problem is relaxable i. e. their algorithm only evaluates the integer variables at integer points. The most surprising result of this article is that the number of function evaluations[1] is less than the number of function calls needed for solving the corresponding relaxed problem without integer variables.

3 Discrete Structures, Algorithms, and Applications

The formal framework of this text is that of a cumulative Habilitation Thesis. This means that the following chapters are sparsely connected research articles. The key question connecting all these is:

"How can discrete (optimization-) problems be solved?"

Clearly this question is rather vague and comprising so that we can deal only with some aspects. Vague and comprising as the title is, that is what this thesis is all about: discrete structures, algorithms, and applications.

There exists a broad variety of discrete structures. In this thesis we only deal with a few selected discrete structures, including polyominoes, integral point sets, and directed graphs.

Optimizing over discrete structures can have several characteristics. In some cases an optimum can be determined analytically or we can derive some properties of the optimal discrete structures. In other cases we can design algorithms that determine optimal solutions in reasonable time or we can design algorithms that locate good solutions including a worst case guarantee.

Our research is motivated from applications of such different areas as social dynamics or allocation optimization of a fashion discounter.

4 Polyominoes

A polyomino is a connected interior-disjoint union of axis-aligned unit squares joined edge-to-edge. In other words, it is an edge-connected union of cells in the planar square lattice. There are at least three ways to define two polyominoes as equivalent, namely factoring out just translations (fixed polyominoes), rotations and translations (chiral polyominoes), or reflections, rotations and translations (free polyominoes). In literature polyominoes are sometimes named animals or one speaks of the cell-growth problem [15, 36].

For the origin of polyominoes we quote Klarner [16]: "Polyominoes have a long history, going back to the start of the 20th century, but they were popularized in the present era initially by Solomon Golomb, then by Martin Gardner in his *Scientific American* columns." To give an illustration

[1] In some applications functions are not given analytically but based on a possible time-consuming simulation so that the number of function evaluations may dominate the overall runtime of the optimization algorithm.

of polyominoes Figure 1 depicts the free polyominoes consisting of at most 5 unit squares.

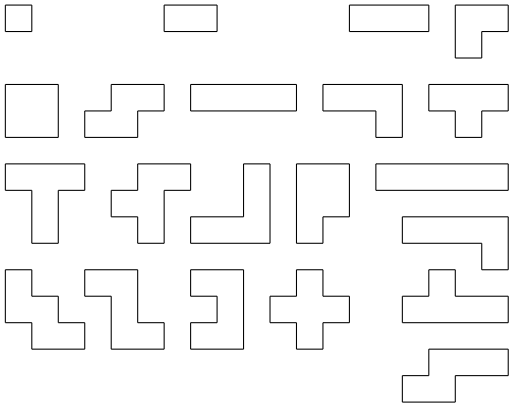


Figure 1: Polyominoes with at most 5 squares.

We would like to mention a few applications and problems for polyominoes. The term cell-growth problem certainly suggests applications in medicine and biology. Polyominoes are for instance used in the *Ising Model* [4] and for modeling neural networks, flocking birds, beating heart cells, atoms, protein folds, biological membrane, social behavior, etc. Further applications of polyominoes are in the fields of chemistry and physics. For problems concerning polyominoes one might mention counting polyominoes, generating polyominoes, achievement games and extremal animals. Since polyominoes are a very broad and still active field of research we would like to refer the interested reader to the MSC-code 05B50 instead of citing a large list of articles.

There are several generalizations of polyominoes, e. g. polyiamonds (edge-to-edge unions of unit equilateral triangles), polyhexes (edge-to-edge unions of unit regular hexagons), polyabolos (edge-to-edge unions of unit right isosceles triangles), polycubes (face-to-face unions of unit cubes), etc. One can also define polyominoes as connected systems of cells on Archimedean tessellations. Besides tessellations and higher dimensions the most natural generalization of polyominoes is to define k-polyominoes as non-intersecting connected systems of regular unit k-gons [17].

Let us move over to concrete problems for polyominoes. In [1] the authors consider three-dimensional polyominoes of minimal surface area and of volume n. The minimal surface area is explicitly computed and yields a discrete isoperimetric inequality. These variational problems are the key to finding the path of escape from the metastable state for the three-dimensional Ising model at very low temperatures. The two-dimensional analogue are polyominoes with minimum perimeter $p(n)$ consisting of n unit squares. The exact value of $p(n)$ was determined in [8] to be $2\left\lceil 2\sqrt{n}\right\rceil$. We describe and enumerate the complete set of polyominoes attaining this bound in Chapter 2 and [23].

Another discrete optimization problem is to ask for the maximum area of the convex hull of a polyomino consisting of n unit squares. Again the maximum value $n + \left\lfloor\frac{n-1}{2}\right\rfloor\left\lfloor\frac{n}{2}\right\rfloor$ was already proved, see [3]. In Chapter 3 and [22] we describe and enumerate the complete set of polyominoes attaining this bound. Additionally we determine the exact maximum value for the general d-dimensional case to be

$$\sum_{I\subseteq\{1,\ldots,d\}} \frac{1}{|I|!}\prod_{i\in I}\left\lfloor\frac{n-2+i}{d}\right\rfloor,$$

which was a conjecture of [3].

5 Integral point sets

Since the time of the Pythagoreans, mathematicians have considered geometrical objects with integral sides. An integral point set $\mathcal{P}$ is a set of n points in the m-dimensional Euclidean space $\mathbb{E}^m$ where all pairwise distances are integral. If we denote the largest occurring distance of a point set as its diameter then the question for the minimum possible diameter $d(m,n)$ arises. For dimension $m = 2$ we have depicted the plane integral point sets with minimum diameter up to $n = 9$ in Figure 2. In [9] some imaginable applications of integral point sets are summarized: radio astronomy (wave lengths), chemistry (molecules), physics (energy quantums), robotics, and architecture.

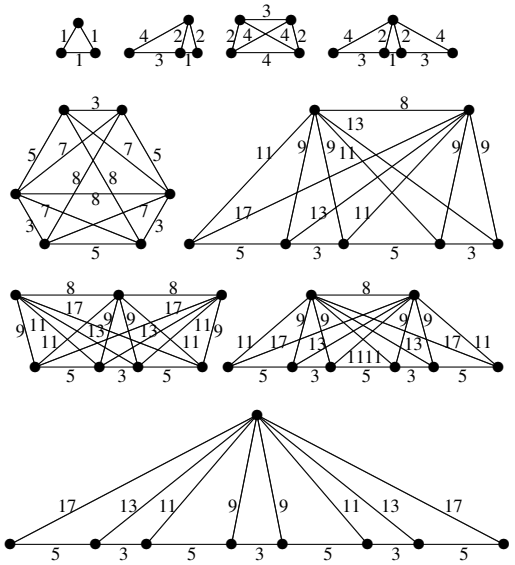


Figure 2: Integral point sets with minimum diameter for $3 \leqslant n \leqslant 9$.

Here we will only detail the first mentioned application. Consider a system of antennas like the VERY LARGE ARRAY (VLA) on the Plains of San Agustin fifty miles west of Socorro, New Mexico, see Figure 3. If the distance between two such antennas is not an integral multiple of the used wave lengths, interference occurs. Since the used wave lengths at the VLA range from 0.7 cm to 400 cm, it is technically possible to place two antennas such that no interference occurs. A complete configuration of antennas without loses due to interference directly corresponds to an integral point set. We would like to mention that in practice there

are a lot of other constraints and objectives beside the need to minimize interference.

Figure 3: Very Large Array in New Mexico, USA (Image courtesy of NRAO/AUI).

The exact values of the minimum possible diameter $d(2,n)$ for $n \leqslant 122$ were determined in [26, 35]. It turned out that the plain integral point sets which attain the lower bound $d(2,n)$ for $9 \leqslant n \leqslant 122$ each consist of $n-1$ points on a line and one point apart. There is a nice correspondence between these integral point sets and factorizations of a certain number, see [26, 35] for details.

Due to the Heron formula $A_\Delta(a,b,c) =$

$$\frac{\sqrt{(a+b+c)(a+b-c)(a-b+c)(-a+b+c)}}{4}$$

for the area A_Δ of a triangle with side lengths a, b, c we can write the area of a non-degenerate triangle with integral side lengths uniquely as $A_\Delta = q\sqrt{k}$ with a rational number q and a squarefree integer k. The number k is called the characteristic $\operatorname{char}(\Delta)$ of the triangle. The following theorem allows us to talk about the characteristic $\operatorname{char}(\mathcal{P})$ of a point set $\mathcal{P}$.

Theorem 5.1 *Each non-degenerate triangle Δ in a plane integral point $\mathcal{P}$ set has the same characteristic* $\operatorname{char}(\mathcal{P}) = \operatorname{char}(\Delta)$.

A proof of Theorem 5.1 can for instance be found in [13]. It is essential for fast exhaustive generation of plane integral point sets, see [26, 35]. In [27] Theorem 5.1 is generalized to arbitrary dimension m.

The inarguably simplest m-dimensional integral point set is an m-dimensional simplex with integral side lengths. So in order to generate m-dimensional integral points sets in general one has to generate m-dimensional integral simplices. In [7] a nice bijection between m-dimensional integral simplices with side lengths in $\{1,2\}$ and the partitions of $m+1$ turns up. In Chapter 4 and [24] we determine the numbers of integral tetrahedra with diameter d up to isomorphism for all $d \leqslant 1000$ via computer enumeration. Therefore we give an algorithm that enumerates the $\Omega\left(d^6\right)$ integral tetrahedra with diameter at most d in $O\left(d^5\right)$ time and an algorithm that can check the canonicity of a given integral tetrahedron with at most 6 integer comparisons. For results on m-dimensional integral point sets for $m \geqslant 3$ we refer the interested reader to [26, 30]. If also the coordinates of the points of an integral point set have to be integral then all simplices of such a set must have characteristic one. A fast generation algorithm for these simplices, especially for dimension $m = 2$, is given in [28].

Although integral point sets have been studied for a long time very few rigorous results are known. A reason might be that in this problem a lot of number theory and geometry is involved. The usual way to deal with a difficult (optimization-) problem is to relax it. So in Chapter 5 and [19, 20] we relax the ring of integers $\mathbb{Z}$ to the finite ring $\mathbb{Z}_n = \mathbb{Z}/\mathbb{Z}n$ and consider integral point sets over $\mathbb{Z}_n^m$. This approach is continued in Chapter 7 and [25] where we specialize to n being a prime or more generally consider two-dimensional integral point sets over finite fields. The higher-dimensional case is treated in Chapter 8 and [31].

During the study of integral point sets over finite fields or more generally over finite rings a lot of similarities to the original problem turned up. Integral point sets with many points seem to consist of large collinear subsets. If we forbid three points to be collinear, then integral point sets with many points seem to be attracted by circles. If we further forbid four points to be situated on a circle then we meet a well known question from discrete geometry:

"Are there seven points in the plane, no three on a line, no four on a circle with pairwise integral distances?"

In Chapter 7 and [25] we present such examples over $\mathbb{Z}_{50}^2$ and over $\mathbb{Z}_{61}^2$. These discoveries motivated us to search more extensively for such an example in the Euclidean plane $\mathbb{E}^2$, which leads to a discovery of two examples in the Euclidean plane $\mathbb{E}^2$, see Chapter 6 and [21].

Beside integral point sets with maximum cardinality over finite fields one is also interested in inclusion-maximal integral point sets. We give the related results in Chapter 9 and [14]. Again the results motivate some research on integral point sets in the Euclidean plane $\mathbb{E}^2$. Also in this space there exist inclusion-maximal integral point sets. The smallest inclusion-maximal integral triangle with integral coordinates is determined in [18]. A more comprising exploration of inclusion-maximal integral point sets over the integer grid $\mathbb{Z}^2$ is outlined in Chapter 10 and [2].

6 Extremal graph theory

The mathematical branch of extremal graph theory studies the graphs which are *extremal* among graphs with a certain property. In Chapter 11 and [29] we consider orientations of undirected graphs such that the resulting diameter of the directed graph becomes minimal. An application is e. g. gossiping and broadcasting in communication networks, see [10].

7 Allocation optimization of a fashion discounter

Although research in applied mathematics is not restricted to real-world business cases our research in allocation op-

timization was initiated by a request of a current industry partner. Consider a fashion discounter with many branches, no replenishment of products, and very small sales numbers per product. The first task is to estimate the future branch-dependent demand from historic sales data. Demand forecasting for never-out-of-stock items is a well-studied topic both in research and practice – the literature is overboarding. For promotional items and other items with single, very short life cycles, however, we did not find any suitable demand forecasting methods.

In Chapter 12 and [32] we introduce a new index to robustly measure the deviation between historic supply and demand for sparse data. The results of a field study applying this index in practice are outlined in Chapter 13 and [34]

branch	S	M	L	XL
1	1.23	2.32	3.21	0.71
2	3.71	6.52	7.79	2.50
3	0.38	1.47	1.63	0.41
4	1.73	3.18	3.08	1.68
5	0.81	1.94	4.32	1.13
6	1.57	3.08	2.94	1.45
7	1.21	2.31	3.22	0.72
8	1.25	2.27	3.35	0.83
9	3.41	5.79	6.37	3.21

Table 1: List of demand vectors.

Once we have good estimates for the future branch-dependent demand at hand, see Table 1 for an example, we have to face another task. Due to cost reasons the delivery of branches is lot-based. This means that we have a finite number of lot-types like e. g.

$$\begin{pmatrix}1\\2\\2\\1\end{pmatrix}, \begin{pmatrix}1\\2\\3\\1\end{pmatrix}, \begin{pmatrix}1\\1\\2\\1\end{pmatrix}, \begin{pmatrix}1\\2\\4\\1\end{pmatrix},$$

meaning that the first lot-type consists of 1 item of size S, 2 items of size M, 2 items of size L, and 1 item of size XL.

Every branch is delivered by an integral multiple of exactly one lot-type. To reduce the complexity in storehouse logistics the number of totally used lot-types per product is bounded by a small number κ. In Table 2 we have given an assignment and multiplicities of lot-types for $\kappa = 2$. To measure the deviation between demand and planned supply we have utilized the sum of absolute differences $\|\cdot\|_1$.

So here we have a discrete optimization problem: select a small number κ of lot-types out of a given list of useable lot-types and assign to each branch lot-type and multiplicity such that the sum of deviations between (fractional) demand and planned supply is minimized. Additionally we have another restriction from practice: the number of items which are delivered in total has to be in a given interval $\left[\underline{I}, \overline{I}\right]$. If we choose the interval $[100, 120]$ in our example the assignment of Table 2 does not remain valid. A valid assignment is given in Table 3.

branch	$\begin{pmatrix}1\\2\\2\\1\end{pmatrix}$	$\begin{pmatrix}1\\2\\3\\1\end{pmatrix}$	$\begin{pmatrix}1\\1\\2\\1\end{pmatrix}$	$\begin{pmatrix}1\\2\\4\\1\end{pmatrix}$	$\|\cdot\|_1$
1		1			1.05
2		3			2.94
3	1				2.11
4	2				2.33
5		1			1.70
6		1			2.16
7		1			1.02
8		1			1.04
9	3				1.20
$\sum$					15.55

Table 2: Assignment and multiplicity of lot-types.

branch	$\begin{pmatrix}1\\2\\3\\1\end{pmatrix}$	$\begin{pmatrix}1\\2\\2\\1\end{pmatrix}$	$\|\cdot\|_1$
1		1	2.05
2	3		2.94
3		1	2.11
4		2	2.33
5	1		1.70
6	1		2.16
7	1		1.02
8	1		1.04
9		4	5.22
$\sum$			20.57

Table 3: Assignment and multiplicity of lot-types respecting a cardinality condition.

In Chapter 14 and [6] we describe an exact algorithm and a fast heuristic which is capable to solve this optimization problem in practice.

8 The bounded confidence model

Another motivation for our studies comes from social sciences. Here one e. g. is interested in the social evolution of opinions and therefore inspects opinion dynamics. One model to formalize an opinion dynamic is the so called bounded confidence model. In the one dimensional case every individual i at time t has an opinion $x_i^{(t)}$. Other individuals do affect the opinion of individual i. The key assumption is that we are only influenced by opinions which are in some sense *near* to our opinion and ignore extreme opinions. One way to formalize this is to introduce a confidence interval $I\left(x_i^{(t)}\right) = \left[x_i^{(t)} - \varepsilon, x_i^{(t)} + \varepsilon\right]$ for a constant $\varepsilon \in \mathbb{R}$. With this we can define the opinion $x_i^{(t+1)}$ of individual i at time $t+1$ as the arithmetic mean of all opinions $x_j^{(t)}$ which are

contained in the confidence interval $I\left(x_i^{(t)}\right)$ of individual i.

In [12] it is shown that in such a model after a finite number of time steps we have a consensus of the individuals. This means that there are some groups of individuals having the same opinion which remains stable in the future time steps.

A more philosophical question asks which premises are sufficient that one can find *the truth*. In some sense scientists are truth seekers. To analyze this question one can enhance the bounded confidence model of opinion dynamics by introducing a truth $h \in \mathbb{R}$. So far a consensus of our individuals is completely unrelated to the location of the truth h. We may assume that there do exist some truth seekers which are attracted by a positive factor α towards the truth. This means that for a truth seeker i the truth influences his opinion by a fixed weight α in the arithmetical mean update of his new opinion. Hegselmann and Krause have conjectured that in any such configuration the truth seekers converge to the truth. In Chapter 4 and [33] we prove this conjecture.

Apart from social science there is another interest in opinion dynamics coming from marketing. The key target of marketing is to influence individuals in their opinion towards a specific product or products of a specific company. To factor in an underlying opinion dynamics is a very natural step. So now we enhance our basic bounded confidence model by the possibility to place some opinions at various time steps. This can be advertisements or in the context of an election campaign communicators. So we obtain another optimization problem which asks for an optimal control - where to place advertisements or communicators in time and opinion space.

This topic may become a new research field, see [11] for an initial introduction.

9 Place of publication

For formal reasons in Table 4 we give an overview on the place of publication of the research articles corresponding to the chapters of this thesis. All chapters coincide besides some minor corrections, reformulations, and changes in the layout with the submissions of the corresponding research articles.

C.	research article
2	S. Kurz. Counting polyominoes with minimum perimeter. *Ars Combin.*, 14 p., (to appear).
3	S. Kurz. Convex hulls of polyominoes. *Beiträge Algebra Geom.*, 49(1):125–136, 2008.
4	S. Kurz. Enumeration of integral tetrahedra. *J. Integer Seq.*, 10:Article 07.9.3, 12 p., 2007.
5	A. Kohnert and S. Kurz. Integral point sets over $\mathbb{Z}_n^m$. *Discrete Appl. Math.*, 20 p., (to appear).
6	T. Kreisel and S. Kurz. There are integral heptagons, no three points on a line, no four on a circle. *Discrete Comput. Geom.*, 4 p., (to appear).
7	S. Kurz. Integral point sets over finite fields, 22 p., (submitted).
8	S. Kurz and H. Meyer. Integral point sets in higher dimensional affine spaces over finite fields, 22 p., (submitted).
9	M. Kiermaier and S. Kurz. Inclusion-maximal integral point sets in affine planes over finite fields, 15 p., (submitted).
10	A. R. Antonov and S. Kurz. Maximal integral point sets over $\mathbb{Z}^2$, 25 p., (submitted).
11	S. Kurz and M. Lätsch. Bounds for the minimum oriented diameter, 21 p., (submitted).
12	S. Kurz and J. Rambau. Demand forecasting for for companies with many branches, low sales numbers per product, and non-recurring orderings. In *Proceedings of the Seventh International Conference on Intelligent Systems Design and Applications, 22-24.10.2007, Rio de Janeiro, Brazil*, pages 196–201, 2007.
13	S. Kurz, J. Rambau, J. Schlüchtermann, and R. Wolf. The top-dog index: A new measurement for the demand consistency of the size distribution in pre-pack orders for a fashion discounter with many small branches, 22 p., (submitted)
14	C. Gaul, S. Kurz, and J. Rambau. Lotsize optimization leading to a p-median problem with cardinalities, 14 p., (submitted)
15	S. Kurz and J. Rambau. On the Hegselmann-Krause conjecture in opinion dynamics, 13 p., (in preparation).

Table 4: Chapter and place of publication.

Bibliography

[1] L. Alonso and R. Cerf, *The three dimensional polyominoes of minimal area*, Electron. J. Combin. **3** (1996), 39 p.

[2] A. R. Antonov and S. Kurz, *Maximal integral point sets over* $\mathbb{Z}^2$, (submitted).

[3] K. Bezdek, P. Braß, and H. Harborth, *Maximum convex hulls of connected systems of segments and of polyominoes*, Beiträge Algebra Geom. **35** (1994), no. 1, 37–43, Festschrift on the occasion of the 65th birthday of Otto Krötenheerdt.

[4] B. A. Cibra, *An introduction to the ising model*, Amer. Math. Monthly **94** (1987), 937–957.

[5] O. Exler and K. Schittkowksi, *A trust region SQP algorithm for mixed-integer nonlinear programming*, Optimization Letters **1**, no. 3, 269–280.

[6] C. Gaul, S. Kurz, and J. Rambau, *Lotsize optimization leading to a p-median problem with cardinalities*, (submitted).

[7] C. Haase and S. Kurz, *A bijection between the* d*-dimensional simplices with distances in* $\{1,2\}$ *and*

the partitions of $\mathrm{d}+1$, J. Combin. Theory Ser. A **113** (2006), no. 4, 736–738.

[8] F. Harary and H. Harborth, *Extremal animals*, J. Combin. Inform. System Sci. **1** (1976), 1–8.

[9] H. Harborth, *Integral distances in point sets*, Butzer, P. L. (ed.) et al., Karl der Grosse und sein Nachwirken. 1200 Jahre Kultur und Wissenschaft in Europa. Band 2: Mathematisches Wissen. Turnhout: Brepols, 1998, 213–224.

[10] S. M. Hedetniemi, S. T. Hedetniemi, and A. L. Liestman, *Survey of Gossiping and Broadcasting in Communication Networks*, Networks **18** (1988), 319–349.

[11] R. Hegselmann, S. König, S. Kurz, C. Niemann, and J. Rambau, *Optimal Control in Opinion Dynamics: Concepts, Structures, Pitfalls, Algorithms*, (in preparation).

[12] R. Hegselmann and U. Krause, *Opinion dynamics and bounded confidence: models, analysis and simulation*, Journal of Artificial Societies and Social Simulation (JASSS) **5** (2002), no. 3.

[13] A. Kemnitz, *Punktmengen mit ganzzahligen Abständen*, Habilitationsschrift, TU Braunschweig, 1988.

[14] M. Kiermaier and S. Kurz, *Inclusion-maximal integral point sets in affine planes over finite fields*, (submitted).

[15] D. A. Klarner, *Cell growth problems*, Canad. J. Math. **19** (1967), 851–863.

[16] D. A. Klarner, *Polyominoes*, Handbook of Discrete and Computational Geometry (Jacob E. Goodman and Joseph O'Rourke, eds.), CRC Press LLC, 1997, 225–242.

[17] M. Koch and S. Kurz, *Enumeration of generalized polyominoes*, (submitted).

[18] A. Kohnert and S. Kurz, *A note on Erdös-Diophantine graphs and Diophantine carpets*, Math. Balkanica (N.S.) **21** 2007, no. 1-2, 1–5.

[19] A. Kohnert and S. Kurz, *Integral point sets over* $\mathbb{Z}_n^m$, Discrete Appl. Math., (to appear).

[20] A. Kohnert and S. Kurz, *Integral point sets over* $\mathbb{Z}_n^m$, Electron. Notes Discrete Math. **27** (2006), 65–66.

[21] T. Kreisel and S. Kurz, *There are integral heptagons, no three points on a line, no four on a circle*, Discrete Comput. Geom., (to appear).

[22] S. Kurz, *Convex hulls of polyominoes*, Beiträge Algebra Geom. **49** (2008), no. 1, 125–136.

[23] S. Kurz, *Counting polyominoes with minimum perimeter*, Ars Combin., (to appear).

[24] S. Kurz, *Enumeration of integral tetrahedra*, J. Integer Seq. **10** (2007), Article 07.9.3.

[25] S. Kurz, *Integral point sets over finite fields*, (submitted).

[26] S. Kurz, *Konstruktion und Eigenschaften ganzzahliger Punktmengen*, Ph.D. thesis, Bayreuth. Math. Schr. 76. Universität Bayreuth, 2006.

[27] S. Kurz, *On the characteristic of integral point sets in* $\mathbb{E}^m$, Australas. J. Combin. **36** (2006), 241–248.

[28] S. Kurz, *On the generation of Heronian triangles*, Serdica Journal of Computing, (to appear).

[29] S. Kurz and M. Lätsch, *Bounds for the minimum oriented diameter*, (submitted).

[30] S. Kurz and R. Laue, *Bounds for the minimum diameter of integral point sets*, Australas. J. Combin. **39** (2007), 233–240.

[31] S. Kurz and H. Meyer, *Integral point sets in higher dimensional affine spaces over finite fields*, (submitted).

[32] S. Kurz and J. Rambau, *Demand forecasting for companies with many branches, low sales numbers per product, and non-recurring orderings*, Proceedings of the Seventh International Conference on Intelligent Systems Design and Applications, 2007, 196–201.

[33] S. Kurz and J. Rambau, *On the Hegselmann-Krause conjecture in opinion dynamics*, (in preparation).

[34] S. Kurz, J. Rambau, J. Schlüchtermann, and R. Wolf, *The top-dog index: A new measurement for the demand consistency of the size distribution in prepack orders for a fashion discounter with many small branches*, (submitted).

[35] S. Kurz and A. Wassermann, *On the minimum diameter of plane integral point sets*, Ars Combin., (to appear).

[36] R. C. Read, *Contributions to the cell growth problem*, Canad. J. Math. **14** (1962), 1–20.

Chapter 2

Counting polyominoes with minimum perimeter

SASCHA KURZ[1]

ABSTRACT. The number of essentially different square polyominoes of order n and minimum perimeter $p(n)$ is enumerated.

2000 MSC: 05B50; 05C30.
Key words and phrases: polyominoes, enumeration.

1 Introduction

Suppose we are given n unit squares. What is the best way to arrange them side by side to gain the minimum perimeter $p(n)$? In [4] F. Harary and H. Harborth proved that $p(n) = 2\left\lceil 2\sqrt{n}\right\rceil$. They constructed an example where the cells grow up cell by cell like spirals for these extremal polyominoes (see Figure 1). In general, this is not the only possibility to reach the minimum perimeter.

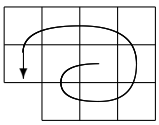


Figure 1: Spiral construction.

Thus the question arises to determine the number $e(n)$ of different square polyominoes of order n and with minimum perimeter $p(n)$ where we regard two polyominoes as equal if they can be mapped onto each other by translations, rotations, and reflections.

We will show that these extremal polyominoes can be obtained by deleting squares at the corners of rectangular polyominoes with the minimum perimeter $p(n)$ and with at least n squares. The process of deletion of squares ends if n squares remain forming a desired extremal polyomino. This process leads to an enumeration of the polyominoes with minimum perimeter $p(n)$.

[1]Sascha Kurz, University of Bayreuth, Department of Mathematics, D-95440 Bayreuth, Germany.
E-mail adress: sascha.kurz@uni-bayreuth.de

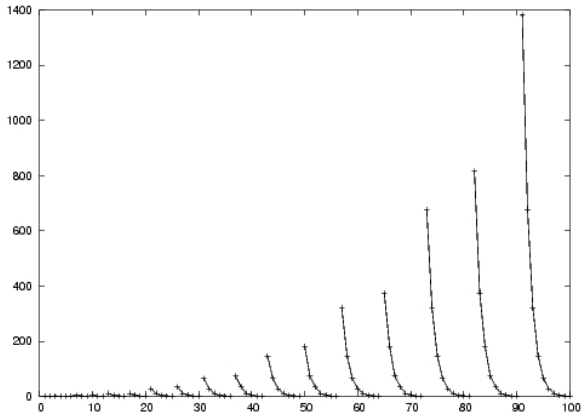


Figure 2: $e(n)$ for $n \leqslant 100$.

Theorem 1.1 *The number $e(n)$ of polyominoes with n squares and minimum perimeter $p(n)$ is given by* $e(n) =$

$$\begin{cases} 1 & \text{if } n = s^2, \\ \sum\limits_{c=0}^{\left\lfloor -\frac{1}{2}+\frac{1}{2}\sqrt{1+4s-4t}\right\rfloor} r_{s-c-c^2-t} & \text{if } n = s^2+t, \\ & \quad 0 < t < s, \\ 1 & \text{if } n = s^2+s, \\ q_{s+1-t} + \sum\limits_{c=1}^{\left\lfloor \sqrt{s+1-t}\right\rfloor} r_{s+1-c^2-t} & \text{if } n = s^2+s+t, \\ & \quad 0 < t \leqslant s, \end{cases}$$

with $s = \lfloor\sqrt{n}\rfloor$, and with r_k, q_k being the coefficient of x^k in the following generating function $r(x)$ and $q(x)$, respectively. The two generating functions

$$s(x) = 1 + \sum_{k=1}^{\infty} x^{k^2} \prod_{j=1}^{k} \frac{1}{1-x^{2j}}$$

and

$$a(x) = \prod_{j=1}^{\infty} \frac{1}{1-x^j}$$

are used in the definition of

$$r(x) = \frac{1}{4}\left(a(x)^4 + 3a\left(x^2\right)^2\right)$$

and

$$q(x) = \frac{1}{8}\left(a(x)^4 + 3a\left(x^2\right)^2 + 2s(x)^2 a\left(x^2\right) + 2a\left(x^4\right)\right).$$

The behavior of $e(n)$ is illustrated in Figure 2. It has a local maximum at $n = s^2 + 1$ and $n = s^2 + s + 1$ for $s \geqslant 1$. Then $e(n)$ decreases to $e(n) = 1$ at $n = s^2$ and $s = s^2 + s$. In the following we give lists of the values of $e(n)$ for $n \leqslant 144$ and of the two maximum cases $e(s^2 + 1)$ and $e(s^2 + s + 1)$ for $s \leqslant 49$,

$e(n) = 1, 1, 2, 1, 1, 1, 4, 2, 1, 6, 1, 1, 11, 4, 2, 1, 11, 6, 1, 1, 28, 11, 4, 2, 1, 35, 11, 6, 1, 1, 65, 28, 11, 4, 2, 1, 73, 35, 11, 6, 1, 1, 147, 65, 28, 11, 4, 2, 1, 182, 73, 35, 11, 6, 1, 1, 321, 147, 65, 28, 11, 4, 2, 1, 374, 182, 73, 35, 11, 6, 1, 1, 678, 321, 147, 65, 28, 11, 4, 2, 1, 816, 374, 182, 73, 35, 11, 6, 1, 1, 1382, 678, 321, 147, 65, 28, 11, 4, 2, 1, 1615, 816, 374, 182, 73, 35, 11, 6, 1, 1, 2738, 1382, 678, 321, 147, 65, 28, 11, 4, 2, 1, 3244, 1615, 816, 374, 182, 73, 35, 11, 6, 1, 1, 5289, 2738, 1382, 678, 321, 147, 65, 28, 11, 4, 2, 1,$

$e(s^2 + 1) = 1, 1, 6, 11, 35, 73, 182, 374, 816, 1615, 3244, 6160, 11678, 21353, 38742, 68541, 120082, 206448, 351386, 589237, 978626, 1605582, 2610694, 4201319, 6705559, 10607058, 16652362, 25937765, 40122446, 61629301, 94066442, 142668403, 215124896, 322514429, 480921808, 713356789, 1052884464, 1546475040, 2261006940, 3290837242, 4769203920, 6882855246, 9893497078, 14165630358, 20206501603, 28718344953, 40672085930, 57404156326, 80751193346,$

$e(s^2 + s + 1) = 2, 4, 11, 28, 65, 147, 321, 678, 1382, 2738, 5289, 9985, 18452, 33455, 59616, 104556, 180690, 308058, 518648, 863037, 1420480, 2314170, 3734063, 5970888, 9466452, 14887746, 23235296, 36000876, 55395893, 84680624, 128636339, 194239572, 291620864, 435422540, 646713658, 955680734, 1405394420, 2057063947, 2997341230, 4348440733, 6282115350, 9038897722, 12954509822, 18496005656, 26311093101, 37295254695, 52682844248, 74170401088, 104083151128.$

2 Proof of the main theorem

The perimeter cannot be a minimum if the polyomino is disconnected or if it has holes. For connected polyominoes without holes the property of having the minimum perimeter is equivalent to the property of having the maximum number of common edges since an edge which does not belong to two squares is part of the perimeter. The maximum number of common edges $B(n)$ is determined in [4] to be

$$B(n) = 2n - \left\lceil 2\sqrt{n} \right\rceil. \tag{1}$$

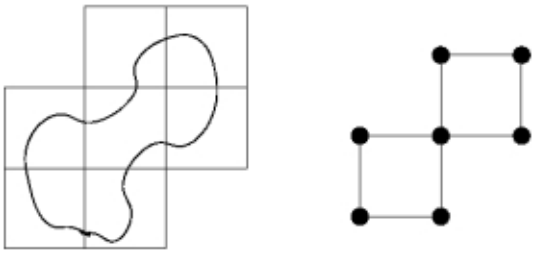


Figure 3: An example of a circle.

Denote the degree of a square by the number of its edge-to-edge neighbors. There is a closed walk trough all edge-to-edge neighboring squares of the perimeter. Now we use the terms of graph theory [3] and consider the squares as vertices. So we can define H to be the cycle $x_1x_2\ldots x_kx_1$ where the x_i are the squares of the above defined closed walk. For short we will set $|H| = k$ in the following lemmas. We would like to mention that $x_i = x_j$ with $i \neq j$ is possible in this definition. An example is depicted in Figure 3 together with the corresponding graph of H. Let furthermore h_i denote the number of squares x_j in H having degree i in the given polyomino. So

$$|H| = h_1 + h_2 + h_3 + h_4.$$

If a polyomino with minimum perimeter $p(n)$ contains a square of degree 1 (i. e. $h_1 > 0$) then $B(n) - B(n-1) = 1$. Considering Equation (1) for $B(n)$, this is equivalent to $n = s^2 + 1$ or $n = s^2 + s + 1$ so that we can assume $h_1 = 0$ in general. In the following two lemmas we prove a connection between the number of common edges of a polyomino and $|H|$.

Lemma 2.1 *If* $h_1 = 0$ *then* $h_2 = h_4 + 4$.

PROOF. Consider the polygon connecting the centers of the squares of H. For $2 \leqslant i \leqslant 4$ there is an inner angle of $\frac{(i-1)\pi}{2}$ in a square of degree i. The sum of the angles of an $|H|$-gon is $(|H| - 2)\pi$. Thus

$$(h_2 + h_3 + h_4 - 2)\pi = h_2\frac{\pi}{2} + h_3\pi + h_4\frac{3\pi}{2}$$

implies the desired equation. □

Lemma 2.2 *If* $h_1 = 0$ *then the number* m *of common edges of squares of the polyomino is*

$$m = 2n - \frac{|H|}{2} - 2.$$

PROOF. Every inner square of the polyomino has four neighbors. Counting the common edges twice yields

$$2m = 4(n - |H|) + 2h_2 + 3h_3 + 4h_4.$$

From Lemma 2.1 we obtain

$$2m = 4n - 4|H| + 3(h_2 + h_3 + h_4) - 4 = 4n - |H| - 4.$$

□

In the next lemma we use the knowledge of $|H|$ to bound the number of squares n of a polyomino.

Lemma 2.3 *For the maximum area $A(|H|)$ of a polyomino with boundary H and $h_1 = 0$ we have*

$$A(|H|) = \begin{cases} \left(\frac{|H|+4}{4}\right)^2 & \text{if } |H| \equiv 0 \pmod 4, \\ \left(\frac{|H|+4}{4}\right)^2 - \frac{1}{4} & \text{if } |H| \equiv 2 \pmod 4. \end{cases}$$

PROOF. Because of Lemma 2.2 the integer $|H|$ has to be an even number. Consider the smallest rectangle surrounding a polyomino and denote the side lengths by a and b. Using the fact that the cardinality of the boundary H of a polyomino is at least the cardinality of the boundary of its smallest surrounding rectangle we conclude $|H| \geqslant 2a + 2b - 4$. The maximum area of the rectangle with given perimeter is obtained if the integers a and b are as equal as possible. Thus $a = \left\lceil \frac{|H|+4}{4} \right\rceil$ and $b = \left\lfloor \frac{|H|+4}{4} \right\rfloor$. The product yields the asserted formula. □

Now we use the fact that we deal with polyominoes with minimum perimeter $p(n)$ and compute $|H|$ as a function of n.

Lemma 2.4 *For a polyomino with $h_1 = 0$ and with minimum perimeter $p(n)$ we have $|H| = 2\left\lceil 2\sqrt{n} \right\rceil - 4$.*

PROOF. Since for connected polyominoes without holes the property of having minimum perimeter $p(n)$ is equivalent to the property of having the maximum number $B(n)$ of common edges, we can use $B(n) = 2n - \left\lceil 2\sqrt{n} \right\rceil$ and Lemma 2.2. □

After providing those technical lemmas we give a strategy to construct all polyominoes with minimum perimeter.

Lemma 2.5 *Each polyomino with $h_1 = 0$ and minimum perimeter $p(n)$ can be obtained by deleting squares of a rectangular polyomino with perimeter $p(n)$ consisting of at least n squares.*

PROOF. Consider a polyomino P with boundary H and minimum perimeter $p(n)$. Denote its smallest surrounding rectangle by R. If the cardinality of the boundary of R is less than $|H|$ then P does not have the minimum perimeter due to Lemma 2.2 and due to the fact that $m = B(n)$ is increasing. Thus $|H|$ equals the cardinality of the boundary of R and P can be obtained by deleting squares from a rectangular polyomino with perimeter $p(n)$ and with an area at least n. Only squares of degree two can be deleted successively if the perimeter does not change. □

For the following classes of n with $s = \lfloor \sqrt{n} \rfloor$ we now characterize all rectangles being appropriate for a deletion process to obtain P with minimum perimeter $p(n)$.

(i) $n = s^2$.
From Lemmas 2.3 and 2.4 we know that the unique polyomino with minimum perimeter $p(n)$ is indeed the $s \times s$ square.

(ii) $n = s^2 + t$, $0 < t < s$.
Since

$$s^2 < n < \left(s + \frac{1}{2}\right)^2 = s^2 + s + \frac{1}{4}$$

Lemma 2.4 yields $|H| = 4s-2$. Denote the side lengths of the surrounding rectangle by a and b. With $2a + 2b - 4 = |H| = 4s - 2$ we let $a = s + 1 + c$ and $b = s - c$ with an integer $c \geqslant 0$. Since at least n squares are needed for the deletion process we have $ab \geqslant n$, yielding

$$0 \leqslant c \leqslant \left\lfloor -\frac{1}{2} + \frac{1}{2}\sqrt{1 + 4s - 4t} \right\rfloor.$$

(iii) $n = s^2 + s$.
The $s \times (s+1)$ rectangle is the unique polyomino with minimum perimeter $p(n)$ due to Lemmas 2.3 and 2.4.

(iv) $n = s^2 + s + t$, $0 < t \leqslant s$.
Since

$$\left(s + \frac{1}{2}\right)^2 = s^2 + s + \frac{1}{4} < n < (s+1)^2 = s^2 + 2s + 1$$

Lemma 2.4 yields $|H| = 4s$. Again a and b denote the side lengths of the surrounding rectangle and we let $a = s + 1 + c$ and $b = s + 1 - c$ with an integer $c \geqslant 0$. The condition $ab \geqslant n$ now yields

$$0 \leqslant c \leqslant \left\lfloor \sqrt{1 + s - t} \right\rfloor.$$

We remark that the deletion process does not change the smallest surrounding rectangle since $ab - n < b$, that is the number of deleted squares is less than the number of squares of the smallest side of this rectangle.

In Lemmas 2.1, 2.2, 2.4, and 2.5 we have required $h_1 = 0$. We now argue that all polyominoes with $h_1 > 0$ and with minimum perimeter $p(n)$ are covered by the deletion process described above ((i)-(iv)).

Lemma 2.6 *The construction of Lemma 2.5 also yields all polyominoes with minimum perimeter $p(n)$ when $h_1 > 0$.*

PROOF. Any square of degree one determines two cases, $n = s^2 + 1$ or $n = s^2 + s + 1$. (See the remark preceding Lemma 2.1.) The deletion of this square leaves a polyomino P with minimum perimeter $p(n-1)$.

In the first case P has the shape of the $s \times s$ square as in (i). Thus we get the original polyomino by deleting $s - 1$ squares from the $s \times (s+1)$ rectangle and this is covered in (ii).

In the second case P has the shape of the $s \times (s+1)$ rectangle as in (iii). Thus we get the original polyomino by deleting $s - 1$ squares from the $s \times (s+2)$ rectangle or by deleting s squares from the $(s+1) \times (s+1)$ square, and this is covered in (iv). □

So far we have described those rectangles from which squares of degree two are removed. Now we examine the process of deleting squares from a rectangular polyomino. Squares of degree two can only be located in the corners of the polyomino. What shape has the set of deleted squares at any corner? There is a maximum square of squares at the corner, the so called "Durfee square", together with squares in rows and columns of decreasing length from outside to the interior part of the polyomino, see Figure 4.

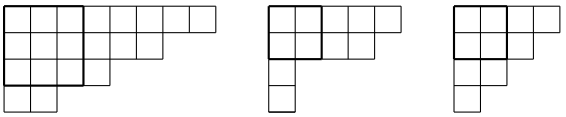


Figure 4: Shape of the deleted squares at the corners.

To count the different possibilities of the sets of deleted squares with respect to the number of the deleted squares we use the concept of a generating function $f(x) = \sum_{i=0}^{\infty} f_i x^i$. Here the coefficient f_i gives the number of different ways to use i squares. Since the rows and columns are ordered by their lengths they form Ferrer's diagrams with generating function $\prod_{j=1}^{\infty} \frac{1}{1-x^j}$ each [2]. So the generating function for the sets of deleted squares in a single corner is given by

$$a(x) = \prod_{j=1}^{\infty} \frac{1}{1-x^j}.$$

Later we will also need the generating function $s(x)$ for the sets of deleted squares being symmetric with respect to the diagonal of the corner square. Since such a symmetric set of deleted squares consists of a square of k^2 squares and the two mirror images of a Ferrer's diagrams with height or width at most k we get

$$s(x) = 1 + \sum_{k=1}^{\infty} x^{k^2} \prod_{j=1}^{k} \frac{1}{1-x^{2j}}.$$

We now consider the whole rectangle. Because of different sets of symmetry axes we distinguish between squares and rectangles. We define generating functions $q(x)$ and $r(x)$ so that the coefficient of x^k in $q(x)$ and $r(x)$ is the number of ways to remove k squares from all four corners of a square or a rectangle, respectively. We mention that the coefficient of x^k gives the desired number only if k is smaller than the small side of the rectangle.

Since we want to count polyominoes with minimum perimeter up to translation, rotation, and reflection, we have to factor out these symmetries. Here the general tool is the lemma of Cauchy-Frobenius, see e. g. [5]. We remark that we do not have to consider translations because we describe the polyominoes without coordinates.

Lemma 2.7 *(Cauchy-Frobenius, weighted form)*
Given a group action of a finite group G on a set S and a map $w : S \longrightarrow R$ from S into a commutative ring R containing $\mathbb{Q}$ as a subring. If w is constant on the orbits of G on S, then we have, for any transversal $\mathcal{T}$ of the orbits:

$$\sum_{t \in \mathcal{T}} w(t) = \frac{1}{|G|} \sum_{g \in G} \sum_{s \in S_g} w(s)$$

where S_g denotes the elements of S being fixed by g, i. e.

$$S_g = \{s \in S | s = gs\}.$$

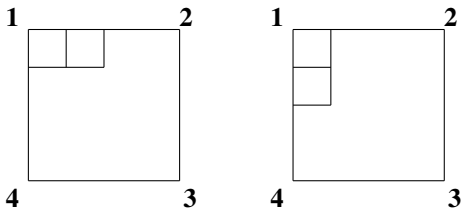


Figure 5: Labeling of the corners.

For G we take the symmetry group of a square or a rectangle, respectively, for S we take the sets of deleted squares on all four corners, and for the weight $w(s)$ we take x^k, where k is the number of squares in s. Here we will only describe in detail the application of this lemma for a determination of $q(x)$. We label the four corners of the square by 1, 2, 3, and 4, see Figure 5. In Table 1 we list the 8 permutations g of the symmetry group of a square, the dihedral group on four points, together with the corresponding generating functions for the sets S_g being fixed by g.

$(1)(2)(3)(4)$	$a(x)^4$
$(1,2,3,4)$	$a\left(x^4\right)$
$(1,3)(2,4)$	$a\left(x^2\right)^2$
$(1,4,3,2)$	$a\left(x^4\right)$
$(1,2)(3,4)$	$a\left(x^2\right)^2$
$(1,4)(2,3)$	$a\left(x^2\right)^2$
$(1,3)(2)(4)$	$s(x)^2 a\left(x^2\right)$
$(1)(2,4)(3)$	$s(x)^2 a\left(x^2\right)$

Table 1: Permutations of the symmetry group of a square together with the corresponding generating functions of S_g.

The generating function of the set of deleted squares on a corner is $a(x)$. If we consider the configurations being fixed by the identity element $(1)(2)(3)(4)$ we see that the sets of deleted squares at the four corners are independent and so $\left|S_{(1)(2)(3)(4)}\right| = a(x)^4$. In the case when $g = (1,2,3,4)$ the sets of deleted squares have to be the same for all 4 corners and we have $\left|S_{(1,2,3,4)}\right| = a\left(x^4\right)$. For the double transposition $(1,2)(3,4)$ the sets of deleted squares at corners 1 and 2, and the sets of deleted squares at corners 3 and 4 have to be equal. Because the sets of deleted squares at corner points 1 and 3 are independent we get $\left|S_{(1,2)(3,4)}\right| = a\left(x^2\right)^2$. Next we consider $g = (1)(2,4)(3)$. The sets of deleted squares at corners 2 and 4 have to be equal. If we apply g on the polyomino of the left hand side of Figure 5 we receive the polyomino on the right hand side and we see

that in general the sets of deleted squares at corners 1 and 3 have to be symmetric. Thus $\left|S_{(1)(2,4)(3)}\right| = s(x)^2 a\left(x^2\right)$. The other cases are left to the reader. Summing up and a division by 8 yields

$$q(x) = \frac{1}{8}\left(a(x)^4 + 3a\left(x^2\right)^2 + 2s(x)^2 a\left(x^2\right) + 2a\left(x^4\right)\right).$$

For the symmetry group of a rectangle we analogously obtain

$$r(x) = \frac{1}{4}\left(a(x)^4 + 3a\left(x^2\right)^2\right).$$

With Lemma 2.6, the preceding characterization of rectangles being appropriate for a deletion process and the formulas for $a(x)$, $s(x)$, $q(x)$, and $r(x)$ we have the proof of Theorem 1.1 at hand.

We would like to close with the first entries of a complete list of polyominoes with minimum perimeter $p(n)$, see Figure 6.

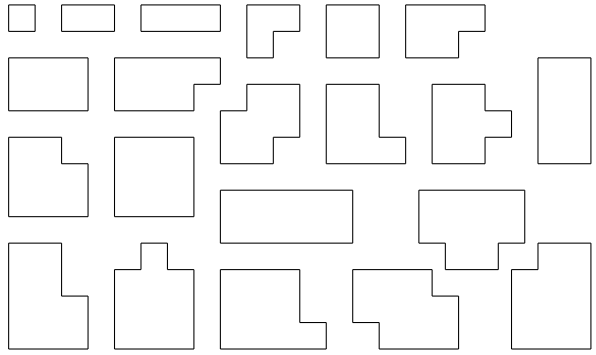


Figure 6: Polyominoes with minimum perimeter $p(n)$ for $n \leqslant 11$.

3 Acknowledgments

I thank Ralf Gugisch, Heiko Harborth, Adalbert Kerber, and Axel Kohnert for many fruitful discussions during the preparation of this article.

Bibliography

[1] L. Alonso and R. Cerf, *The three dimensional polyominoes of minimal area*, Electron. J. Combin. **3** (1996), 39p.

[2] G. E. Andrews, *The theory of partitions*, Encyclopedia of Mathematics and its Applications, vol. 2, Cambridge University Press, 1984.

[3] R. Diestel, *Graph theory*, 2nd. ed., Graduate Texts in Mathematics. 173, Springer, 2000.

[4] F. Harary and H. Harborth, *Extremal animals*, J. Combin. Inform. System Sci. **1** (1976), 1–8.

[5] A. Kerber, *Applied finite group actions*, 2nd ed., Algorithms and Combinatorics, vol. 19, Springer, Berlin, 1999.

Chapter 3

Convex hulls of polyominoes

SASCHA KURZ[1]

ABSTRACT. In this article we prove a conjecture of Bezdek, Braß, and Harborth concerning the maximum volume of the convex hull of any facet-to-facet connected system of n unit hypercubes in $\mathbb{R}^d$ [4]. For $d = 2$ we enumerate the extremal polyominoes and determine the set of possible areas of the convex hull for each n.

2000 MSC: 05B50; 05D99, 52C99.
Key words and phrases: polyominoes, convex hull, dido-type problem, isoperimetric inequality.

1 Introduction

In the legend [1] of the founding of Carthage, Queen Dido purchased the right to get as much land as she could enclose with the skin of an ox. She splitted the skin into thin stripes and tied them together. Using the natural boundary of the sea and by constructing a giant semicircle she enclosed more land than the seller could have ever imagined.

Dido-type problems have been treated by many authors e. g. [2, 4, 5, 6, 9], here we consider the maximum volume of a union of unit hypercubes. A d-dimensional polyomino is a facet-to-facet connected system of d-dimensional unit hypercubes. Examples for 2-dimensional polyominoes are the pieces of the computer game Tetris.

In 1994 Bezdek, Braß, and Harborth conjectured that the maximum volume of the convex hull of a d-dimensional polyomino consisting of n hypercubes is given by

$$\sum_{I\subseteq\{1,\dots,d\}} \frac{1}{|I|!}\prod_{i\in I}\left\lfloor\frac{n-2+i}{d}\right\rfloor,$$

but were only able to prove it for $d = 2$. In Section 3 we prove this conjecture. They also asked for the number $c_2(n)$ of different polyominoes with n cells and maximum area $n + \left\lfloor\frac{n-1}{2}\right\rfloor\left\lfloor\frac{n}{2}\right\rfloor$. In Section 2 we prove

Theorem 1.1

$$c_2(n) = \begin{cases} \frac{n^3-2n^2+4n}{16} & \text{if } n \equiv 0 \mod 4,\\ \frac{n^3-2n^2+13n+20}{32} & \text{if } n \equiv 1 \mod 4,\\ \frac{n^3-2n^2+4n+8}{16} & \text{if } n \equiv 2 \mod 4,\\ \frac{n^3-2n^2+5n+8}{32} & \text{if } n \equiv 3 \mod 4. \end{cases}$$

Besides the maximum area $n + \left\lfloor\frac{n-1}{2}\right\rfloor\left\lfloor\frac{n}{2}\right\rfloor$ and the minimum area n of the convex hull of polyominoes with n cells several other values may be attained. For each n we characterize the corresponding sets.

Theorem 1.2 *A polyomino consisting of n cells with area $\alpha = n + \frac{m}{2}$ of the convex hull exists if and only if $m \in \mathbb{N}_0$, $0 \leqslant m \leqslant \left\lfloor\frac{n-1}{2}\right\rfloor\left\lfloor\frac{n}{2}\right\rfloor$, and $m \neq 1$ if $n+1$ is a prime.*

2 The planar case

An example which attains the upper bound $n + \frac{1}{2}\left\lfloor\frac{n-1}{2}\right\rfloor\left\lfloor\frac{n}{2}\right\rfloor$ for the area of the convex hull of a polyomino with n cells is quite obvious, see Figure 1. Instead of proving this upper bound by induction over n we specify polyominoes by further parameters and then apply an induction argument.

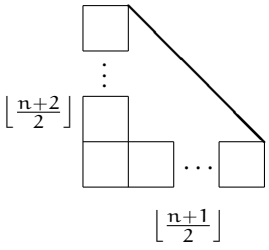


Figure 1: 2-dimensional polyomino with maximum area of the convex hull.

We describe these parameters for the more general d-dimensional case and therefore denote the standard coordinate axes of $\mathbb{R}^d$ by $1,\dots,d$. Every d-dimensional polyomino has a smallest surrounding box with side lengths $l_1,\dots,l_d$, where l_i is the length in direction i. If we build up a polyomino cell by cell then after adding a cell one of the l_i will increase by 1 or none of the l_i will increase. In the second case we increase v_i by 1, where the

[1]Sascha Kurz, University of Bayreuth, Department of Mathematics, D-95440 Bayreuth, Germany.
E-mail adress: sascha.kurz@uni-bayreuth.de

new hypercube has a facet-neighbor in direction of axis i. If M is the set of axis-directions of facet-neighbors of the new hypercube, then we will increase ν_i by 1 for only one $i \in M$. Since at this position there is the possibility to choose, we must face the fact that there might be different tuples $(l_1, \ldots, l_d, \nu_1, \ldots, \nu_d)$ for the same polyomino. We define $\nu_1 = \cdots = \nu_d = 0$ for the polyomino consisting of a single hypercube. This definition of the l_i and the ν_i leads to

$$n = 1 + \sum_{i=1}^{d} (l_i - 1) + \sum_{i=1}^{d} \nu_i. \tag{1}$$

Example 2.1 *The possible tuples describing a rectangular* 2×3*-polyomino are* $(2,3,2,0)$*,* $(2,3,1,1)$*, and* $(2,3,0,2)$*.*

Definition 2.2

$$\begin{aligned} f_2(l_1,l_2,\nu_1,\nu_2) &= 1 + (l_1-1) + (l_2-1) \\ &\quad + \frac{(l_1-1)(l_2-1)}{2} + \nu_1 + \nu_2 \\ &\quad + \frac{\nu_1(l_2-1)}{2} + \frac{\nu_2(l_1-1)}{2} + \frac{\nu_1\nu_2}{2}. \end{aligned}$$

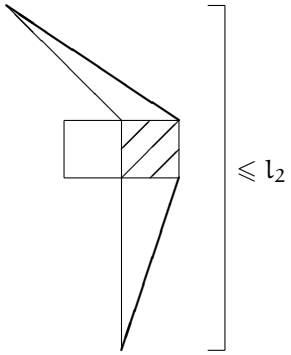


Figure 2: Increasing l_1.

Lemma 2.3 *The area of the convex hull of a 2-dimensional polyomino with tuple* (l_1, l_2, ν_1, ν_2) *is at most* $f_2(l_1, l_2, \nu_1, \nu_2)$*.*

PROOF. We prove the statement by induction on n, using Equation (1). For $n = 1$ only $l_1 = l_2 = 1, \nu_1 = \nu_2 = 0$ is possible. With $f_2(1,1,0,0) = 1$ the induction base is done. Now we assume that the statement is true for all possible tuples (l_1, l_2, ν_1, ν_2) with

$$1 + \sum_{i=1}^{2} (l_i - 1) + \sum_{i=1}^{d} \nu_i = n - 1.$$

Due to symmetry we consider only the growth of l_1 or ν_1, and the area a of the convex hull by adding the n-th square.

(i) l_1 increases by one: We depict (see Figure 2) the new square by 3 diagonal lines. Since l_1 increases the new square must have a left or a right neighbor. Without loss of generality we assume that it has a left neighbor. The new square contributes at most 2 (thick) lines to the convex hull of the polyomino. By drawing lines from the neighbor square to the endpoints of the new lines we see that the growth is at most $1 + \frac{l_2-1}{2}$, a growth of 1 for the new square and the rest for the triangles. Since $f_2(l_1+1, l_2, \nu_1, \nu_2) - f_2(l_1, l_2, \nu_1, \nu_2) = 1 + \frac{l_2-1}{2} + \frac{\nu_2}{2}$ the induction step follows.

(ii) ν_1 increases by one: In Figure 3 we depict the new square by 3 diagonal lines. Without loss of generality we assume that the new square has a left neighbor, and contributes at most 2 lines to the convex hull of the polyomino. As l_1 is not increased there must be a square in the same column as the new square. Similar to (i) we draw lines from the neighbor square to the endpoints of the new lines and see that the growth of the area of the convex hull is less than $1 + \frac{l_2-1}{2}$.

□

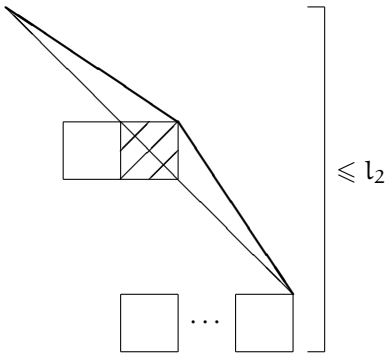


Figure 3: Increasing ν_1.

Theorem 2.4 *The area of the convex hull of a 2-dimensional polyomino with* n *unit squares is at most* $n + \frac{1}{2}\left\lfloor \frac{n-1}{2} \right\rfloor \left\lfloor \frac{n}{2} \right\rfloor$*.*

PROOF. For given n we determine the maximum of $f_2(l_1, l_2, \nu_1, \nu_2)$. Since $f_2(l_1+1, l_2, \nu_1-1, \nu_2) - f_2(l_1, l_2, \nu_1, \nu_2) = 0$ and due to symmetry we assume $\nu_1 = \nu_2 = 0$ and $l_1 \leqslant l_2$. With

$$f_2(l_1+1, l_2-1, 0, 0) - f_2(l_1, l_2, 0, 0) = \frac{l_2 - l_1 - 1}{2} > 0$$

we conclude $0 \leqslant l_2 - l_1 \leqslant 1$. Using Equation (1) gives $l_1 = \left\lfloor \frac{n+1}{2} \right\rfloor$ and $l_2 = \left\lfloor \frac{n+2}{2} \right\rfloor$. Inserting in Lemma 2.3 yields $f_2(l_1, l_2, \nu_1, \nu_2) \leqslant n + \frac{1}{2}\left\lfloor \frac{n-1}{2} \right\rfloor \left\lfloor \frac{n}{2} \right\rfloor$. This maximum is attained for example by the polyomino in Figure 1. □

In the next lemma we describe the shape of the 2-dimensional polyominoes with maximum area of the convex hull in order to determine their number $c_2(n)$.

Lemma 2.5 *Every 2-dimensional polyomino with parameters* l_1*,* l_2*,* ν_1*,* ν_2*, and with the maximum area* $n + \frac{1}{2}\left\lfloor \frac{n-1}{2} \right\rfloor \left\lfloor \frac{n}{2} \right\rfloor$ *of the convex hull consists of a linear strip*

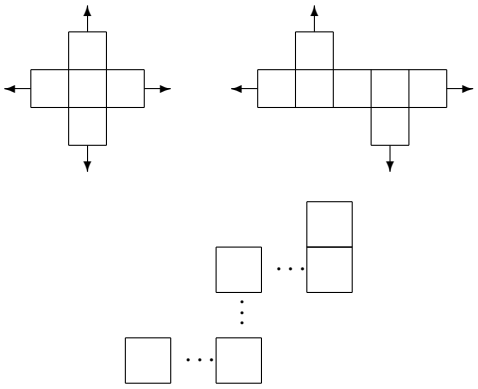


Figure 4: The two shapes of polyominoes with maximum area of the convex hull and a forbidden sub-polyomino.

with at most one orthogonal linear strip on each side (see the upper two pictures in Figure 4). Additionally we have $v_1 = v_2 = 0$ *and the area of the convex hull is given by* $f_2(l_1, l_2, v_1, v_2)$.

PROOF. From the proof of Lemma 2.3 we deduce $v_1 = v_2 = 0$ and that every sub-polyomino has also the maximum area of the convex hull. Since the area of the polyomino on the lower side of Figure 4 has an area of the convex hull which is less than $f_2(l_1, l_2, 0, 0)$ it is a forbidden sub-polyomino and only the described shapes remain. All these polyominoes attain the maximum $f_2(l_1, l_2, 0, 0)$. □

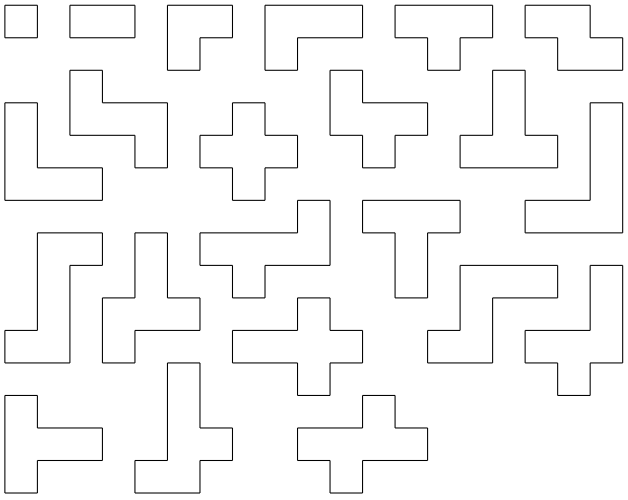


Figure 5: Complete set of extremal polyominoes with $n \leqslant 6$ cells.

Theorem 1.1

$$c_2(n) = \begin{cases} \frac{n^3-2n^2+4n}{16} & \text{if } n \equiv 0 \mod 4, \\ \frac{n^3-2n^2+13n+20}{32} & \text{if } n \equiv 1 \mod 4, \\ \frac{n^3-2n^2+4n+8}{16} & \text{if } n \equiv 2 \mod 4, \\ \frac{n^3-2n^2+5n+8}{32} & \text{if } n \equiv 3 \mod 4. \end{cases}$$

PROOF OF THEOREM 1.1. (Formula for $c_2(n)$.)

We use Lemma 2.5 and do a short calculation applying the lemma of Cauchy-Frobenius. □

Corollary 2.6 *The ordinary generating function for* $c_2(n)$ *is given by*

$$\frac{1+x-x^2-x^3+2x^5+8x^6+2x^7+4x^8+2x^9-x^{10}+x^{12}}{(1-x^2)^2(1-x^4)^2}.$$

We have depicted the polyominoes with at most 6 cells and maximum area of the convex hull in Figure 5. For more cells we give only a few concrete numbers:

$$(c_2(n))_{n=1,\ldots} = 1, 1, 1, 3, 5, 11, 9, 26, 22, 53, 36, 93, 64, 151, 94, 228, 143, 329, 195, 455, 271, 611, 351, 798, 460, 1021, 574, 1281, 722, 1583, 876, 1928, 1069, 2321, 1269, 2763, 1513, 3259, 1765, 3810, 2066, 4421, 2376, 5093, \ldots$$

This is sequence A122133 in the *Online-Encyclopedia of Integer Sequences* [10].

Besides the maximum area $n + \frac{1}{2}\left\lfloor\frac{n-1}{2}\right\rfloor\left\lfloor\frac{n}{2}\right\rfloor$ and the minimum area n of the convex hull of polyominoes with n cells several other values may be attained. In Theorem 1.2 we have completely characterized the set of areas of the convex hull of polyominoes with n cells.

PROOF OF THEOREM 1.2. Since the vertex points of the convex hull of a polyomino are lattice points on an integer grid the area of the convex hull is an integral multiple of $\frac{1}{2}$. Using Theorem 2.4 we conclude that the desired set is a subset of

$$S = \left\{ n + \frac{m}{2} \mid m \leqslant \left\lfloor\frac{n-1}{2}\right\rfloor\left\lfloor\frac{n}{2}\right\rfloor,\ m \in \mathbb{N}_0 \right\}.$$

A polyomino P consisting of n cells with area $n + \frac{1}{2}$ of the convex hull must contain a triangle of area $\frac{1}{2}$. If we extend the triangle to a square we get a convex polyomino P' consisting of $n+1$ cells. Thus P' is an rectangular $s \times t$-polyomino with $s \cdot t = n + 1$ and $s, t \in \mathbb{N}$. If $n + 1$ is a prime there exists only the $1 \times (n + 1)$-polyomino where deleting a square yields an area of n for the convex hull. So we have to exclude this case in the above set S and receive the proposed set.

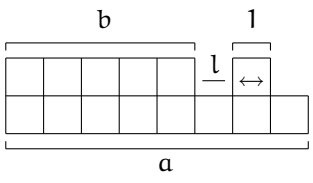


Figure 6: Construction 1: $2 \leqslant m \leqslant 2n - 8$.

For the other direction we give some constructions. For $m = 0$ we have the rectangular $1 \times n$-polyomino as an example. The above consideration $m = 1$ yields a construction if $n + 1$ is a composite number. Now we consider Construction 1 depicted in Figure 6. We choose $n = a + b + 1$, $\left\lceil\frac{n}{2}\right\rceil \leqslant a \leqslant n - 2$, and $0 \leqslant l \leqslant a - b - 1 = 2a - n$. Thus

$a \geqslant b+1$ and Construction 1 is possible. If we run through the possible values of a and l we obtain examples for

$$\begin{aligned} m \in \quad & \{0\},\{2,3,4\},\ldots,\{2a-n,\ldots,4a-2n\}, \\ & \ldots,\{n-4,\ldots,2n-8\} \\ = \quad & \{0,2,3,\ldots,2n-8\} \end{aligned}$$

if $n \equiv 0 \mod 2$ and for

$$\begin{aligned} m \in \quad & \{1,2\},\{3,4,5,6\},\ldots,\{2a-n,\ldots,4a-2n\}, \\ & \ldots,\{n-4,\ldots,2n-8\} \\ = \quad & \{1,2,\ldots,2n-8\} \end{aligned}$$

if $n \equiv 1 \mod 2$.

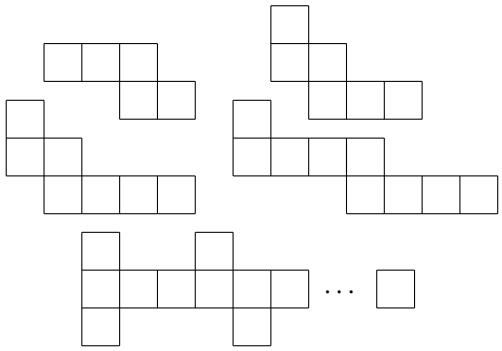


Figure 7: Construction 2: $m = 2n-7$.

In Figure 7 we give a construction for $m = 2n-7$ and in Figure 8 we give by the upper picture a construction for $2n-6 \leqslant m \leqslant \left\lceil \frac{n^2-4n}{4} \right\rceil$ with parameters k_1, k_2, and b. The conditions for these parameters are $0 \leqslant k_1, k_2 \leqslant n-2b-2$ and $n-2b-2 \geqslant b$. With given k_1, k_2, b, n we have $m = bn-2b^2-2b+k_1+k_2(b-1)$. Since we can vary k_1 at least between 0 and $b-1$ we can produce for a fix b all values between $b(n-2b-2)$ and $2b(n-2b-2)$ by varying k_1 and k_2. Now we want to combine those intervals for successive values for b. The assumption that the intervals leave a gap is equivalent to $2(b-1)(n-2(b-1)-2) < b(n-2b-2)$, that is, $n < 2b\frac{b-3}{b-2}$. We choose $2 \leqslant b \leqslant \lfloor \frac{n}{4} \rfloor$ and receive constructions for

$$m \in \left\{ 2n-6, 2n-5, \ldots, \left\lceil \frac{n^2-4n}{4} \right\rceil \right\}.$$

The lower picture of Figure 8 gives a construction for $n \geqslant 5$ and

$$m \in \left\{ \left\lfloor \frac{n^2-4n}{4} \right\rfloor, \ldots, \left\lfloor \frac{n^2-2n-8}{4} \right\rfloor \right\}.$$

□

3 Dimensions $d \geqslant 3$

To prove the conjecture of Bezdek, Braß, and Harborth for dimensions $d \geqslant 3$ we proceed similar as in Section 2.

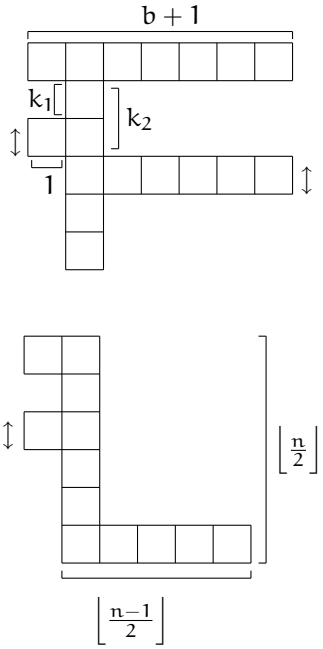


Figure 8: Construction 3 and Construction 4.

Definition 3.1

$$f_d(l_1,\ldots,l_d,v_1,\ldots,v_d) = \sum_{I\subseteq\{1,\ldots,d\}} \frac{1}{|I|!2^{d-|I|}} \sum_{b=0}^{2^d-1} \prod_{i\in I} q_{b,i}$$

with $d \geqslant 1$ *and* $b = \sum_{j=1}^{d} b_j 2^{j-1}$, $b_j \in \{0,1\}$, $q_{b,i} = \begin{cases} l_i - 1 & \text{for } b_i = 0, \\ v_i & \text{for } b_i = 1. \end{cases}$

Example 3.2

$$\begin{aligned} f_3(l_1,l_2,l_3,v_1,v_2,v_3) = {} & 1 + \overline{l_1} + \overline{l_2} + \overline{l_3} + \frac{\overline{l_1 l_2}}{2} + \\ & \frac{\overline{l_1 l_3}}{2} + \frac{\overline{l_2 l_3}}{2} + \frac{\overline{l_1 l_2 l_3}}{6} + \frac{v_1\overline{l_2}}{2} + \frac{v_1\overline{l_3}}{2} + \frac{v_2\overline{l_1}}{2} + \frac{v_2\overline{l_3}}{2} + \\ & \frac{v_3\overline{l_1}}{2} + \frac{v_3\overline{l_2}}{2} + \frac{v_1\overline{l_2 l_3}}{6} + \frac{v_2\overline{l_1 l_3}}{6} + \frac{v_3\overline{l_1 l_2}}{6} + \frac{v_1 v_2\overline{l_3}}{6} + \\ & \frac{v_1 v_3\overline{l_2}}{6} + \frac{v_2 v_3\overline{l_1}}{6} + v_1 + v_2 + v_3 + \frac{v_1 v_2}{2} + \frac{v_1 v_3}{2} + \\ & \frac{v_2 v_3}{2} + \frac{v_1 v_2 v_3}{6} \end{aligned}$$

with $\overline{l_i} = l_i - 1$.

Lemma 3.3 *The* d*-dimensional volume of the convex hull of a polyomino with* n *unit hypercubes is at most* $f_d(l_1,\ldots,l_d,v_1,\ldots,v_d)$.

PROOF. We prove the statement by double induction on d and n, using Equation (1). Since the case $d = 2$ is already done in Theorem 2.4 we assume that the lemma is proven for the $\overline{d} < d$. Since for $n = 1$ only $l_i = 1$, $v_i = 0$, $i \in$

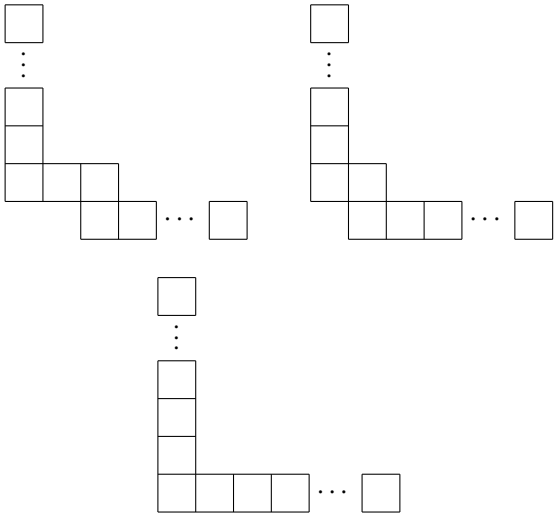


Figure 9: Construction 5: $\left\lfloor \frac{n^2-2n-6}{4} \right\rfloor \leqslant m \leqslant \left\lfloor \frac{n^2-2n+2}{4} \right\rfloor$.

Constructions for the remaining values

$$m \in \left\{ \left\lfloor \frac{n^2-2n-6}{4} \right\rfloor, \left\lfloor \frac{n^2-2n-2}{4} \right\rfloor, \left\lfloor \frac{n^2-2n+2}{4} \right\rfloor = \left\lfloor \frac{n-1}{2} \right\rfloor \left\lfloor \frac{n}{2} \right\rfloor \right\}$$

are given in Figure 9.

$\{1,\dots,d\}$ is possible and $f_d(1,\dots,1,0,\dots,0) = 1$ the induction base for n is done. Now we assume that the lemma is proven for all possible tuples $(l_1,\dots,l_d,v_1,\dots,v_d)$ with $1 + \sum_{i=1}^{d}(l_i-1) + \sum_{i=1}^{d} v_i = n-1$. Due to symmetry we consider only the growth of l_1 or v_1, and the volume of the convex hull by adding the n-th hypercube.

(i) l_1 increases by one:

As in the proof of Lemma 2.3 we draw lines of the convex hull of the n-th cube and its neighbor cube N, see Figure 10 for a 3-dimensional example. To be more precisely each line of the new convex hull has a corner point X of the upper face of the n-th cube as an endpoint. We will denote the second endpoint of this line by Y. In direction of axis 1 there is a corner point $\overline{X}$ of the bottom face of the n-th cube. Since $\overline{X}$ is also a corner point of N the line $\overline{X}Y$ is part of the old convex hull if Y is part of the old convex hull. In this case we draw the line $\overline{X}Y$. In the other case Y is also a corner point of the upper face of the new cube and we draw the line $\overline{XY}$ where $\overline{Y}$ is similar defined as $\overline{X}$. Additionally we draw all lines XY and $X\overline{X}$.

Doing this we have constructed a geometrical body which contains the increase of the convex hull and is subdivided into nice geometrical objects O_i with volume $\frac{\text{base} \times \text{height}}{k_i}$, for some $k_i \in \{1,\dots,d\}$ each. For dimension $d = 3$ the cases $k_i = 1$, $k_i = 2$, or $k_i = 3$ correspond to a box, a prism, or a tetrahedron.

We project the convex hull of the whole polyomino into the hyperplane orthogonal to axis direction 1 and

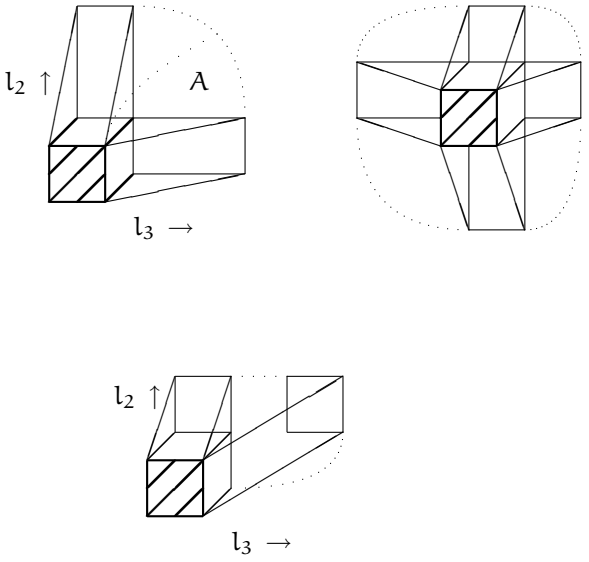


Figure 10: Increasing l_1 in the 3-dimensional case.

receive a hypervolume A. This is the convex hull of a $(d-1)$-dimensional polyomino with parameters $\bar{l}_2,\dots,\bar{l}_d,\bar{v}_2,\dots,\bar{v}_d$ where $\bar{l}_i \leqslant l_i$ and $\bar{v}_i \leqslant v_i$. From the induction hypothesis we know $A \leqslant f_{d-1}(l_2,\dots,l_d,v_2,\dots,v_d)$. We apply the same projection to the O_i and objects A_i. Due to the construction the A_i are non overlapping and we have $\sum A_i \leqslant A$. Using Cavalieri's theorem we determine the volume of O_i to be $\frac{A_i \times 1}{k_i}$. More precisely, we choose lines of the form $\overline{X}X$ as height and lift the old base up until it is orthogonal to axis direction 1. Thus we may assign a factor $\frac{1}{k}$ to each piece of A to bound the growth of the volume of the convex hull. We estimate the parts in a way that the parts with the higher factors are as big as theoretical possible.

For every $0 \leqslant r \leqslant d-1$ we consider the sets $\{i_1,i_2,\dots,i_r\}$ with $1 \neq i_a \neq i_b$ for $a \neq b$. Let Z be such a set. Define $\overline{Z} = \{j_1,\dots,j_{d-r-1}\}$ by $Z \cap \overline{Z} = \{\}$ and $Z \cup \overline{Z} = \{2,\dots,d\}$. So the vector space spanned by the axis directions of Z and the vector space spanned by the axis directions of $\overline{Z}$ are orthogonal. If we project the convex hull in the vector space spanned by $\overline{Z}$ the resulting volume is at most $f_{d-r-1}(l_{j_1},\dots,l_{j_{d-r-1}},v_{j_1},\dots,v_{j_{d-r-1}})$ since it is the convex hull of a $(d-r-1)$-dimensional polyomino. Since $\overline{Z}$ has cardinality $d-r-1$ the set Z yields a contribution of $\frac{1}{d-r} f_{d-r-1}(l_{j_1},\dots,l_{j_{d-r-1}},v_{j_1},\dots,v_{j_{d-r-1}})$ to the volume of the convex hull. With the notations from Definition 3.1 this is

$$\frac{1}{d-r} \sum_{I \subseteq \{j_1,\dots,j_{r-d-1}\}} \frac{1}{|I|!\,2^{d-r-1-|I|}} \sum_{b=0}^{2^{d-r-1}-1} \prod_{i \in I} q_{b,i}\,.$$

Our aim is to assign the maximum possible factor to

each part of A. For that reason we count for Z a maximum contribution of

$$\frac{1}{d-r}\frac{1}{|d-r-1|!}\sum_{b=0}^{2^{d-r-1}-1}\prod_{i\in\overline{Y}} q_{b,i}$$

to the volume of the convex hull.

If we do so for all possible sets Z we have assigned a factor between 1 and $\frac{1}{d}$ to every summand of $f_{d-1}(l_2,\dots,l_d,v_2,\dots,v_d)$. To get the induction step now we have to remark that the above described sum with its factors is exactly the difference between $f_d(l_1+1,\dots,l_d,v_1,\dots,v_d)$ and $f_d(l_1,\dots,l_d,v_1,\dots,v_d)$.

(ii) v_1 increases by one:

Due to symmetry of the l_i and v_i in Definition 3.1 this is similar to case (i). Additionally we remark that the maximum cannot be achieved in this case since we double count a part of the contribution of the new cube to the volume of the convex hull in our estimations.

□

Theorem 3.4 *The d-dimensional volume of the convex hull of any facet-to-facet connected system of n unit hypercubes is at most*

$$\sum_{I\subseteq\{1,\dots,d\}}\frac{1}{|I|!}\prod_{i\in I}\left\lfloor\frac{n-2+i}{d}\right\rfloor.$$

PROOF. For given n we determine the maximum of $f_d(l_1,\dots,f_d,v_1,\dots,v_d)$. Due to

$$f_d(l_1+1,l_2,\dots,l_d,v_1-1,v_2,\dots,v_d)-$$
$$f_d(l_1,l_2,\dots,l_d,v_1,v_2,\dots,v_d)=0$$

and due to symmetry we assume $v_1=\dots=v_d=0$ and $l_1\leqslant l_2\leqslant\dots\leqslant l_d$. Since for $l_d-l_1\geqslant 2$

$$f_d(l_1+1,l_2,\dots,l_{d-1},l_d-1,0,0,\dots,0)-$$
$$f_d(l_1,l_2,\dots,l_d,0,0,\dots,0)>0 \qquad (2)$$

holds, we have $0\leqslant l_d-l_1\leqslant 1$. Inequality (2) is valid due to the following consideration. If a summand of $f_d(\dots)$ contains the term l_1 and does not contain l_d then there will be a corresponding summand with l_1 replaced by l_d, so those terms equalize each other in the above difference. Clearly the summands containing none of the terms l_1 or l_d equalize each other in the difference. So there are left only the summands with both terms l_1 and l_d. Since $(l_1+1-1)(l_d-1-1)-(l_1-1)(l_d-1)=l_d-l_1-1>0$ Inequality (2) is valid.

Combining Equation (1) with $0\leqslant l_d-l_1\leqslant 1$ and $l_1\leqslant l_2\leqslant\dots\leqslant l_d$ gives $l_i=\left\lfloor\frac{n-2+i+d}{d}\right\rfloor$. Thus by inserting in Lemma 3.3 we receive the upper bound. The maximum is attained for example by a polyomino consisting of d pairwise orthogonal linear arms with $\left\lfloor\frac{n-2+i}{d}\right\rfloor$ cubes ($i=1\dots d$) joined to a central cube. □

Conjecture 3.5 *Every d-dimensional polyomino P with parameters $l_1,\dots,l_d,v_1,\dots,v_d$ and maximum volume of the convex hull fulfills $v_1=\dots=v_d=0$ and contains a sub polyomino P' fulfilling:*

(i) P' has height 1 in direction of axis i,

(ii) the projection of P' along i has also maximal volume of the convex hull and parameters $l_1,\dots,l_{i-1},l_{i+1},\dots,l_d$,

(iii) P can be decomposed into P' and up to two orthogonal linear arms.

We remark that $v_1=\dots=v_d=0$ and the maximality of the volume of the convex hull of sub polyominoes and projections of P can be concluded from the proof of Theorem 3.4.

Lemma 3.6 *If there exists a d-dimensional polyomino with n cells and volume v of the convex hull, then $v\in V_{d,n}$ with*

$$V_{d,n}=\left\{n+\frac{m}{d!}\middle|\, m\leqslant\sum_{I\subseteq\{1,\dots,d\}}\frac{d!}{|I|!}\prod_{i\in I}\left\lfloor\frac{n-2+i}{d}\right\rfloor\ m\in\mathbb{N}_0\right\}.$$

PROOF. For the determination of the volume of the convex hull of a d-dimensional polyomino we only have to consider the set of S corner points of its hypercubes which lie on an integer grid. We can decompose the convex hull into d-dimensional simplices with the volume

$$\frac{1}{d!}\begin{vmatrix} x_{1,1} & \dots & x_{1,d} & 1\\ \vdots & \ddots & \vdots & \vdots\\ x_{d+1,1} & \dots & x_{d+1,d} & 1\end{vmatrix}$$

where the coordinates of the $d+1$ points are given by $(x_{i,1},\dots,x_{i,d})\in\mathbb{Z}^d$. Thus the volume of the convex hull is an integer multiple of $\frac{1}{d!}$. The lower bound $n\leqslant v$ is obvious and the upper bound is given by Theorem 3.4. □

4 Remarks

We leave the description and the enumeration of the polyominoes with maximum convex hull for dimension $d\geqslant 3$ as a task for the interested reader. It would also be nice to see a version of Theorem 1.2 for higher dimensions.

The authors of [4] mention another class of problems which are related to the problems in [3] and [11]: What is the maximum area of the convex hull of all connected edge-to-edge packing's of n congruent regular k-gons (also denoted as k-polyominoes, see [7]) in the plane. The methods of Section 2 might be applicable for these problems.

Conjecture 4.1 *The area of the convex hull of any edge-to-edge connected system of regular unit hexagons is at most $\frac{1}{6}\left\lfloor n^2+\frac{14}{3}n+1\right\rfloor$.*

Bibliography

[1] http://en.wikipedia.org/wiki/Byrsa.

[2] A. Bezdek and K. Bezdek, *On a discrete Dido-type question*, Elem. Math. **44** (1989), no. 4, 92–100.

[3] K. Bezdek, *Connected arrangements of finitely many circles*, Mitt. Math. Sem. Giessen (to appear).

[4] K. Bezdek, P. Braß, and H. Harborth, *Maximum convex hulls of connected systems of segments and of polyominoes*, Beiträge Algebra Geom. **35** (1994), no. 1, 37–43, Festschrift on the occasion of the 65th birthday of Otto Krötenheerdt.

[5] L. Fejes Tóth, *Über das Didosche Problem*, Elem. Math. **23** (1968), 97–101.

[6] L. Fejes Tóth, *Research Problem No. 6*, Period. Math. Hungar. **4** (1973), 231–232.

[7] M. Koch and S. Kurz, *Enumeration of generalized polyominoes*, (submitted).

[8] S. Kurz, *Polyominoes with maximum convex hull*, Master's thesis, Bayreuth, 2004.

[9] A. Siegel, *A Dido problem as modernized by Fejes Tóth*, Discrete Comput. Geom. **27** (2002), no. 2, 227–238.

[10] N. J. A. Sloane, *The on-line encyclopedia of integer sequences*, published electronically at http://www.research.att.com/ njas/sequences/, 2006.

[11] H. Harborth und S. Jäger, *Konvexe Hüllen von kantenbenachbarten regulären Vielecken*, Math. Semesterber. **38** (1991), 126–134.

Chapter 4

Enumeration of integral tetrahedra

SASCHA KURZ[1]

ABSTRACT. We determine the numbers of integral tetrahedra with diameter d up to isomorphism for all $d \leqslant 1000$ via computer enumeration. Therefore we give an algorithm that enumerates the integral tetrahedra with diameter at most d in $O(d^5)$ time and an algorithm that can check the canonicity of a given integral tetrahedron with at most 6 integer comparisons. For the number of isomorphism classes of integral 4×4 matrices with diameter d fulfilling the triangle inequalities we derive an exact formula.

2000 MSC: 33F05; 05A15.

Key words and phrases: implicit enumeration, integral tetrahedra, geometric probability, Euclidean metric, orderly generation, canonicity check.

1 Introduction

Geometrical objects with integral side lengths have fascinated mathematicians for ages. A very simple geometric object is an m-dimensional simplex. Recently an intriguing bijection between m-dimensional simplices with edge lengths in $\{1,2\}$ and the partitions of $m+1$ was discovered [2]. So far, for m-dimensional simplices with edge lengths in $\{1,2,3\}$ no formula is known and exact numbers are obtained only up to $m = 13$ [9]. Let us more generally denote by $\alpha(m,d)$ the number of non-isomorphic m-dimensional simplices with edge lengths in $\{1,\dots,d\}$ where at least one edge has length d. We also call d the diameter of the simplex. The known results, see e. g. [9], are, besides some exact numbers,

$$\begin{aligned}\alpha(1,d) &= 1,\\ \alpha(2,d) &= \left\lfloor\frac{d+1}{2}\right\rfloor\left\lfloor\frac{d+2}{2}\right\rfloor = \left\lfloor\frac{(d+1)^2}{4}\right\rfloor,\\ \alpha(m,1) &= 1,\\ \alpha(m,2) &= p(m+1)-1,\end{aligned}$$

where $p(m+1)$ denotes the number of partitions of $m+1$. The aim of this article is the determination of the number of non-isomorphic integral tetrahedra $\alpha(3,d)$.

[1]Sascha Kurz, University of Bayreuth, Department of Mathematics, D-95440 Bayreuth, Germany.
E-mail adress: sascha.kurz@uni-bayreuth.de

Besides an intrinsic interest in integral simplices their study is useful in field of integral point sets. These are sets of n points in the m-dimensional Euclidean space $\mathbb{E}^m$ with pairwise integral distances. Applications for this combinatorial structure involving geometry and number theory are imaginable in radio astronomy (wave lengths), chemistry (molecules), physics (energy quantum's), robotics, architecture, and other fields, see [3] for an overview. We define the largest occurring distance of an integral point set $\mathcal{P}$ as its diameter. From the combinatorial point of view there is a natural interest in the determination of the minimum possible diameter $d(m,n)$ for given parameters m and n [3, 4, 5, 7, 9, 10, 11, 12, 14, 16]. In most cases exact values of $d(m,n)$ are obtained by an exhaustive enumeration of integral point sets with diameter $d \leqslant d(m,n)$. A necessary first step for the enumeration of m-dimensional integral point sets is the enumeration of m-dimensional integral simplices. Hence there is a need for an efficient enumeration algorithm.

Another application of integral tetrahedra concerns geometric probabilities. Suppose you are given a symmetric 3×3 matrix Δ_2 with entries being equi-distributed in $[0,1]$ and zeros on the main diagonal. The probability $\mathcal{P}_2$ that Δ_2 is the distance matrix of a triangle in the Euclidean metric can be easily determined to be $\mathcal{P}_2 = \frac{1}{2}$. As a generalization we ask for the probability $\mathcal{P}_m$ of a similar defined $(m+1) \times (m+1)$ matrix Δ_m being the distance matrix of an m-dimensional simplex in the Euclidean metric. To analyze the question for $m = 3$ we consider a discretization and obtain $\mathcal{P}_3 = \lim_{d\to\infty} \frac{4\cdot\alpha(3,d)}{d^5}$.
Our main results are the determination of $\alpha(3,d)$ for $d \leqslant 1000$,

Theorem 1.1 *The number $\hat{\alpha}_{\leqslant}(d,3)$ of symmetric 4×4 matrices with entries in $\{1,\dots,d\}$ fulfilling the triangle inequalities is given by* $\hat{\alpha}_{\leqslant}(d,3) =$

$$\begin{cases} \frac{17d^6+425d^4+1628d^2}{2880} & \text{for } d \equiv 0 \mod 2,\\ \frac{17d^6+425d^4+1763d^2+675}{2880} & \text{for } d \equiv 1 \mod 2.\end{cases}$$

If we additionally request a diameter of exactly d we have $\hat{\alpha}(d,3) =$

$$\begin{cases} \frac{34d^5-85d^4+680d^3-962d^2+1776d-960}{960} & \text{for } d \equiv 0 \mod 2,\\ \frac{34d^5-85d^4+680d^3-908d^2+1722d-483}{960} & \text{for } d \equiv 1 \mod 2,\end{cases}$$

Theorem 1.2

$$0.090 \leqslant \mathcal{P}_3 \leqslant 0.111,$$

and the enumeration algorithms of Section 4 and Section 5, which allow us to enumerate integral tetrahedra with diameter at most d in time $O\left(d^5\right)$ and to check a 4×4-matrix for canonicity using at most 6 integer comparisons.

2 Number of integral tetrahedra

Because a symmetric 4×4-matrix with zeros on the main diagonal has six independent non-zero values there are d^6 labeled integral such matrices with diameter at most d. To obtain the number $\overline{\alpha}_{\leqslant}(d,3)$ of unlabeled matrices we need to apply the following well known Lemma:

Lemma 2.1 *(Cauchy-Frobenius, weighted form)*
Given a group action of a finite group G on a set S and a map $w : S \longrightarrow R$ from S into a commutative ring R containing $\mathbb{Q}$ as a subring. If w is constant on the orbits of G on S, then we have, for any transversal $\mathcal{T}$ of the orbits:

$$\sum_{t \in \mathcal{T}} w(t) = \frac{1}{|G|} \sum_{g \in G} \sum_{s \in S_g} w(s)$$

where S_g denotes the elements of S being fixed by g, i. e.

$$S_g = \{s \in S | s = gs\}.$$

For a proof, notation and some background we refer to [6]. Applying the lemma yields:

Lemma 2.2

$$\overline{\alpha}_{\leqslant}(d,3) = \frac{d^6 + 9d^4 + 14d^2}{24}$$

and

$$\begin{aligned} \overline{\alpha}(d,3) &= \overline{\alpha}_{\leqslant}(d,3) - \overline{\alpha}_{\leqslant}(d-1,3) \\ &= \frac{6d^5 - 15d^4 + 56d^3 - 69d^2 + 70d - 24}{24}. \end{aligned}$$

As geometry is involved in our problem we have to take into account some properties of Euclidean spaces. In the Euclidean plane $\mathbb{E}^2$ the possible occurring triples of distances of triangles are completely characterized by the triangle inequalities. In general there is a set of inequalities using the so called Cayley-Menger determinant to characterize whether a given symmetric $(m+1) \times (m+1)$ matrix with zeros on the main diagonal is a distance matrix of an m-dimensional simplex [13]. For a tetrahedron with distances $\delta_{i,j}$, $0 \leqslant i \leqslant j < 4$, the inequality

$$CMD_3 = \begin{vmatrix} 0 & \delta_{0,1}^2 & \delta_{0,2}^2 & \delta_{0,3}^2 & 1 \\ \delta_{1,0}^2 & 0 & \delta_{1,2}^2 & \delta_{1,3}^2 & 1 \\ \delta_{2,0}^2 & \delta_{2,1}^2 & 0 & \delta_{2,3}^2 & 1 \\ \delta_{3,0}^2 & \delta_{3,1}^2 & \delta_{3,2}^2 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{vmatrix} > 0 \quad (1)$$

has to be fulfilled besides the triangle inequalities.

In a first step we exclusively consider the triangle inequalities for $m = 3$ and count the number $\hat{\alpha}_{\leqslant}(d,3)$ of non-isomorphic symmetric 4×4 matrices with entries in $\{1, \ldots, d\}$ fulfilling the triangle inequalities.

d	$\alpha(d,3)$	d	$\alpha(d,3)$	d	$\alpha(d,3)$
1	1	51	8854161	120	639349793
2	4	52	9756921	140	1382200653
3	16	53	10732329	160	2695280888
4	45	54	11783530	180	4857645442
5	116	55	12916059	200	8227353208
6	254	56	14133630	220	13251404399
7	516	57	15442004	240	20475584436
8	956	58	16845331	260	30554402290
9	1669	59	18349153	280	44260846692
10	2760	60	19957007	300	62496428392
11	4379	61	21678067	320	86300970558
12	6676	62	23514174	340	116862463817
13	9888	63	25473207	360	155526991341
14	14219	64	27560402	380	203808692441
15	19956	65	29783292	400	263399396125
16	27421	66	32145746	420	336178761892
17	37062	67	34657375	440	424224122232
18	49143	68	37322859	460	529820175414
19	64272	69	40149983	480	655468974700
20	82888	70	43145566	500	803900006590
21	105629	71	46318399	520	978079728301
22	133132	72	49673679	540	1181221582297
23	166090	73	53222896	560	1416796092768
24	205223	74	56969822	580	1688540496999
25	251624	75	60926247	600	2000468396580
26	305861	76	65098817	620	2356880503873
27	369247	77	69497725	640	2762373382787
28	442695	78	74130849	660	3221850132593
29	527417	79	79008179	680	3740530243895
30	624483	80	84138170	700	4323958989350
31	735777	81	89532591	720	4978017317882
32	861885	82	95198909	740	5708932993276
33	1005214	83	101149823	760	6523288334629
34	1166797	84	107392867	780	7428031732465
35	1348609	85	113942655	800	8430487428682
36	1552398	86	120807154	820	9538364312059
37	1780198	87	127997826	840	10759766492473
38	2033970	88	135527578	860	12103204603044
39	2315942	89	143409248	880	13577602128303
40	2628138	90	151649489	900	15192308794063
41	2973433	91	160268457	920	16957109053082
42	3353922	92	169272471	940	18882231158104
43	3773027	93	178678811	960	20978358597822
44	4232254	94	188496776	980	23256639532080
45	4735254	95	198743717	1000	25728695195597
46	5285404	96	209427375		
47	5885587	97	220570260		
48	6538543	98	232180129		
49	7249029	99	244275592		
50	8019420	100	256866619		

Table 1: Number $\alpha(d,3)$ of integral tetrahedra with diameter d.

Proof of Theorem 1.1.
Counting labeled symmetric 4×4 matrices with entries in $\{1, \dots, d\}$ fulfilling the triangle inequalities is equivalent to determining integral points in a six-dimensional polyhedron. Prescribing the complete automorphism group results in some further equalities and an application of the inclusion-exclusion principle. Thus, after a lengthy but rather easy computation we can apply Lemma 2.1 and obtain

$$24\hat{\alpha}_{\leqslant}(d,3) = 3 \cdot \left\lceil \frac{4d^4+5d^2}{12} \right\rceil + 8 \cdot \left(d^2 - d\left\lceil \frac{d}{2} \right\rceil + \left\lceil \frac{d}{2} \right\rceil^2 \right) + 6 \cdot \frac{37d^4 - 18d^3 + 20d^2 - 21d + \left(36d^2+42\right)\left\lceil \frac{d}{2} \right\rceil}{96} + \left\lceil \frac{34d^6+55d^4+136d^2}{240} \right\rceil + 6 \cdot \left(d^2 - d\left\lceil \frac{d}{2} \right\rceil + \left\lceil \frac{d}{2} \right\rceil^2 \right),$$

which can be modified to the stated formulas. □

In addition to this proof we have verified the stated formula for $d \leqslant 500$ via a computer enumeration. We remark that $\frac{\hat{\alpha}_{\leqslant}(d,3)}{\overline{\alpha}_{\leqslant}(d,3)}$ and $\frac{\hat{\alpha}(d,3)}{\overline{\alpha}(d,3)}$ tend to $\frac{17}{120} = 0.141\overline{6}$ if $d \to \infty$. Moreover we were able to obtain an exact formula for $\hat{\alpha}(d,3)$ because the Cayley-Menger determinant

$$\mathrm{CMD}_2 = \begin{vmatrix} 0 & \delta_{0,1}^2 & \delta_{0,2}^2 & 1 \\ \delta_{1,0}^2 & 0 & \delta_{1,2}^2 & 1 \\ \delta_{2,0}^2 & \delta_{2,1}^2 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{vmatrix}$$

for dimension $m = 2$ can be written as

$$\begin{aligned} \mathrm{CMD}_2 &= -(\delta_{0,1}+\delta_{0,2}+\delta_{1,2})(\delta_{0,1}+\delta_{0,2}-\delta_{1,2}) \\ &\quad \cdot (\delta_{0,1}-\delta_{0,2}+\delta_{1,2})(-\delta_{0,1}+\delta_{0,2}+\delta_{1,2}). \end{aligned}$$

Thus $\mathrm{CMD}_2 < 0$ is equivalent to the well known linear triangle inequalities $\delta_{0,1} + \delta_{0,2} > \delta_{1,2}$, $\delta_{0,1} + \delta_{1,2} > \delta_{0,2}$ and $\delta_{0,2} + \delta_{1,2} > \delta_{0,1}$. Unfortunately for $m \geqslant 3$ the Cayley-Menger determinant is irreducible [1] and one cannot simplify $(-1)^{m+1}\mathrm{CMD}_m > 0$ into a set of inequalities of lower degree. So we are unable to apply the same method to derive an analytic formula for $\alpha(d,3)$.

Lemma 2.3 *We have* $\alpha(3,d) \in \Theta\left(d^5\right)$ *and* $\alpha_{\leqslant}(3,d) \in \Theta\left(d^6\right)$*, where* $f \in \Theta(g)$ *iff* $f \in O(g)$ *and* $f \in \Omega(g)$.

PROOF. The upper bounds are trivial since they also hold for symmetric matrices with integer values at most d and zeros on the main diagonal. For the lower bounds we consider sixtuples $\delta_{0,1} \in [d, d(1-\varepsilon))$, $\delta_{0,2} \in [d(1-\varepsilon), d(1-2\varepsilon))$, $\delta_{1,2} \in [d(1-2\varepsilon), d(1-3\varepsilon))$, $\delta_{0,3} \in [d(1-3\varepsilon), d(1-4\varepsilon))$, $\delta_{1,3} \in [d(1-4\varepsilon), d(1-5\varepsilon))$, and $\delta_{2,3} \in [d(1-5\varepsilon), d(1-6\varepsilon))$. For each ε there are $\Omega\left(d^6\right)$ non-isomorphic matrices. If ε is suitable small then all these matrices fulfill the triangle conditions and Inequality (1). □

In general we have $\alpha_{\leqslant}(m,d) \in \Theta\left(d^{m(m+1)/2}\right)$ and $\alpha(m,d) \in \Theta\left(d^{m(m+1)/2-1}\right)$.

In Section 4 and Section 5 we give an algorithm to obtain $\alpha(d,3)$ via implicit computer enumeration. Some of these computed values are given in Table 1. For a complete list of $\alpha(d,3)$ for $d \leqslant 1000$ we refer to [8]. This amounts to

$$\alpha_{\leqslant}(1000,3) = 4299974867606266 \approx 4.3 \cdot 10^{15}.$$

3 Bounds for $\mathcal{P}_3$

In this section we give bounds for the probability $\mathcal{P}_3$ that Δ_3 is the distance matrix of a tetrahedron in the 3-dimensional Euclidean space $\mathbb{E}^3$, where Δ_3 is a symmetric 4×4 matrix with zeros on the main diagonal and the remaining entries being equi-distributed in $[0,1]$. Therefore we consider a discretization. Let d be a fixed number. We consider the d^6 six-dimensional cubes $\mathcal{C}_{i_1,\dots,i_6} := \times_{j=1}^{6} \left[\frac{i_j}{d}, \frac{i_j+1}{d}\right] \subseteq [0,1]^6$. For every cube $\mathcal{C}$ it is easy to decide whether every point of $\mathcal{C}$ fulfills the triangle conditions, no points of $\mathcal{C}$ fulfill the triangle conditions, or both cases occur. For Inequality (1) we have no explicit test but we are able to compute a lower bound $\underline{\mathrm{CMD}_3}(\mathcal{C})$ and an upper bound $\overline{\mathrm{CMD}_3}(\mathcal{C})$, so that we have

$$\underline{\mathrm{CMD}_3}(\mathcal{C}) \leqslant \mathrm{CMD}_3(x) \leqslant \overline{\mathrm{CMD}_3}(\mathcal{C}) \text{ for all } x \in \mathcal{C}.$$

Thus for some cubes $\mathcal{C}$ we can decide that all $x \in \mathcal{C}$ correspond to a tetrahedron. We denote this case by $\Xi(\mathcal{C}) = 1$. If no $x \in \mathcal{C}$ corresponds to a tetrahedron we set $\Xi(\mathcal{C}) = -1$. In all other cases we define $\Xi(\mathcal{C}) = 0$. With this we obtain for all $d \in \mathbb{N}$ the following bounds:

Lemma 3.1

$$\sum_{\mathcal{C}:\Xi(\mathcal{C})=1} \frac{1}{d^6} \leqslant \mathcal{P}_3 \leqslant 1 - \sum_{\mathcal{C}:\Xi(\mathcal{C})=-1} \frac{1}{d^6}.$$

Thus we have a method to obtain bounds on $\mathcal{P}_3$ using computer calculations. For the actual computation we use two further speed ups. We can take advantage of symmetries and use an adaptive strategy: We start with a small value of d and subdivide cubes $\mathcal{C}$ with $\Xi(\mathcal{C}) = 0$ recursively into 8 smaller cubes. After a computer calculation we obtain

$$0.090 \leqslant \mathcal{P}_3 \leqslant 0.111,$$

which proves Theorem 1.2. Clearly Theorem 1.2 can be improved by simply letting the computers work for a longer time or by using a computing cluster, but the convergence of our approach seems to be rather slow. An enhanced check whether a cube C fulfills Inequality (1) would be very useful.

Good estimates for $\mathcal{P}_3$ can be obtained by considering the values $\alpha(3,d)$ in the following way. At first we consider the probability $\widetilde{\mathcal{P}}_3$ being defined as $\mathcal{P}_3$ where additionally $\delta_{0,1} = 1$.

Lemma 3.2

$$\widetilde{\mathcal{P}}_3 = \mathcal{P}_3.$$

PROOF. The problem of determining $\mathcal{P}_3$ or $\widetilde{\mathcal{P}}_3$ is an integration problem. Due to symmetry we only need to consider the domain where $\delta_{0,1}$ is the maximum of the six side lengths. For every $\delta_{0,1} \in (0,1]$ there is a probability $p(\delta_{0,1})$ that $\delta_{0,1}, \dots, \delta_{2,3}$ are distances of a tetrahedron where $\delta_{0,2}, \dots, \delta_{2,3} \in (0, \delta_{0,1}]$ are equi-distributed random variables. Since $p(\delta_{0,1})$ is constant we can conclude the stated equation. □

Lemma 3.3

$$\mathcal{P}_3 = \lim_{d\to\infty} \frac{4\cdot\alpha(d,3)}{d^5}.$$

PROOF. We consider a modified version of the algorithm described above to obtain exact bounds on $\widetilde{\mathcal{P}}_3$. As already mentioned, the triangle inequalities alone define a five-dimensional polyhedron. Since determinants are continuous $\mathrm{CMD}_3 = 0$ defines a smooth surface and so the volume of all cubes $\mathcal{C}$ with $\Xi(\mathcal{C}) = 0$ converges to zero. Thus substituting $\Xi(\mathcal{C})$ by the evaluation of Ξ in an arbitrary corner of $\mathcal{C}$ yields the correct value for $\widetilde{\mathcal{P}}_3 = \mathcal{P}_3$ for $d \to \infty$. Since there are at most $O\left(d^4\right)$ six-tuples $(d, i_2, \ldots, i_6)$, $i_j \in \{1, \ldots, d\}$ with non-trivial automorphism group we can factor out symmetry and conclude the stated result. □

Using Lemma 2.2 and Theorem 1.1 we can modify this to

$$\mathcal{P}_3 = \lim_{d\to\infty} \frac{\alpha(d,3)}{\overline{\alpha}(d,3)} \leqslant \lim_{d\to\infty} \frac{\hat{\alpha}(d,3)}{\overline{\alpha}(d,3)} = \frac{17}{120} = 0.141\overline{6}.$$

Heuristically we observe that the finite sequence $\left(\frac{\alpha(d,3)}{\overline{\alpha}(d,3)}\right)_{1\leqslant d\leqslant 1000}$ is strictly decreasing. So the following values might be seen as a good estimate for $\mathcal{P}_3$:

$$\frac{\alpha(600,3)}{\overline{\alpha}(600,3)} = \frac{2000468396580}{19359502966749} \approx 0.103333,$$
$$\frac{\alpha(800,3)}{\overline{\alpha}(800,3)} = \frac{8430487428682}{81665192828999} \approx 0.103232, \text{ and}$$
$$\frac{\alpha(1000,3)}{\overline{\alpha}(1000,3)} = \frac{25728695195597}{249377330461249} \approx 0.103172.$$

4 Orderly generation of integral tetrahedra

Our strategy to enumerate integral tetrahedra is to merge two triangles along a common side. In Figure 1 we have depicted the two possibilities in the plane to join two triangles $(0,1,2)$ and $(0,1,3)$ along the side $\overline{01}$. If we rotate the triangle $(0,1,3)$ in the 3-dimensional space from the position on the left in Figure 1 to the position on the right we obtain tetrahedra and the distance $\delta_{2,3}$ forms an interval $[l,u]$. The restriction to integral tetrahedra is fairly easy.

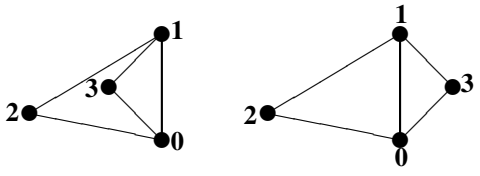


Figure 1: Joining two triangles.

Let us consider the example $\delta_{0,1} = 6$, $\delta_{0,2} = \delta_{1,2} = 5$, $\delta_{0,3} = 4$, and $\delta_{1,3} = 3$. Solving $\mathrm{CMD}_3 = 0$ over the positive real numbers yields that the configuration is a tetrahedron iff $\delta_{2,3} \in \left(\frac{\sqrt{702-24\sqrt{455}}}{6}, \frac{\sqrt{702+24\sqrt{455}}}{6}\right) \approx (2.297719304, 5.806934304)$. Thus there are integral tetrahedra for $\delta_{2,3} \in \{3,4,5\}$. In general we denote such a set of tetrahedra by

$$\delta_{0,1}, \delta_{0,2}, \delta_{1,2}, \delta_{0,3}, \delta_{1,3}, \delta_{2,3} \in [l,r].$$

This notation permits to implicitly list $\Omega\left(d^6\right)$ integral tetrahedra in $O\left(d^5\right)$ time.

All integral tetrahedra can be obtained in this manner. So an enumeration method is to loop over all suitable pairs of integral triangles and to combine them. We will go into detail in a while. Before that we have to face the fact that our enumeration method may construct pairs of isomorphic tetrahedra. Looking at Table 1 we see that storing all along the way constructed non-isomorphic integral tetrahedra in a hash table is infeasible. Here we use the concept of orderly generation [15] which allows us to decide independently for each single constructed discrete structure if we have to take or to reject it. Therefore we have to define a canonical form of an integral tetrahedron. Here we say that a tetrahedron $\mathcal{T}$ with side lengths $\delta_{i,j}$ is canonical if for the lexicographic ordering of vectors $\succeq$,

$$(\delta_{0,1}, \delta_{0,2}, \delta_{1,2}, \delta_{0,3}, \delta_{1,3}, \delta_{2,3}) \succeq (\delta_{\tau(0),\tau(1)}, \ldots, \delta_{\tau(2),\tau(3)})$$

holds for all permutation τ of the points $0,1,2,3$. We describe the algorithmic treatment of a canonicity function $\chi(\mathcal{T}) \mapsto \{\mathtt{true}, \mathtt{false}\}$ which decides whether a given integral tetrahedron $\mathcal{T}$ is canonical in Section 5. We have the following obvious lemma:

Lemma 4.1 *If* $\chi(\delta_{0,1}, \delta_{0,2}, \delta_{1,2}, \delta_{0,3}, \delta_{1,3}, \delta_{2,3}) = \mathtt{true}$ *and* $\chi(\delta_{0,1}, \delta_{0,2}, \delta_{1,2}, \delta_{0,3}, \delta_{1,3}, \delta_{2,3}+1) = \mathtt{false}$ *then* $\chi(\delta_{0,1}, \delta_{0,2}, \delta_{1,2}, \delta_{0,3}, \delta_{1,3}, \delta_{2,3}+n) = \mathtt{false}$ *for all* $n \geqslant 1$. *If* $\chi(\delta_{0,1}, \delta_{0,2}, \delta_{1,2}, \delta_{0,3}, \delta_{1,3}, \delta_{2,3}) = \mathtt{true}$ *and* $\chi(\delta_{0,1}, \delta_{0,2}, \delta_{1,2}, \delta_{0,3}, \delta_{1,3}, \delta_{2,3}-1) = \mathtt{false}$ *then* $\chi(\delta_{0,1}, \delta_{0,2}, \delta_{1,2}, \delta_{0,3}, \delta_{1,3}, \delta_{2,3}-n) = \mathtt{false}$ *for all* $1 \leqslant n \leqslant \delta_{2,3}$.

Thus for given $\delta_{0,1}$, $\delta_{0,2}$, $\delta_{1,2}$, $\delta_{0,3}$, and $\delta_{1,3}$ the possible values for $\delta_{2,3}$ which correspond to a canonical tetrahedron form an interval $\left[\hat{l}, \hat{u}\right]$. Clearly, the value of $\chi(\delta_{0,1}, \delta_{0,2}, \delta_{1,2}, \delta_{0,3}, \delta_{1,3}, \delta_{2,3})$ has to be evaluated for $\delta_{2,3} \in \left\{\delta_{i,j} - 1, \delta_{i,j}, \delta_{i,j} + 1 \mid (i,j) \in \{(0,1),(0,2),(1,2),(0,3),(1,3)\}\right\}$ only. Thus we can determine the interval $\left[\hat{l}, \hat{u}\right]$ using $O(1)$ evaluations of $\chi(\mathcal{T})$.

Algorithm 4.2 Orderly generation of integral tetrahedra
Input: Diameter d
Output: A complete list of canonical integral tetrahedra with diameter d
begin
  $\delta_{0,1} = d$
  for $\delta_{0,2}$ **from** $\left\lfloor \frac{d+2}{2} \right\rfloor$ **to** d **do**
    for $\delta_{1,2}$ **from** $d + 1 - \delta_{0,2}$ **to** $\delta_{0,2}$ **do**
      for $\delta_{0,3}$ **from** $d + 1 - \delta_{0,2}$ **to** $\delta_{0,2}$ **do**

```
            for δ_{1,3} from d + 1 − δ_{0,3} to δ_{0,2} do
               Determine the interval [l̂, û] for δ_{2,3}
               print δ_{0,1}, δ_{0,2}, δ_{1,2}, δ_{0,3}, δ_{1,3}, [l̂, û]
            end
         end
      end
   end
end
```

We leave it as an exercise for the reader to prove the correctness of Algorithm 4.2 (see [15] for the necessary and sufficient conditions of an orderly generation algorithm). Since we will see in the next section that we can perform $\chi(\mathcal{T})$ in $O(1)$ the runtime of Algorithm 4.2 is $O\left(d^4\right)$. By an obvious modification Algorithm 4.2 returns a complete list of all canonical integral tetrahedra with diameter at most d in $O\left(d^5\right)$ time.

We remark that we have implemented Algorithm 4.2 using Algorithm 5.1 for the canonicity check. For the computation of $\alpha(800,3)$ our computer needs only 3.3 hours which is really fast compared to the nearly 3 hours needed for a simple loop from 1 to $\alpha(800,3)$ on the same machine. Due to the complexity of $O\left(d^4\right)$ for suitable large d the determination of $\alpha(d,3)$ will need less time than the simple loop from 1 to $\alpha(d,3)$ on every machine.

5 Canonicity check

In the previous section we have used the canonicity check $\chi(\mathcal{T})$ as a black box. The straight forward approach to implement it as an algorithm is to run over all permutations $\tau \in S_4$ and to check whether $(\delta_{0,1},\ldots,\delta_{2,3}) \succeq \left(\delta_{\tau(0),\tau(1)},\ldots,\delta_{\tau(2),\tau(3)}\right)$. This clearly leads to running time $O(1)$ but is too slow for our purpose. It may be implemented using $24\cdot 6 = 144$ integer comparisons. Here we can use the fact that the tetrahedra are generated by Algorithm 4.2. So if we arrange the comparisons as in Algorithm 5.1 at most 6 integer comparisons suffice.

Algorithm 5.1 Canonicity check for integral tetrahedra generated by Algorithm 4.2
Input: $\delta_{0,1},\delta_{0,2},\delta_{1,2},\delta_{0,3},\delta_{1,3},\delta_{2,3}$
Output: $\chi(\delta_{0,1},\delta_{0,2},\delta_{1,2},\delta_{0,3},\delta_{1,3},\delta_{2,3})$

```
begin
   if δ_{0,1} = δ_{0,2} then
      if δ_{0,2} = δ_{1,2} then
         if δ_{0,3} < δ_{1,3} then return false
         else
            if δ_{1,3} < δ_{2,3} then return false
            else return true end
         end
      else
         if δ_{1,3} < δ_{2,3} then return false
         else
            if δ_{0,1} = δ_{0,3} then
               if δ_{1,2} < δ_{1,3} then return false
               else return true end
            else
               if δ_{0,1} > δ_{1,3} then return true
               else
                  if δ_{1,2} < δ_{0,3} then return false
                  else return true end
               end
            end
         end
      end
   else
      if δ_{0,2} = δ_{1,2} then
         if δ_{0,1} < δ_{2,3} or δ_{0,3} < δ_{1,3} then return false
         else return true end
      else
         if δ_{0,2} = δ_{1,3} then
            if δ_{0,3} > δ_{1,2} or δ_{0,1} < δ_{2,3} then return false
            else return true end
         else
            if δ_{0,2} = δ_{0,3} then
               if δ_{1,2} < δ_{1,3} or δ_{0,1} ⩽ δ_{2,3}
               then return false
               else return true end
            else
               if δ_{0,3} > δ_{1,2} then
                  if δ_{0,1} ⩽ δ_{2,3} then return true
                  else return false end
               else
                  if δ_{0,1} < δ_{2,3} then return true
                  else return false end
               end
            end
         end
      end
   end
end
```

To prove the correctness of Algorithm 5.1 we consider all vectors $(\delta_{0,1},\ldots,\delta_{2,3})$ with $\delta_{i,j} \in \{0,\ldots,5\}$, $\delta_{0,1} \geqslant \delta_{0,2} \geqslant \delta_{1,2}$, $\delta_{0,2} \geqslant \delta_{0,3}$, and $\delta_{0,2} \geqslant \delta_{1,3}$. It suffices to show that Algorithm 5.1 returns the correct value for this finite set of vectors since these inequalities are fulfilled by Algorithm 4.2 and also necessary for $\chi(\mathcal{T}) = \texttt{true}$. Algorithm 5.1 can be considered as a binary decision tree. It might be a task to optimize this type of binary decision tree in the worst or in the average case.

6 Dimensions m $\geqslant$ 4

Clearly the question for bounds for P_m arises also for $m \geqslant 4$. But non-trivial answers seem out of reach by our approach. So far we have no efficient equivalent of Algorithm 5.1 at hand and the number $\alpha(4,d)$ of integral 4-dimensional simplices with diameter d is $\Omega\left(d^9\right)$. We give the known values of $\alpha(4,d)$ in Table 6.

Bibliography

[1] C. d'Andrea and M. Sombra, *The Cayley-Menger determinant is irreducible for* $n \geqslant 3$, Siberian Math.

d	$\alpha(d,4)$	d	$\alpha(d,4)$
1	1	27	4716186332
2	6	28	6541418450
3	56	29	8970194384
4	336	30	12168243592
5	1840	31	16344856064
6	7925	32	21748894367
7	29183	33	28688094208
8	91621	34	37529184064
9	256546	35	48713293955
10	648697	36	62769489452
11	1508107	37	80321260053
12	3267671	38	102108730634
13	6679409	39	128999562925
14	12957976	40	162007000505
15	24015317	41	202323976907
16	42810244	42	251321436143
17	73793984	43	310607982160
18	123240964	44	382002253424
19	200260099	45	467627887530
20	317487746	46	569910996879
21	492199068	47	691631229557
22	747720800	48	835911697430
23	1115115145	49	1006370948735
24	1634875673	50	1207047969441
25	2360312092	51	1442539675756
26	3358519981	52	1718015775541

Table 2: Number $\alpha(d,4)$ of integral 4-dimensional simplices with diameter $1 \leqslant d \leqslant 52$.

J. **46** (2005), 90–97.

[2] C. Haase and S. Kurz, *A bijection between the* d*-dimensional simplices with distances in* $\{1,2\}$ *and the partitions of* d+*1*, J. Combin. Theory Ser. A **113** (2006), no. 4, 736–738.

[3] H. Harborth, *Integral distances in point sets*, Karl der Grosse und sein Nachwirken. 1200 Jahre Kultur und Wissenschaft in Europa. Band 2: Mathematisches Wissen. Turnhout: Brepols (P. L. Butzer et al., eds.), 1998, 213–224.

[4] H. Harborth, A. Kemnitz, and M. Möller, *An upper bound for the minimum diameter of integral point sets*, Discrete Comput. Geom. **9** (1993), no. 4, 427–432.

[5] A. Kemnitz, *Punktmengen mit ganzzahligen Abständen*, Habilitationsschrift, TU Braunschweig, 1988.

[6] A. Kerber, *Applied finite group actions. 2nd, rev. and exp. ed.*, Algorithms and Combinatorics. 19. Berlin: Springer. xxvi, 454 p., 1999.

[7] A. Kohnert and S. Kurz, *A note on Erdös-Diophantine graphs and Diophantine carpets*, Math. Balkanica (N.S.) **21** (2007), no. 1-2, 1–5.

[8] S. Kurz, http://www.wm.uni-bayreuth.de/index.php?id=252.

[9] S. Kurz, *Konstruktion und Eigenschaften ganzzahliger Punktmengen*, Ph.D. thesis, Bayreuth. Math. Schr. 76. Universität Bayreuth, 2006.

[10] S. Kurz, *On the characteristic of integral point sets in* $\mathbb{E}^m$, Australas. J. Combin. **36** (2006), 241–248.

[11] S. Kurz and R. Laue, *Bounds for the minimum diameter of integral point sets*, Australas. J. Combin. **39** (2007), 233–240.

[12] S. Kurz and A. Wassermann, *On the minimum diameter of plane integral point sets*, Ars Combin., (to appear).

[13] K. Menger, *Untersuchungen über allgemeine Metrik*, Math. Ann. **100** (1928), 75–163.

[14] L. Piepmeyer, *Räumliche ganzzahlige Punktmengen*, Master's thesis, TU Braunschweig, 1988.

[15] R. C. Read, *Every one a winner or how to avoid isomorphism search when cataloguing combinatorial configurations*, Ann. Discrete Math. **2** (1978), 107–120.

[16] J. Solymosi, *Note on integral distances*, Discrete Comput. Geom. **30** (2003), no. 2, 337–342.

Chapter 5

Integral point sets over $\mathbb{Z}_n^m$

AXEL KOHNERT[1] AND SASCHA KURZ[2]

ABSTRACT. There are many papers studying properties of point sets in the Euclidean space $\mathbb{E}^m$ or on integer grids $\mathbb{Z}^m$, with pairwise integral or rational distances. In this article we consider the distances or coordinates of the point sets which instead of being integers are elements of $\mathbb{Z}/\mathbb{Z}n$, and study the properties of the resulting combinatorial structures.

2000 MSC: 52C10; 51E99.
Key words and phrases: integral distances, exhaustive search, finite rings, orderly generation.

1 Introduction

There are many papers studying properties of point sets in the Euclidean space $\mathbb{E}^m$, with pairwise integral or rational distances (for short integral point sets or rational point sets, respectively), see [17] for an overview and applications. A recent collection of some classical open problems is given in [6, Section 5.11]. Some authors also require that the points are located on an integer grid $\mathbb{Z}^m$ [11, 31]. In this paper we modify the underlying space and study instead of $\mathbb{Z}$ the integers modulo n, which we denote by $\mathbb{Z}_n$ or $\mathbb{Z}/\mathbb{Z}n$. This was a suggestion of S. Dimiev. Our motivation was to gain some insight for the original problem in $\mathbb{Z}^m$ and $\mathbb{E}^m$. In the next subsection we shortly repeat the basic facts and questions about integral point sets in $\mathbb{Z}^m$ and $\mathbb{E}^m$.

1.1 Integral point sets in $\mathbb{Z}^m$ and $\mathbb{E}^m$

So let us now consider integral point sets in $\mathbb{E}^m$. If we denote the largest distance of an integral point set, consisting of n points, as its diameter, the natural question for the minimum possible diameter $d(n,m)$ arises, see Figure 1 for an example. Obviously we have $d(n,1) = n-1$. To avoid the corresponding trivial 1-dimensional configuration in higher dimensions, it is common to request that an m-dimensional integral point set is not contained in a hyperplane of $\mathbb{E}^m$. We call a set of $m+1$ points in $\mathbb{Z}^m$ or $\mathbb{E}^m$ degenerated, if the points are indeed contained in a hyperplane. There are quite a lot of constructions which show that $d(n,m)$ exists for $n+1 \geqslant m$, see e. g. [18]. Some exact values are determined in [21, 24, 27, 28, 33]. The best known upper bound $d(n,m) \in O\left(e^{c\sqrt{\log(n-m)\log\log(n-m)}}\right)$ is given in [18]. For $m = 2$ Solymosi [36] gives the best known lower bound $d(n,2) \geqslant cn$. For $m = 2$ and $n \geqslant 9$ the shape of the examples with minimum diameter is conjectured to consist of $n-1$ collinear points and one point apart [28], see Figure 1 for an example with $n = 9$. We would like to remark that this conjecture is confirmed for $n \leqslant 122$ by an exhaustive search [28]. If for a fix $\rho > 0$, we have a sequence of plane integral point set $\mathcal{P}_i$, each containing a collinear subset of cardinality least n^ρ, then the diameters of the $\mathcal{P}_i$ are in $\Omega\left(e^{c\sqrt{\log n\log\log n}}\right)$ [24, 28]. For $m \geqslant 3$ we refer to [24, 27], where some bounds and exact numbers are given.

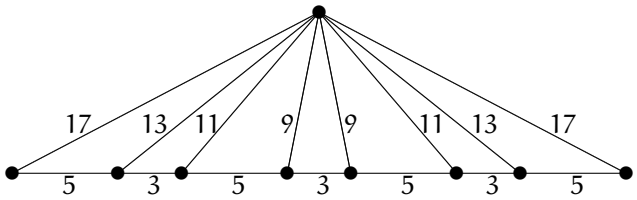


Figure 1: A 2-dimensional integral point set with $n = 9$ and diameter 29.

Some authors require integral point sets to fulfill certain further conditions. The two classical conditions are, that no $m+1$ points are contained in an $(m-1)$-dimensional hyperplane, and that no $m+2$ points are located on an $(m-1)$-dimensional hypersphere. For ease of notation we speak of semi-general position in the first case and of general position if both conditions are fulfilled. We denote the minimum diameter of integral point sets in semi-general position by $\overline{d}(n,m)$ and of integral point sets in general position by $\dot{d}(n,m)$. For some small parameters the exact values have been determined in [21, 23, 24, 28, 33]. We would like to remark that for dimension $m = 2$ and $3 \leqslant n \leqslant 36$ points, the examples with minimum possible diameter $\overline{d}(n,2)$, consist of points on a circle [24, 28].

A famous question of Erdős asks for point sets in the plane with seven points in general position (e. g. no three

[1]Axel Kohnert, University of Bayreuth, Department of Mathematics, D-95440 Bayreuth, Germany.
E-mail adress: axel.kohnert@uni-bayreuth.de
[2]Sascha Kurz, University of Bayreuth, Department of Mathematics, D-95440 Bayreuth, Germany.
E-mail adress: sascha.kurz@uni-bayreuth.de

on a line and no four on a circle) with pairwise integral distances. Actually he first asked for such a set with five points, which was answered by Harborth [15, 16], then for a set with six points, which was answered by Kemnitz [21]. Kemnitz even gives a construction for infinitely many such sets with coprime distances. For a long time no example consisting of seven points was known. Very recently one of the authors has discovered two such examples with diameters 22270 and 66810 [23]. For dimensions $m \geqslant 3$ we refer to [17, 24].

As a specialization, integral point sets in general position, with all n points on an integer grid $\mathbb{Z}^m$, are called n_m-clusters. Noll and Bell have found n_m-clusters for $m \leqslant 5$ and $n \leqslant m+4$ but have no example for $n \geqslant m+5$ [31]. For $m \geqslant 3$ even no integral point set in semi-general position with at least $m+5$ points is known.

Conjecture 1.1 *(Erdős and Noll) For any $m > 1$, $n > 1$, there exists either none or an infinite number of non-isomorphic n_m-clusters.*

An important invariant of an integral point set is its characteristic, which is defined as follows:

Definition 1.2 *Let $\mathcal{S}$ be a non-degenerated integral point set of $m+1$ points in the m-dimensional Euclidean space $\mathbb{E}^m$. By V_m we denote the m-dimensional volume of the simplex being formed by the convex hull of $\mathcal{S}$. Since the pairwise distances of $\mathcal{S}$ are integral and $\mathcal{S}$ is not degenerated we have $(V_m)^2 \in \mathbb{N}\backslash\{0\}$. Thus V_m can be uniquely written as $V_m = q\sqrt{c}$ with $q \in \mathbb{Q}$ and a squarefree integer c. This integer c is called the **characteristic** char$(\mathcal{S})$ of an integral simplex $\mathcal{S}$.*

The following theorem allows us to define the characteristic of an integral point set.

Theorem 1.3 *In an m-dimensional integral point set $\mathcal{P}$ each non-degenerate integral simplex $\mathcal{S}$ has the same characteristic char$(\mathcal{S})$.*

Definition 1.4 *Let $\mathcal{P}$ be an m-dimensional integral point set and $\mathcal{S} \subseteq \mathcal{P}$ be an arbitrary m-dimensional non-degenerate integral sub-simplex of $\mathcal{P}$. The **characteristic** char$(\mathcal{P})$ of $\mathcal{P}$ is given by char$(\mathcal{P})$ = char$(\mathcal{S})$.*

For dimension $m = 2$ Theorem 1.3 can be traced back at least to Kummer [21], for $m \geqslant 3$ we refer to [25]. We would like to remark that if we are in the special case, where also the coordinates of an m-dimensional integral point set $\mathcal{P}$ are integral, every subset $\mathcal{S}$ of $\mathcal{P}$, consisting of $m+1$ points, has an integral volume. In our notation this means, that for an integral point set $\mathcal{P}$ in $\mathbb{Z}^m$ we have char$(\mathcal{P}) = 1$. So all n_m-clusters have characteristic one.

From [13, 25] we know, that if $\mathcal{P}$ is an m-dimensional integral point set in $\mathbb{E}^m$ with characteristic char$(\mathcal{P}) = 1$, then there exists an embedding of $\mathcal{P}$ in $\mathbb{E}^m$ using only rational coordinates. The existence of an embedding using only integral coordinates is an interesting open conjecture of [13].

2 Integral point sets over $\mathbb{Z}_n^m$

In the previous section we have seen, that almost certainly there is a lot of hidden structure in the set of integral point sets which attain the minimum possible diameter and fulfill certain further conditions. Although the problem of integral point sets is a very classical one, not much progress has been achieved towards structure results or tight bounds on the minimum diameter. The idea of this paper is to study similar problems, which might be easier to handle, but may give some insight in the original problem. At first we want to consider the study of integral point sets in $\mathbb{Z}^m$ as our *original problem* and relate it to some problem of point sets in $\mathbb{Z}_n^m$.

So let $\mathcal{P}'$ be an integral point set over $\mathbb{Z}^m$. To relate $\mathcal{P}'$ to a set $\mathcal{P}$ of points in $\mathbb{Z}_n^m$ we consider the canonical mapping $\phi_n : \mathbb{Z} \to \mathbb{Z}_n$, $x \mapsto x+\mathbb{Z}n = \overline{x}$, which maps coordinates in $\mathbb{Z}^m$ to coordinates in $\mathbb{Z}_n^m$. If n is suitably large no two points of $\mathcal{P}'$ will be mapped onto the same point in $\mathcal{P}$. To be able to translate results in $\mathbb{Z}_n^m$ back to $\mathbb{Z}^m$, we define the inverse mapping $\Psi_n : \mathbb{Z}_n \to \{0, \ldots, n-1\}$ by $\Psi(\phi_n(x)) = x$ for $x \in \{0, \ldots, n-1\}$. As an abbreviation we set $\Psi_n(x) = \widehat{x}$ and $\phi_n(x) = \overline{x}$, whenever the value of n is clear from the context. Since points in $\mathcal{P}'$ have integral distances in $\mathbb{Z}^m$ we need a similar definition of integral distances in $\mathbb{Z}_n^m$. The most natural way to define an integral distance over $\mathbb{Z}_n^m$ is:

Definition 2.1 *Two points $(u_1, \ldots, u_m), (v_1, \ldots, v_m) \in \mathbb{Z}_n^m$ are at **integral distance**, if there exists a number $d \in \mathbb{Z}_n$ with*

$$\sum_{i=1}^{m} (u_i - v_i)^2 = d^2.$$

With this definition an integral point set $\mathcal{P}'$ over $\mathbb{Z}^m$ is mapped via ϕ_n onto an integral point set $\mathcal{P}$ over $\mathbb{Z}_n^m$. Since ϕ_n may map some point set $\mathcal{P}'$ over $\mathbb{Z}^m$, which is not contained in a hyperplane of $\mathbb{Z}^m$, onto a point set $\mathcal{P}'$, where all points are contained in a hyperplane of $\mathbb{Z}_n^m$, we do not make any requirements on the distribution of the points in an integral point set over $\mathbb{Z}_n^m$ in the first run. The next definition to translate from $\mathbb{Z}^m$ or $\mathbb{E}^m$ to $\mathbb{Z}_n^m$ is the minimum diameter. In $\mathbb{Z}^m$ and $\mathbb{E}^m$ we need the concept of a minimum diameter to get a finite space, whereas $\mathbb{Z}_n^m$ is finite for itself. So we find it natural to consider the maximum number of integral points.

Definition 2.2 *By $\mathcal{I}(n, m)$ we denote the maximum number of points in $\mathbb{Z}_n^m$ with pairwise integral distances.*

Theorem 2.3 *$\mathcal{I}(n, 1) = n$, $\mathcal{I}(1, m) = 1$, and $\mathcal{I}(2, m) = 2^m$.*

PROOF. Because there are only n^m different elements in $\mathbb{Z}_n^m$ we have the trivial upper bound $\mathcal{I}(n, m) \leqslant n^m$. This upper bound is only attained if $m = 1$ or $n \leqslant 2$, since $\mathbb{Z}_n$ has at least one quadratic non residue for $n \geqslant 3$. □

For $n \geqslant 3$ we so far were not able to derive explicit formulas for $\mathcal{I}(n, m)$ and so we give in Table 1 some values

$m\backslash n$	3	4	5	7	8	9
2	3	8	5	7	16	27
3	4	16	25	8	64	81
4	9	32	25	49	512	324
5	27	128	125	343	2048	$\geqslant 893$
6	33	256	$\geqslant 125$		$\geqslant 15296$	
7	$\geqslant 35$	1024			$\geqslant 81792$	

$m\backslash n$	11	13	16	17
2	11	13	64	17
3	11	169	256	289
4	121	$\geqslant 169$	1024	
5	$\geqslant 1331$	$\geqslant 2197$		

Table 1: Values of $\mathcal{I}(n, m)$ for small parameters n and m.

for small parameters n and m, obtained by exhaustive enumeration via clique search, which we will describe in the next subsection. Further exact values or lower bounds can be determined using Theorem 2.3 and 2.7 of Subsection 2.2.

2.1 Exhaustive enumeration of integral point sets over $\mathbb{Z}_n^m$ via clique search

In this subsection we describe how the exact values $\mathcal{I}(n, m)$ of Table 1 were obtained. We model our problem as a graph $\mathcal{G}$, so that the cliques (e. g. complete subgraphs) of $\mathcal{G}$ are in bijection to integral point sets over $\mathbb{Z}_n^m$. Therefore we choose the elements of $\mathbb{Z}_n^m$ as vertices and connect $x, y \in \mathbb{Z}_n^m$ via an edge, if and only if x and y are at integral distance.

To determine $\mathcal{I}(n, m)$, we only have to determine the maximum cardinality of a clique of $\mathcal{G}$. Unfortunately this is an $\mathcal{NP}$-hard problem in general, but practically this approach was also successful in the case of integral point sets over $\mathbb{E}^m$ [24, 28], due to good heuristic maximum-clique algorithms. Besides an implementation of the Bron-Kerbosch algorithm [7] written by ourself we use the software package CLIQUER [30, 32] of Niskanen and Östergård.

By prescribing points or distances of an integral point set $\mathcal{P}$, it is possible to reduce the complexity for the clique-search algorithm. The first variant is, that due to symmetry we can assume that the point $0 = (\overline{0}, \ldots, \overline{0}) \in \mathbb{Z}_n^m$ is part of $\mathcal{P}$. As vertices of $\mathcal{G}$ we choose the points in $\mathbb{Z}_n^m \backslash \{0\}$, which have an integral distance to 0. Again two vertices $x, y \in \mathcal{G}$ are joined by an edge, if the corresponding points are at integral distance.

For the second variant we consider the set $D_{n,m}$ of all points $d = (d_1, \ldots, d_m) \in \mathbb{Z}_n^m$, which have an integral distance to 0 and which fulfill $\widehat{d}_i \leqslant \lfloor \frac{n}{2} \rfloor$, for all $1 \leqslant i \leqslant m$. So for every two points $u = (u_1, \ldots, u_m) \neq v = (v_1, \ldots, v_d) \in \mathbb{Z}_n^m$, having an integral distance, the tuple

$$\delta_n(u, v) = \Big(\overline{\min\left(|\widehat{u}_1 - \widehat{v}_1|, n - |\widehat{u}_1 - \widehat{v}_1|\right)}, \ldots, \overline{\min\left(|\widehat{u}_m - \widehat{v}_m|, n - |\widehat{u}_m - \widehat{v}_m|\right)} \Big)$$

is an element of $D_{n,m}$. Actually we consider the vector of the Lee weights [34] of the coordinates of the difference $u - v$. Now we choose an arbitrary numbering of this set $D_{n,m}\backslash\{0\} = \{e_0, \ldots, e_{|D_{n,m}|-2}\}$ and consider the graphs $\mathcal{G}_i$, which consist of the points of $\mathbb{Z}_n^m \backslash \{0, e_i\}$, with integral distances to 0 and e_i, as vertices. Two vertices $x \neq y \in \mathcal{G}$ are joined by an edge if and only if the corresponding points fullfill $\delta_n(x, y) = e_j$ with $i \leqslant j$. Again one can show, that an integral point set in $\mathbb{Z}_n^m$ corresponds to a clique in some graph $\mathcal{G}_i$ and vice versa. For some values of n and m it is worth to put some effort in a suitable choice of the numbering of $D_{n,m}\backslash\{0\}$.

2.2 Hamming spaces and homomorphisms

In this subsection we want to relate the problem of integral point sets over $\mathbb{Z}_n^m$ to problems in Hamming spaces. In coding theory the Hamming distance $h(u, v)$ of two vectors $u = (u_1, \ldots, u_m)$, $v = (v_1, \ldots, v_m) \in \mathbb{Z}_n^m$ is the number of positions i where u_i and v_i differ. Normally one is interested in large subsets of $\mathbb{Z}_n^m$ where all the Hamming distances are either 0 or larger than a given constant c. In our subject, we are interested in large subsets of $\mathbb{Z}_n^m$, where all the Hamming distances are taken from a specific proper subset of $\{0, 1, \ldots, m\}$. This point of view has been proven useful e. g. also in the $0/1$-Borsuk problem in low dimensions, see [37]. Here we also want to mention the study of two-weight codes, see e. g. [9, 22].

So let us go back to the determinantion of $\mathcal{I}(n, m)$. As there are trivial formulas for $\mathcal{I}(1, m)$ and $\mathcal{I}(2, m)$, the next open case for fixed ring order n is the determination of $\mathcal{I}(3, m)$. Due to $1^2 \equiv 2^2 \equiv 1 \mod 3$, integral point sets over $\mathbb{Z}_3^m$ correspond to sets of $\mathbb{Z}_3^m$ with Hamming distances $h(u, v) \not\equiv 2 \mod 3$. So this is our first example of a selection problem in a Hamming space.

For the determination of $\mathcal{I}(2n, m)$ we can utilize homomorphisms to make the problem easier. Therefore we need some definitions.

Definition 2.4 *For an integer n we define the mapping $\widetilde{\varphi}_{2n} : \mathbb{Z}_{2n} \to \mathbb{Z}_n$, $x \mapsto \widehat{x} + \mathbb{Z}n$, and by $\varphi_{2n,m}$ we denote its extensions to $\mathbb{Z}_{2n}^m$.*

Definition 2.5 *The weight function $\widetilde{w}_{2n} : \mathbb{Z}_n^2 \to \mathbb{Z}_{2n}$ is defined by $(u_i, v_i) \mapsto (\widehat{u}_i - \widehat{v}_i)^2 + \mathbb{Z} \cdot 2n$.*

$$\mathbb{H}_{2n}^m := \left\{ S \subseteq \mathbb{Z}_n^m \mid \forall s_1, s_2 \in S : \exists d \in \mathbb{Z}_{2n} : d^2 = w(s_1, s_2) \right\},$$

where $w_{2n,m} : (\mathbb{Z}_n^m)^2 \to \mathbb{Z}_{2n}$ is given by $((u_1, \ldots, u_m), (v_1, \ldots, v_m)) \mapsto \sum_{i=1}^{m} \widetilde{w}_{2n}(u_i, v_i)$. By $\mathbb{I}_n^m$ we denote the set of integral point sets in $\mathbb{Z}_n^m$.

Lemma 2.6

$$2^m \mid \mathcal{I}(2n, m).$$

PROOF. We consider the ring homomorphism $\varphi_{2n,m}$ and restrict it to $\varphi'_{2n,m} : \mathbb{I}_{2n}^m \to \mathbb{H}_{2n}^m$. If $\mathcal{P}$ is an element of $\mathbb{H}_{2n}^m$ then the preimage $\varphi_{2n,m}^{-1}(\mathcal{P})$ is an integral point set, due to $(x + n)^2 \equiv x^2 + n \mod 2n$ for odd n and $(x + n)^2 \equiv x^2 \mod 2n$ for even n. For all $x \in \mathbb{Z}_n^m$ we have $\left|\varphi_{2n,m}^{-1}(x)\right| = 2^m$. □

So for the determination of $\mathcal{I}(2n,m)$, it suffices to determine the maximum cardinality of the elements of $\mathbb{H}_{2n}^m$, which actually are subsets of $\mathbb{Z}_n^m$.

$$\mathcal{I}(2n,m) = 2^m \cdot \max_{S \in \mathbb{H}_{2n}^m} |S|$$

As an example we want to apply this result for $n = 2$. Here $w_{4,m}$ is exactly the Hamming distance in $\mathbb{Z}_2^m$. Since the squares of $\mathbb{Z}_4$ are given by $\{0,1\}$, we conclude that $\mathbb{H}_4^m$ is the set of all subsets of $\mathbb{Z}_2^m$, with Hamming distance congruent to 0 or 1 modulo 4. With the mapping $\varphi'_{4,m}$ at hand, we can exhaustively generate the maximal sets in $\mathbb{H}_4^m$, via a clique search, to extend Table 1:

$$\begin{aligned}(\mathcal{I}(4,m))_{m\leqslant 12} &= 4, 8, 16, 32, 128, 256, 1024, 4096,\\ &\quad 16384, 32768, 65536, 131072.\end{aligned}$$

The next theorem shows, that it suffices to determine $\mathcal{I}(a,m)$ for prime powers $a = p^r$.

Theorem 2.7 *For two coprime integers* a *and* b *we have* $\mathcal{I}(a\cdot b,m) = \mathcal{I}(a,m)\cdot\mathcal{I}(b,m)$.

PROOF. Since a and b are coprime we have $\mathbb{Z}_{ab} \simeq \mathbb{Z}_a \times \mathbb{Z}_b$. If $\mathcal{P}$ is an integral point set in $\mathbb{Z}_a \times \mathbb{Z}_b$, then the projections into $\mathbb{Z}_a$ and $\mathbb{Z}_b$ are also integral point sets. If on the other hand, $\mathcal{P}_1$ and $\mathcal{P}_2$ are integral point sets over $\mathbb{Z}_a$ and $\mathbb{Z}_b$, respectively, then $\mathcal{P} := \mathcal{P}_1 \times \mathcal{P}_2$ is an integral point set over $\mathbb{Z}_a \times \mathbb{Z}_b$, due to a straight forward calculation. □

If we drop the condition that a and b are coprime Theorem 2.7 does not remain valid in general. One can see this by looking at the example $\mathcal{I}(2,3)\cdot\mathcal{I}(4,3) > \mathcal{I}(8,3)$ in table 1. Also $\mathcal{I}(a,m) \mid \mathcal{I}(a\cdot b,m)$ does not hold in general, as on can see by a look at the example $\mathcal{I}(3,3) \nmid \mathcal{I}(9,3)$. We would like to mention, that in a recent preprint [26] the exact values of $\mathcal{I}(p,2)$ and $\mathcal{I}\left(p^2,2\right)$ have been determined.

Theorem 2.8 *For a prime* $p \geqslant 3$ *we have*

$$\mathcal{I}(p,2) = p \quad \textit{and} \quad \mathcal{I}\left(p^2,2\right) = p^3.$$

2.3 Integral point sets over the plane $\mathbb{Z}_n^2$

In Theorem 2.3 we have given an exact formula for $\mathcal{I}(n,1)$. So, if we fix the dimension m, the next case is the determination of $\mathcal{I}(n,2)$. At first we give two constructions to obtain lower bounds for $\mathcal{I}(n,2)$.

Lemma 2.9 *If the prime factorization of* n *is given by* $n = \prod_{i=1}^{s} p_i^{r_i}$, *with pairwise different primes* p_i, *we have*

$$\mathcal{I}(n,2) \geqslant n \cdot \prod_{i=1}^{s} p_i^{\lfloor \frac{r_i}{2} \rfloor}.$$

PROOF. We choose the points $\left(u_i, v_j\overline{k}\right)$, where $u_i, v_j \in \mathbb{Z}_n$ and $k = \prod_{i=1}^{s} p_i^{\lceil \frac{r_i}{2} \rceil}$. Due to

$$(u_{i_1} - u_{i_2})^2 + \left(v_{j_1}\overline{k} - v_{j_2}\overline{k}\right)^2 = (u_{i_1} - u_{i_2})^2,$$

all occurring distances are integral. □

An example of the construction of Lemma 2.9 is given in Figure 2, for $n = 12 = 2^2 \cdot 3$.

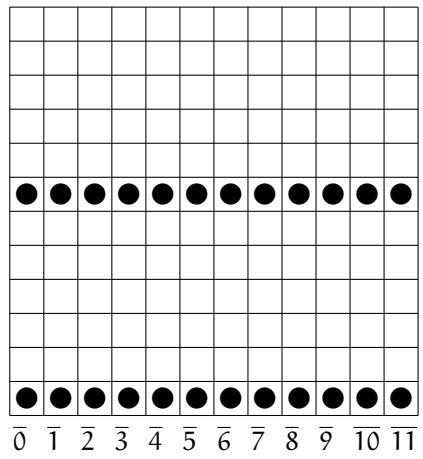


Figure 2: An integral pointset over $\mathbb{Z}_{12}^2$ constructed via Lemma 2.9.

In the case of $n = 2 \mod 4$ we can improve the above lemma:

Lemma 2.10 *If the prime factorization of* n *is given by* $n = 2 \cdot \prod_{i=2}^{s} p_i^{r_i}$, *with pairwise different primes* $p_i \neq 2$ *we have*

$$\mathcal{I}(n,2) \geqslant 2n \cdot \prod_{i=2}^{s} p_i^{\lfloor \frac{r_i}{2} \rfloor}.$$

PROOF. We choose the points $\left(u_i, v_j\overline{k}\right)$, where $u_i, v_j \in \mathbb{Z}_n$ and $k = \prod_{i=2}^{s} p_i^{\lceil \frac{r_i}{2} \rceil}$. Since $2k^2 \equiv 0 \mod n$ and

$$\begin{aligned}&(u_{i_1} - u_{i_2})^2 + \left(v_{j_1}\overline{k} - v_{j_2}\overline{k}\right)^2\\ = \;&(u_{i_1} - u_{i_2})^2 + \left(v_{j_1}^2 + v_{j_2}^2\right)\overline{k^2}\end{aligned}$$

either

$$(u_{i_1} - u_{i_2})^2 + \left(v_{j_1}\overline{k} - v_{j_2}\overline{k}\right)^2 = (u_{i_1} - u_{i_2})^2$$

or

$$(u_{i_1} - u_{i_2})^2 + \left(v_{j_1}\overline{k} - v_{j_2}\overline{k}\right)^2 = \left(u_{i_1} - u_{i_2} + \overline{k^2}\right)^2$$

holds. □

Conjecture 2.11 *For all* $n \in \mathbb{N}$ *either the lower bound of Lemma 2.9 or the lower bound of Lemma 2.10 is tight.*

Remark 2.12 *By Theorem 2.7 and an exhaustive enumeration of integral point sets over* $\mathbb{Z}_n^2$, *via clique search, we have verified Conjecture 2.11 up to* $n = 307$.

If n is squarefree and 2 does not divide n, then our constructions from Lemma 2.9 and Lemma 2.10 yield point sets of the form $\mathcal{P} = \{(u,0) \mid u \in \mathbb{Z}_n\}$. This is somewhat similar to the situation in $\mathbb{E}^2$, where integral collinear point sets with small diameter can consist of many points. Since we also want to speak of collinear point sets in $\mathbb{Z}_n^2$ we give:

Definition 2.13 *A set of* r *points* $(u_i, v_i) \in \mathbb{Z}_n^2$ *is collinear, if there are* $a, b, t_1, t_2, w_i \in \mathbb{Z}_n$ *with*

$$a + w_i t_1 = u_i \quad \text{and} \quad b + w_i t_2 = v_i.$$

Let us first look at collinearity from the algorithmic point of view. Checking three points for being collinear, by running through the possible values of $a, b, t_1, t_2, w_i \in \mathbb{Z}_n$, would cost $\mathcal{O}\left(n^7\right)$ time. Setting, w.l.o.g., $a = u_1$, $b = v_1$, $w_1 = \overline{0}$ reduces this to $\mathcal{O}\left(n^4\right)$. If n is prime, then we are working in a field, and there is an easy and well known way to check, whether three points are collinear, in $\mathcal{O}(1)$ time:

Lemma 2.14 *For a prime* n *the points* $(u_1, v_1), (u_2, v_2), (u_3, v_3) \in \mathbb{Z}_n^2$ *are collinear, if and only if*

$$\begin{vmatrix} u_1 & v_1 & \overline{1} \\ u_2 & v_2 & \overline{1} \\ u_3 & v_3 & \overline{1} \end{vmatrix} = \overline{0}. \tag{1}$$

We remark that in $\mathbb{Z}_8$ the points $(\overline{0}, \overline{0})$, $(\overline{2}, \overline{4})$, $(\overline{4}, \overline{4})$ fulfill Equation (1), but are not collinear with respect to Definition 2.13. So in general Equation (1) is necessary but not sufficient for three points to be collinear. We would like to remark that there exists a fast algorithm, which checks three points in $\mathbb{Z}_n^2$ for being collinear, in $O\left(\frac{\log n}{\log \log n}\right)$ time. We do not go into detail here, since in practice one simply determines for each pair $x, y \in \mathbb{Z}_n^2$, whether the triple $0, x, y$ is collinear or not, in a precalculation.

The study of collinear point sets is motivated by the situation in the case of non-modular point sets. Due to a theorem of Erdős each integral point set in $\mathbb{E}^2$, with infinitely many points, is located on a line [1, 12]. And, as already mentioned in the introduction the, non-collinear integral point sets in $\mathbb{E}^2$ with minimum diameter, are conjectured to consist of $n - 1$ collinear points and one point apart.

In this context we would like to mention a theorem, which was recently proven in [26].

Theorem 2.15 *For* p *being a prime, with* $p \equiv 3 \mod 4$*, each integral point set over* $\mathbb{Z}_p^2$*, consisting of* p *points, is collinear.*

For primes p, of the form $p \equiv 1 \mod 4$, also a different type of integral point sets occurs. To describe these sets, we need some new notation. For a prime $p \equiv 1 \mod 4$, there is a unique element $\omega(p) \in \mathbb{N}$, with $\omega(p) < \frac{p}{2}$ and $\omega^2(p) \equiv -1 \mod p$. By $\square_n = \left\{i^2 \mid i \in \mathbb{Z}_n\right\}$ we denote the set of squares in $\mathbb{Z}_n$.

Lemma 2.16 *For a prime* $p \geqslant 3$*, the set* $\mathcal{P} = (1, \pm\omega(p)) \cdot \square_p$ *is a non-collinear integral point set over* $\mathbb{Z}_p^2$ *with cardinality* p.

PROOF. For an odd prime p we have exactly $\frac{p+1}{2}$ squares in $\mathbb{Z}_p$. Since $(0, 0)$, $(1, \omega(p))$, and $(1, -\omega(p))$ are elements of $\mathcal{P}$, the point set is clearly non-collinear. For the property of pairwise integral distances we consider two arbitrary elements $q, q' \in \square_p$ and the corresponding distances

$$\begin{aligned} (q - q')^2 + \omega^2(p)\,(q - q')^2 &= \overline{0}, \\ (q - q')^2 + \omega^2(p)\,(q + q')^2 &= (2\omega(p))^2\, qq', \\ (q + q')^2 + \omega^2(p)\,(q - q')^2 &= 2^2 qq', \\ (q + q')^2 + \omega^2(p)\,(q + q')^2 &= \overline{0}. \end{aligned}$$

□

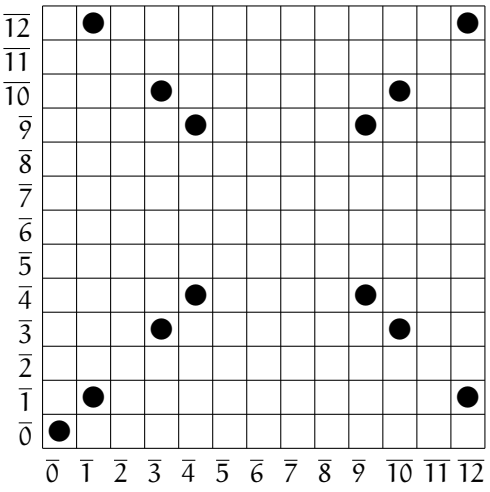


Figure 3: The integral point set $\mathcal{P} = (1, \omega(p)) \cdot \square_p$ for $p = 13$.

In Figure 3 we have depicted an integral point set, being constructed as described in Lemma 2.16 for $p = 13$. We remark that recently in [26] it was proven, that integral point sets $\mathcal{P}$ over $\mathbb{Z}_p^2$, with cardinality $p \geqslant 3$, are either collinear or a translated version of the integral point set constructed in Lemma 2.16.

2.4 Integral point sets over $\mathbb{Z}_n^2$ with further conditions

In the last subsection we have recognized, that integral point sets over $\mathbb{Z}_n^2$ are, similar to integral point sets over $\mathbb{E}^2$, somewhat attracted by collinear sets. So we investigate in this subsection integral point sets $\mathcal{P}$ over $\mathbb{Z}_n^2$, where no three points are collinear.

Definition 2.17 *By* $\overline{\jmath}(n, m)$ *we denote the maximum number of points in semi-general position over* $\mathbb{Z}_n^m$*, where all pairwise distances are integral.*

If we drop the condition of pairwise integral distances, our studied objects become very familiar discrete structures. In the case of affine finite geometries (classical [19] in the case of $\mathbb{Z}_n$ with n a prime, Hjelmslev geometries [8] in the other cases) point sets in semi-general position, with arbitrary pairwise distances, are called arcs in the case of planes or caps [3] in the three dimensional case. With the results

from Subsection 2.2 in mind, we would like to mention the connection of these objects to linear coding theory, see e. g. [4] for the details.
In Table 2 we give some values of $\overline{\mathcal{I}}(n,2)$ for small n, obtained by Algorithm 2.22 described later on.

n	$\overline{\mathcal{I}}(n,2)$	n	$\overline{\mathcal{I}}(n,2)$	n	$\overline{\mathcal{I}}(n,2)$
1	1	21	4	41	20
2	4	22	8	42	6
3	2	23	12	43	22
4	4	24	6	44	10
5	4	25	10	45	11
6	4	26	10	46	14
7	4	27	10	47	24
8	6	28	8	48	8
9	6	29	14	49	$\geqslant 18$
10	6	30	6	50	$\geqslant 17$
11	6	31	16	51	8
12	4	32	14	52	12
13	6	33	6	53	26
14	6	34	10	54	$\geqslant 13$
15	4	35	6	55	8
16	8	36	12	56	10
17	8	37	18	57	10
18	10	38	12	58	$\geqslant 16$
19	10	39	6	59	30
20	8	40	10	60	8

Table 2: Values of $\overline{\mathcal{I}}(n,2)$ for small parameters n.

Now we want to derive an upper bound for $\overline{\mathcal{I}}(n,2)$, by relaxing the condition of pairwise integral distances. Let $\mathcal{P}$ be a point set over $\mathbb{Z}_n^2$ in semi-general position. We consider the lines $\{(i,j) \mid j \in \mathbb{Z}_n\}$ for $i \in \mathbb{Z}_n$. Since these n lines form a partition of $\mathbb{Z}_n^2$ and each line can contain at most two points of $\mathcal{P}$, we obtain the trivial upper bound $\overline{\mathcal{I}}(n,2) \leqslant 2n$. This is connected to a famous open problem in number theory [14, sec. F4], where people work on an upper bound for the *no-three-in-a-line* problem. Considering all lines in $\mathbb{Z}_n^2$ we receive

$$\overline{\mathcal{I}}(p,2) \leqslant p+1$$

for odd primes p [5] and

$$\overline{\mathcal{I}}(n,2) \leqslant n \cdot \left(1 + p^{-\lceil \frac{a+1}{2} \rceil} + p^{-a}\right)$$

where $p^a \mid n$ and $p^{a+1} \nmid n$ for a prime p [20].

Very recently for the case of odd primes p, tight bounds on $\overline{\mathcal{I}}(p,2)$ are proven [26]:

Theorem 2.18 *For* $p \equiv 3 \mod 4$ *we have*

$$\overline{\mathcal{I}}(2,p) = \frac{p+1}{2}$$

and for $p \equiv 1 \mod 4$ *we have*

$$\frac{p-1}{2} \leqslant \overline{\mathcal{I}}(2,p) \leqslant \frac{p+3}{2}.$$

We would like to remark that the known construction uses half of the points of the circle $\left\{(a,b) \in \mathbb{Z}_p^2 \mid a^2+b^2 = \overline{1}\right\}$, see [26] for the details. For $p \equiv 1 \mod 4$, $p \neq 5$ we conjecture $\overline{\mathcal{I}}(p,2) = \frac{p-1}{2}$.
By a look at the situation in $\mathbb{E}^2$ and with the famous question of Erdős in mind. it seems interesting to investigate integral point sets over $\mathbb{Z}_n^2$, where no three points are collinear and no four points are situated on a circle.

Definition 2.19 *Four points* $p_i = (x_i, y_i)$ *in* $\mathbb{Z}_n^2$ *are said to be situated on a circle if there exist* $a, b \in \mathbb{Z}_n$, $r \in \mathbb{Z}_n \setminus \{\overline{0}\}$ *with*

$$(x_i - a)^2 + (y_i - b)^2 = r^2$$

for all i.

We have the following necessary condition:

Lemma 2.20 *Four points* $p_i = (x_i, y_i)$ *in* $\mathbb{Z}_n^2$ *being situated on a circle fulfill*

$$\begin{vmatrix} x_1^2+y_1^2 & x_1 & y_1 & \overline{1} \\ x_2^2+y_2^2 & x_2 & y_2 & \overline{1} \\ x_3^2+y_3^2 & x_3 & y_3 & \overline{1} \\ x_4^2+y_4^2 & x_4 & y_4 & \overline{1} \end{vmatrix} = \overline{0}. \qquad (2)$$

Definition 2.21 *By* $\dot{\mathcal{I}}(n,m)$ *we denote the maximum number of points in* $\mathbb{Z}_n^m$ *with pairwise integral distances, where no three points are collinear and no four points are situated on a circle. Here we also talk of general position.*

n	$\dot{\mathcal{I}}(n,2)$	n	$\dot{\mathcal{I}}(n,2)$	n	$\dot{\mathcal{I}}(n,2)$	n	$\dot{\mathcal{I}}(n,2)$
1	1	21	4	41	9	61	$\geqslant 9$
2	4	22	8	42	6	62	$\geqslant 11$
3	2	23	5	43	8	63	8
4	4	24	4	44	8	64	$\geqslant 10$
5	4	25	6	45	8	65	7
6	4	26	8	46	10	66	8
7	3	27	7	47	7	67	$\geqslant 9$
8	4	28	6	48	8	68	$\geqslant 10$
9	4	29	7	49	$\geqslant 11$	69	7
10	6	30	6	50	$\geqslant 12$	70	$\geqslant 9$
11	4	31	6	51	7		
12	4	32	8	52	$\geqslant 9$		
13	5	33	4	53	$\geqslant 9$		
14	6	34	10	54	$\geqslant 11$		
15	4	35	5	55	6		
16	6	36	$\geqslant 10$	56	6		
17	5	37	7	57	6		
18	8	38	8	58	$\geqslant 11$		
19	5	39	6	59	$\geqslant 9$		
20	6	40	6	60	8		

Table 3: Values of $\dot{\mathcal{I}}(n,2)$ for small parameters n.

Trivially we have $\dot{\mathcal{I}}(n,2) \leqslant \overline{\mathcal{I}}(n,2)$. In Table 3 we give some exact values of $\dot{\mathcal{I}}(n,2)$, obtained by Algorithm 2.22

described later on. One might conjecture that $\dot{\jmath}(n,2)$ is unbounded.

Because semi-general position or general position is a property of three or four points, respectively, we cannot apply our approach via clique search for the determination of $\overline{\jmath}(n,2)$ and $\dot{\jmath}(n,2)$ directly. Instead of going over to hypergraphs we use a variant of orderly generation [35], which glues two integral point sets consisting of r points, having $r-1$ points in common, to obtain recursively integral point sets of $r+1$ points. The used variant of orderly generation was introduced, and applied for the determination of the minimum distance $\dot{d}(n,2)$ of integral point sets in general position in $\mathbb{E}^2$, in [24, 28].

Now we go into detail. To describe integral point sets over $\mathbb{Z}_n^2$, we utilize the set $D_{n,2}$, where the coordinates of the points are *reduced* with respect to the Lee weight via

$$\delta_n((x_1,y_1),(x_2,y_2)) = \Big(\min(|\hat{x}_1-\hat{x}_2|, n-|\hat{x}_1-\hat{x}_2|), \min(|\hat{y}_1-\hat{y}_2|, n-|\hat{y}_1-\hat{y}_2|)\Big).$$

By $\mathcal{B} = \{b_0, b_1, \ldots, b_t\}$ we denote the subset of $D_{n,2} = \left\{\delta_n(0,x) \mid x \in \mathbb{Z}_n^2\right\}$, where the points x are at integral distance to 0. We define $b_0 = \left(\overline{0},\overline{0}\right)$. The numbering of the remaining b_i is arbitrary but fix. Each integral point set $\mathcal{P} = \{p_1, \ldots, p_r\}$ over $\mathbb{Z}_n^2$ is, up to translations and reflections, completely described by a matrix

$$\Delta_n(\mathcal{P}) = \Big(\iota(\delta_n(p_i,p_j))\Big)_{i,j},$$

where we set $\delta_n(p_i,p_i) = b_0$ and $\iota : \mathcal{B} \to \mathbb{N}$, $b_i \mapsto i$. We use these matrices as a data structure for integral point sets over $\mathbb{Z}_n^2$. Next we extend the natural order $\leqslant$ on $\mathbb{N}$ to $\preceq$ for symmetric matrices, with zeros on the main diagonal as Δ_n, by using a column-lexicographical order of the upper right matrix. A matrix Δ_n is said to be *canonical* if $\Delta_n \geqslant \pi(\Delta_n)$ for every permutation $\pi \in S_r$ acting on the rows and columns of Δ_n. If $\downarrow\Delta_n$ denotes the removal of the last column and last row of a matrix Δ_n, then Δ_n is said to be *semi-canonical* if $\downarrow\Delta_n \geqslant \downarrow\pi(\Delta_n)$ for every permutation $\pi \in S_r$. The function Γ_r does the glueing of two integral point sets over $\mathbb{Z}_n^2$ consisting of r points having $r-1$ points in common. The result of the function Γ_r is an, with respect to $\preceq$, ordered list of integral point sets consisting of $r+1$ points. By $\mathcal{L}_r$ we denote the ordered list of all semi-canonical matrices Δ_n, with respect to $\preceq$, which correspond to integral point sets over $\mathbb{Z}_n^2$. It can be figured out easily that Γ_r produces a list with at most two integral point sets. With these definitions we can state:

Algorithm 2.22
Input: $\mathcal{L}_r$
Output: $\mathcal{L}_{r+1}$
begin
  $\mathcal{L}_{r+1} = \emptyset$
  loop over $x_1 \in \mathcal{L}_r$, x_1 *is canonical* **do**
    loop over $x_2 \in \mathcal{L}_r$, $x_2 \preceq x_1$, $\downarrow x_1 = \downarrow x_2$ **do**
      loop over $y \in \Gamma_r(x_1,x_2)$
        if y *is semi-canonical* **then** *add* y *to* $\mathcal{L}_{r+1}$ **end**
      end
    end
  end
end

A starting list $\mathcal{L}_3$ of the integral triangles can be generated by a nested loop. In order to apply Algorithm 2.22 for the determination of $\overline{\jmath}(n,2)$ or $\dot{\jmath}(n,2)$, we only have to modify it in that way, that it only accepts integral point sets in semi-general or general position, respectively, for the lists $\mathcal{L}_r$.

3 Integral point sets over $(\mathbb{R}/\mathbb{Z}n)^2$

In the previous section we have required also the coordinates of the point sets to be *integral*. This corresponds somewhat to integral point sets in $\mathbb{Z}^m$. In this section we try to develop a setting for an analogous treatment of integral point sets in $\mathbb{E}^m$ over the ring $\mathbb{Z}_n$ instead of $\mathbb{Z}$ for the distances. We start with $n = p$ being an odd prime.

Let p be an odd prime, then $\mathbb{Z}_p$ is a finite field. Given three elements $a,b,c \in \mathbb{Z}_p \setminus \left\{\overline{0}\right\}$, which we consider as edge lengths of a triangle. Then we can determine a coordinate represention, given by three points (x_1,y_1), (x_2,y_2), (x_3,y_3) in $(\mathbb{R}/\mathbb{Z}p)^2$, as follows. Due to translations, rotations and reflections we can assume $(x_1,y_1) = \left(\overline{0},\overline{0}\right)$ and $(x_2,y_2) = \left(a,\overline{0}\right)$. For the third point (x_3,y_3) we get the system of equations

$$\begin{aligned} x_3^2 + y_3^2 &= b^2, \\ (x_3-a)^2 + y_3^2 &= c^2. \end{aligned}$$

Solving this system yields

$$\begin{aligned} x_3 &= \frac{b^2-c^2+a^2}{2a}, \\ y_3^2 &= \frac{(a+b+c)(a+b-c)(a-b+c)(-a+b+c)}{(2a)^2}, \end{aligned}$$

which is defined in $\mathbb{Z}_p$ because of $2a \neq \overline{0}$. By $\alpha(p)$ we denote the smallest quadratic non-residue in $\mathbb{Z}_p$. With the above system of equations it can be seen that $x_3 \in \mathbb{Z}_p$ and y_3 is either also in $\mathbb{Z}_p$ or in $\mathbb{Z}_p \cdot \sqrt{\alpha(p)}$. Since this is similar to the case in $\mathbb{E}^m$, see [24, 25], we define the characteristic of an integral triangle similarly.

Definition 3.1 *For an odd prime* p *the characteristic of three side lengths* $a,b,c \in \mathbb{Z}_p$ *with* $V^2 = (a+b+c)(a+b-c)(a-b+c)(-a+b+c) \neq \overline{0}$ *is defined as* $\overline{1}$ *if* V^2 *is a quadratic residue in* $\mathbb{Z}_p$ *and as* $\alpha(p)$ *otherwise.*

For the ease of notation we associate $\mathbb{E}_p^m$ with $(\mathbb{R}/\mathbb{Z}p)^m$. We remark that the three points are collinear exactly if V^2 equals $\overline{0}$. So, similarly to the case in $\mathbb{E}^2$ [29], we have the following lemma, where the determinant equals V^2, if we associate $a = \delta(v_1,v_2)$, $b = \delta(v_1,v_3)$, and $c = \delta(v_2,v_3)$.

Lemma 3.2 *Points $v_1, v_2, v_3 \in \mathbb{E}_p^2$ are collinear if and only if their Euclidean distances $\delta(v_i, v_j)$ fulfill*

$$\begin{vmatrix} \delta^2(v_1,v_1) & \delta^2(v_1,v_2) & \delta^2(v_1,v_3) & \overline{1} \\ \delta^2(v_2,v_1) & \delta^2(v_2,v_2) & \delta^2(v_2,v_3) & \overline{1} \\ \delta^2(v_3,v_1) & \delta^2(v_3,v_2) & \delta^2(v_3,v_3) & \overline{1} \\ \overline{1} & \overline{1} & \overline{1} & \overline{0} \end{vmatrix} = \overline{0}.$$

Our definition of the characteristic of an integral triangle in $\mathbb{Z}_p$ is properly chosen in the sense that we have the following theorem.

Theorem 3.3 *In an integral point set over $\mathbb{E}_p^2$ where p is an odd prime the characteristic of each non-degenerated triangle is equal.*

PROOF. Without loss of generality we assume that the two triangles have two points in common and the points are given by the coordinates $(\overline{0},\overline{0})$, $(\overline{0},a)$, $(x, y\sqrt{c})$, $\left(x', y'\sqrt{c'}\right)$, where a, x, x', y, y' are elements of $\mathbb{Z}_p$ and c, c' are the characteristics. The squared distance of the last two points is given by

$$\begin{aligned} & (x-x')^2 + \left(y\sqrt{c} - y'\sqrt{c'}\right)^2 \\ = \;& (x-x')^2 + y^2c - 2yy'\sqrt{cc'} + y'^2c'. \end{aligned}$$

Because this number must be an element of $\mathbb{Z}_p$ we have that cc' is a quadratic residue in $\mathbb{Z}_p$ yielding $c = c'$. □

As we have proceeded completely analogous to the case in $\mathbb{E}^m$ we can generalize Definition 3.1 and Theorem 3.3.

Definition 3.4 *For an odd prime p the characteristic of an integral point set with $m+1$ points in $\mathbb{E}_p^m$ given by its distances $\delta_{i,j}$ is 1 if V_m^2 is a quadratic residue in $\mathbb{Z}_p$ and $\alpha(p)$ otherwise, where*

$$V_m^2 = \begin{vmatrix} \delta_{1,1}^2 & \cdots & \delta_{1,m+1}^2 & \overline{1} \\ \vdots & \ddots & \ddots & \vdots \\ \delta_{m+1,1}^2 & \cdots & \delta_{m+1,m+1}^2 & \overline{1} \\ \overline{1} & \cdots & \overline{1} & \overline{0} \end{vmatrix}.$$

Theorem 3.5 *In an integral pointset over $\mathbb{E}_p^m$ where p is an odd prime the characteristic of each non-degenerated simplex is the same.*

PROOF. We do the corresponding calculations as in [25] over $\mathbb{Z}_p$ instead of $\mathbb{Q}$. □

For completeness we give a necessary coordinatefree criterion for $m+2$ points being situated on an m-dimensional sphere.

Lemma 3.6 *If $m+2$ points in $\mathbb{E}_n^m$ described by their distances $\delta_{i,j}$ are situated on an m-dimensional sphere then*

$$\begin{vmatrix} \delta_{1,1}^2 & \cdots & \delta_{1,m+1}^2 \\ \vdots & \ddots & \vdots \\ \delta_{m+1,1}^2 & \cdots & \delta_{m+1,m+1}^2 \end{vmatrix} = \overline{0}.$$

So far we have transferred the theory of integral point sets in $\mathbb{E}^m$ to integral point sets over $\mathbb{E}_p^m$ for odd primes p. For general n instead of p there are some twists if we use coordinates. The most natural approach to settle these would be, with respect to the situation in $\mathbb{E}^m$, to leave out coordinates and use Mengers characterization of embedable distance matrices [29] and replace the conditions over $\mathbb{Z}$ by conditions over $\mathbb{Z}_n$.

Definition 3.7 *An integral point set $\mathcal{P}$ over $\mathbb{E}_n^m$ is a set of $r \geqslant m+1$ points with distances $\delta_{i,j} \in \mathbb{Z}_n \setminus \{\overline{0}\}$ for $1 \leqslant i \neq j \leqslant r$ which fulfill*

$$V_{t-1}^2(\{i_1,\ldots,i_t\}) = \begin{vmatrix} \delta_{i_1,i_1}^2 & \cdots & \delta_{i_1,i_t}^2 & \overline{1} \\ \vdots & \ddots & \ddots & \vdots \\ \delta_{i_t,i_1}^2 & \cdots & \delta_{i_t,i_t}^2 & \overline{1} \\ \overline{1} & \cdots & \overline{1} & \overline{0} \end{vmatrix} = \overline{0}$$

for each subset of points $\{i_1,\ldots,i_t\}$ of cardinality $t = m+2$ and $t = m+3$, and there exists a subset $\{\widetilde{i_1},\ldots,\widetilde{i_t}\}$ of cardinality $t = m+1$ with $V_{t-1}^2(\{\widetilde{i_1},\ldots,\widetilde{i_t}\}) \neq \overline{0}$.

To model the extra conditions we could define that $\mathcal{P}$ is in semi-general position if for every $m+1$ points $\{i_1,\ldots,i_{m+1}\}$ we have $V_{m+1}^2(\{i_1,\ldots,i_{m+1}\}) \neq \overline{0}$ and that $\mathcal{P}$ is in general position if the condition of Lemma 3.6 is fulfilled. We remark that for $m = 2$ the determinant of Lemma 3.6 can be factorized to

$$\begin{aligned} & -(\delta_{1,2}\delta_{3,4} + \delta_{1,3}\delta_{2,4} + \delta_{1,4}\delta_{2,3}) \\ & \cdot(\delta_{1,2}\delta_{3,4} + \delta_{1,3}\delta_{2,4} - \delta_{1,4}\delta_{2,3}) \\ & \cdot(\delta_{1,2}\delta_{3,4} - \delta_{1,3}\delta_{2,4} + \delta_{1,4}\delta_{2,3}) \\ & \cdot(-\delta_{1,2}\delta_{3,4} + \delta_{1,3}\delta_{2,4} + \delta_{1,4}\delta_{2,3}). \end{aligned}$$

For $m = 2$ we also have

$$\begin{aligned} V_2^2(\{1,2,3\}) \;=\; & (\delta_{1,2}+\delta_{1,3}+\delta_{2,3})(\delta_{1,2}+\delta_{1,3}-\delta_{2,3}) \cdot \\ & (\delta_{1,2}-\delta_{1,3}+\delta_{2,3})(-\delta_{1,2}+\delta_{1,3}+\delta_{2,3}). \end{aligned}$$

So one may leave out the first factor and request that one of the remaining factors equals $\overline{0}$ instead of the condition in Definition 3.7 and the condition in Lemma 3.6, respectively. For $m \geqslant 3$ the two corresponding determinants are irreducible [10].

Another way to generalize integral point sets is to consider the edge lengths and coordinates as elements in a finite field $\mathbb{F}_{p^k}$ or a commutative ring $\mathcal{R}$ instead of $\mathbb{F}_p = \mathbb{Z}_p$. For some results we refer to [2, 26]. Here we only give a very general definition of an integral point set over an commutative ring $\mathcal{R}$:

Definition 3.8 *For a commutative ring $\mathcal{R}$ a set $\mathcal{P}$ of n points in $\mathcal{R}^m$ is called an integral point set if for each $(x_1,\ldots,x_m), (y_1,\ldots,y_m) \in \mathcal{R}^m$ there exists an element $d \in \mathcal{R}$ fulfilling*

$$\sum_{i=1}^{m} (x_i - y_i)^2 = d^2.$$

4 Conclusion

We have generalized the theory of integral point sets over $\mathbb{Z}^m$ to integral point sets over $\mathbb{Z}_n^m$. Some exact values $\mathcal{I}(n,m)$ of the maximal cardinality of a set with pairwise integral distances in $\mathbb{Z}_n^m$ with or without further conditions on the position are given together with algorithms to determine them.

There are two connections to coding theory, first via the special case of arcs and caps, secondly by the observation that $\mathcal{I}(n,m)$ leads to a class of codes where the Hamming distances of the codewords have to fulfill certain modular restrictions.

For odd primes p the theory of integral point sets in $\mathbb{E}^m$ is transferred to a theory of integral point sets over $\mathbb{E}_p^m$ including the fundamental theorem about the characteristic of an integral simplex.

There are some open questions left and the given results motivate for further research on integral point sets over $\mathbb{Z}_n^m$ and $\mathbb{E}_n^m$, as they seem to be interesting combinatorial structures.

Bibliography

[1] N. H. Anning and P. Erdős, *Integral distances*, Bull. Amer. Math. Soc. **51** (1945), 598–600.

[2] A. Antonov and M. Brancheva, *Algorithm for finding maximal Diophantine figures*, Spring Conference 2007 of the Union of Bulgarian Mathematicians, 2007.

[3] J. Bierbrauer, *Large caps*, J. Geom. **76** (2003), no. 1-2, 16–51.

[4] J. Bierbrauer, *Introduction to coding theory*, Discrete Mathematics and its Applications. Boca Raton, FL: Chapman & Hall/CRC. xxiii, 390 p., 2005.

[5] J. Bierbrauer and Y. Edel, *Bounds on affine caps*, J. Combin. Des. **10** (2002), no. 2, 115–115.

[6] P. Brass, W. Moser, and J. Pach, *Research problems in discrete geometry*, New York, NY: Springer, 2005.

[7] C. Bron and J. Kerbosch, *Finding all cliques of an undirected graph*, Commun. ACM **16** (1973), 575–577.

[8] F. (ed.) Buekenhout, *Handbook of incidence geometry: buildings and foundations*, Amsterdam: North-Holland. xi, 1420 p., 1995.

[9] R. Calderbank and W. M. Kantor, *The geometry of two-weight codes*, Bull. Lond. Math. Soc. **18** (1986), 97–122.

[10] C. d'Andrea and M. Sombra, *The Cayley-Menger determinant is irreducible for* $n \geqslant 3$, Siberian Math. J. **46** (2005), 90–97.

[11] S. Dimiev, *A setting for a Diophantine distance geometry*, Tensor (N.S.) **66** (2005), no. 3, 275–283. MR MR2189847

[12] P. Erdős, *Integral distances*, Bull. Amer. Math. Soc. **51** (1945), 996.

[13] J. Fricke, *On Heron simplices and integer embeding*, preprint (2002).

[14] R. K. Guy, *Unsolved problems in number theory. 3rd ed.*, Problem Books in Mathematics. New York, NY: Springer-Verlag., 2004.

[15] H. Harborth, *On the problem of P. Erdös concerning points with integral distances*, Ann. New York Acad. Sci. **175** (1970), 206–207.

[16] H. Harborth, *Antwort auf eine Frage von P. Erdös nach fünf Punkten mit ganzzahligen Abständen. (Answer to a question of P. Erdös for five points with integer distances)*, Elem. Math. **26** (1971), 112–113.

[17] H. Harborth, *Integral distances in point sets*, Butzer, P. L. (ed.) et al., Karl der Grosse und sein Nachwirken. 1200 Jahre Kultur und Wissenschaft in Europa. Band 2: Mathematisches Wissen. Turnhout: Brepols, 1998, 213–224.

[18] H. Harborth, A. Kemnitz, and M. Möller, *An upper bound for the minimum diameter of integral point sets*, Discrete Comput. Geom. **9** (1993), no. 4, 427–432.

[19] J. W. P. Hirschfeld, *Projective geometries over finite fields. 2nd ed.*, Oxford Mathematical Monographs. Oxford: Clarendon Press, 1998.

[20] J. Huizenga, *The maximum size of caps in* $(\mathbb{Z}/n\mathbb{Z})^2$, preprint (2005).

[21] A. Kemnitz, *Punktmengen mit ganzzahligen Abständen*, Habilitationsschrift, TU Braunschweig, 1988.

[22] A. Kohnert, *Constructing two-weight codes with prescribed groups of automorphisms*, Discrete Appl. Math. **155** (2007), no. 11, 1451–1457.

[23] T. Kreisel and S. Kurz, *There are integral heptagons, no three points on a line, no four on a circle*, Discrete Comput. Geom. (to appear).

[24] S. Kurz, *Konstruktion und Eigenschaften ganzzahliger Punktmengen*, Ph.D. thesis, Bayreuth. Math. Schr. 76. Universität Bayreuth, 2006.

[25] S. Kurz, *On the characteristic of integral point sets in* $\mathbb{E}^m$, Australas. J. Combin. **36** (2006), 241–248.

[26] S. Kurz, *Integral point sets over finite fields*, (submitted).

[27] S. Kurz and R. Laue, *Bounds for the minimum diameter of integral point sets*, Australas. J. Combin. **39** (2007), 233–240.

[28] S. Kurz and A. Wassermann, *On the minimum diameter of plane integral point sets*, Ars Combin. (to appear).

[29] K. Menger, *Untersuchungen über allgemeine Metrik*, Math. Ann. **100** (1928), 75–163.

[30] S. Niskanen and P. R. J. Östergård, *Cliquer user's guide, version 1.0*, Tech. Report T48, Communications Laboratory, Helsinki University of Technology, Espoo, Finland, 2003.

[31] L. C. Noll and D. I. Bell, n*-clusters for* $1 < n < 7$, Math. Comp. **53** (1989), no. 187, 439–444.

[32] P. R. J. Östergård, *A fast algorithm for the maximum clique problem*, Discrete Appl. Math. **120** (2002), no. 1-3, 197–207.

[33] L. Piepmeyer, *Räumliche ganzzahlige Punktmengen*, Master's thesis, TU Braunschweig, 1988.

[34] V. S. (ed.) Pless and W. C. (ed.) Huffman, *Handbook of coding theory. Vol. 1. Part 1: Algebraic coding. Vol. 2. Part 2: Connections, Part 3: Applications*, Amsterdam: Elsevier. 2169 p., 1998.

[35] R. C. Read, *Every one a winner or how to avoid isomorphism search when cataloguing combinatorial configurations*, Ann. Discrete Math. **2** (1978), 107–120.

[36] J. Solymosi, *Note on integral distances*, Discrete Comput. Geom. **30** (2003), no. 2, 337–342.

[37] G. M. Ziegler, *Coloring Hamming graphs, optimal binary codes, and the 0/1-Borsuk problem in low dimensions*, Alt, Helmut (ed.), Computational discrete mathematics. Advanced lectures. Berlin: Springer. Lect. Notes Comput. Sci. 2122, 159-171 (2001), 2001.

Chapter 6

There are integral heptagons, no three points on a line, no four on a circle

TOBIAS KREISEL[1] AND SASCHA KURZ[2]

ABSTRACT. We give two configurations of seven points in the plane, no three points on a line, no four points on a circle with pairwise integral distances. This answers a famous question of Paul Erdős.

2000 MSC: 52C10; 52C35, 52-04, 52A99, 51K99.
Key words and phrases: integral distances, exhaustive search, orderly generation, solution to an Erdős problem.

1 Introduction

A famous open problem of P. Erdős asks for seven points in the plane, no three on a line, no four on a circle with pairwise rational or integral distances [1, 3]. For six points parameter solutions for infinite families of such point sets are known, see e. g. [6]. Since for finite point sets we can multiply the occurring distances with their denominators' smallest common multiple we confine ourselves to considering integral distances only. From the combinatorial point of view the question for the smallest possible diameter $\dot{d}(2,n)$ of n points arises, where the diameter is the largest occurring distance in a point set. So far

$$\left(\dot{d}(2,n)\right)_{n=3,\dots,6} = 1, 8, 73, 174$$

are known [4]. By exhaustive search the bound $\dot{d}(2,7) \geqslant 20000$ could be determined [10, 11]. Up to diameter 20000 there are only few integral point sets consisting of 6 points, no three on a line, no four on a circle with pairwise integral distances, see [8] for a complete list. Some attempts to show that no integral point set in general position consisting of more than six points can exist are known [5], but the suggested proofs turned out to be incorrect. So there was little hope to discover such a point set. But then by a suggestion of S. Dimiev [2] we considered integral point sets over $\mathbb{Z}_n^2$ [7].

Definition 1.1 *Two points* $(u_1,\dots,u_m),(v_1,\dots,v_m) \in \mathbb{Z}_n^m := (\mathbb{Z}\backslash\mathbb{Z}n)^m$ *are at* ***integral distance*** *if there exists a number* $d \in \mathbb{Z}_n$ *with* $\sum_{i=1}^{m}(u_i-v_i)^2 = d^2$.

So, an integral point set in $\mathbb{Z}_n^2$ is defined as a subset of $\mathbb{Z}_n^2$ where all pairs of points are at integral distance. To have an analogue to the "no three on a line and no four on a circle" restriction we need two further definitions.

Definition 1.2 *A set of* r *points* $(u_i,v_i) \in \mathbb{Z}_n^2$ *is collinear if there are* $a,b,t_1,t_2,w_i \in \mathbb{Z}_n$ *with* $a+w_it_1 = u_i$ *and* $b+w_it_2 = v_i$.

Definition 1.3 *Four points* $p_i = (x_i,y_i)$ *in* $\mathbb{Z}_n^2$ *are said to be situated on a circle if there exist* $a,b \in \mathbb{Z}_n$, $r \in \mathbb{Z}_n\backslash\{\overline{0}\}$ *with* $(x_i-a)^2+(y_i-b)^2 = r^2\ \forall i$.

By $\dot{j}(n,2)$ we denote the maximum number of points in $\mathbb{Z}_n^2$ with pairwise integral distances where no three are collinear and no four points are situated on a circle. By combinatorial search techniques—see [7] for the details—we found two point sets proving $\dot{j}(50,2) \geqslant 12$ and $\dot{j}(61,2) \geqslant 9$. Surely this does not imply the existence of an integral point set over the real plane in general position, i. e. no three points on a line, no four points on a circle, however it did give us a fresh impetus to continue our search.

2 Integral heptagons in general position

The results for the "relaxed" problem over $\mathbb{Z}_n^2$ motivated us to maintain our approach of exhaustive generation of all plane integral point sets in general position up to a given diameter by a variant of orderly generation, see [10, 11] for details. Also, without changing our approach but simply by harnessing more computational power we were lucky

[1]Tobias Kreisel, Fakultät für Mathematik, Physik und Informatik, Universität Bayreuth, Germany.
E-mail adress: tobias.kreisel@uni-bayreuth.de
[2]Sascha Kurz, Fakultät für Mathematik, Physik und Informatik, Universität Bayreuth, Germany.
E-mail adress: sascha.kurz@uni-bayreuth.de

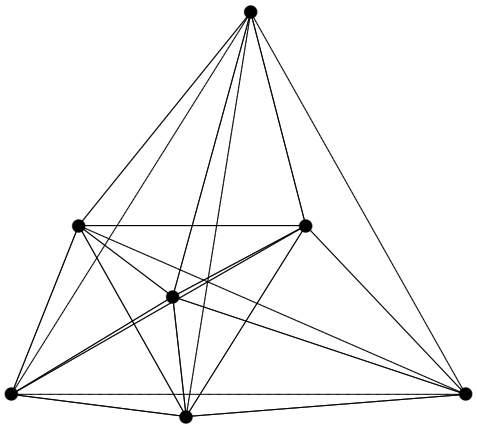
enough to discover the following distance matrix

$$\begin{pmatrix} 0 & 22270 & 22098 & 16637 & 9248 & 8908 & 8636 \\ 22270 & 0 & 21488 & 11397 & 15138 & 20698 & 13746 \\ 22098 & 21488 & 0 & 10795 & 14450 & 13430 & 20066 \\ 16637 & 11397 & 10795 & 0 & 7395 & 11135 & 11049 \\ 9248 & 15138 & 14450 & 7395 & 0 & 5780 & 5916 \\ 8908 & 20698 & 13430 & 11135 & 5780 & 0 & 10744 \\ 8636 & 13746 & 20066 & 11049 & 5916 & 10744 & 0 \end{pmatrix} \quad (1)$$

corresponding to a plane integral point set in general position with diameter 22270 consisting of seven points. So this answers Erdős's question positively. Since we applied an exhaustive search we receive:

Theorem 2.1 $\dot{d}(2,7) = 22270$.

To avoid duplicated listings of isomorphic point sets we give all point sets in the following canonical form. Consider the vector $v(\Delta)$ formed by the columns of the upper right triangle of a distance matrix Δ. A certain distance matrix Δ of a point set $\mathcal{P}$ (induced by a labeling of the points) is said to be canonical or maximal if its vector $v(\Delta)$ is the largest one in the set of all vectors of distance matrices of $\mathcal{P}$ with respect to the lexicographic order.



$$\left(\frac{0}{1}, \frac{0}{1}\sqrt{2002}\right)$$
$$\left(\frac{22270}{1}, \frac{0}{1}\sqrt{2002}\right)$$
$$\left(\frac{26127018}{2227}, \frac{932064}{2227}\sqrt{2002}\right)$$
$$\left(\frac{245363}{17}, \frac{3144}{17}\sqrt{2002}\right)$$
$$\left(\frac{17615968}{2227}, \frac{238464}{2227}\sqrt{2002}\right)$$
$$\left(\frac{56068}{17}, \frac{3144}{17}\sqrt{2002}\right)$$
$$\left(\frac{19079044}{2227}, -\frac{54168}{2227}\sqrt{2002}\right)$$

Figure 1: First example of an integral heptagon in general position.

In Figure 1 we give an embedding of distance matrix (1) in the plane and an exact coordinate representation. Discovering this point set clearly motivates to search for further examples to get ideas how to construct an infinite family of examples. Unfortunately this point set is the only example with at most 30000 in diameter. For diameters greater than 30000 our approach of exhaustive search requires too much computational power so that we decided to skip to a restricted search. To describe the details of our restriction of the search space we need:

Definition 2.2 *The **characteristic** of an integral triangle with side lengths $a, b, c \in \mathbb{Z}$ is the square free part of $(a+b+c)(a+b-c)(a-b+c)(-a+b+c)$.*

Theorem 2.3 *Each non degenerated triangle in a plane integral point set has equal characteristic.*

In point set (1) the characteristic is given by $2002 = 2 \cdot 7 \cdot 11 \cdot 13$ which explains the shape of the y-coordinates, see Figure 1 and [9]. We notice that the characteristic of point set (1) is composed of relatively small prime factors. By a look at our list of integral hexagons in general position [8] we see that this seems to be a phenomenon that holds for a great part of the known examples. This phenomenon seems to hold for similar problems also. By determing the minimum diameter $d(2,n)$ of plane integral point sets without further restrictions up to $n = 122$ points [11] we could check that the known minimal examples also have a characteristic composed of small prime factors. If additionally no three points are allowed to be collinear we denote the corresponding minimum diameter by $\overline{d}(n,2)$. By determing all those minimal integral point sets with up to $n = 36$ points [10, 11] we could check that the same phenomenon also occurs in this case. So it seemed worth a try to exhaustively construct all plane integral point sets in general position with given diameter of at most 70000 and the characteristic being a divisor of $6469693230 = 2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 \cdot 13 \cdot 17 \cdot 19 \cdot 23 \cdot 29$. The outcome was yet another example:

$$\begin{pmatrix} 0 & 66810 & 66555 & 66294 & 49928 & 41238 & 40290 \\ 66810 & 0 & 32385 & 64464 & 32258 & 25908 & 52020 \\ 66555 & 32385 & 0 & 34191 & 16637 & 33147 & 33405 \\ 66294 & 64464 & 34191 & 0 & 34322 & 53244 & 26724 \\ 49928 & 32258 & 16637 & 34322 & 0 & 20066 & 20698 \\ 41238 & 25908 & 33147 & 53244 & 20066 & 0 & 32232 \\ 40290 & 52020 & 33405 & 26724 & 20698 & 32232 & 0 \end{pmatrix} \quad (2)$$

Unfortunately the discovery of further examples is currently beyond our means since the algorithm we use is of running time $\Omega\left(d^3\right)$ for the search for plane integral point sets in general position with diameter at most d. Though the restriction on the characteristic did accelerate computations significantly the theoretic lower bound for the complexity remains. (There are $O\left(d^3\right)$ integral triangles with diameter at most d.)

3 Open problems

Clearly, one can ask for further examples or an infinite family of integral heptagons in general position. Since our two given examples are in non convex position it would be interesting to see a convex example. As a further restriction Bell and Noll [12] also required the coordinates of the point sets to be integral. Such point sets are commonly called $\mathfrak{n}_\mathfrak{m}$-clusters, where $\mathfrak{n}$ is the number of points and $\mathfrak{m}$ the dimension. In general the set of $\mathfrak{n}_2$-cluster equals the set of plane integral point sets in general position with characteristic 1. So far no 7_2-cluster is known and even its existence is unclear. The smallest 6_2-cluster has diameter 1886. At first sight it seems that we have answered Erdős question completely, but from a realistic point of view we have only pushed the frontier a step further. Originally P. Erdős asked for five points in the plain, no three on a line, no for on a circle with pairwise integral distances. When such a set was found he asked for 6-set then for a seven set. So now we ask as a substitute:

„Are there eight points in the plane, no three on a line, no four on a circle with pairwise integral distances?“

Bibliography

[1] P. Brass, W. Moser, and J. Pach, *Research problems in discrete geometry*, Springer, 2005.

[2] S. Dimiev, *A setting for a Diophantine distance geometry*, Tensor (N.S.) **66** (2005), no. 3, 275–283. MR MR2189847

[3] R. K. Guy, *Unsolved problems in number theory. 2nd ed.*, Springer, 1994.

[4] H. Harborth, *Integral distances in point sets*, Butzer, P. L. (ed.) et al., Karl der Grosse und sein Nachwirken. 1200 Jahre Kultur und Wissenschaft in Europa. Band 2: Mathematisches Wissen. Turnhout: Brepols, 1998, 213–224.

[5] H. Harborth, 2005, personal communication.

[6] A. Kemnitz, *Punktmengen mit ganzzahligen Abständen*, Habilitationsschrift, TU Braunschweig, 1988.

[7] A. Kohnert and S. Kurz, *Integral point sets over* $\mathbb{Z}_\mathfrak{n}^\mathfrak{m}$, Discrete Appl. Math., (to appear).

[8] T. Kreisel and S. Kurz, *List of integral hexagons in general position*, 2006, http://www.wm.uni-bayreuth.de/index.php?id=erdoes.

[9] S. Kurz, *On the characteristic of integral point sets in* $\mathbb{E}^\mathfrak{m}$, Australas. J. Combin. **36** (2006), 241–248.

[10] S. Kurz, *Konstruktion und Eigenschaften ganzzahliger Punktmengen*, Ph.D. thesis, Bayreuth. Math. Schr. 76, Universität Bayreuth, 2006.

[11] S. Kurz and A. Wassermann, *On the minimum diameter of plane integral point sets*, Ars Combin., (to appear).

[12] L. C. Noll and D. I. Bell, $\mathfrak{n}$*-clusters for* $1 < \mathfrak{n} < 7$, Math. Comp. **53** (1989), no. 187, 439–444.

Chapter 7

Integral point sets over finite fields

SASCHA KURZ[1]

ABSTRACT. We consider point sets in the affine plane $\mathbb{F}_q^2$ where each Euclidean distance of two points is an element of $\mathbb{F}_q$. These sets are called integral point sets and were originally defined in m-dimensional Euclidean spaces $\mathbb{E}^m$. We determine their maximal cardinality $\mathcal{I}(\mathbb{F}_q, 2)$. For arbitrary commutative rings $\mathcal{R}$ instead of $\mathbb{F}_q$ or for further restrictions as no three points on a line or no four points on a circle we give partial results. Additionally we study the geometric structure of the examples with maximum cardinality.

2000 MSC: 51E20; 05B25.
Key words and phrases: finite geometry, point configurations, integral point sets, universal geometry.

1 Introduction

Originally integral point sets were defined in m-dimensional Euclidean spaces $\mathbb{E}^m$ as a set of n points with pairwise integral distances in the Euclidean metric, see [10, 14, 16, 17] for a overview on the most recent results. Here we transfer the concept of an integral point set to modules $\mathcal{R}^m$ of a commutative ring with 1. We equip those spaces with a squared distance

$$d^2(u,v) := \sum_{i=1}^{m} (u_i - v_i)^2 \quad \in \mathcal{R}.$$

for any two points $u = (u_1, \ldots, u_m)$, $v = (v_1, \ldots, v_m)$ in $\mathcal{R}^m$ and say that they are at integral distance if $d^2(u,v)$ is contained in the set $\square_{\mathcal{R}} := \{r^2 \mid r \in \mathcal{R}\}$ consisting of the squares in $\mathcal{R}$. A set of points $\mathcal{P}$ is called an integral point set if every pair of points is at integral distance.

The concept of integral point sets over finite fields is not brand-new. There are some recent papers and preprints [27, 28, 29, 30] by L. A. Vinh dealing with quadrance graphs. These are in the authors definition point sets in the affine plane $\mathbb{F}_q^2$ where the squared distances, there called quadrances, are elements of $\square_{\mathbb{F}_q} \setminus \{0\}$. So for $q \equiv 3 \mod 4$ quadrance graphs coincide with integral point sets over $\mathbb{F}_q^2$. For $q \equiv 1 \mod 4$ we have the small difference that $0 = 0^2$ is not considered as an integral distance. So e. g. the points $(0,0)$ and $(2,3)$ in $\mathbb{F}_{13}$ are not considered to be at an integral distance since $d^2((0,0),(2,3)) = 2^2 + 3^2 = 0$. We would like to mention that quadrance graphs and so integral point sets over finite fields are isomorphic to strongly regular graphs and that there are some connections to other branches of Combinatorics including Ramsey theory and association schemes [23, 24, 31]. The origin of quadrance graphs lies in the more general concept of rational trigonometry and universal geometry by N. J. Wildberger, see [32] for more background.

Some related results on integral point sets over commutative rings can be found in [1, 8, 13].

A somewhat older topic of the literature is also strongly connected to integral point sets over finite fields. The Paley graph $\mathcal{PG}_q$ has the elements of the finite field $\mathbb{F}_q$ as its vertices. Two vertices u and v are connected via an edge if and only if their difference is a non-zero square in $\mathbb{F}_q$. For $q = q'^2$ with $q' \equiv 3 \mod 4$ we have a coincidence between the Paley graph $\mathcal{PG}_q$ and integral point sets over $\mathcal{PG}_{q'}^2$ or quadrance graphs. It is somewhat interesting that these one-dimensional and two-dimensional geometrical objects are so strongly connected. See e. g. [2, 28] for a detailed description and proof of this connection. Actually one uses the natural embedding of $\mathbb{F}_{q^2}$ in $\mathbb{F}_q^2$.

So what are the interesting questions about integral point sets over finite fields? From the combinatorial point of view one could ask for the maximum cardinality $\mathcal{I}(\mathcal{R}, m)$ of those point sets in $\mathcal{R}^m$. For $\mathcal{R} = \mathbb{F}_q$ with $q \equiv 3 \mod 4$ and $m = 2$ this is a classical question about maximum cliques of Paley graphs of square order, where the complete answer is given in [3]. See also [26] for some generalizations. A geometer might ask for the geometric structure of the maximal examples. Clearly the case where $\mathcal{R}$ is a finite field $\mathbb{F}_q$ is the most interesting one.

1.1 Our contribution

For primes p we completely classify maximal integral point sets in the affine planes $\mathbb{F}_p^2$ and for prime powers $q = p^r$ we give partial results. Since in an integral point set not all directions can occur we can apply some Rédei-type results in this context. Although these results are not at hand in general we can derive some results for arbitrary rings $\mathcal{R}$ and

[1]Sascha Kurz, University of Bayreuth, Department of Mathematics, D-95440 Bayreuth, Germany.
E-mail adress: sascha.kurz@uni-bayreuth.de

special cases like $\mathcal{R} = \mathbb{Z}_{p^2}$ or rings with characteristic two.

It will turn out that most maximal examples or constructions in the plane consist of only very few lines. So it is interesting to consider the case where we forbid three points to be collinear. This means that we look at 2-arcs with the additional integrality condition. Here we denote the maximal cardinality by $\bar{\mathcal{I}}(\mathcal{R}, m)$ where we in general forbid that $m+1$ points are contained in a hyperplane. We give a construction and a conjecture for the case $\mathcal{R} = \mathbb{F}_q$, $2 \nmid q$, and $m = 2$ using point sets on circles.

Being even more restrictive we also forbid $m+2$ points to be situated on a hypersphere and denote the corresponding maximal cardinality by $\dot{\mathcal{I}}(\mathcal{R}, m)$. Although in this case we have almost no theoretical insight so far, this is the most interesting situation when we look from the viewpoint of integral point sets in $\mathbb{E}^m$. As a motivation for further research the following open problem of P. Erdős and C. Noll [20] may serve:

> Are there seven points in the plane, no three on a line, no four on a circle with integral coordinates and pairwise integral distances?

If we drop the condition of integral coordinates the problem was recently solved in [14]. As a connection to our problem one may use the ring homomorphism $\mathbb{Z}^m \to \mathbb{Z}_n^m$, $x \mapsto x + (n\mathbb{Z})^m$, which preserves integral distances and coordinates. For lines and circles the situation is a bit more complicated. We give some examples for various primes p showing $\dot{\mathcal{I}}(\mathbb{Z}_p, 2) \geqslant 7$ and determine some exact numbers. Perhaps in the future an application of the Chinese remainder theorem helps to construct the desired example in $\mathbb{Z}^2$.

1.2 Organization of the paper

The paper is arranged as follows. In Section 2 we give the basic definitions and facts on integral point sets over commutative rings $\mathcal{R}$. In Section 3 we determine the automorphism group of the affine plane $\mathbb{F}_q^2$ with respect to Δ. For $q \equiv 3 \mod 4$ it is the well known automorphism group of the Paley graph $\mathcal{PG}_{q^2}$ which is isomorphic to a subgroup of $\mathrm{PG\Gamma}(1, q^2)$ of index 2, see e. g. [6, 12, 25]. For $q \equiv 1 \mod 4$ the automorphism group was not known. We give a proof for both cases and prove some lemmas on integral point sets over finite fields which will be useful in the following sections. Most of the automorphisms also exist in some sense for arbitrary commutative rings $\mathcal{R}$. In Section 4 we determine the maximum cardinality $\mathcal{I}(\mathbb{F}_q, 2)$ of an integral point set over $\mathbb{F}_q^2$ and classify the maximal examples up to isomorphism in some cases. Here we use a result of Blokhuis et al. on point sets with a restricted number of directions. In Section 5 we give some results on $\mathcal{I}(\mathbb{Z}_n, 2)$ and give some constructions which reach this upper bound. In Section 6 we determine the maximum cardinality $\bar{\mathcal{I}}(\mathbb{F}_q, 2)$ of integral point sets over $\mathbb{F}_q$ where no three points are collinear for $q \equiv 3 \mod 4$. For $q \equiv 1 \mod 4$ we give lower and upper bounds which are only two apart. In Section 7 we consider the maximum cardinality $\dot{\mathcal{I}}(\mathbb{F}_q, 2)$ of integral point sets over $\mathbb{F}_q^2$ where no three points are collinear and no four points are situated on a circle. We determine some exact values via an exhaustive combinatorial search and list some maximum examples.

2 Integral point sets

If not stated otherwise we assume that $\mathcal{R}$ is a commutative ring with 1 and consider sets of elements of the $\mathcal{R}$-module $\mathcal{R}^m$. We speak of these elements as points with a geometric interpretation in mind. For our purpose we equip the module $\mathcal{R}^m$ with something similar to an Euclidean metric:

Definition 2.1 *For two points $u = (u_1, \ldots, u_m)$, $v = (v_1, \ldots, v_m)$ in $\mathcal{R}^m$ we define the* ***squared distance*** *as*

$$d^2(u, v) := \sum_{i=1}^{m} (u_i - v_i)^2 \quad \in \mathcal{R}.$$

We are interested in those cases where $d^2(u, v)$ is contained in the set $\square_{\mathcal{R}} := \{r^2 \mid r \in \mathcal{R}\}$ of squares of $\mathcal{R}$.

Definition 2.2 *Two points $u = (u_1, \ldots, u_m)$, $v = (v_1, \ldots, v_m)$ in $\mathcal{R}^m$ are at* ***integral distance*** *if there exists an element r in $\mathcal{R}$ with $d^2(u, v) = r^2$. As a shorthand we define $\Delta : \mathcal{R}^m \times \mathcal{R}^m \to \{0, 1\}$,*

$$(u, v) \mapsto \begin{cases} 1 & \text{if } u \text{ and } v \text{ are at integral distance,} \\ 0 & \text{otherwise.} \end{cases}$$

A set $\mathcal{P}$ of points in $\mathcal{R}^m$ is called an ***integral point set*** *if all pairs of points are at integral distance.*

If $\mathcal{R}$ is a finite ring it makes sense to ask for the maximum cardinality of an integral point set in $\mathcal{R}^m$.

Definition 2.3 *By $\mathcal{I}(\mathcal{R}, m)$ we denote the maximum cardinality of an integral point set in $\mathcal{R}^m$.*

Lemma 2.4

$$|\mathcal{R}| \leqslant \mathcal{I}(\mathcal{R}, m) \leqslant |\mathcal{R}|^m.$$

PROOF. For the lower bound we consider the *line* $\mathcal{P} = \{(r, 0, \ldots, 0) \mid r \in \mathcal{R}\}$. □

Lemma 2.5 *If $\mathcal{R}$ has characteristic 2, meaning that $1 + 1 = 0$ holds, then we have $\mathcal{I}(\mathcal{R}, m) = |\mathcal{R}|^m$.*

PROOF. For two points $u = (u_1, \ldots, u_m)$, $v = (v_1, \ldots, v_m)$ in $\mathcal{R}^m$ we have

$$d^2(u, v) = \sum_{i=1}^{m} (u_i - v_i)^2 = \underbrace{\left(\sum_{i=1}^{r} u_i + v_i \right)^2}_{\in \mathcal{R}}.$$

□

So in the remaining part of this article we consider only rings with characteristic not equal to two. If a ring $\mathcal{R}$ is the Cartesian product of two rings $\mathcal{R}_1, \mathcal{R}_2$, where we define the operations componentwise, then we have the following theorem:

Theorem 2.6

$$\mathcal{I}(\mathcal{R}_1 \times \mathcal{R}_2, m) = \mathcal{I}(\mathcal{R}_1, m) \cdot \mathcal{I}(\mathcal{R}_2, m).$$

PROOF. If $\mathcal{P}$ is an integral point set in $\mathcal{R}_1 \times \mathcal{R}_2$ then the projections into $\mathcal{R}_1$ and $\mathcal{R}_2$ are also integral point sets. If on the other hand $\mathcal{P}_1$ and $\mathcal{P}_2$ are integral point sets over $\mathcal{R}_1$ and $\mathcal{R}_2$, respectively, then $\mathcal{P} := \mathcal{P}_1 \times \mathcal{P}_2$ is an integral point set over $\mathcal{R}_1 \times \mathcal{R}_2$. □

Lemma 2.7 *If* N *is an additive subgroup of* $\{n \in \mathcal{R} \mid n^2 = 0\}$ *or* $\{n \in \mathcal{R} \mid 2n^2 = 0 \wedge n^2 = n^4\}$ *then we have for* $m \geqslant 2$

$$|N|^{m-1} \cdot |\mathcal{R}| \leqslant \mathcal{I}(\mathcal{R}, m) \leqslant |\mathcal{R}|^m.$$

PROOF. We can take the integral point set $\mathcal{P} = \{(r, n_1, \ldots, n_{m-1}) \mid r \in \mathcal{R},\ n_i \in N\}$ and have $r^2 + \sum_{i=1}^{m-1} n_i^2 = r^2$ or $r^2 + \sum_{i=1}^{m-1} n_i^2 = \left(r + \sum_{i=1}^{m-1} n_i^2\right)^2$. □

If we specialize these general results to rings of the from $\mathcal{R} = \mathbb{Z}/\mathbb{Z}n =: \mathbb{Z}_n$ then we have the following corollaries:

Corollary 2.8

$$\mathcal{I}(\mathbb{Z}_n, 1) = n \text{ and } \mathcal{I}(\mathbb{Z}_2, m) = 2^m.$$

Corollary 2.9 *For coprime integers* a *and* b *we have* $\mathcal{I}(\mathbb{Z}_{ab}, m) = \mathcal{I}(\mathbb{Z}_a, m) \cdot \mathcal{I}(\mathbb{Z}_b, m)$.

Corollary 2.10 *For a prime* $p > 2$ *we have*

$$\mathcal{I}(\mathbb{Z}_{p^r}, m) \geqslant p^r \cdot p^{m-1\left\lfloor \frac{r}{2} \right\rfloor}.$$

To be able to do some algebraic calculations later on we denote the set of invertible elements of $\mathcal{R}$ by $\mathcal{R}^*$ and derive a ring $\mathcal{R}'$ from the module $\mathcal{R}^2$.

Definition 2.11

$$\mathcal{R}' := \mathcal{R}[x]/\left(x^2 + 1\right).$$

With i being a root of $x^2 + 1$ we have the following bijection

$$\rho : \mathcal{R}^2 \to \mathcal{R}',\ (a, b) \mapsto a + bi.$$

The big advantage of the ring $\mathcal{R}'$ is that we naturally have an addition and multiplication. The construction of the ring is somewhat a reverse engineering of the connection between Paley graphs of square order and integral point sets over the affine plane $\mathbb{F}_q^2$ for $q \equiv 3 \mod 4$. With the similar construction of the complex numbers in mind we define:

Definition 2.12

$$\overline{a + bi} = a - bi.$$

Lemma 2.13 *For* $p, p_1, p_2 \in \mathcal{R}'$ *we have*

1. $d^2(p_1, p_2) = (p_1 - p_2) \cdot \overline{(p_1 - p_2)}$,
2. $p\overline{p} \in \mathcal{R}$,
3. $\overline{p_1 + p_2} = \overline{p_1} + \overline{p_2}$,
4. $\overline{p_1 \cdot p_2} = \overline{p_1} \cdot \overline{p_2}$, *and*
5. $\overline{\overline{p}} = p$.

3 Automorphism group of the plane $\mathcal{R}^2$

Since we want to classify maximal integral point sets up to isomorphism we have to define what we consider as an automorphism.

Definition 3.1 *An automorphism of* $\mathcal{R}'$ *with respect to* Δ *is a bijective mapping* φ *of* $\mathcal{R}'$ *with*

(1) $\Delta(a + bi, c + di) = \Delta(\varphi(a + bi), \varphi(c + di))$ *and*

(2) there exist $a', b', c', d' \in \mathcal{R}$ *such that*
$\{\varphi(a + bi + r(c + di)) \mid r \in \mathcal{R}\} = \{a' + b'i + r(c' + d'i) \mid r \in \mathcal{R}\}$

for all a, b, c, d *in* $\mathcal{R}$.

In words this definition says that φ has to map points to points, lines to lines, and has to preserve the integral distance property. There is a natural similar definition for $\mathcal{R}^2$ instead of $\mathcal{R}'$.

Lemma 3.2 *We have the following examples of automorphisms:*

(1) $\varphi_s(r) = r + s$ *for* $s \in \mathcal{R}'$,

(2) $\widetilde{\varphi}(a + bi) = b + ai$,

(3) $\widetilde{\varphi}_y(r) = ry$ *for* $y \in \mathcal{R}'^*$ *with* $\exists r' \in \mathcal{R}^* : y\overline{y} = r'^2$, *and*

(4) $\widehat{\varphi}_j(a + bi) = a^{p^j} + b^{p^j}i$ *for* $j \in \mathbb{N}$ *and* p *being the characteristic of a field* $\mathcal{R}$.

PROOF. The first two cases are easy to check. For the third case we consider

$$\begin{aligned} d^2(r_1 y, r_2 y) &= (r_1 y - r_2 y) \cdot \overline{(r_1 y - r_2 y)}, \\ &= (r_1 - r_2) \cdot \overline{(r_1 - r_2)} y\overline{y}, \\ &= d^2(r_1, r_2) \cdot y\overline{y}. \end{aligned}$$

For the fourth case we have $d^2(\widehat{\varphi}_j(a_1 + b_1 i), \widehat{\varphi}_j(a_2 + b_2 i))$

$$\begin{aligned} &= \left(a_1^{p^j} - a_2^{p^j}\right)^2 + \left(b_1^{p^j} - b_2^{p^j}\right)^2, \\ &= (a_1 - a_2)^{p^j \cdot 2} + (b_1 - b_2)^{p^j \cdot 2}, \\ &= \left((a_1 - a_2)^2 + (b_1 - b_2)^2\right)^{p^j}, \\ &= d^2(a_1 + b_1 i, a_2 + b_2 i)^{p^j} \end{aligned}$$

Thus integral point sets are mapped onto integral point sets. That lines are mapped onto lines can be checked immediately. Since we have requested that $\mathcal{R}$ is a field in the forth case the mappings are injective. □

After this general definition of automorphisms we specialize to the case $\mathcal{R} = \mathbb{F}_q$ with $2 \nmid q$. As a shorthand we use $\square_q := \square_{\mathbb{F}_q}$. We remark that the case (4) of Lemma 3.2 is the set of Frobenius automorphisms of the field $\mathbb{F}_q$ which is a cyclic group of order r for $q = p^r$.

Theorem 3.3 *For $q = p^r$, $p \neq 2$, $q \neq 5, 9$ the automorphisms of $\mathbb{F}'_q$ with respect to Δ are completely described in Lemma 3.2.*

For $q \equiv 3 \mod 4$ this is a well known result on the automorphism group of Paley graphs as mentioned in the introduction. If we consider the set of automorphisms from Lemma 3.2 in $\mathbb{F}_q^2$ instead of $\mathbb{F}'_q$ then they form a group with its elements being compositions of the following four mappings:

1. $\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} a \\ b \end{pmatrix}$ where $a, b \in \mathbb{F}_q$,
2. $\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} a & b \\ -b & a \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix}$ where $a, b \in \mathbb{F}_q$, $a^2 + b^2 \in \square_q \backslash \{0\}$,
3. $\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix}$, and
4. $\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x^p \\ y^p \end{pmatrix}$.

In the remaining part of this section we will prove Theorem 3.3. For the sake of completeness we also give the proof for $q \equiv 3 \mod 4$. If we forget about respecting Δ then the automorphism group of $\mathbb{F}_q^2$ is the well known group $A\Gamma L(2, \mathbb{F}_q)$. It is a semi-direct product of the translation group, the Frobenius group $\mathrm{Aut}(\mathbb{F}_q)$, and $\mathrm{GL}(2, \mathbb{F}_q)$, the group of multiplications with invertible 2×2 matrices over $\mathbb{F}_q$. So if G' is the automorphism group of $\mathbb{F}_q^2$ with respect to Δ it suffices to determine the group $G := G' \cap \mathrm{GL}(2, \mathbb{F}_q)$ because every translation and every element in $\mathrm{Aut}(\mathbb{F}_q)$ respects Δ. So all elements of G can be written as $\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x & y \end{pmatrix} \cdot M$ with M being an invertible 2×2-matrix. As a shorthand we say that M is an element of the automorphism group G.

Lemma 3.4 *If $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is an element of the automorphism group G then we have $ad - bc \neq 0$ and $a^2 + b^2$, $a^2 + c^2$, $b^2 + d^2$, $c^2 + d^2 \in \square_q$.*

PROOF. Since M is also an element of $\mathrm{GL}(2, \mathbb{F}_q)$ its determinant does not vanish. By considering the points $(0, 0)$ and $(0, y)$ which are at an integral distance we obtain that $b^2 + d^2$ must be a square in $\mathbb{F}_q$. Similarly we obtain that $a^2 + c^2$, $a^2 + b^2$, and $c^2 + d^2$ must be squares in $\mathbb{F}_q$. □

To go on we need some facts about roots in $\mathbb{F}_q$ and the set of solutions of quadratic equations in $\mathbb{F}_q$.

Definition 3.5 *For $p^r \equiv 1 \mod 4$ we denote by ω_q an element with $\omega_q^2 = -1$.*

Lemma 3.6 *For a finite field $\mathbb{F}_q$ with $q = p^r$ and $p \neq 2$ we have $-1 \in \square_q$ iff $q \equiv 1 \mod 4$, $\omega_q \in \square_q$ iff $q \equiv 1 \mod 8$, and $2 \in \square_q$ iff $q \equiv \pm 1 \mod 8$.*

PROOF. The multiplicative group of the units $\mathbb{F}_q^*$ is cyclic of order $q-1$. Elements of order 4 are exactly those elements x with $x^2 = -1$. A similar argument holds for the fourth roots of -1. The last statement is the second Ergänzungssatz of the quadratic reciprocity law generalized to $\mathbb{F}_q$. For a proof we may consider the situation in $\mathbb{F}_p$ and adjungate x modulo the ideal $(x^2 - 2)$. □

Lemma 3.7 *For a fix $c \neq 0$ and $2 \nmid q$ the equation $a^2 + b^2 = c^2$ in $\mathbb{F}_q$ has exactly $q + 1$ different solutions if $-1 \notin \square_q$ and exactly $q - 1$ different solutions if $-1 \in \square_q$.*

PROOF. If $b = 0$ then we have $a = \pm c$. Otherwise

$$a^2 + b^2 = c^2 \quad \Leftrightarrow \quad \frac{a - c}{b} \cdot \frac{a + c}{b} = -1.$$

Here we set $t := \frac{a+c}{b} \in \mathbb{F}_q^*$ ($t = 0$ corresponds to $b = 0$). We obtain

$$2\frac{a}{b} = t - t^{-1}, \quad 2\frac{c}{b} = t + t^{-1} \neq 0,$$

yielding

$$t^2 \neq -1, \quad b = \frac{2c}{t + t^{-1}}, \text{ and } a = c \cdot \frac{t - t^{-1}}{t + t^{-1}}.$$

If t and t' yield an equal b then we have $t' = t^{-1}$. For $t \neq t^{-1}$ we have different values for a in these cases. Summing up the different solutions proves the stated result. □

Lemma 3.8 *In $\mathbb{F}'_q$ the set $C = \{z \in \mathbb{F}'_q \mid z\bar{z} = 1\}$ forms a cyclic multiplicative group.*

PROOF. If $-1 \notin \square_q$ then $\mathbb{F}'_q$ is a field and thus C must be cyclic. For the case $-1 \in \square_q$ we utilize the bijection

$$\rho_q : \mathbb{F}_q^* \to G, \quad t \mapsto \frac{1 + t^2}{2t} + \omega_q \frac{1 - t^2}{2t} x.$$

Now we only have to check that the mapping is a group isomorphism, namely

$$\rho_q(i \cdot j) = \rho_q(i) \cdot \rho_q(j).$$

□

Our next ingredient is a classification of the subgroups of the projective special linear group $\mathrm{PSL}(2, q)$.

Theorem 3.9 *(Dickson [7]) The subgroups of $PSL(2, p^r)$ are isomorphic to one of the following families of groups:*

(1) elementary abelian p-groups,

(2) cyclic groups of order z, where z is a divisor of $\frac{p^r \pm 1}{k}$ and $k = \gcd(p^r - 1, 2)$,

(3) dihedral groups of order $2z$*, where* z *is defined as in (2),*

(4) alternating group A_4 *(this can occur only for* $p > 2$ *or when* $p = 2$ *and* $r \equiv 0 \mod 12$*),*

(5) symmetric group S_4 *(this can only occur if* $p^{2r} \equiv 1 \mod 16$*),*

(6) alternating group A_5 *(for* $p = 5$ *or* $p^{2r} \equiv 1 \mod 5$*),*

(7) a semidirect product of an elementary abelian group of order p^m *with a cyclic group of order* t*, where* t *is a divisor of* $p^m - 1$ *and of* $p^r - 1$*, or*

(8) the group PSL$(2, p^m)$ *for* m *a divisor of* r*, or the group PGL*$(2, p^m)$ *for* $2m$ *a divisor of* r*.*

By $Z := \pm E$ we denote the center of $SL(2, q)$, where E is the identity matrix. Our strategy is to consider $H := (G \cap SL(2, q))/Z = G \cap PSL(2, q)$ and to prove $H \simeq H'$ for $q \geqslant 13$ where H' is the group of those automorphisms of Lemma 3.2 which are also elements of $PSL(2, q)$. For $-1 \notin \square_q$ we set $\widetilde{H} := \left\{ \begin{pmatrix} a & b \\ -b & a \end{pmatrix} \mid a^2 + b^2 = 1 \right\}$ and for $-1 \in \square_q$ we set $\widetilde{H} := \left\{ \begin{pmatrix} a & b \\ -b & a \end{pmatrix} \mid a^2 + b^2 = 1 \right\} \cup \left\{ \begin{pmatrix} -b & a \\ a & b \end{pmatrix} \mid a^2 + b^2 = -1 \right\}$.

Lemma 3.10 *For* $q \equiv 3 \mod 4$ *we have* $\widetilde{H} \simeq \mathbb{Z}_{q+1}$ *and for* $q \equiv 1 \mod 4$ *we have* $\widetilde{H} \simeq D_{q-1}$*, where* D_{q-1} *is the dihedral group of order* $2(q-1)$*.*

PROOF. Utilizing Lemma 3.7 and checking that both sets are groups we get

$$\left|\widetilde{H}\right| = \begin{cases} q+1 & \text{if } q \equiv 3 \mod 4, \\ 2(q-1) & \text{if } q \equiv 1 \mod 4. \end{cases}$$

In the first case the group is cyclic due to Lemma 3.8. In the second case it contains a cyclic subgroup of order $q-1$. By checking the defining relations of a dihedral group we can conclude $\widetilde{H} \simeq D_{q-1}$ for $q \equiv 1 \mod 4$. □

Now we define $H' := \widetilde{H}/Z$.

Lemma 3.11 *For* $q \geqslant 13$*,* $q \equiv 3 \mod 4$ *we have* $H' \simeq \mathbb{Z}_{\frac{q+1}{2}}$ *and for* $q \geqslant 13$*,* $q \equiv 1 \mod 4$ *we have* $H' \simeq D_{\frac{q-1}{2}}$*.*

PROOF. We have $|H'| = \frac{|\widetilde{H}|}{2}$. It remains to show that H' is not abelian for $q \equiv 1 \mod 4$. Therefore we may consider the sets $\{\pm M_1\}$ and $\{\pm M_2\}$ where a, b, c, d are elements of $\mathbb{F}_q^*$ with $a^2 + b^2 = 1$, $c^2 + d^2 = -1$ and where

$$M_1 = \begin{pmatrix} a & b \\ -b & a \end{pmatrix} \quad \text{and} \quad M_2 = \begin{pmatrix} -d & c \\ c & d \end{pmatrix}.$$

□

Lemma 3.12 *For* $q \geqslant 13$ *we have* $H \simeq H'$*.*

PROOF. Since H is a subgroup of $PSL(2, q)$ we can utilize Theorem 3.9. We run through the subgroups of $PSL(2, q)$, identify H' and show that H is none of the subgroups of $PSL(2, q)$ containing H' as a proper subgroup. With the numbering from the theorem we have the following case distinctions. We remark that for $q \equiv 1 \mod 4$ the group H' is the group of case (3) and for $q \equiv 3 \mod 4$ the group H' is the group of case (2)

(1) H is not an elementary abelian p-group since $|H'|$ is not a p-power.

(2) For $q \equiv 1 \mod 4$ the order of H' is larger than $\frac{p^r \pm 1}{2}$ and for $q \equiv 3 \mod 4$ the characterized group must be H' itself.

(3) For $q \equiv 1 \mod 4$ the characterized group must be H' itself due to the order of the groups. For $q \equiv 3 \mod 4$ we must have a look at the elements of order 2 in $PSL(2, q)$. These are elements $M \cdot Z$ where $M = \begin{pmatrix} a & b \\ c & b \end{pmatrix}$ with $ad - bc = 1$ and $M^2 = E$ or $M^2 = -E$. Solving this equation system yields $M = \pm E$ which corresponds to an element of H' and $M = \begin{pmatrix} a & b \\ -\frac{a^2+1}{b} & -a \end{pmatrix}$ where $a \in \mathbb{F}_q$ and $b \in \mathbb{F}_q^*$. Now we choose a matrix $N = \begin{pmatrix} u & v \\ -v & u \end{pmatrix}$ with $u^2 + v^2 = 1$ and $u, v \neq 0$. So $N \cdot Z = \{\pm N\} \in H'$ and since $\langle H', N \rangle$ would be a dihedral group we have the following relation

$$MZ \cdot NZ \cdot MZ = N^{-1}Z$$

$$\Leftrightarrow \{\pm M\} \cdot \{\pm N\} \cdot \{\pm M\} = \left\{\pm N^{-1}\right\} = \left\{\pm \begin{pmatrix} u & -v \\ v & u \end{pmatrix}\right\}$$

$$\Leftrightarrow \left\{\pm \begin{pmatrix} \frac{-ab^2v - a^3v - av - bu}{b} & -v\left(a^2+b^2\right) \\ \frac{v\left(a^2b^2 + a^4 + 2a^2 + 1\right)}{b^2} & \frac{-bu + ab^2v + a^3v + av}{b} \end{pmatrix}\right\} = \left\{\pm \begin{pmatrix} u & -v \\ v & u \end{pmatrix}\right\}.$$

By comparing the diagonal elements we get $av\left(a^2 + b^2 + 1\right) = 0$ and $v\left(b^4 - a^4 - 2a^2 - 1\right) = 0$. Due to $v \neq 0$ this is equivalent to $a\left(a^2 + b^2 + 1\right) = 0$ and $\left(a^2 + b^2 + 1\right) \cdot \left(a^2 - b^2 + 1\right) = 0$. Together with $a^2 + b^2 \in \square_q$ we conclude $a = 0$ and $b = \pm 1$. Since these solutions correspond to an element of H' we derive that case (3) is not possible for $q \equiv 3 \mod 4$.

(4) If $H' < H \leqslant A_4$ then H' must be contained in a maximal subgroup of A_4. Since the order of a maximal subgroup of A_4 is at most 4 and $q \geqslant 13$ this case can not occur.

(5) Since we have $q \geqslant 13$ and the maximal subgroups of the S_4 are isomorphic to A_4, D_4, and S_3, this case can not occur.

(6) The maximal subgroups of A_5 are isomorphic to D_5, S_3, and A_4. So this case can not occur for $q \geqslant 13$.

(7) We have that $|H|$ divides $(q-1)\cdot p^m$. Since $\gcd\left(\frac{q+1}{2}, (q-1)\cdot p^m\right) \leqslant 2$ and $|H'|$ divides $|H|$, only $q \equiv 1 \mod 4$, $|H'| = q-1$, $t = q-1$, and $r|m$ is possible. If $m \geqslant 2r$ then $|H| \geqslant q^2(q-1) > |\mathrm{PSL}(2,q)| = \frac{1}{2}\left(q^2-1\right)q$, which is a contradiction. So only $m = r$ is possible and H must be a semidirect product of an abelian group of order q and a cyclic group of order $q-1$. Using Zassenhaus' theorem [11, I.18.3] we can deduce that all subgroups of order $q-1$ of H are conjugates and so isomorphic. Since H' is not abelian (for $q \equiv 1 \mod 4$) it is not cyclic and so at the end case (7) of Theorem 3.9 is impossible.

(8) Clearly $H \ncong \mathrm{PSL}(2,q)$. Since $|H'|$ does not divide $|\mathrm{PSL}(2,p^m)| = \frac{\left(p^{2m}-1\right)p^m}{2}$ only the second possibility is left. Since $|H'|$ divides $|\mathrm{PGL}(2,p^m)| = \left(p^{2m}-1\right)\left(p^{2m}-p^m\right)$ we have $2m = r$, $p^m = \sqrt{q}$, and $q \equiv 1 \mod 4$. But for $q \geqslant 13$ we have $D_{\frac{q-1}{2}} \nleqslant \mathrm{PGL}\left(2,\sqrt{q}\right)$, see e. g. [5], thus case (8) is also not possible.

□

To finish the proof of the characterization of the automorphisms of $\mathbb{F}_q^2$ with respect to Δ we need as a last ingredient a result on the number of solutions of an elliptic curve in $\mathbb{F}_q$.

Theorem 3.13 *(Hasse, e. g. [22]) Let f be a polynomial of degree 3 in $\mathbb{F}_q$ without repeated factors then we have for the number N of different solutions of $f(t) = s^2$ in $\mathbb{F}_q^2$ the inequality $|N - q - 1| \leqslant 2\sqrt{q}$.*

PROOF OF THEOREM 3.3. For the cases $q = 3, 7, 11$ we utilize a computer to check that there are no other automorphisms. So we can assume $q \geqslant 13$.

If $M \in G$ is an automorphism for $q \equiv 3 \mod 4$ then there exists an element $x \in \mathbb{F}_q^*$ so that either $x \cdot M$ or $x \cdot M \cdot \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ has determinant 1. Thus with the help of Lemma 3.12 and Lemma 3.2 the theorem is proven for $q \equiv 3 \mod 4$. With the same argument we can show that for $q \equiv 1 \mod 4$ any possible further automorphism which is not contained in the list of Lemma 3.2 must have a determinant which is a non-square in $\mathbb{F}_q$. Let $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ be an element of G with $\det(M) = ad - bc \notin \square_q$. So $M^2 = \begin{pmatrix} a^2+bc & b(a+d) \\ c(a+d) & bc+d^2 \end{pmatrix}$ is also an element of G. Since we have $\det\left(M^2\right) = \det(M)^2 \in \square_q$ we have $a^2+bc = bc+d^2$, $b(a+d) = -c(a+d)$ or $a^2+bc = -\left(bc+d^2\right)$, $b(a+d) = c(a+d)$ due to Lemma 3.12. This leads to the four cases

(1) $a = d$, $b = -c$,

(2) $a = d = 0$,

(3) $a = -d$, and

(4) $b = c$, $a^2 + d^2 = -2b^2$.

Now we consider the derived matrix $M' := M \cdot \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} b & a \\ d & c \end{pmatrix}$ with $\det(M') \notin \square_q$ which must be also an automorphism. So each of the matrices M and M' must be in one of the four cases. From this we can conclude some equations and derive a contradiction for each possibility. W.l.o.g. we assume that the number of the case of M' is at least the number of the case of M.

(1) M as in (1): With the help of Lemma 3.4 we get $\det(M) = a^2 + b^2 \in \square_q$, which is a contradiction.

(2) M as in (2): Since $\det(M) \notin \square_q$ the only possibility for M' is case (4). Thus we have $b^2 + c^2 = 0 \Leftrightarrow b = \pm\omega_q c$, where we can assume $c = 1$ and $b = \omega_q$ without loss of generality. Since $\det(M')$ must be a non-square in $\mathbb{F}_q$ we have $q \equiv 5 \mod 8$. If we apply M' onto the points $(0,0)$ and $(1,1)$ then we can conclude that 2 must be a square in $\mathbb{F}_q$, which is not the case if $q \equiv 5 \mod 8$.

(3) M as in (3): Due to $\det(M) \notin \square_q$ the matrix M' must be in case (4). So we have $a = d = 0$, a situation already treated in case (2).

(4) M as in (4): Thus also M' has to be in case (4). Here we have $a = d$, $b = c$, $2a^2 = -2b^2$. Without loss of generality we can assume $a = 1$ and $b = \omega_q$. Due to $\det(M) = 2 \notin \square_q$ we have $q \equiv 5 \mod 8$. For two elements $x, y \in \mathbb{F}_q$ with $x^2 + y^2$ being a square we have that also $\widetilde{M} := \begin{pmatrix} 1 & \omega_q \\ \omega_q & 1 \end{pmatrix} \cdot \begin{pmatrix} x & y \\ -y & x \end{pmatrix} = \begin{pmatrix} x - \omega_q x & x\omega_q + y \\ x\omega_q - y & x + y\omega_q \end{pmatrix}$ is an automorphism. Thus with Lemma 3.4 we get that $(x\omega_q + y)^2 + (x + y\omega_q)^2 = 2^2 xy\omega_q$ must be a square in $\mathbb{F}_q$ for all possible values $x, y \neq 0$. So for $q \equiv 5 \mod 8$ for all possible x, y the product $xy \neq 0$ must be a non-square. We specialize to $x^2 + y^2 = 1^2$ and so can get with the help of Lemma 3.7 that $x = \frac{2}{t+t^{-1}}$ and $y = \frac{t-t^{-1}}{t+t^{-1}}$ with $t^2 \neq -1$, $t \neq 0$. If we require $t^4 \neq 1$ instead of $t^2 \neq -1$ we get $x, y \neq 0$. Thus $xy = \frac{2\left(t-t^{-1}\right)}{\left(t+t^{-1}\right)^2}$ must be a non-square for all $t \in \mathbb{F}_q^*$ with $t^4 \neq 1$. Since 2 is a non-square we have that $t - t^{-1}$ and so also $t^3 - t = t(t-1)(t+1)$ must be a square for all $t \in \mathbb{F}_q^*$ with $t^4 \neq 1$. By checking the five excluded values we see that $f(t) := t(t-1)(t+1)$ must be a square for all $t \in \mathbb{F}_q$. So $f(t) = s^2$ has exactly $N := 2q - 3$ solutions in $\mathbb{F}_q$. Since f has not repeated factors and degree 3 we can apply Theorem 3.13 to get a contradiction to $q \geqslant 13$.

□

Lemma 3.14 *For two points $p_1 \neq p_2 \in \mathbb{F}'_q$ at integral distance there exists an isomorphism φ with either $\varphi(p_1) = 0$, $\varphi(p_2) = 1$ or $\varphi(p_1) = 0$, $\varphi(p_2) = 1 + \omega_q i$.*

PROOF. Without loss of generality we assume $p_1 = 0$. Since the points p_1 and p_2 are at integral distance there exists an element $r \in \mathbb{F}_q$ with $p_2\overline{p_2} = r^2$ and since $p_2 \neq p_1$ we have $p_2 \in {\mathbb{F}'_q}^*$. If $p_2\overline{p_2} \neq 0$ we choose $\cdot p_2^{-1}$ as the isomorphism φ. Otherwise we have $p_2 = a + bi$ with $a^2 + b^2 = 0$ where $a, b \neq 0$. Thus $\left(\frac{b}{a}\right)^2 = -1$ and $\varphi = \cdot a^{-1}$. □

We remark that Lemma 3.14 can be sharpened a bit. For three pairwise different non-collinear points $p_1, p_2, p_3 \in \mathbb{F}'_q$ with pairwise integral distances there exists an isomorphism φ with $\{0, 1\} \subset \{\varphi(p_1), \varphi(p_2), \varphi(p_3)\}$.

Via a computer calculation we can determine the automorphism groups of the missing cases $q = 5, 9$.

Lemma 3.15 *For $q = 5$ the group $G \leqslant GL(2, \mathbb{F}_5)$ is given by*

$$\left\{ M = \begin{pmatrix} a & b \\ \pm b & \pm a \end{pmatrix} \;\middle|\; a, b \in \mathbb{F}_5, \; a^2 + b^2 \in \square_5, \det(M) \neq 0 \right\}$$

where the two signs can be chosen independently.

Lemma 3.16 *For $q = 9$ the group $G \leqslant GL(2, \mathbb{F}_9)$ is given by*

$$\left\langle \begin{pmatrix} 1 & 0 \\ 0 & y^2 \end{pmatrix}, \left\{ M = \begin{pmatrix} a & b \\ \pm b & \pm a \end{pmatrix} \;\middle|\; a, b \in \mathbb{F}_9, \; a^2 + b^2 \in \square_9, \det(M) \neq 0 \right\} \right\rangle$$

where the two signs can be chosen independently and where y is a primitive root in $\mathbb{F}_9^*$.

For $q = 5$ there are exactly 32 such matrices and for $q = 9$ there are exactly 192 such matrices. For $q = 5, 9$ Lemma 3.14 can be sharpend. Here the automorphism group acts transitively on the pairs of points with integral distance, as for $q \equiv 3 \mod 4$.

We would like to remark that also for $q \equiv 3 \mod 4$ the automorphism group of $\mathbb{F}_q^2$ with respect to Δ is isomorphic to the automorphism group of the quadrance graph over $\mathbb{F}_q^2$. This can easily be verified be going over the proof of Theorem 3.3 again and by checking the small cases using a computer.

4 Maximal integral point sets in the plane $\mathbb{F}_q^2$

Very nice rings are those which are integral domains. These are in the case of finite commutative rings exactly the finite fields $\mathbb{F}_q$ where $q = p^r$ is a prime power. So far we only have the lower bound $\mathcal{I}(\mathbb{F}_q, 2) \geqslant q$. In this section we will prove $\mathcal{I}(\mathbb{F}_q, 2) = q$ for $q > 2$. In the case of $\mathbb{F}_p$ we will even classify the maximum integral point sets up to isomorphism. One way to prove $\mathcal{I}(\mathbb{F}_q, 2) = q$ for $2 \nmid q$ is to consider the graph $\mathcal{G}_q$ with the elements of $\mathbb{F}_q$ as its vertices and pairs of points at integral distance as edges. For $q \equiv 3 \mod 4$ the graph $\mathcal{G}_q$ is isomorphic to the Paley graph of order q^2. From [3] we know that in this case a maximum clique of $\mathcal{G}_q$ has size q and is isomorphic to a line. Also for $q \equiv 1 \mod 4$ the graph $\mathcal{G}_q$ is a strongly regular graph. So we can apply a result from [18, 19] on cliques of strongly regular graphs. It turns out that a maximum clique has size q and that every clique $\mathcal{C}$ of size q is *regular*, in the sense of [18, 19], this means in our special case that every point not in $\mathcal{C}$ is adjacent to $\frac{q+1}{2}$ points in $\mathcal{C}$. To start with our classification of maximum integral point sets over $\mathbb{F}_q$ we need the concept of directions.

Definition 4.1 *For a point $p = a + bi \in \mathbb{F}'_q$ the quotient $\frac{b}{a} \in \mathbb{F}_q \cup \{\infty\}$ is called the **direction** of p. For two points $p_1 = a_1 + b_1 i$, $p_2 = a_2 + b_2 i$ the direction is defined as $\frac{b_1 - b_2}{a_1 - a_2} \in \mathbb{F}_q \cup \{\infty\}$. We call an direction d **integral** if two points p_1, p_2 with direction d have an integral distance.*

Point sets of cardinality q in $\mathbb{F}_q^2$ with at most $\frac{q+3}{2}$ directions are more or less completely classified:

Theorem 4.2 *(Ball, Blokhuis, Brouwer, Storme, Szőnyi, [4]) Let $f : \mathbb{F}_q \to \mathbb{F}_q$, where $q = p^n$, p prime, $f(0) = 0$. Let $N = |D_f|$, where D_f is the set of directions determined by the function f. Let e (with $0 \leqslant e \leqslant n$) be the largest integer such that each line with slope in D_f meets the graph of f in a multiple of p^e points. Then we have the following:*

1. *$e = 0$ and $\frac{q+3}{2} \leqslant N \leqslant q + 1$,*
2. *$e = 1$, $p = 2$, and $\frac{q+5}{3} \leqslant N \leqslant q - 1$,*
3. *$p^e > 2$, $e|n$, and $\frac{q}{p^e} + 1 \leqslant N \leqslant \frac{q-1}{p^e - 1}$,*
4. *$e = n$ and $N = 1$.*

Moreover, if $p^e > 3$ or ($p^e = 3$ and $N = \frac{q}{3} + 1$), then f is a linear map on $\mathbb{F}_q$ viewed as a vector space over $\mathbb{F}_{p^e}$. (All possibilities for N can be determined in principle.)

Here a function $f : \mathbb{F}_q \to \mathbb{F}_q$ determines a point set $\mathcal{P} = \{(x, f(x)) \mid x \in \mathbb{F}_q\}$ of cardinality q. In the case $N = 1$ the point set is a line. In the case $e = 0$ and $N = \frac{q+3}{2}$ then $\mathcal{P}$ is affine equivalent to the point set corresponding to $x \mapsto x^{\frac{q+1}{2}}$.

We remark that affine equivalence is a bit more than our equivalence because we have to respect Δ. The next thing to prove is that integral point sets can not determine too many directions.

Lemma 4.3 *For $2 \nmid q$ an integral point set over $\mathbb{F}_q^2$ determines at most $\frac{q+3}{2}$ different directions if $-1 \in \square_q$ and at most $\frac{q+1}{2}$ different directions if $-1 \notin \square_q$.*

PROOF. We consider the points $p = a + bi$ at integral distance to 0. There exists an element $c' \in \mathbb{F}_q$ with $a^2 + b^2 = c'^2$. In the case $a = 0$ we obtain the direction ∞. Otherwise we set $d := \frac{b}{a}$ and $c := \frac{c'}{a}$, yielding $1 = c^2 - d^2 = (c-d)(c+d)$, where d is the direction of the point. Now we set $c + d =: t \in \mathbb{F}_q^*$ yielding $c = \frac{t+t^{-1}}{2}$, $d = \frac{t-t^{-1}}{2}$. The two values t and $-t^{-1}$ produce an equal direction. Since $t = -t^{-1} \Leftrightarrow t^2 = -1$ we get the desired bounds. □

We need a further lemma on the number of points on a line in a non-collinear integral point set:

Lemma 4.4 *If $2 \nmid q$ and $\mathcal{P}$ is a non-collinear integral point set over $\mathbb{F}_q^2$, then each line l contains at most $\frac{q-1}{2}$ points for $-1 \notin \square_q$ and at most $\frac{q+1}{2}$ points for $-1 \in \square_q$.*

PROOF. If l is a line with an integral pair of points on it, then its slope is an integral direction. Now we consider the intersections of lines with integral directions containing a point $p \notin l$, with l. □

We remark that there would be only $\frac{q-1}{2}$ integral directions for $q \equiv 1 \mod 4$ if we would not consider 0 as a square as for quadrance graphs. In this case there could be at most $\frac{q-3}{2}$ points on l for $q \equiv 1 \mod 4$ in Lemma 4.4.

To completely classify maximum integral point sets over $\mathbb{F}_q'$ we need the point set $\mathcal{P}_q := (1 \pm \omega_q i)\,\square_q$.

Lemma 4.5 *$\mathcal{P}_q$ is an integral point set of cardinality q.*

PROOF.

$$\begin{aligned} d^2\left(r_1^2 + r_1^2\omega_q i, r_2^2 + r_2^2\omega_q i\right) &= 0^2, \\ d^2\left(r_1^2 + r_1^2\omega_q i, r_2^2 - r_2^2\omega_q i\right) &= \left(2\omega_q r_1 r_2\right)^2, \\ d^2\left(r_1^2 - r_1^2\omega_q i, r_2^2 - r_2^2\omega_q i\right) &= 0^2. \end{aligned}$$

□

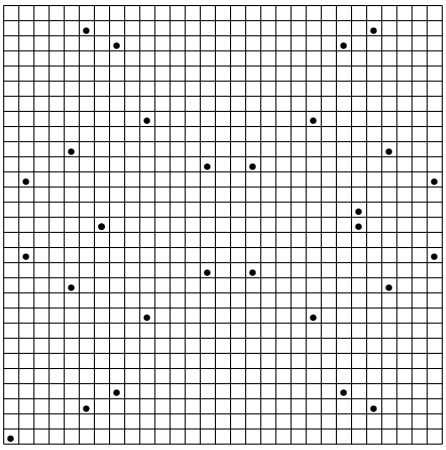


Figure 1: The maximum integral point set $\mathcal{P}_{29}$.

In Figure 1 we have depicted $\mathcal{P}_{29}$ as an example. By construction the points of $\mathcal{P}_q$ are located on the two lines $(1, \omega_q) \cdot \mathbb{F}_q$ and $(1, -\omega_q) \cdot \mathbb{F}_q$ which intersect in $(0,0)$ with an *angle* of 90 degree, but this fact seems not that obvious by looking at Figure 1. We remark that this construction of $\mathcal{P}_q$ works in any commutative ring $\mathcal{R}$ where $-1 \in \square_{\mathcal{R}}$ and that none of these point sets corresponds to a quadrance graph. If we apply this construction on $\mathcal{R} = \mathbb{Z}_{p^r}$ we obtain an integral point set of cardinality $\phi(p^r) + 1 = (p-1) \cdot p^{r-1} + 1$, where ϕ is the Euler-function defined by $\phi(n) = |\mathbb{Z}_n^*|$.

Lemma 4.6 *For $2 \mid r$ the point set*

$$\mathcal{P} := \left\{(a, b) \mid a, b \in \mathbb{F}_{\sqrt{q}}\right\}$$

is an integral point set.

PROOF. We have $\mathbb{F}_{\sqrt{q}} \subset \square_q$. □

We remark that for $\sqrt{q} \equiv 1 \mod 4$ also the point set $\mathcal{P} := \left\{(a, \omega_q b) \mid a, b \in \mathbb{F}_{\sqrt{q}}\right\}$ is integral.

We say that an integral point set is maximal if we can not add a further point without destroying the property *integral point set*. All given examples of integral point sets of size q are maximal. This could be proved be applying results on cliques of strongly regular graphs or in the following way.

Lemma 4.7 *The lines $1 \cdot \mathbb{F}_q$ and $(1 + \omega_q i) \cdot \mathbb{F}_q$ are maximal.*

PROOF. We apply Lemma 4.4. □

Lemma 4.8 *The integral point set $\mathcal{P} = (1 \pm \omega_q i) \cdot \square_q$ is maximal.*

PROOF. Let us assume there is a further point $(a + bi) \notin \mathcal{P}$ with $a, b \in \mathbb{F}_q$ such that $\mathcal{P} \cup \{(a + bi)\}$ is also an integral point set. We know that $(a + bi)$ can not lie on one of the lines $(1 + \omega_q i) \cdot \mathbb{F}_q$ or $(1 - \omega_q i) \cdot \mathbb{F}_q$. Thus $a^2 + b^2 \neq 0$. The points of $\mathcal{P}$ are given by $(1 + \omega_q i)\, r_1^2$ and $(1 - \omega_q i)\, r_2^2$ for arbitrary $r_1, r_2 \in \mathbb{F}_q$. We define functions $f_1, f_2 : \mathbb{F}_q \to \mathbb{F}_q$ via

$$\begin{aligned} f_1(r_1) &= \left(a - r_1^2\right)^2 + \left(b - r_1^2\omega_q\right)^2 \\ &= a^2 + b^2 - 2r_1^2(a + b\omega_q), \\ f_2(r_2) &= \left(a - r_2^2\right)^2 + \left(b + r_2^2\omega_q\right)^2 \\ &= a^2 + b^2 - 2r_2^2(a - b\omega_q). \end{aligned}$$

Since these are exactly the squared distances of the points of $\mathcal{P}$ to the point $(a + bi)$ we have $\mathrm{Bi}(f_1), \mathrm{Bi}(f_2) \subseteq \square_q$. Using a counting argument we have $\mathrm{Bi}(f_1), \mathrm{Bi}(f_2) = \square_q$. The term $-2(a + b\omega_q)$ is a fix number. Let us assume that it is a square. Then for each square r^2 and $c = a^2 + b^2 \neq 0$ the difference $r^2 - c$ must be a square. But the equation $r^2 - c = h^2$ has $\frac{q+1}{2} < q$ solutions for r, which is a contradiction. Thus $-2(a + b\omega_q)$ and $-2(a - b\omega_q)$ are non-squares. But $r^2 - c \notin \square_q$ has $\frac{q-1}{2}$ solutions, thus we have a contradiction □

Theorem 4.9 *For $q = p^r > 9$ with $p \neq 2$, $r = 1$ or $q \equiv 3 \mod 4$ an integral point set of cardinality q is isomorphic to one of the stated examples.*

PROOF. We consider a point set $\mathcal{P}$ of $\mathbb{F}_q$ of cardinality q with at most $\frac{q+3}{2}$ directions and utilize Theorem 4.2. If $e = r$ and $N = 1$ then $\mathcal{P}$ is a line. If $e = 1$ then $\mathcal{P}$ is affine equivalent to $X := \left\{\left(x, x^{\frac{q+1}{2}}\right) \mid x \in \mathbb{F}_q\right\}$. This is only possible for $q \equiv 1 \mod 4$. The set X consists of two orthogonal lines. Since there are only two types of non-isomorphic integral lines in $\mathbb{F}_q^2$ and each point p not on a line l is at integral distance to $\frac{q+1}{2}$ points on l we have two unique candidates of integral point sets of this type. One is given by $(1 \pm \omega_q i) \cdot \square_q$. For the other possibility we may assume that $(0,0), (1,0) \in \mathcal{P}$. Thus $(0, \pm\omega_q) \in \mathcal{P}$, $(-1,0), (\pm\omega_q, 0), (0, \pm 1) \in \mathcal{P}$. So $\mathcal{P}$ must be symmetric in the following sense: There exists a set $S \subset \mathbb{F}_q^*$ such that $\mathcal{P} = (0,0) \cup \{(0,a), (a,0) \mid a \in S\}$. The elements s of S must fullfill $s \in \mathbb{F}_q^*$, $s^2 + 1 \in \square_q$ and $s^2 - 1 \in \square_q$. Each condition alone has only $\frac{q-1}{2}$ solutions. Fulfilling both conditions, meaning $|S| = \frac{q-1}{2}$ is possible only for $q \leqslant 9$. For $q = 5, 9$ there are such examples. For $q \equiv 3 \mod 4$ we refer to [3]. □

We remark that there may be further examples of integral point sets of cardinality q for $q = p^r \equiv 1 \mod 4$ and $r > 1$. Those examples would correspond to case (3) of Theorem 4.2.

Theorem 4.10 *For $q = p^r$ with $p \neq 2$ we have $\mathfrak{I}(\mathbb{F}_q, 2) = q$.*

PROOF. Let $\mathcal{P}$ be an arbitrary integral point set of cardinality q. Now we show that $\mathcal{P}$ is maximal. If we assume that there is another integral point set $\mathcal{P}'$ with $\mathcal{P} \subset \mathcal{P}'$ and $|\mathcal{P}'| = q+1$ then we can delete a point of $\mathcal{P}'$ in such a way that we obtain an integral point set $\mathcal{P}''$ with $e = 1$ in the notation of 4.2. Thus $\mathcal{P}'' \simeq (1 \pm \omega_q i) \cdot \square_q$. Since $\mathcal{P}''$ is maximal due to Lemma 4 we have a contradiction. □

5 Maximal integral point sets in the plane $\mathbb{Z}_n^2$

Due to Theorem 2.6 for the determination of $\mathfrak{I}(\mathbb{Z}_n, 2)$ we only need to consider the cases $n = p^r$.

Lemma 5.1

$$\mathfrak{I}\left(\mathbb{Z}_{p^{r+1}}, 2\right) \leqslant p^2 \cdot \mathfrak{I}\left(\mathbb{Z}_{p^r}, 2\right).$$

PROOF. We consider the natural ring epimorphism $\nu : \mathbb{Z}_{p^{r+1}} \to \mathbb{Z}_{p^r}$. If $\mathcal{P}$ is an integral point set in $\mathbb{Z}_{p^{r+1}}^2$ then $\nu(\mathcal{P})$ is an integral point set in $\mathbb{Z}_{p^r}^2$. □

For $p \geqslant 3$ we have the following examples of integral point sets in $\mathbb{Z}_{p^r}^2$ with big cardinality (with some abuse of notation in the third case).

$$\left\{\left(i, j \cdot p^{\lceil \frac{r}{2} \rceil}\right) \mid i, j \in \mathbb{Z}_{p^r}\right\},$$
$$\left\{\left(i, i\omega_{\mathbb{Z}_{p^r}} + j \cdot p^{\lceil \frac{r}{2} \rceil}\right) \mid i, j \in \mathbb{Z}_{p^r}\right\}, \text{ and}$$
$$(1, \pm\omega_{\mathbb{Z}_p}) \cdot \square_{\mathbb{Z}_p} + \{(p \cdot a, p \cdot b) \mid a, b \in \mathbb{Z}_{p^r}\} \text{ for } r = 2.$$

Each of these examples has cardinality $p^r \cdot p^{\lfloor \frac{r}{2} \rfloor}$.

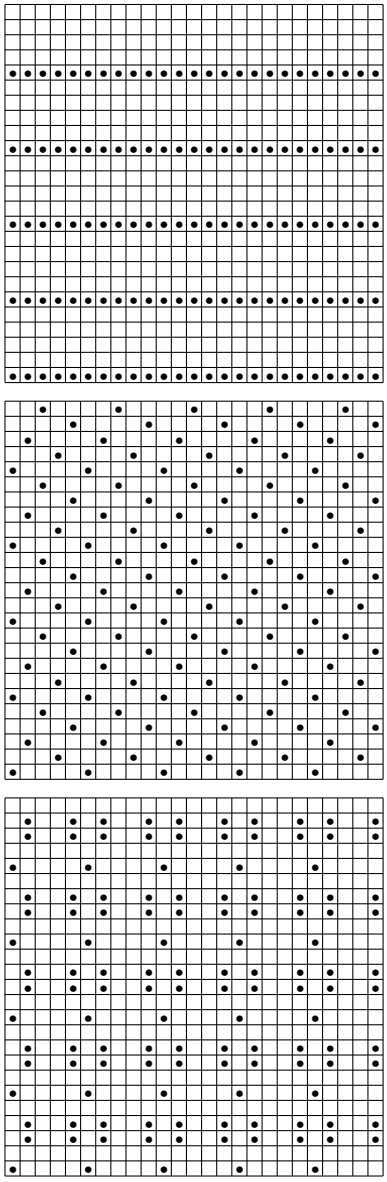


Figure 2: Three maximal integral point sets over $\mathbb{Z}_{25}^2$ of cardinality 125.

Conjecture 5.2 *The above list is the complete list of maximum integral point sets in $\mathbb{Z}_{p^r}^2$ up to isomorphism.*

So far we do not even know the automorphism group of $\mathbb{Z}_n^2$ with respect to Δ. But with Definition 3.1 Conjecture 5.2 is well defined. Using Lemma 3.2 we know at least a subgroup of the automorphism group. If there are any further

automorphisms is an open question which has to be analyzed in the future.

Theorem 5.3 *For $p \geqslant 3$ we have $\mathcal{I}\left(\mathbb{Z}_{p^2}, 2\right) = p^3$ and the above list of extremal examples is complete.*

PROOF. With $\mathcal{I}(\mathbb{Z}_p, 2) = p$, Lemma 5.1 and the examples we get $\mathcal{I}\left(\mathbb{Z}_{p^2}, 2\right) = p^3$. Let $\mathcal{P}$ be a maximum integral point set in $\mathbb{Z}_{p^2}$. By S denote the lower left $p \times p$-square of $\mathbb{Z}_{p^2}$

$$S := \left\{(i,j) + \mathbb{Z}_{p^2}^2 \mid 0 \leqslant i,j \leqslant p-1,\ i,j \in \mathbb{Z}\right\}.$$

Using Theorem 4.10 and Lemma 5.1 we can deduce that for each $(u,v) \in \mathbb{Z}_{p^2}^2$ we have

$$|\mathcal{P} \cap ((u,v) + S)| \leqslant p.$$

Since we can tile $\mathbb{Z}_{p^2}$ with p^2 such sets (including $S + (u,v)$) equality must hold. After a transformation we can assume that $\mathcal{P} \cap S$ equals one of the three following possibilities

1. $\{(i,0) \mid 0 \leqslant i \leqslant p-1\}$,
2. $\left\{\left(i, \omega_{\mathbb{Z}_p} i\right) \mid 0 \leqslant i \leqslant p-1\right\}$, or
3. $\left(1, \pm\omega_{\mathbb{Z}_p}\right) \cdot \square_{\mathbb{Z}_p}$.

In the first case we consider $\mathcal{P} \cap (S + (1,0))$. With Lemma 4.4 we get $(p,0) \in \mathcal{P}$ and iteratively we get $(i,0) \in \mathcal{P}$ for all $i \in \mathbb{Z}_{p^2}$. Now we consider $\mathcal{P} \cap (S + (0,1))$ and conclude $\mathcal{P} = \left\{(i, j \cdot p) \mid i,j \in \mathbb{Z}_{p^2}\right\}$. With the same argument we can derive $\mathcal{P} = \left\{\left(i, i\omega_{\mathbb{Z}_p} + j \cdot p\right) \mid i,j \in \mathbb{Z}_{p^2}\right\}$ in the second case and $\mathcal{P} = \left(1, \pm\omega_{\mathbb{Z}_p}\right) \cdot \square_{\mathbb{Z}_p} + \{(p \cdot a, p \cdot b) \mid a,b \in \mathbb{Z}_{p^r}\}$ in the third case. □

6 Maximal integral point sets without three collinear points

In this and the next section we study the interplay between the integrality condition for a point set and further common restrictions for lines and circles.

Definition 6.1 *A set of r points $(u_i, v_i) \in \mathcal{R}^2$ is said to be **collinear** if there are $a, b, t_1, t_2, w_i \in \mathcal{R}$ with*

$$a + w_i t_1 = u_i \quad \text{and} \quad b + w_i t_2 = v_i.$$

There is an easy necessary criterion to decide whether three points are collinear.

Lemma 6.2 *If three points (u_1, v_1), (u_2, v_2), and $(u_3, v_3) \in \mathcal{R}^2$ are collinear then it holds*

$$\left|\begin{pmatrix} u_1 & v_1 & 1 \\ u_2 & v_2 & 1 \\ u_3 & v_3 & 1 \end{pmatrix}\right| = 0.$$

If $\mathcal{R}$ is an integral domain the above criterion is also sufficient. The proof is easy and left to the reader.

Definition 6.3 *By $\overline{\mathcal{I}}(\mathcal{R}, 2)$ we denote the maximum cardinality of an integral point set with no three collinear points.*

Lemma 6.4

$$\overline{\mathcal{I}}(\mathcal{R}, 2) \leqslant 2 \cdot |\mathcal{R}|.$$

PROOF. We ignore the integrality condition and consider the lines $l_i = \{(i,r) \mid r \in \mathcal{R}\}$ for all $i \in \mathcal{R}$. □

Lemma 6.5 *If $-1 \in \square_q$ we have $\overline{\mathcal{I}}(\mathbb{F}_q, 2) \leqslant \frac{q+3}{2}$ and for $-1 \notin \square_q$ we have $\overline{\mathcal{I}}(\mathbb{F}_q, 2) \leqslant \frac{q+1}{2}$.*

PROOF. Let $\mathcal{P}$ be an integral point set over $\mathbb{F}_q$ without a collinear triple. We choose a point $p \in \mathcal{P}$. The directions of p to the other points p' of $\mathcal{P}$ are pairwise different. Since there are at most $\frac{q+3}{2}$ or $\frac{q+1}{2}$ different directions in an integral point set over $\mathbb{F}_q$ (Lemma 4.3), we obtain $|\mathcal{P}| \leqslant \frac{q+5}{2}$ for $-1 \in \square_q$ and $|\mathcal{P}| \leqslant \frac{q+3}{2}$ for $-1 \notin \square_q$. Suppose that this upper bound is achieved. So all points must have exactly one *neighbor* in direction 0 and one in direction ∞. Thus $|\mathcal{P}|$ must be even in this case, which is a contradiction due to Lemma 3.6. □

Using an element $z \in \mathcal{R}'$ with $z\bar{z} = 1$ we can describe a good construction for lower bounds. Actually this equation describes something like a circle with radius one. An example for $q = 31$ is depicted in Figure 3.

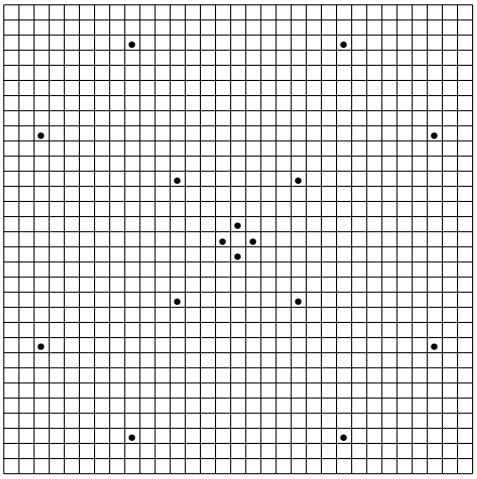


Figure 3: Integral point set corresponding to the construction from Lemma 6.6 for $q = 31$.

Lemma 6.6 *For $z \in \mathcal{R}'$ with $z\bar{z} = 1$ the set $\mathcal{P} = \left\{z^{2i} \mid i \in \mathbb{N}\right\}$ is an integral point set.*

PROOF. With $c := a - b$ we have

$$\begin{aligned} d\left(z^{2a}, z^{2b}\right) &= \left(z^{2a} - z^{2b}\right) \cdot \overline{\left(z^{2a} - z^{2b}\right)} \\ &= \left(z^{2c} - 1\right) \cdot \overline{z^{2c} - 1} \\ &= 2 - z^{2c} 2 - \overline{z^{2c}} \\ &= \left(\underbrace{z^c i - \overline{z^c} i}_{\in \mathcal{R}}\right)^2. \end{aligned}$$

□

We remark that the set $\mathcal{P}' = \left\{z^{2i+1} \mid i \in \mathbb{N}\right\}$ is an isomorphic integral point set. The set of solutions of $z\bar{z} = 1$ forms a cyclic multiplicative group G due to Lemma 3.8. From Lemma 3.7 we know that G has size $q+1$ for $-1 \notin \square_q$ and size $q-1$ if $-1 \in \square_q$. So by Lemma 6.6 we get a construction of an integral point set in $\mathbb{F}_q$ which is near the upper bound of Lemma 6.5. We only have to prove that our construction does not produce three collinear points in $\mathbb{F}_q$.

Lemma 6.7 *For $\mathcal{R} = \mathbb{F}_q$ with $2 \nmid q$ the point set from Lemma 6.6 contains no collinear triple.*

PROOF. We assume that we have three pairwise different points p_1, p_2, p_3 in $\mathcal{R}'$ which are collinear. So there exist a, b, c, d, t_1, t_2, and t_3 in $\mathcal{R}$ fullfilling

$$\begin{aligned} p_1 &= a + bt_1 + (c + dt_1)\,i, \\ p_2 &= a + bt_2 + (c + dt_2)\,i, \\ p_3 &= a + bt_3 + (c + dt_3)\,i, \end{aligned}$$

and $t_i \neq t_j$ for $i \neq j$. Since $p_i\overline{p_i} = 1$ we have

$$\begin{aligned} a^2 + 2abt_1 + b^2t_1^2 + c^2 + 2cdt_1 + d^2t_1^2 &= 1, \\ a^2 + 2abt_2 + b^2t_2^2 + c^2 + 2cdt_2 + d^2t_2^2 &= 1, \\ a^2 + 2abt_3 + b^2t_3^2 + c^2 + 2cdt_3 + d^2t_3^2 &= 1. \end{aligned}$$

Subtracting the first two and the last two equations yields

$$\begin{aligned} 2ab\,(t_1 - t_2) + b^2\,(t_1 - t_2)\,(t_1 + t_2) & \\ +2cd\,(t_1 - t_2) + d^2\,(t_1 - t_2)\,(t_1 + t_2) &= 0, \\ 2ab\,(t_2 - t_3) + b^2\,(t_2 - t_3)\,(t_2 + t_3) & \\ +2cd\,(t_2 - t_3) + d^2\,(t_2 - t_3)\,(t_2 + t_3) &= 0. \end{aligned}$$

Because $t_1 \neq t_2$, $t_2 \neq t_3$ and $\mathcal{R}$ is an integral domain we obtain

$$\begin{aligned} 2ab + b^2\,(t_1 + t_2) + 2cd + d^2\,(t_1 + t_2) &= 0, \\ 2ab + b^2\,(t_2 + t_3) + 2cd + d^2\,(t_2 + t_3) &= 0. \end{aligned}$$

Another subtraction yields

$$b^2\,(t_1 - t_3) + d^2\,(t_1 - t_3) = 0 \quad \Rightarrow \quad b^2 + d^2 = 0.$$

Inserting yields

$$2ab + 2cd = 0 \quad \Leftrightarrow \quad 2ab = -2cd$$

and

$$a^2 + c^2 = 1.$$

Thus

$$4a^2b^2 = 4c^2d^2 \;\Leftrightarrow\; \left(a^2 + c^2\right)4b^2 = 0 \;\Leftrightarrow\; b = 0.$$

In the same way we obtain $d = 0$ and so $p_1 = p_2 = p_3$, which is a contradiction. □

Corollary 6.8 *For $-1 \notin \square_q$ we have $\overline{\mathcal{I}}\,(\mathbb{F}_q, 2) = \frac{q+1}{2}$ and for $-1 \in \square_q$ we have $\frac{q-1}{2} \leqslant \overline{\mathcal{I}}\,(\mathbb{F}_q, 2) \leqslant \frac{q+3}{2}$.*

Conjecture 6.9 *For $-1 \in \square_q$ we have $\overline{\mathcal{I}}\,(\mathbb{F}_q, 2) = \frac{q-1}{2}$.*

We remark that Conjecture 6.9 would be true for quadrance graphs. Following the proof of Lemma 6.5 we would get $\frac{q-1}{2}$ as an upper bound for $q \equiv 1 \mod 4$. Since $z^c - \overline{z^c} = 0$ would imply $2c = q - 1$ the construction from Lemma 6.6 does not contain a pair of points with squared distance 0.

7 Integral point sets in general position

Our best construction for integral point sets where no three points are collinear consists of points on a *circle*. So it is interesting to study integral point sets where additionally no 4 points are allowed to be situated on a *circle*.

Definition 7.1 *Points $p_i = (x_i, y_i)$ in $\mathcal{R}^2$ are said to be situated on a circle if there exist $a, b, r \in \mathcal{R}$ with $(x_i - a)^2 + (y_i - b)^2 = r$ for all i.*

We have the following condition:

Lemma 7.2 *Four distinct points $p_i = (x_i, y_i)$ in $\mathbb{F}_q^2$ which contain no collinear triple are situated on a circle if and only if*

$$\left|\begin{pmatrix} x_1 & y_1 & x_1^2 + y_1^2 & 1 \\ x_2 & y_2 & x_2^2 + y_2^2 & 1 \\ x_3 & y_3 & x_3^2 + y_3^2 & 1 \\ x_4 & y_4 & x_4^2 + y_4^2 & 1 \end{pmatrix}\right| = 0.$$

PROOF. If there exist $a, b, r \in \mathbb{F}_q$ with $(x_i - a)^2 + (y_i - b)^2 = r$ for all $1 \leqslant i \leqslant 4$ then the determinant clearly vanishes since $r = (x_i - a)^2 + (y_i - b)^2 = \left(x_i^2 + y_i^2\right) - 2a \cdot x_i - 2b \cdot y_i + \left(a^2 + b^2\right)$. For the other direction we consider the unique circle $\mathcal{C}$ through the points (x_1, y_1), (x_2, y_2), (x_3, y_3) described by the parameters $a, b, r \in \mathbb{F}_q$. With the same idea as before we get

$$\left|\begin{pmatrix} x_1 & y_1 & 0 & 1 \\ x_2 & y_2 & 0 & 1 \\ x_3 & y_3 & 0 & 1 \\ x_4 & y_4 & (x_4 - a)^2 + (y_4 - b)^2 - r & 1 \end{pmatrix}\right| = 0.$$

If (x_4, y_4) is not on the circle $\mathcal{C}$ then we can develop the determinant after the third column and obtain

$$\left|\begin{pmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{pmatrix}\right| = 0.$$

which is a contradiction to the fact that (x_1, y_1), (x_2, y_2), and (x_3, y_3) are not collinear, see Lemma 6.2. □

We remark that for arbitrary commutative rings $\mathcal{R}$ the determinant criterion from Lemma 7.2 is a necessary condition.

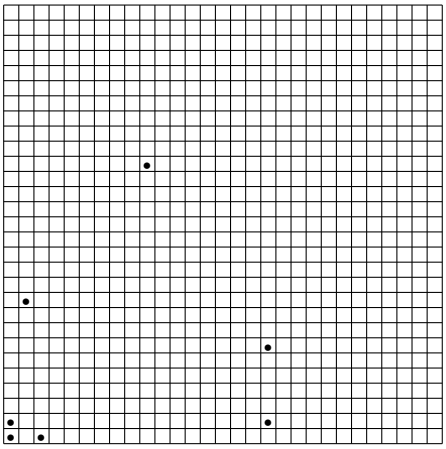


Figure 4: A maximum integral point set in general position over $\mathbb{F}_{29}^2$.

Definition 7.3 *By $\dot{\jmath}(\mathcal{R},2)$ we denote the maximum cardinality of an integral point set in $\mathcal{R}^2$ which is in general position, this means that it contains no collinear triple and no four points on a circle.*

As a shorthand for the conditions of Definition 7.3 we also say that the points are in general position. An example of seven points over $\mathbb{F}_{29}^2$ in general position which pairwise integral distances is depicted in Figure 4. As trivial upper bound we have $\dot{\jmath}(\mathcal{R},2) \leqslant \bar{\jmath}(\mathcal{R},2)$. By applying the automorphisms of $\mathbb{F}_q^2$ with respect to Δ we see that they conserve circles.

p	$\dot{\jmath}$	p	$\dot{\jmath}$	p	$\dot{\jmath}$	p	$\dot{\jmath}$	p	$\dot{\jmath}$
2	4	17	5	41	9	67	9	97	11
3	2	19	5	43	8	71	11	101	13
5	4	23	5	47	7	73	10	103	11
7	3	29	7	53	9	79	11	107	11
11	4	31	6	59	9	83	11	109	12
13	5	37	7	61	10	89	11	113	12

Table 1: Values of $\dot{\jmath}(\mathbb{F}_p,2) = \dot{\jmath}(\mathbb{Z}_p,2)$ for small primes p.

Via an exhaustive combinatorial search we have determined $\dot{\jmath}(\mathbb{F}_p,2)$ for small values of p, see Table 1. Since it is a non-trivial task to determine these numbers exactly, at least for $p \geqslant 100$, we give an outline of our used algorithm.

Algorithm 7.4 ***(Generation of integral point sets in general position over $\mathbb{F}_q$)***
Input: q
Output: Integral point sets $\mathcal{P} \subset \mathbb{F}_q$ in general position
begin
  $\mathcal{P} = [(0,0),(0,1)]$
  $\mathtt{blocked}[(0,0)] = \mathtt{blocked}[(0,1)] = \mathtt{true}$
  loop over $d \in \mathbb{F}_q$ $\mathcal{L}_d = []$ **end**
  loop over $x \in \mathbb{F}_q^2 \setminus \{(0,0),(0,1)\}$
    blocked$[x] = \mathtt{false}$
    if $\Delta((0,0),x) = 0$ **or** $\Delta((0,1),x) = 0$
    then *blocked*$[x] = \mathtt{true}$ **end**
    if $\mathtt{collinear}((0,0),(0,1),x)$
    then *blocked*$[x] = \mathtt{true}$ **end**
    if *blocked*$[x] = \mathtt{true}$
    then $\mathcal{L}_{\mathtt{get\_direction}(x)}.\mathtt{append}(x)$ **end**
  end
  $\mathtt{add\_point}(\mathcal{P},0)$
end

So far almost nothing is done. We restrict our search to integral point sets $\mathcal{P}$ of cardinality at least 3. So we may assume that $\mathcal{P}$ contains the points $(0,0)$ and $(0,1)$. For each $x \in \mathbb{F}_q^2$ the variable $\mathtt{blocked}[x]$ says whether x can be appended to $\mathcal{P}$ without destroying the property integral point set or general position. The lists $\mathcal{L}_d$ cluster the points of $\mathbb{F}_q^2$ according to their direction. The fact that $\mathcal{P}$ can contain besides $(0,0)$ and $(0,1)$ at most one member from each $\mathcal{L}_d$ can be used to prune the search tree if one searches only for integral point sets with maximum cardinality.

Algorithm 7.5 ***(add_point)***
Input: Lower bound l on the direction and an integral point set $\mathcal{P}$
Output: Integral point sets $\mathcal{P} \subset \mathbb{F}_q$ in general position
begin
  loop over $d \in \mathbb{F}_q$ **with** $d \geqslant l$
    loop over $x \in \mathcal{L}_d$ **with** $\mathtt{blocked}[x] = \mathtt{false}$
      if $\mathtt{canon\_check}(\mathcal{P},x) = \mathtt{true}$ **then**
        $\mathcal{P}.\mathtt{append}(x)$
        block all $y \in \mathbb{F}_q^2$ *where* $\Delta(y,x) = 0$
        or $\mathtt{collinear}(p_1,x,y) = \mathtt{true}$
        or $\mathtt{on\_circle}(p_1,p_2,x,y) = \mathtt{true}$
        for $p_1,p_2 \in \mathcal{P}$
        output $\mathcal{P}$
        $\mathtt{add\_point}(\mathcal{P},d+1)$
        unblock
        $\mathcal{P}.\mathtt{remove}(x)$
      end
    end
  end
end

The subroutine $\mathtt{add\_point}$ simply adds another point to the point set $\mathcal{P}$ and maintains the set of further candidates for adding a further point. Some lookahead is possible to implement. Since the automorphism group of $\mathbb{F}_q^2$ with respect to Δ is very large we would obtain lots of isomorphic integral point sets if we do without isomorphism pruning. With the framework of orderly generation, see e. g. [21], it is possible to write a subroutine $\mathtt{canon\_check}$ that let our algorithm output a complete list of pairwise non-isomorphic integral point sets in general position. For our purpose it suffices to have a subroutine $\mathtt{canon\_check}$ that rejects the majority of isomorphic copies but as a return has a good performance. Let $m : \mathbb{F}_q^2 \to \mathbb{F}_q^2$, $(x,y) \mapsto (-x,y)$ the automorphism that mirrors at the y-axis and let $\preceq$ be a total ordering on the points of $\mathbb{F}_q^2$ if $u \prec v$ for $\mathtt{direction}(u) < \mathtt{direction}(v)$. For the latter comparison we use an arbitrary but fix total ordering of $\mathbb{F}_q$, where 0 is the smallest element and which

is also used for the looping over $\mathbb{F}_q$. By $\mathcal{P}[2]$ we denote the third point of a list $\mathcal{P}$.

Algorithm 7.6 ***(canon_check)***
Input: An integral point set $\mathcal{P}$
Output: Returns false if $\mathcal{P}$ should be rejected due to isomorphism pruning
begin
  loop over *some disjoint triples* $(u, v, w) \in \mathcal{P} \times \mathcal{P} \times \mathcal{P}$
  with $\delta^2(u, v) \neq 0$
    determine an automorphism α *with* $\alpha(u) = (0, 0)$
    and $\alpha(v) = (0, 1)$
    if $\alpha(w) \prec \mathcal{P}[2]$ **or** $m(\alpha(w)) \prec \mathcal{P}[2]$
    then return `false` **end**
  end
  return `true`
end

For further examples we refer to [15] where we list the coordinates of one extremal example for $p \leqslant 113$.

A formal proof of the correctness of the proposed algorithm is not difficult but a bit technical and so left to the reader. We remark that there are several non-isomorphic integral point sets in general position which achieve the upper bound $\dot{\jmath}(\mathbb{Z}_n, 2)$. So far we have no insight in their structure or in the asymptotic behavior of $\dot{\jmath}(\mathbb{Z}_n, 2)$. It seems that we have $\dot{\jmath}(\mathbb{Z}_p, 2) \geqslant 7$ for all sufficiently large primes p. This is interesting because the question whether $\dot{\jmath}(\mathbb{Z}, 2) \geqslant 7$ is unsolved so far. In other words, there is no known 7_2-cluster [9]. This is a set of seven points in the plane, no three points on a line, no four points on a circle, where the coordinates and the pairwise distances are integral.

Conjecture 7.7 *For each l there is a p' so that for all $p \geqslant p'$ we have $\dot{\jmath}(\mathbb{Z}_p, 2) \geqslant l$.*

8 Conclusion and outlook

In this paper we have considered sets of points $\mathcal{P}$ in the affine plane $\mathrm{AG}(2, q)$ with pairwise integral distances. We have presented several connections to other discrete structures and problems. Some questions concerning maximum cardinalities and complete classifications of extremal examples remain open. Clearly similar questions could be asked in $\mathrm{AG}(3, q)$ or higher dimensional spaces.

Acknowledgment

I am thankful to Aart Blokhuis, Stancho Diemiev, Michael Kiermaier, Harald Meyer, and Ivo Radloff whose comments were very helpful during the preparation of this article.

Bibliography

[1] A. Antonov and M. Brancheva, *Algorithm for finding maximal Diophantine figures*, Spring Conference 2007 of the Union of Bulgarian Mathematicians, 2007.

[2] R. D. Baker, G. L. Ebert, J. Hemmeter, and A. Woldar, *Maximal cliques in the Paley graph of square order*, J. Statist. Plann. Inference **56** (1996), no. 1, 33–38.

[3] A. Blokhuis, *On subsets of* $\mathrm{GF}(q^2)$ *with square differences*, Indag. Math. **46** (1984), 369–372.

[4] A. Blokhuis, S. Ball, A. E. Brouwer, L. Storme, and T. Szőnyi, *On the number of slopes of the graph of a function defined on a finite field*, J. Combin. Theory Ser. A **86** (1999), no. 1, 187–196.

[5] P. J. Cameron, G. R. Omidi, and B. Tayfeh-Rezaie, 3-*designs from PGL*$(2, q)$, Electron. J. Combin. **13** (2006), no. 1.

[6] L. Carlitz, *A theorem on permutations in a finite field*, Proc. Amer. Math. Soc. **11** (1960), 456–459. MR MR0117223 (22 #8005)

[7] L. E. Dickson, *Linear groups. With an exposition of the Galois field theory. With an introduction by Wilhelm Magnus. Unabridged and unaltered republ. of the first ed.*, New York: Dover Publications, Inc. XVI, 312 p., 1958.

[8] S. Dimiev, *A setting for a Diophantine distance geometry*, Tensor (N.S.) **66** (2005), no. 3, 275–283. MR MR2189847

[9] R. K. Guy, *Unsolved problems in number theory. 2nd ed.*, Unsolved Problems in Intuitive Mathematics. 1. New York, NY: Springer- Verlag. xvi, 285 p., 1994.

[10] H. Harborth, *Integral distances in point sets*, Butzer, P. L. (ed.) et al., Karl der Grosse und sein Nachwirken. 1200 Jahre Kultur und Wissenschaft in Europa. Band 2: Mathematisches Wissen. Turnhout: Brepols, 1998, 213–224.

[11] B. Huppert, *Endliche Gruppen. I*, Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen. 134. Berlin-Heidelberg-New York: Springer-Verlag. XII, 793 S. mit 15 Abb., 1967.

[12] T. Khoon Lim and C. E. Praeger, *On generalised Paley graphs and their automorphism groups*, Michigan Math. J. (to appear).

[13] A. Kohnert and S. Kurz, *Integral point sets over* $\mathbb{Z}_n^m$, Electron. Notes Discrete Math. **27** (2006), 65–66.

[14] T. Kreisel and S. Kurz, *There are integral heptagons, no three points on a line, no four on a circle*, Discrete Comput. Geom., (to appear).

[15] S. Kurz, *Coordinates of maximal integral point sets over* $\mathbf{F}_q^2$ *in general position*, http://www.wm.uni-bayreuth.de/index.php?id=322.

[16] S. Kurz, *Konstruktion und Eigenschaften ganzzahliger Punktmengen*, Ph.D. thesis, Bayreuth. Math. Schr. 76. Universität Bayreuth, 2006.

[17] S. Kurz and A. Wassermann, *On the minimum diameter of plane integral point sets*, Ars Combin, (to appear).

[18] A. Neumaier, *Cliques and claws in edge-transitive strongly regular graphs*, Math. Z. **174** (1980), 197–202.

[19] A. Neumaier, *Regular cliques in graphs and special 1,5-designs*, Finite geometries and designs, Proc. 2nd Isle of Thorns Conf. 1980, Lond. Math. Soc. Lect. Note Ser. 49, 244–259, 1981.

[20] L. C. Noll and D. I. Bell, n-*clusters for* $1 < n < 7$, Math. Comp. **53** (1989), no. 187, 439–444.

[21] R. C. Read, *Every one a winner or how to avoid isomorphism search when cataloguing combinatorial configurations*, Ann. Discrete Math. **2** (1978), 107–120.

[22] W. M. Schmidt, *Equations over finite fields. An elementary approach. 2nd ed.*, Heber City, UT: Kendrick Press. x, 333 p., 2004.

[23] J. Sheehan, *Finite Ramsey theory is hard*, Combinatorial mathematics VIII, Proc. 8th Aust. Conf., Geelong/ Aust. 1980, Lect. Notes Math. 884, 99-106 (1981).

[24] S. Y. Song, *Commutative association schemes whose symmetrizations have two classes*, J. Algebraic Combin. **5** (1996), no. 1, 47–55.

[25] D. B. Surowski, *Automorphism groups of certain unstable graphs*, Math. Slovaca **53** (2003), no. 3, 215–232.

[26] P. Sziklai, *On subsets of* $GF(q^2)$ *with* d*th power differences*, Discrete Math. **208-209** (1999), 547–555.

[27] L. A. Vinh, *On chromatic number of unit-quadrance graphs (finite euclidean graphs)*, ArXiv Mathematics math/0510092 (2005).

[28] L. A. Vinh, *Quadrance polygons, association schemes and strongly regular graphs*, ArXiv Mathematics math/0509598 (2005).

[29] L. A. Vinh, *Quadrance graphs*, Austral. Math. Soc. Gaz. **33** (2006), no. 5, 330–332.

[30] L. A. Vinh, *Some colouring problems for unit-quadrance graphs*, ArXiv Mathematics math/0606482 (2006).

[31] N. Wage, *Character sums and Ramsey properties of generalized Paley graphs*, Integers **6** (2006).

[32] N. J. Wildberger, *Divine proportions. Rational trigonometry to universal geometry*, Kingsford: Wild Egg. xx, 300 p., 2005.

Chapter 8

Integral point sets in higher dimensional affine spaces over finite fields

SASCHA KURZ[1] AND HARALD MEYER[2]

ABSTRACT. We consider point sets in m-dimensional affine space $\mathbb{F}_q^m$ where each squared Euclidean distance of two points is a square in $\mathbb{F}_q$ and determine the automorphism group of these spaces, which preserves integral distances. For some small parameters m and q we determine the maximum cardinality $\mathfrak{I}(\mathbb{F}_q, m)$ of integral point sets in $\mathbb{F}_q^m$. For $m = 3$ we give upper bounds and for general m we give some lower bounds on $\mathfrak{I}(\mathbb{F}_q, m)$. If we map integral distances to edges in a graph, we can define a graph $\mathfrak{G}_{m,q}$ with vertex set $\mathbb{F}_q^m$. It turns out that $\mathfrak{G}_{m,q}$ arises from a 3-class association scheme and even is strongly regular for some cases.

2000 MSC: 51E15;05D99,05B25,05E30,20B25.
Key words and phrases: finite geometry, integral distances, integral point sets, automorphism group, association schemes, strongly regular graphs.

1 Introduction and Notation

Let p be a prime and let q be a power of p. We write $\mathbb{F}_q$ for the field with q elements and $\mathbb{F}_q^* := \mathbb{F}_q \backslash \{0\}$ for the units of $\mathbb{F}_q$. Our notation $\mathrm{GL}(m, \mathbb{F}_q)$ for the general linear group is standard. By

$$\mathrm{O}(m, \mathbb{F}_q) := \left\{ A \in \mathrm{GL}(m, \mathbb{F}_q) \mid A^T A = A A^T = E_m \right\},$$

where E_m is the $m \times m$ identity matrix, we denote the orthogonal group in dimension m. We remark that for even dimension $2n$ the orthogonal group comes in two types $\mathrm{O}^+(2n, \mathbb{F}_q)$ and $\mathrm{O}^-(2n, \mathbb{F}_q)$, and the group we defined above is isomorphic to $\mathrm{O}^+(2n, \mathbb{F}_q)$ in this case. By $A\Gamma\mathrm{L}(m, \mathbb{F}_q)$ we denote the affine general semilinear group over $\mathbb{F}_q$. It is well known that the center $Z := Z(\mathrm{GL}(m, \mathbb{F}_q))$ consists of the diagonal matrices with equal entries at the diagonal, i. e. the corresponding linear map is just a multiplication with an element $d \in \mathbb{F}_q^*$.

[1]Sascha Kurz, University of Bayreuth, Department of Mathematics, 95440 Bayreuth, Germany.
E-mail adress: sascha.kurz@uni-bayreuth.de
[2]Harald Meyer, University of Bayreuth, Department of Mathematics, 95440 Bayreuth, Germany.
E-mail adress: harald.meyer@uni-bayreuth.de

Originally integral point sets were defined in m-dimensional Euclidean spaces $\mathbb{E}^m$ as sets of n points with pairwise integral distances in the Euclidean metric, see e. g. [6, 11, 12, 14, 15] for an overview on the most recent results. Here we consider integral point sets in the affine spaces $\mathbb{F}_q^m$. We equip those spaces with a bilinear form

$$\langle u, v \rangle := u^T v = \sum_{i=1}^{m} u_i v_i$$

and a squared distance

$$\begin{aligned} d^2(u, v) &:= \langle u - v, u - v \rangle \\ &= (u - v)^T (u - v) \\ &= \sum_{i=1}^{m} (u_i - v_i)^2 \quad \in \mathbb{F}_q \end{aligned}$$

for any two points $u = (u_1, \ldots, u_m)$, $v = (v_1, \ldots, v_m)$ in $\mathbb{F}_q^m$. We say that two points $u, v \in \mathbb{F}_q^m$ are at integral distance if $d^2(u, v)$ is contained in the set $\square_{\mathbb{F}_q} := \left\{ r^2 \mid r \in \mathbb{F}_q \right\}$ consisting of the squares in $\mathbb{F}_q$. As in the Euclidian space we define the cross product of two vectors $u, v \in \mathbb{F}_q^3$ by

$$u \times v := (u_2 v_3 - u_3 v_2, -u_1 v_3 + u_3 v_1, u_1 v_2 - u_2 v_1) \in \mathbb{F}_q^3.$$

The proofs of some of the common formulas for the cross product do not depend on any specific attributes of the Euclidian space, so they still hold in $\mathbb{F}_q^3$. Especially this is true for the formulas

$$\langle u \times v, u \rangle = \langle u \times v, v \rangle = 0$$

and

$$\langle u \times v, u \times v \rangle = \langle u, u \rangle \cdot \langle v, v \rangle - \langle u, v \rangle^2$$

we will use later on. The notation

$$U^\perp := \left\{ v \in \mathbb{F}_q^m \mid \langle u, v \rangle = 0 \text{ for all } u \in U \right\}$$

for a subspace $U \leqslant \mathbb{F}_q^m$ is also inspired by the notation for the Euclidian space.

Using the Kronecker delta $\delta_{ij} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases}$ we can state the equivalence

$$\begin{aligned} & A = (a_1 | \ldots | a_m) \in O(m, \mathbb{F}_q) \\ \iff \quad & \langle a_i, a_j \rangle = \delta_{ij} \quad \text{for all } 1 \leqslant i, j \leqslant m \end{aligned}$$

and we have the equation

$$\langle Av, Aw \rangle = (Av)^T (Aw) = v^T A^T A w = v^T w = \langle v, w \rangle$$

for all $v, w \in \mathbb{F}_q^m$ and all $A \in O(m, \mathbb{F}_q)$. If we have a matrix $A \in GL(m, \mathbb{F}_q)$ with $\langle Av, Aw \rangle = \langle v, w \rangle$ for all $v, w \in \mathbb{F}_q^m$, then we have

$$v^T A^T A w = (Av)^T (Aw) = \langle Av, Aw \rangle = \langle v, w \rangle = v^T w$$

for all $v, w \in \mathbb{F}_q^m$, i. e. $A^T A = E_m$ and so A is an element of $O(m, \mathbb{F}_q)$.

1.1 Related work

There has been done a lot of work on integral point sets in Euclidean spaces, see e. g. [6, 11, 12, 14, 15]. Some authors also consider other spaces, e. g. Banach spaces [5], integral point sets over rings [10], or integral point sets over finite fields [1, 4, 8, 13]. In [13] one of the authors of this article determines the automorphism group for dimension $m = 2$ and in [8] integral point sets over $\mathbb{F}_q^2$ which are maximal with respect to inclusion were classified for $q \leqslant 47$. For $m = 2$ and $q \equiv 3 \mod 4$ the graphs $\mathfrak{G}_{m,q}$ from Section 4 are isomorphic to Paley graphs of square order. In [3] Blokhuis has determined the structure of cliques of maximal size in Paley graphs of square order.

1.2 Our contribution

We determine the automorphism group of $\mathbb{F}_q^m$ with respect to integral distances and analyze the graphs of integral distances $\mathfrak{G}_{m,q}$ for $m \geqslant 3$, $2 \nmid q$. They arise from 3-class association schemes. The determination of some of their parameters let us conjecture that they are strongly regular for even dimensions m. We give some new exact numbers $\mathcal{I}(\mathbb{F}_q, 3)$ and upper bounds for the maximum cardinality of integral point sets over $\mathbb{F}_q^3$ for dimension $m = 3$. For general dimension n we give some constructions yielding lower bounds.

1.3 Organization of the paper

After the short introduction and the basic notation in this section we give the basic facts on integral point sets over commutative rings in Section 2. In Section 3 we completely determine the automorphism group of $\mathbb{F}_q^m$ with respect to integral distances and analyze its operation on $\mathbb{F}_q^m$. We introduce and analyze the graphs of integral distances $\mathfrak{G}_{m,q}$ in Section 4. In Section 5 we consider the maximum cardinality $\mathcal{I}(\mathbb{F}_q, m)$ of integral point sets over $\mathbb{F}_q^m$. We finish with a conclusion and an outlook in Section 6.

2 Integral point sets

From a more general point of view one can define integral point sets over commutative rings with 1. If not stated otherwise we assume that $\mathcal{R}$ is a commutative ring with 1 and consider sets of elements of the $\mathcal{R}$-module $\mathcal{R}^m$. We speak of these elements as points with a geometric interpretation in mind. For our purpose we equip the module $\mathcal{R}^m$ with something similar to an Euclidean metric:

Definition 2.1 *For two points $u = (u_1, \ldots, u_m)$, $v = (v_1, \ldots, v_m)$ in $\mathcal{R}^m$ we define the **squared distance** as*

$$d^2(u, v) := \sum_{i=1}^{m} (u_i - v_i)^2 \quad \in \mathcal{R}.$$

We are interested in those cases where $d^2(u, v)$ is contained in the set $\square_{\mathcal{R}} := \{r^2 \mid r \in \mathcal{R}\}$ of squares of $\mathcal{R}$.

Definition 2.2 *Two points $u = (u_1, \ldots, u_m)$, $v = (v_1, \ldots, v_m)$ in $\mathcal{R}^m$ are at **integral distance** if there exists an element r in $\mathcal{R}$ with $d^2(u, v) = r^2$. As a shorthand we define $\Delta : \mathcal{R}^m \times \mathcal{R}^m \to \{0, 1\}$,*

$$(u, v) \mapsto \begin{cases} 1 & \text{if } u \text{ and } v \text{ are at integral distance}, \\ 0 & \text{otherwise}. \end{cases}$$

*A set $\mathcal{P}$ of points in $\mathcal{R}^m$ is called an **integral point set** if all pairs of points are at integral distance.*

If $\mathcal{R}$ is a finite ring it makes sense to ask for the maximum cardinality of an integral point set in $\mathcal{R}^m$.

Definition 2.3 *By $\mathcal{I}(\mathcal{R}, m)$ we denote the maximum cardinality of an integral point set in $\mathcal{R}^m$.*

Lemma 2.4

$$|\mathcal{R}| \leqslant \mathcal{I}(\mathcal{R}, m) \leqslant |\mathcal{R}|^m.$$

PROOF. For the lower bound we consider the *line* $\mathcal{P} = \{(r, 0, \ldots, 0) \mid r \in \mathcal{R}\}$. □

Lemma 2.5 *If $\mathcal{R}$ has characteristic 2, meaning that $1 + 1 = 0$ holds, then we have $\mathcal{I}(\mathcal{R}, m) = |\mathcal{R}|^m$.*

PROOF. For two points $u = (u_1, \ldots, u_m)$, $v = (v_1, \ldots, v_m)$ in $\mathcal{R}^m$ we have

$$d^2(u, v) = \sum_{i=1}^{m} (u_i - v_i)^2 = \underbrace{\left(\sum_{i=1}^{r} u_i + v_i \right)^2}_{\in \mathcal{R}}.$$

□

So in the remaining part of this article we consider only $\mathcal{R} = \mathbb{F}_q$ where two does not divide q.

Also the lower bound of Lemma 2.4 is attained in some cases, see e. g. [13] for a proof:

Theorem 2.6

$$\mathcal{I}(\mathbb{F}_q, 2) = q \text{ for } 2 \nmid q.$$

3 Automorphisms preserving integral distances

As shorthand we use $\square_q := \square_{\mathbb{F}_q}$. We need some facts about roots in $\mathbb{F}_q$ and the set of solutions of quadratic equations in $\mathbb{F}_q$.

Definition 3.1 *For* $q \equiv 1 \mod 4$ *we denote by* ω_q *an element with* $\omega_q^2 = -1$.

Lemma 3.2 *For a finite field* $\mathbb{F}_q$ *with* $q = p^r$ *and* $p \neq 2$ *we have* $-1 \in \square_q$ *iff* $q \equiv 1 \mod 4$, $\omega_q \in \square_q$ *iff* $q \equiv 1 \mod 8$, *and* $2 \in \square_q$ *iff* $q \equiv \pm 1 \mod 8$.

PROOF. The multiplicative group of the units $\mathbb{F}_q^*$ is cyclic of order $q-1$. Elements of order 4 are exactly those elements x with $x^2 = -1$. A similar argument holds for the the fourth roots of -1. For the last statement we have to generalize the second Ergänzungssatz of the quadratic reciprocity law to $\mathbb{F}_q$: If $q = p$ is prime then the statement is true by the second Ergänzungssatz. If 2 is a square in $\mathbb{F}_p$ then 2 is also a square in $\mathbb{F}_{p^r}$ and from $p \equiv \pm 1 \mod 8$ we get $q = p^r \equiv \pm 1 \mod 8$, i. e. the statement is true in this case. If 2 is not a square in $\mathbb{F}_p$ then the polynomial $x^2 - 2 \in \mathbb{F}_p[x]$ is irreducible and 2 is a square in $\mathbb{F}_p[x]/\left(x^2-2\right) \cong \mathbb{F}_{p^2}$. Hence 2 is a square in $\mathbb{F}_{p^k}$ iff $2 \mid k$. As p is an odd prime with $p \equiv \pm 3 \mod 8$ we obtain $p^2 \equiv 1 \mod 8$ and also $p^{2m} \equiv 1 \mod 8$ as well as $p^{2m+1} \equiv \pm 3$ in this case. □

Definition 3.3 *A triple* (a,b,c) *is called* Pythagorean triple over $\mathbb{F}_q$ *if* $a^2 + b^2 = c^2$.

In the following it will be useful to have a parametric representation of the Pythagorean triples over $\mathbb{F}_q$.

Lemma 3.4 *For* $2 \nmid q$ *let* $c \in \mathbb{F}_q$ *and let* P_c *be the set of Pythagorean triples* (a,b,c) *over* $\mathbb{F}_q$.

(a) If $c = 0$ *then*

$$P_0 = \begin{cases} \{(t, \pm t\omega_q, 0) \mid t \in \mathbb{F}_q\} & \text{if } q \equiv 1 \mod 4 \\ \{(0,0,0)\} & \text{if } q \equiv 3 \mod 4 \end{cases}$$

and

$$|P_0| = \begin{cases} 2q-1 & \text{if } q \equiv 1 \mod 4 \\ 1 & \text{if } q \equiv 3 \mod 4. \end{cases}$$

(b) If $c \neq 0$ *then*

$$P_c = \{(\pm c, 0, c)\} \cup \{(0, \pm c, c)\} \cup \left\{ \left(\tfrac{t^2-1}{t^2+1} \cdot c, \tfrac{2t}{t^2+1} \cdot c, c \right) \mid t \in \mathbb{F}_q^*, t^2 \neq \pm 1 \right\}$$

and

$$|P_c| = \begin{cases} q-1 & \text{if } q \equiv 1 \mod 4 \\ q+1 & \text{if } q \equiv 3 \mod 4. \end{cases}$$

(c) There are exactly q^2 *Pythagorean Triples over* $\mathbb{F}_q$.

PROOF. Part (a) is easy to verify. For part (b) there are 4 solutions with $ab = 0$, these are $\{(0, \pm c, c), (\pm c, 0, c)\}$. For $ab \neq 0$ we get:

$$a^2 + b^2 = c^2 \quad \Leftrightarrow \quad \frac{c-a}{b} \cdot \frac{c+a}{b} = 1.$$

Setting $t := \frac{c+a}{b} \in \mathbb{F}_q^*$ we obtain $t^{-1} = \frac{c-a}{b}$, hence

$$\frac{a}{b} = \frac{t - t^{-1}}{2} \quad \text{and} \quad \frac{c}{b} = \frac{t + t^{-1}}{2}.$$

Because of $a \neq 0$, $c \neq 0$ we have $t \neq \pm t^{-1}$, i. e. $t^2 \notin \{-1, 1\}$. It follows

$$a = \frac{t - t^{-1}}{t + t^{-1}} \cdot c \quad \text{and} \quad b = \frac{2}{t + t^{-1}} \cdot c.$$

It is easily checked that for all admissible values of t, the resulting triples (a,b,c) are pairwise different Pythagorean triples.

The expression for the number of solutions follows because -1 is a square in $\mathbb{F}_q$ exactly if $q \equiv 1 \mod 4$.

With part (a) and part (b) we get the number of Pythagorean triples over $\mathbb{F}_q$ as

$$\sum_{c \in \mathbb{F}_q} |P_c| = |P_0| + (q-1)|P_1| = q^2$$

So also part (c) is shown. □

From this lemma we can deduce the following corollary.

Corollary 3.5 *If* $Q_c := \left\{(a,b) \mid a^2 + b^2 = c\right\}$ *then we have*

$$|Q_0| = \begin{cases} 2q-1 & \text{if } q \equiv 1 \mod 4 \\ 1 & \text{if } q \equiv 3 \mod 4 \end{cases}$$

and

$$|Q_c| = \begin{cases} q-1 & \text{if } q \equiv 1 \mod 4 \\ q+1 & \text{if } q \equiv 3 \mod 4 \end{cases}$$

for $c \neq 0$.

PROOF. For $c \in \square_q$ (this includes $c = 0$) the formulas were proven in Lemma 3.4. So let $c \in \mathbb{F}_q$ be a non-square. As the squares $\square_q \backslash \{0\}$ form a subgroup of $\mathbb{F}_q^*$, the non-squares have the form $c \cdot d^2$ with $d^2 \in \square_q \backslash \{0\}$. If $a^2 + b^2 = c$ then $(ad)^2 + (bd)^2 = cd^2$. Therefore the number of solutions (a,b) is the same for all non-squares and we can determine the number of solutions by counting: There are q^2 pairs $(a,b) \in \mathbb{F}_q \times \mathbb{F}_q$. As there are $\frac{q-1}{2}$ squares and non-squares in $\mathbb{F}_q^*$ we obtain

$$\begin{aligned} \frac{q-1}{2}|Q_c| &= q^2 - (2q-1) - \frac{q-1}{2} \cdot (q-1) \\ &= \frac{1}{2}(q-1)^2 \end{aligned}$$

for $q \equiv 1 \mod 4$ and

$$\begin{aligned} \frac{q-1}{2}|Q_c| &= q^2 - 1 - \frac{q-1}{2} \cdot (q+1) \\ &= \frac{1}{2}(q-1)(q+1) \end{aligned}$$

for $q \equiv 3 \mod 4$, which gives our statement. □

Since later on we want to study the automorphism group of $\mathbb{F}_q^m$ with respect to Δ we have to define what we consider as an automorphism.

Definition 3.6 *An automorphism of $\mathbb{F}_q^m$ with respect to Δ is a bijective mapping $\varphi \in A\Gamma L(\mathbb{F}_q, m)$ with*

$$\begin{aligned} & \Delta(u_1, \ldots, u_m, v_1, \ldots, v_m) \\ = \; & \Delta(\varphi(u_1, \ldots, u_m), \varphi(v_1, \ldots, v_m)) \end{aligned}$$

for all $(u_1, \ldots, u_m), (v_1, \ldots, v_m) \in \mathbb{F}_q^m$. The group of automorphisms with respect to Δ is denoted by $\mathrm{Aut}(\mathbb{F}_q^m, \Delta)$.

In other words this definition says that φ has to map affine subspaces, like points, lines, or hyperplanes, to subspaces with equal dimension, and has to preserve the integral distance property.

Lemma 3.7 *We have the following examples of automorphisms:*

1. $\varphi_{(a_1,\ldots,a_m)}(x_1, \ldots, x_m) = (x_1 + a_1, \ldots, x_m + a_m)$ *for* $(a_1, \ldots, a_m) \in \mathbb{F}_q^m$,
2. $\widetilde{\varphi}_\lambda(x_1, \ldots, x_m) = (\lambda x_1, \ldots, \lambda x_m)$ *for* $\lambda \in \mathbb{F}_q^*$,
3. $\widetilde{\varphi}_A(x_1, \ldots, x_m) = A \cdot (x_1, \ldots, x_m)^T$ *for* $A \in O(m, \mathbb{F}_q)$, *and*
4. $\widehat{\varphi}_j(x_1, \ldots, x_m) = \left(x_1^{p^j}, \ldots, x_m^{p^j}\right)$ *for* $j \in \mathbb{N}$ *and* p *being the characteristic of* $\mathbb{F}_q$.

PROOF. The first two cases are easy to check. For the third case we consider $d^2(Au, Av)$

$$\begin{aligned} & = \langle A(u-v), A(u-v)\rangle = (u-v)^T A^T A(u-v) \\ & = (u-v)^T(u-v) = \langle u-v, u-v\rangle = d^2(u,v) \end{aligned}$$

and for the fourth case we have
$d^2\left(\widehat{\varphi}_j(0, \ldots, 0), \widehat{\varphi}_j(x_1, \ldots, x_m)\right)$

$$\begin{aligned} & = \sum_{i=1}^{m} \left(x_i^{p^j}\right)^2 = \sum_{i=1}^{m} \left(x_i^2\right)^{p^j} \\ & = \left(\sum_{i=1}^{m} x_i^2\right)^{p^j} = d^2(0, \ldots, 0, x_1, \ldots, x_m)^{p^j} \end{aligned}$$

□

We would like to remark the orders of the groups $O(m, \mathbb{F}_q)$, $GL(m, \mathbb{F}_q)$, and $\langle O(m, \mathbb{F}_q), Z\rangle$:

(1) $|GL(m, \mathbb{F}_q)| = \prod_{i=0}^{m-1} \left(q^m - q^i\right)$ for all $m \in \mathbb{N}$.

(2) $|O(2n+1, \mathbb{F}_q)| = 2q^n \cdot \prod_{i=0}^{n-1} \left(q^{2n} - q^{2i}\right)$ for $n \in \mathbb{N}$.

(3) $|O(2n, \mathbb{F}_q)| = 2(q^n - 1) \cdot \prod_{i=1}^{n-1} \left(q^{2n} - q^{2i}\right)$ for $n \in \mathbb{N}$ and $-1 \in \square_q$.

(4) $|O(2n, \mathbb{F}_q)| = 2\left(q^n + (-1)^{n+1}\right) \cdot \prod_{i=1}^{n-1} \left(q^{2n} - q^{2i}\right)$ for $n \in \mathbb{N}$ and $-1 \notin \square_q$.

(5) $|\langle O(m, \mathbb{F}_q), Z\rangle| = \frac{q-1}{2} \cdot |O(m, \mathbb{F}_q)|$ for all $m \in \mathbb{N}\backslash\{1\}$.

For dimension $m = 3$ we have $|O(3, \mathbb{F}_q)| = 2(q-1)q(q+1)$, $|GL(3, \mathbb{F}_q)| = q^3(q-1)^3(q+1)\left(q^2+q+1\right)$, and $|\langle O(3, \mathbb{F}_q), Z\rangle| = (q-1)^2 q(q+1)$.

Since the Frobenius homomorphisms and the translations are automorphisms with respect to Δ it suffices to determine the matrix group $\mathrm{Aut}(\mathbb{F}_q^m, \Delta) \cap GL(m, \mathbb{F}_q)$ of all matrices which are automorphisms with respect to Δ in order to determine the whole automorphism group. Due to Lemma 3.7 we have $\langle O(m, \mathbb{F}_q), Z\rangle \leqslant \mathrm{Aut}(\mathbb{F}_q^m, \Delta) \cap GL(m, \mathbb{F}_q)$. Thus for dimension $m = 3$ we have $(q-1)^2 q(q+1) \mid \left|\mathrm{Aut}(\mathbb{F}_q^3, \Delta) \cap GL(3, \mathbb{F}_q)\right|$. We will prove later on that $\langle O(3, \mathbb{F}_q), Z\rangle$ already is isomorphic to $\mathrm{Aut}(\mathbb{F}_q^3, \Delta) \cap GL(3, \mathbb{F}_q)$.

At first we summarize our knowledge on $\mathrm{Aut}(\mathbb{F}_q^m, \Delta)$:

Theorem 3.8 *We have*

(1) $\mathrm{Aut}(\mathbb{F}_q^m, \Delta) = A\Gamma L(m, \mathbb{F}_q)$ *for* $2 \mid q$,

(2) $\mathrm{Aut}(\mathbb{F}_q^1, \Delta) = A\Gamma L(1, \mathbb{F}_q)$,

(3) $\mathrm{Aut}(\mathbb{F}_q^2, \Delta) \cap GL(2, \mathbb{F}_q) = \langle O(2, \mathbb{F}_q), Z\rangle$ *for* $2 \nmid q$, $q \notin \{5, 9\}$,

(4) $\mathrm{Aut}(\mathbb{F}_5^2, \Delta) \cap GL(2, \mathbb{F}_5) > \langle O(2, \mathbb{F}_5), Z\rangle$, $\frac{|\mathrm{Aut}(\mathbb{F}_5^2, \Delta) \cap GL(2, \mathbb{F}_5)|}{|\langle O(2, \mathbb{F}_5), Z\rangle|} = 2$, *and*

(5) $\mathrm{Aut}(\mathbb{F}_9^2, \Delta) \cap GL(2, \mathbb{F}_9) > \langle O(2, \mathbb{F}_9), Z\rangle$, $\frac{|\mathrm{Aut}(\mathbb{F}_9^2, \Delta) \cap GL(2, \mathbb{F}_9)|}{|\langle O(2, \mathbb{F}_9), Z\rangle|} = 3$.

PROOF. (1) and (2) hold since for $m = 1$ or $2 \mid q$ all distances are integral. So in general we assume dimension $m \geqslant 2$ and odd characteristic $2 \nmid q$ if not stated otherwise in the rest of this article. For the proof of (3), (4), and (5) we refer to [13]. □

Next we prove some results on the orbits of $\mathbb{F}_q^m$ under the groups $O(m, \mathbb{F}_q)$ and $\langle O(m, \mathbb{F}_q), Z\rangle$. Therefore we need:

Definition 3.9 *By $\mathcal{P}_k(\mathbb{F}_q^m)$ we denote the set $\left\{v \in \mathbb{F}_q^m \backslash \{0\} \mid d^2(0, v) = k\right\}$. Whenever q and m are clear from the context we use $\mathcal{P}_k$ instead of $\mathcal{P}_k(\mathbb{F}_q^m)$.*

Lemma 3.10 *For every $k \in \mathbb{F}_q \backslash \{0\}$ the group $O(2, \mathbb{F}_q)$ acts transitively on $\mathcal{P}_k$.*

PROOF. Let (a, b) and (c, d) be two points in $\mathbb{F}_q^2$ with $a^2 + b^2 = c^2 + d^2 \neq 0$. With $x = \frac{ac+bd}{a^2+b^2}$ and $y = \frac{bc-ad}{a^2+b^2}$ we have $x^2 + y^2 = \frac{(a^2+b^2)\cdot(c^2+d^2)}{(a^2+b^2)^2} = 1$. Thus the matrix $A = \begin{pmatrix} x & y \\ -y & x \end{pmatrix}$ is an element of $O(2, \mathbb{F}_q)$ which maps (a, b) to (c, d). □

Lemma 3.11 *The group* $O(2,\mathbb{F}_q)$ *acts transitively on* $\mathcal{P}_0$.

PROOF. If $-1 \notin \square_q$ then we have $|\mathcal{P}_0| = 0$. Thus we may assume $-1 \in \square_q$. For $a, b \in \mathbb{F}_q$ with $a^2 + b^2 = 0$ we have either $a = b = 0$ or $a, b \neq 0$. In the latter case we have $\left(\frac{a}{b}\right)^2 = -1$, which has two solutions $\frac{a}{b} = \omega$ and $\frac{a}{b} = -\omega$, where ω is a root of -1. Thus we can write all elements of $\mathcal{P}_0$ as $\begin{pmatrix} z & \pm z\omega \end{pmatrix}^T$ with $z \in \mathbb{F}_q^*$. Now we apply all matrices of the form $M := \begin{pmatrix} x & y \\ -y & x \end{pmatrix}$ with $x^2 + y^2 = 1$ to the vector $\begin{pmatrix} 1 & \omega \end{pmatrix}^T$. By definition these matrices are elements of $O(2,\mathbb{F}_q)$. If we parametrize x and y as in Lemma 3.4 we get $M\begin{pmatrix} 1 & \omega \end{pmatrix}^T = \begin{pmatrix} z & z\omega \end{pmatrix}^T$, where $z = \frac{t-t^{-1}}{t+t^{-1}} + \frac{2}{t+t^{-1}} \cdot \omega$. By a small computation we check that

$$\begin{aligned} f: \quad & \mathbb{F}_q \backslash \{0, \omega, -\omega\} \to \mathbb{F} \backslash \{-1, 0, 1\}, \\ & t \mapsto \frac{t - t^{-1}}{t + t^{-1}} + \frac{2}{t + t^{-1}} \cdot \omega \end{aligned}$$

is well defined and injective. Thus all points $\begin{pmatrix} z & z\omega \end{pmatrix}^T$ are in the same orbit as $\begin{pmatrix} 1 & \omega \end{pmatrix}^T$ under the action of $O(2,\mathbb{F}_q)$. Since $\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \in O(2,\mathbb{F}_q)$, the points $\begin{pmatrix} z & -z\omega \end{pmatrix}^T$ are also contained in this orbit and the proposed statement holds. □

Lemma 3.12 *The group* $O(3,\mathbb{F}_q)$ *acts transitively on* $\mathcal{P}_0$.

PROOF. See e. g. [17, Theorem 11.6] for a proof. □

Lemma 3.13 *For each* $v_1 := (x_1, y_1, z_1) \in \mathbb{F}_q^3$ *with* $x_1^2 + y_1^2 + z_1^2 = 1$ *there exist* $v_2 := (x_2, y_2, z_2)$, $v_3 := (x_3, y_3, z_3) \in \mathbb{F}_q^3$ *fulfilling* $x_i^2 + y_i^2 + z_i^2 = 1$ *for* $i = 2, 3$, *such that* $\langle v_i, v_j \rangle = 0$ *for* $i \neq j$ *and* $i, j = 1, 2, 3$.

PROOF. The set $\langle v_1 \rangle^\perp$ of vectors v solving the linear equation $\langle v, v_1 \rangle = 0$ forms a 2-dimensional vector space. Let $w \in \langle v_1 \rangle^\perp$ be an arbitrary element with $w \neq 0$. Then $\langle w \rangle^\perp$ is also a 2-dimensional vector space and certainly $v_1 \in \langle w \rangle^\perp$. Thus we get $\langle w \rangle^\perp \neq \langle v_1 \rangle^\perp$. Therefore the orthogonal vector space $\langle v_1 \rangle^\perp$ is non-degenerate in the sense of [7, II.10.1]. By [7, II.10.2 b)] there is an orthogonal basis $\{w_2, w_3\}$ of $\langle v_1 \rangle^\perp$, i. e. we have $\langle w_2, w_3 \rangle = 0$ and $\langle w_i, w_i \rangle \neq 0$ for $i = 2, 3$. For $\lambda, \mu \in \mathbb{F}_q$ we obtain

$$\langle \lambda w_2 + \mu w_3, \lambda w_2 + \mu w_3 \rangle = \lambda^2 \langle w_2, w_2 \rangle + \mu^2 \langle w_3, w_3 \rangle .$$

As $\langle w_2, w_2 \rangle, \langle w_3, w_3 \rangle \in \mathbb{F}_q^*$ there exist $\lambda, \mu \in \mathbb{F}_q$ such that

$$\langle \lambda w_2 + \mu w_3, \lambda w_2 + \mu w_3 \rangle = 1$$

by [17, Lemma 11.1]. Therefore there is a vector $v_2 \in \mathbb{F}_q^3$ such that $\langle v_1, v_2 \rangle = 0$ and $\langle v_2, v_2 \rangle = 1$. Now v_3 can easily be constructed: The cross product $v_3 := v_1 \times v_2$ is a vector with $\langle v_1, v_3 \rangle = \langle v_2, v_3 \rangle = 0$ and

$$\langle v_3, v_3 \rangle = \langle v_1, v_1 \rangle \cdot \langle v_2, v_2 \rangle - \langle v_1, v_2 \rangle^2 = 1.$$

□

From the previous lemma we can easily deduce:

Lemma 3.14 *The group* $O(3,\mathbb{F}_q)$ *acts transitively on* $\mathcal{P}_k$ *for all* $k \in \square_q$.

PROOF. Due to Lemma 3.12 we only have to consider the case $k \neq 0$. Let $v \in \mathbb{F}_q^3$ such that $\langle v, v \rangle = a^2 \neq 0$. We put $v_1 := a^{-1} v$. Then $\langle v_1, v_1 \rangle = 1$ and by Lemma 3.13 there exist $v_2, v_3 \in \mathbb{F}_q^3$ such that $A = (v_1 | v_2 | v_3)$ is an orthogonal matrix. Thus the vectors $(1, 0, 0)$ and v_1 are in the same orbit of $O(3,\mathbb{F}_q)$. Thus all v with $\langle v, v \rangle = a^2 \neq 0$ and the vectors $(\pm a, 0, 0)$ are in the same orbit. □

Lemma 3.15 *The group* $O(3,\mathbb{F}_q)$ *acts transitively on* $\mathcal{P}_k$ *for all* $k \notin \square_q$.

PROOF. Let $v = (v_1, v_2, v_3) \in \mathcal{P}_k$ be a arbitrary vector. We show that there exists an element $A \in O(3,\mathbb{F}_q)$ such that the third coordinate of Av equals zero. This reduces the problem to the 2-dimensional case where we can apply Lemma 3.10 or Lemma 3.11, as we can extend a 2-dimensional Matrix $A' \in O(2,\mathbb{F}_q)$ to a matrix $A \in O(3,\mathbb{F}_q)$ by adding a third row and a third column consisting of a one in the diagonal and zeros elsewhere.

If $v_2^2 + v_3^2 = l^2 \neq 0$ then due to Lemma 3.10 there exists an element $A' \in O(2,\mathbb{F}_q)$ which maps (v_2, v_3) to $(l, 0)$. Thus we can extend A' to a desired matrix $A \in O(3,\mathbb{F}_q)$ such that the third coordinate of Av equals zero. Since $v_1^2 + v_2^2 + v_3^2 \notin \square_q$ we can not have $v_i^2 + v_j^2 = 0$ for $i \neq j$. So we can assume $v_i^2 + v_j^2 \notin \square_q$ for $i \neq j$.

For the remaining cases we use another technique. We set

$$\mathcal{P}_{f,k} := \left\{ (v_1, v_2, v_3) \in \mathbb{F}_q^3 \backslash \{0\} \mid v_1 = f,\ v_1^2 + v_2^2 + v_3^2 = k \right\}.$$

By Lemma 3.10 all points of $\mathcal{P}_{f,k}$ are contained in the same orbit under $O(3,\mathbb{F}_q)$. From Corollary 3.5 we deduce $|\mathcal{P}_{f,k}| = q - 1$ for $q \equiv 1 \mod 4$ and $|\mathcal{P}_{f,k}| = q + 1$ for $q \equiv 3 \mod 4$. Hence we have

$$|\mathcal{P}_k| = \sum_{f \in \mathbb{F}_p} |\mathcal{P}_{f,k}| = q \cdot |\mathcal{P}_{0,k}| .$$

Now let us consider an arbitrary point $v = (v_1, v_2, v_3) \in \mathcal{P}_{f,k}$ and set $l = v_1^2 + v_2^2$. Since $v_1^2 + v_2^2 + v_3^2 = k \notin \square_q$ we have $l \neq 0$. Due to Lemma 3.10 all points (u_1, u_2, v_3) with $u_1^2 + u_2^2 = l$ lie in the same orbit as v under $O(3,\mathbb{F}_q)$.

Due to Corollary 3.5 we have at least $\frac{q+1}{2}$ solutions u_1 of the equation $u_1^2 + u_2^2 = l$ for $q \equiv 3 \mod 4$. This means that every point in $\mathcal{P}_k$ lies in an orbit with at least $\frac{q+1}{2} \cdot |\mathcal{P}_{u_1,k}| = \frac{(q+1)^2}{2} > \frac{|\mathcal{P}_k|}{2} = \frac{q(q+1)}{2}$ points. Thus there can only be one orbit.

For $q \equiv 1 \mod 4$ we similarly conclude that every point in $\mathcal{P}_k$ lies in an orbit with at least $\frac{(q-1)^2}{2}$ points. Since $|\mathcal{P}_k| = (q-1)q$ and $|\mathcal{P}_{f,k}| = q - 1$ for all $f \in \mathbb{F}_p$ there can exist at most two orbits and the length of every orbit has to be divisible by $|\mathcal{P}_{f,k}| = q - 1$. If there exist exactly two orbits $\mathcal{B}_1, \mathcal{B}_2$ then we have w.l.o.g. $|\mathcal{B}_1| = \frac{q-1}{2} \cdot (q-1)$ and $|\mathcal{B}_2| = \frac{q+1}{2} \cdot (q-1)$. Due to $|\mathcal{B}_1| \,\Big|\, |O(3,\mathbb{F}_q)|$ we have $(q-$

$1)^2 \mid 4\cdot(q-1)q(q+1)$. Using $\gcd(q-1, q) = 1$ we conclude $q-1 \mid 4(q+1)$. Thus we have $q-1 \mid 8$, which is equivalent to $q \in \{3, 5, 9\}$. As $3 \not\equiv 1 \mod 4$ we only have to consider the cases $q = 9$ and $q = 5$. In $\mathbb{F}_9 \simeq \mathbb{F}_3[x]/\left(x^2+1\right)$ we have $\square_9 = \{0, 1, 2, x, 2x\}$. Since we have either $v_i = 0$ for some i or $|\{v_1, v_2, v_3\} \cap \{1, 2\}| \geqslant 2$ or $|\{v_1, v_2, v_3\} \cap \{x, 2x\}| \geqslant 2$ there exist i, j with $v_i^2 + v_j^2 \in \square_9$ in this case and we can apply our reduction to the 2-dimensional case.

For $q = 5$, $k = 2$ we have

$$\mathcal{B}_1 = \left\{ (v_1, v_2, v_3) \mid v_1, v_2, v_3 \in \{2, 3\} \right\},$$
$$\mathcal{B}_2 = \left\{ (0, v_1, v_2), (v_1, 0, v_2), (v_1, v_2, 0) \mid v_1, v_2 \in \{1, 4\} \right\},$$

and for $q = 5$, $k = 3$ we have

$$\mathcal{B}_1 = \left\{ (v_1, v_2, v_3) \mid v_1, v_2, v_3 \in \{1, 4\} \right\},$$
$$\mathcal{B}_2 = \left\{ (0, v_1, v_2), (v_1, 0, v_2) (v_1, v_2, 0) \mid v_1, v_2 \in \{2, 3\} \right\}.$$

By considering the matrix $M_1 = \begin{pmatrix} 1 & 2 & 4 \\ 2 & 1 & 4 \\ 1 & 1 & 3 \end{pmatrix}$ in $O(3, \mathbb{F}_q)$ we conclude that in both cases $\mathcal{B}_1$ and $\mathcal{B}_2$ are contained in the same orbit. □

Lemma 3.16 *For dimension $m \geqslant 4$ and $v = (v_1, \ldots, v_m) \in \mathbb{F}_q^m$ there exists an element $A \in O(m, \mathbb{F}_q)$ such that the m-th coordinate of Av equals zero.*

PROOF. If one of the v_i equals zero then there clearly exists such a matrix A. So we assume $v_i \neq 0$ for $1 \leqslant i \leqslant m$.

If $v_h^2 + v_i^2 + v_j^2 = 0$ for all pairwise different $1 \leqslant h, i, j \leqslant m$ then we would have $v = 0$ or $3 \mid q$: As $m \geqslant 4$, there is at least one further index k. If we replace v_h by v_k then $v_k^2 + v_i^2 + v_j^2 = 0$ and $v_h^2 + v_i^2 + v_j^2 = 0$ give $v_h^2 = v_k^2$. Replacing v_i and v_j by v_k leads to $v_h^2 = v_i^2 = v_j^2 = v_k^2$, so we obtain $3v_i^2 = 0$ and thus $v = 0$ if $3 \nmid q$.

For $3 \mid q$ the same computation leads to $v_i = \pm v_j$ for all i, j. W. l. o. g. let $v_1 = 1$. Then we have $v_i^2 + v_j^2 = 2$ for all $i, j > 1$. By Lemma 3.10 the group $O(2, \mathbb{F}_q)$ acts transitively on $\mathcal{P}_2$. As we can extend 2-dimensional orthogonal matrices by ones in the diagonal we can assume $v_i = 1$ for $1 \leqslant i \leqslant m$. As the matrix

$$A = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 2 & 2 \\ 1 & 2 & 1 & 2 \\ 1 & 2 & 2 & 1 \end{pmatrix} \in O(4, \mathbb{F}_q)$$

maps $\begin{pmatrix} 1 & 1 & 1 & 1 \end{pmatrix}^T$ to $\begin{pmatrix} 1 & 0 & 0 & 0 \end{pmatrix}^T$, we can extend A to a matrix in $O(m, \mathbb{F}_q)$ such that Av has a zero at coordinate m.

So we may assume $0 \neq v_{m-2}^2 + v_{m-1}^2 + v_m^2 =: k$. Since there exist $a, b \in \mathbb{F}_q$ with $a^2 + b^2 = k \neq 0$ by [17, Lemma 11.1] we can apply Lemma 3.14 and Lemma 3.15 to deduce that there exists an element $A' \in O(3, \mathbb{F}_q)$ which maps (v_{m-2}, v_{m-1}, v_m) onto $(a, b, 0)$. Clearly we can extend A' to obtain the desired matrix $A \in O(m, \mathbb{F}_q)$ mapping v onto a point with m-th coordinate being equal to zero. □

Lemma 3.17 *For dimension $m \geqslant 2$ the group $O(m, \mathbb{F}_q)$ acts transitively on $\mathcal{P}_k$ for all $k \in \mathbb{F}_q$.*

PROOF. We prove the Lemma by induction and utilize lemmas 3.10, 3.11, 3.12, 3.14, and 3.15 as induction basis. Now let $u, v \in \mathcal{P}_k$ be arbitrary. Due to Lemma 3.16 there exist $A_1, A_2 \in O(m, \mathbb{F}_k)$ such that the m-th coordinate of $u' = A_1 u$ and the m-th coordinate of $v' = A_2 v$ are both equal to zero. Deleting the last coordinate from u' and v' gives two vectors $\tilde{u}$ and $\tilde{v}$ in $\mathcal{P}_k\left(\mathbb{F}_q^{m-1}\right)$, respectively. Due to our induction hypothesis there exists an element $\tilde{A}_3 \in O(m-1, \mathbb{F}_q)$ with $\tilde{A}_3 \tilde{u} = \tilde{v}$. Clearly we can extend $\tilde{A}_3$ to a matrix $A_3 \in O(m, \mathbb{F}_q)$ with $A_3 u' = v'$. With $A = A_2^{-1} A_3 A_1$ we have $A \in O(m, \mathbb{F}_q)$ and $Au = v$. □

Definition 3.18 *We set*

$$\mathcal{P}^+ := \bigcup_{k \in \square_q \setminus \{0\}} \mathcal{P}_k \quad \text{and} \quad \mathcal{P}^- := \bigcup_{k \notin \square_q} \mathcal{P}_k.$$

Lemma 3.19 *For $2 \nmid q$ and $m \geqslant 2$ the orbits of $\mathbb{F}_q^m$ under the group $\langle O(m, \mathbb{F}_q), Z \rangle$ are $\mathcal{P}^+$, $\mathcal{P}_0$, and $\mathcal{P}^-$.*

PROOF. From the previous lemmas we know that $O(m, \mathbb{F}_q)$ acts transitively on $\mathcal{P}_k$ for $2 \nmid q$, $m \geqslant 2$, and $k \in \mathbb{F}_q$. Thus $\langle O(m, \mathbb{F}_q), Z \rangle$ acts transitively on $\mathcal{P}^+$, $\mathcal{P}_0$, and $\mathcal{P}^-$. (For every element $B \in \langle O(m, \mathbb{F}_q), Z \rangle$ there exists an element $k \in \mathbb{F}_q^*$ so that we have $\langle Bv, Bv \rangle = k^2 \langle v, v \rangle$ for all $v \in \mathbb{F}_q^m$.) □

Lemma 3.20 *Let $v = (v_1, v_2, v_3)$, $w = (w_1, w_2, w_3) \in \mathbb{F}_q^3$ with $v, w \neq 0$. If $\langle v, v \rangle = \langle v, w \rangle = 0$ and $\langle v \rangle \neq \langle w \rangle$ then we have $\langle w, w \rangle \in \square_q$ if $q \equiv 1 \mod 4$ and $\langle w, w \rangle \notin \square_q$ if $q \equiv 3 \mod 4$.*

PROOF. If the v_i are non-zero we can assume w.l.o.g. that $v_3 = 1$. From $\langle v, w \rangle = 0$ we conclude $w_3 = -w_1 v_1 - w_2 v_2$. Using $v_1^2 + v_2^2 + 1 = 0$ this gives

$$\begin{aligned} \langle w, w \rangle &= w_1^2 + w_2^2 + w_1^2 v_1^2 + w_2^2 v_2^2 + 2 w_1 w_2 v_1 v_2 \\ &= -v_2^2 w_1^2 - v_1^2 w_2^2 + 2 w_1 w_2 v_1 v_2 \\ &= -(v_2 w_1 - v_1 w_2)^2 . \end{aligned}$$

Since $-1 \notin \square_q$ iff $q \equiv 3 \mod 4$ by Lemma 3.2 we have

$$\begin{array}{ll} \langle w, w \rangle \notin \square_q \setminus \{0\} & \text{for } q \equiv 3 \mod 4 \text{ and} \\ \langle w, w \rangle \in \square_q & \text{for } q \equiv 1 \mod 4 \end{array}$$

in this case.

Let us assume $q \equiv 3 \mod 4$ and $\langle w, w \rangle = 0$ for a moment. Since $-1 \notin \square_q$ we have $v_1, v_2 \neq 0$ using $v_1^2 + v_2^2 + 1 = 0$. Thus we have $w_1 = w_2 \frac{v_1}{v_2}$. Inserting yields $w = \left(w_2 \frac{v_1}{v_2}, w_2, -w_2 \frac{v_1}{v_2} \cdot v_1 - w_2 v_2 \right) = \frac{w_2}{v_2} \cdot v$. Since $v, w \neq 0$ we would have $\langle w \rangle = \langle v \rangle$. Thus we even have $\langle w, w \rangle \notin \square_q$ for $q \equiv 3 \mod 4$.

In the remaining case we assume w.l.o.g. $v_3 = 0$. Since $v_1^2 + v_2^2 = 0$ we have $-1 \in \square_q$, $q \equiv 1 \mod 4$, and $v_1, v_2 \neq 0$. We can further assume w.l.o.g. $v_1 = 1$ and $v_2 = \omega_q$, where $\omega_q^2 = -1$. With this $\langle v, w \rangle = 0$ is equivalent to $w_2 = \omega_q w_1$. Thus we have $\langle w, w \rangle = w_3^2 \in \square_q$. □

Lemma 3.21 *For $2 \nmid q$ and $m \geqslant 3$ the orbits of $\mathbb{F}_q^m$ under the group* $\mathrm{Aut}\left(\mathbb{F}_q^m, \Delta\right) \cap \mathrm{GL}(m, \mathbb{F}_q)$ *are $\mathcal{P}^+$, $\mathcal{P}_0$, and $\mathcal{P}^-$.*

PROOF. Since $\langle \mathrm{O}(m, \mathbb{F}_q), Z\rangle \leqslant \mathrm{Aut}\left(\mathbb{F}_q^m, \Delta\right) \cap \mathrm{GL}(m, \mathbb{F}_q)$ and due to Lemma 3.19 it may only happen that some elements of $\mathcal{P}^+$, $\mathcal{P}_0$, and $\mathcal{P}^-$ are contained in the same orbit. Due to Definition 3.6 and Definition 2.2 $\mathcal{P}^-$ forms its own orbit. Thus only $\mathcal{P}^+$ and $\mathcal{P}_0$ may be contained in the same orbit. Now we show that this is not the case.

In Section 4 we introduce the graph $\mathfrak{G}_{m,q}$ of integral distances corresponding to $\mathbb{F}_q^m$ and its integral distances. Due to Lemma 4.5 for dimension $m = 3$ and $2 \nmid q$ the graph $\mathfrak{G}_{3,q}$ is not strongly regular. Thus $\mathcal{P}^+$ and $\mathcal{P}_0$ are disjoint orbits.

For $m \geqslant 4$ let us assume that there exists an element v in $\mathbb{F}_q^m$ with $\langle v, v\rangle = 0$ and there exists a matrix A in $\left(\mathrm{Aut}\left(\mathbb{F}_q^m, \Delta\right) \cap \mathrm{GL}(m, \mathbb{F}_q)\right)$ with $\left\langle A^{-1}v, A^{-1}v\right\rangle \in \square_q \backslash \{0\}$. W.l.o.g. we assume $A^{-1}v = e_1$, where e_i is a vector consisting of zeros and a single one at coordinate i, this is the i-th unit vector. So we have $Ae_1 = v$ and we set $w_i := Ae_i$, $\mu_i := \langle v, w_i\rangle$ for $2 \leqslant i \leqslant 4$. Now we show that there exists a vector $v' \in \mathbb{F}_q^m$ with $\langle e_1, v'\rangle = 0$, $\langle v', v'\rangle \neq 0$, and $\langle Ae_1, Av'\rangle = 0$. If there exists $2 \leqslant i \leqslant 4$ with $\mu_i = 0$ then we may choose $v' = e_i$. Otherwise we have $\mu_2, \mu_3, \mu_4 \neq 0$. We remark that $\left(\frac{\mu_i}{\mu_j}\right)^2 = -1$ is equivalent to $\left(\frac{\mu_j}{\mu_i}\right)^2 = -1$ for all $2 \leqslant i, j \leqslant 4$. Due to $\left(\frac{\mu_1}{\mu_2}\right)^2 \cdot \left(\frac{\mu_2}{\mu_3}\right)^2 \cdot \left(\frac{\mu_3}{\mu_1}\right)^2 = 1 \neq -1$ there exist i and j with $i \neq j$, $\left(\frac{\mu_i}{\mu_j}\right)^2 \neq -1$. We set $v' := -\mu_j e_i + \mu_i e_j$ which yields

$$\begin{aligned}
\langle e_1, v'\rangle &= 0,\\
\langle v', v'\rangle &= \mu_i^2 + \mu_j^2 \neq 0, \text{ and}\\
\langle Ae_1, Av'\rangle &= \langle v, -\mu_j w_i + \mu_i w_j\rangle\\
&= -\mu_j \langle v, w_i\rangle + \mu_i \langle v, w_j\rangle\\
&= 0.
\end{aligned}$$

Let χ be the characteristic function of $\square_q$, this is $\chi(x) = 1$ for $x \in \square_q$ and $\chi(x) = 0$ for $x \notin \square_q$. We set $k := \langle v', v'\rangle \neq 0$ and $l := \langle Av', Av'\rangle$. For all $\lambda_1, \lambda_2 \in \mathbb{F}_q$ we have

$$\begin{aligned}
d^2(0, \lambda_1 e_1 + \lambda_2 v') &= \langle \lambda_1 e_1 + \lambda_2 v', \lambda_1 e_1 + \lambda_2 v'\rangle\\
&= \lambda_1^2 + k \cdot \lambda_2^2 \text{ and}\\
d^2(0, A(\lambda_1 e_1 + \lambda_2 v')) &= \langle \lambda_1 v + \lambda_2 Av', \lambda_1 v + \lambda_2 Av'\rangle\\
&= l \cdot \lambda_2^2.
\end{aligned}$$

Since $A \in \left(\mathrm{Aut}\left(\mathbb{F}_q^m, \Delta\right) \cap \mathrm{GL}(m, \mathbb{F}_q)\right)$ we have $\chi\left(\lambda_1^2 + k \cdot \lambda_2^2\right) = \chi\left(l \cdot \lambda_2^2\right)$ for all λ_1, λ_2 in $\mathbb{F}_q$. Inserting $\lambda_2 = 1$ yields $\chi(l) = \chi\left(\lambda_1^2 + k\right)$ for all $\lambda_1 \in \mathbb{F}_q$. Due to $\left|\left\{\lambda_1^2 + k \mid \lambda_1 \in \mathbb{F}_q\right\}\right| = \frac{q+1}{2}$ we conclude $\chi(l) = \chi\left(\lambda_1^2 + k\right) = 1$. W.l.o.g. we may assume $k = 1 \in \square_q \backslash \{0\}$. Thus for $q = p^r$ and $x \in \square_q$ we have

$$\begin{aligned}
1 &= \chi(x) = \chi(x+1) = \chi((x+1)+1)\\
&= \chi((x+2)+1) = \cdots = \chi((x+p-2)+1).
\end{aligned}$$

We conclude $p \,\Big|\, |\square_q| = \frac{q+1}{2} = \frac{p^r+1}{2}$, which is a contradiction. □

Theorem 3.22

$$\mathrm{Aut}\left(\mathbb{F}_q^3, \Delta\right) \cap \mathrm{GL}(3, \mathbb{F}_q) = \langle \mathrm{O}(3, \mathbb{F}_q), Z\rangle \text{ for } 2 \nmid q.$$

PROOF. Let $A \in \mathrm{Aut}(\mathbb{F}_q^3, \Delta) \cap \mathrm{GL}(3, \mathbb{F}_q)$ be an automorphism. The idea of the proof is, to use the fact that A takes every vector v of integral norm $\langle v, v\rangle \neq 0$ to another vector of integral norm $\neq 0$ with the aim to construct an automorphism in $\mathrm{Aut}\left(\mathbb{F}_q^2, \Delta\right)$. Using the classification of the 2-dimensional automorphisms in Theorem 3.8, see also [13], we conclude $A \in \langle \mathrm{O}(3, \mathbb{F}_q), Z\rangle$.

Clearly, A is uniquely defined by its images of $e_1 = \begin{pmatrix}1 & 0 & 0\end{pmatrix}^T$, $e_2 = \begin{pmatrix}0 & 1 & 0\end{pmatrix}^T$, and $e_3 = \begin{pmatrix}0 & 0 & 1\end{pmatrix}^T$. Due to Lemma 3.19 and Lemma 3.21 we can assume $A \cdot e_1 = e_1$. We set $A \cdot e_2 =: \begin{pmatrix}a & b & c\end{pmatrix}^T$, where we have

$$a^2 + b^2 + c^2 = d^2 \in \square_q \backslash \{0\}$$

due to Lemma 3.21. Let $\chi : \mathbb{F}_q \to \{0, 1\}$, where $\chi(x) = 1$ iff $x \in \square_q$, be the characteristic function of $\square_q$. By applying A on $\begin{pmatrix}x & y & 0\end{pmatrix}^T$ we obtain

$$\chi\left(x^2 + y^2\right) = \chi\left(x^2 + 2axy + d^2y^2\right) \text{ for all } x, y \in \mathbb{F}_q. \tag{1}$$

Inserting $x = -2a$, $y = 1$ yields $\chi\left(4a^2 + 1\right) = \chi\left(d^2\right) = 1$. Now we prove $\chi\left(a^2 + 1\right) = 1$. Putting $x = 2na$, $y = 1$ for an arbitrary $n \in \mathbb{F}_q$ in Equation (1) we obtain

$$\chi\left(4n^2a^2 + 1\right) = \chi\left(\left(4n^2 + 4n\right)a^2 + d^2\right).$$

Inserting $x = -2(n+1)a$, $y = 1$ yields

$$\chi\left(4(n+1)^2a^2 + 1\right) = \chi\left(\left(4n^2 + 4n\right)a^2 + d^2\right),$$

hence we get

$$\chi\Big((2n)^2a^2 + 1\Big) = \chi\Big((2n+2)^2a^2 + 1\Big)$$

for all $n \in \mathbb{F}_q$. For $n = 1$ we have $\chi\left(4a^2 + 1\right) = 1$, therefore we obtain $\chi\left((2n)^2a^2 + 1\right) = 1$ for all $n \in \mathbb{F}_p$ (but not necessarily for all $n \in \mathbb{F}_q$). As we have $p \neq 2$, we can take $n = 2^{-1} \in \mathbb{F}_p$ and get $\chi\left(a^2 + 1\right) = 1$.

If we insert $x = -a$, $y = 1$ in Equation (1) we obtain

$$\begin{aligned}
1 &= \chi\left(a^2 + 1\right) = \chi\left(x^2 + y^2\right)\\
&= \chi\left(x^2 + 2axy + d^2y^2\right) = \chi\left(d^2 - a^2\right).
\end{aligned}$$

Thus there exists an $e \in \mathbb{F}_q$ with $a^2 + e^2 = d^2$. Let us consider the matrix $M = \begin{pmatrix}1 & a\\ 0 & e\end{pmatrix}$. Since we have $\chi\left(x^2 + y^2\right) = \chi\Big(x^2 + 2axy + \left(a^2 + e^2\right)y^2\Big)$ for all $x, y \in \mathbb{F}_q$ the matrix M is an automorphism for $\mathbb{F}_q^2$ with respect to Δ.

For $q \neq \{5, 9\}$ we can apply Theorem 3.8.(3) and conclude $a = 0$, $b^2 + c^2 = e^2 = d^2 = 1$. Now we set

$A \cdot e_3 =: \begin{pmatrix} \tilde{a} & \tilde{b} & \tilde{c} \end{pmatrix}^T$ and similarly conclude $\tilde{a} = 0$, $\tilde{b}^2 + \tilde{c}^2 = 1$. Therefore A has the form

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & b & \tilde{b} \\ 0 & c & \tilde{c} \end{pmatrix} =: \begin{pmatrix} 1 & 0 & 0 \\ 0 & & \\ 0 & & A' \end{pmatrix}.$$

By applying A on all vectors $\begin{pmatrix} 0 & y & z \end{pmatrix}$ for $y, z \in \mathbb{F}_q$ we see that A' is an element of $\mathrm{Aut}\left(\mathbb{F}_q^2, \Delta\right) \cap \mathrm{GL}(2, \mathbb{F}_q)$. Due to Theorem 3.8.(3) the matrix A' is orthogonal and we conclude $A \in \mathrm{O}(3, \mathbb{F}_q)$.

We deal with the missing cases $q \in \{5, 9\}$ using the the classification of the 2-dimensional automorphism group $\mathrm{Aut}\left(\mathbb{F}_q^2, \Delta\right)$ as follows. Either we utilize the precise classification in [13] or we utilize an exhaustive enumeration of the elements in $\mathrm{GL}(2, \mathbb{F}_q)$ to conclude $a = 0$, $b^2 + c^2 = e^2 = d^2 = 1$ for $q = 5$ and $a = 0$, $b^2 + c^2 = e^2 = d^2 \in \{\pm 1\}$ for $q = 9$. Now we set $A \cdot e_3 =: \begin{pmatrix} \tilde{a} & \tilde{b} & \tilde{c} \end{pmatrix}^T$ and similarly conclude $\tilde{a} = 0$, $\tilde{b}^2 + \tilde{c}^2 = 1$ for $q = 5$ and $\tilde{b}^2 + \tilde{c}^2 \in \{\pm 1\}$ for $q = 9$. Therefore A has the form

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & b & \tilde{b} \\ 0 & c & \tilde{c} \end{pmatrix} =: \begin{pmatrix} 1 & 0 & 0 \\ 0 & & \\ 0 & & A' \end{pmatrix}.$$

Additionally we have $\langle Ae_2, Ae_3 \rangle = 0$ in both cases, where we refer to [13] or an exhaustive enumeration. Next we exclude the case $b^2 + c^2 = -1$ for $q = 9$. We use $\mathbb{F}_9 \simeq \mathbb{F}_3[t]/\left(t^2 + 1\right)$ and assume the contrary $b^2 + c^2 = -1$. Since A is an automorphism of $\mathbb{F}_9^3$ with respect to Δ we have

$$\begin{aligned} &\chi\left(x^2 + y^2 + z^2\right) = \chi\Big(x^2 + \left(by + \tilde{b}z\right)^2 + \left(cy + \tilde{c}z\right)^2\Big) \\ &= \chi\Big(x^2 + \left(b^2 + c^2\right)y^2 + \left(\tilde{b}^2 + \tilde{c}^2\right)z^2\Big) \end{aligned}$$

for all $x, y, z \in \mathbb{F}_9$. Inserting insert $x = 1$, $y = 1$, and $z = t + 2$ yields

$$\chi\left(x^2 + y^2 + z^2\right) = \chi\left(2 + t^2 + 4t + 4\right) = \chi(t + 2) = 0$$

and

$$\begin{aligned} &\chi\Big(x^2 + \left(b^2 + c^2\right)y^2 + \left(\tilde{b}^2 + \tilde{c}^2\right)z^2\Big) \\ &= \chi\Big(\left(\tilde{b}^2 + \tilde{c}^2\right)z^2\Big) = 1, \end{aligned}$$

a contradiction. Thus due to symmetry we have $b^2 + c^2 = \tilde{b}^2 + \tilde{c}^2 = 1$ and $A \in \mathrm{O}\left(3, \mathbb{F}_q^3\right)$ in both cases. □

Remark 3.23 *Rotations in 3-space can be modeled using quaternions. We consider unit quaternions* $a + bi + cj + dk$ *with* $a^2 + b^2 + c^2 + d^2 \neq 0$. *The corresponding rotation can be written as a rightmultiplication with the* 3×3*-matrix* $M =$

$$\begin{pmatrix} a^2 + b^2 - c^2 - d^2 & 2bc - 2ad & 2ac + 2bd \\ 2ad + 2bc & a^2 - b^2 + c^2 - d^2 & 2cd - 2ab \\ 2bd - 2ac & 2ab + 2cd & a^2 - b^2 - c^2 + d^2 \end{pmatrix}.$$

We have $\left\|\begin{pmatrix} x & y & z \end{pmatrix} \cdot M\right\| = \left\|\begin{pmatrix} x & y & z \end{pmatrix} \cdot M^t\right\| = \left(x^2 + y^2 + z^2\right)\left(a^2 + b^2 + c^2 + d^2\right)^2$ *for these matrices. Thus they all correspond to automorphisms with respect to* Δ. *Unfortunately this set is only isomorphic to a proper subset of* $\langle \mathrm{O}(3, \mathbb{F}_q), Z \rangle$ *and does not form a group.*

Theorem 3.24 *For dimension* $m \geqslant 3$ *and* $2 \nmid q$ *we have* $\mathrm{Aut}\left(\mathbb{F}_q^m, \Delta\right) \cap \mathrm{GL}(m, \mathbb{F}_q) = \langle \mathrm{O}(m, \mathbb{F}_q), Z \rangle$.

PROOF. We prove the theorem by induction on the dimension m. For the induction basis we refer to Theorem 3.22. Now let $m \geqslant 4$ and $A \in \mathrm{Aut}\left(\mathbb{F}_q^m, \Delta\right) \cap \mathrm{GL}(m, \mathbb{F}_q)$ be an automorphism. Due to Lemma 3.19 and Lemma 3.21 we can assume $A \cdot e_1 = e_1$. For $2 \leqslant i \leqslant m$ we set $A \cdot e_i =: \begin{pmatrix} v_{1,i} & \dots & v_{m,i} \end{pmatrix}^T$, where we have

$$\sum_{j=1}^{m} v_{j,i}^2 = d_i^2 \in \square_q \backslash \{0\}.$$

Using a similar calculation as in the proof of Theorem 3.22 we obtain $v_{1,i} = 0$ and $\sum_{j=1}^{m} v_{j,i}^2 = 1$ for all $2 \leqslant i \leqslant m$. Therefore A has the form

$$A = \begin{pmatrix} 1 & 0 & 0 & \dots \\ 0 & v_{1,2} & v_{1,3} & \dots \\ 0 & v_{2,2} & v_{2,3} & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} =: \begin{pmatrix} 1 & 0 & \dots \\ 0 & & \\ \vdots & A' & \end{pmatrix}.$$

Since A is an automorphism of $\mathbb{F}_q^m$ with respect to Δ the matrix A' is an automorphism of $\mathbb{F}_q^{m-1}$ with respect to Δ. Due to $\sum_{j=1}^{m} v_{j,i}^2 = 1$ for all $2 \leqslant i \leqslant m$ and the induction hypothesis we have $A' \in \mathrm{O}(m-1, \mathbb{F}_q)$. Thus we have $A \in \mathrm{O}(m, \mathbb{F}_q)$. □

4 Graph of integral distances

It turns out that it is usefull to model integral point sets as cliques of certain graphs. For a given prime power $q = p^r$ and a given dimension m we define a graph $\mathfrak{G}_{m,q}$ with vertex set $\mathbb{F}_q^m$, where two vertices v and w are adjacent if $d^2(v, w) \in \square_q$. In this section we want to study the properties of $\mathfrak{G}_{m,q}$. A motivation for this study is that the graph $\mathfrak{G}_{2,q}$ for dimension $m = 2$ is a strongly regular graph. A graph is strongly regular, if there exist positive integers k, λ, and μ such that every vertex has k neighbors, every adjacent pair of vertices has λ common neighbors, and every nonadjacent pair has μ common neighbors, see e. g. [19]. If we denote the number of vertices by v, our graph $\mathfrak{G}_{2,q}$ has the parameters $(v, k, \lambda, \mu) =$

$$\left(q^2, \frac{q^2 + 2q - 3}{2}, \frac{q^2 + 4q - 9}{4}, \frac{q^2 + 4q + 3}{4}\right)$$

for $q \equiv 1 \mod 4$ and the parameters $(v, k, \lambda, \mu) =$

$$\left(q^2, \frac{q^2 - 1}{2}, \frac{q^2 - 5}{4}, \frac{q^2 - 1}{4}\right)$$

for $q \equiv 3 \mod 4$. See e. g. [8] for this fact, which is easy to prove.

For $2|q$ or $m = 1$ the graph of integral distances $\mathfrak{G}_{m,q}$ is equivalent to a complete graph on q^m vertices. Thus we assume $2 \nmid q$ and $m \geqslant 3$ in the following.

Since the translations of $\mathbb{F}_q^m$ are automorphisms with respect to Δ acting transitively on the points we know that $\mathfrak{G}$ is a regular graph, which means that every vertex u has an equal number, called the degree of u, of neighbors. Thus we can speak of a degree of $\mathfrak{G}_{m,q}$.

Lemma 4.1 *The degree of $\mathfrak{G}_{3,q}$ is given by*

$$\Delta(\mathfrak{G}_{3,q}) = \begin{cases} (q-1)\cdot\frac{(q+2)(q+1)}{2} & \text{if } q \equiv 1 \mod 4, \\ (q-1)\cdot\frac{q^2+q+2}{2} & \text{if } q \equiv 3 \mod 4. \end{cases}$$

PROOF. It suffices to determine the number of vectors $(a,b,c) \neq (0,0,0)$ fulfilling $a^2+b^2+c^2 \in \square_q$. So let $a^2+b^2+c^2 = d^2$. If $d = 0$ then we have $a^2+b^2 = -c^2$. Using Corollary 3.5 we obtain q^2-1 solutions in this case. For $d \neq 0$ we have $\frac{q-1}{2}$ possible values for d^2. Using Corollary 3.5 we obtain the number of solutions (a,b) of the equation $a^2+b^2 = d^2-c^2$ for all possible values of c and d. Summing up everything gives the stated formula. □

To determine the degree $\Delta(\mathfrak{G}_{m,q})$ of the graph of integral distances $\mathfrak{G}_{m,q}$ in arbitrary dimension we define the three functions

$$S_m(q) := \left|\left\{(a_1,\ldots,a_m) \in \mathbb{F}_q^m \;\middle|\; \sum_{i=1}^m a_i^2 \in \square_q \backslash \{0\}\right\}\right|,$$

$$Z_m(q) := \left|\left\{(a_1,\ldots,a_m) \in \mathbb{F}_q^m \;\middle|\; \sum_{i=1}^m a_i^2 = 0\right\}\right|, \text{ and}$$

$$N_m(q) := \left|\left\{(a_1,\ldots,a_m) \in \mathbb{F}_q^m \;\middle|\; \sum_{i=1}^m a_i^2 \notin \square_q\right\}\right|.$$

The first few functions are given by

$$\begin{aligned} S_1(q) &= q-1, \\ Z_1(q) &= 1, \\ N_1(q) &= 0, \\ S_2(q) &= \begin{cases} \frac{(q-1)^2}{2} & \text{if } q \equiv 1 \mod 4, \\ \frac{q^2-1}{2} & \text{if } q \equiv 3 \mod 4, \end{cases} \\ Z_2(q) &= \begin{cases} 2q-1 & \text{if } q \equiv 1 \mod 4, \\ 1 & \text{if } q \equiv 3 \mod 4, \end{cases} \text{ and} \\ N_2(q) &= \begin{cases} \frac{(q-1)^2}{2} & \text{if } q \equiv 1 \mod 4, \\ \frac{q^2-1}{2} & \text{if } q \equiv 3 \mod 4, \end{cases} \end{aligned}$$

see Corollary 3.5. To recursively determine these functions we can utilize:

Lemma 4.2 *Let Q_0 and Q_1 be the sets defined in Corollary 3.5. Then for dimension $m \geqslant 3$ we have*

$$\begin{aligned} Z_m(q) &= Z_{m-2}(q)\cdot|Q_0| \\ &\quad + \left(q^{m-2} - Z_{m-2}(q)\right)\cdot|Q_1|, \\ S_m(q) &= \frac{q-1}{2}\cdot\left(N_{m-2}(q) + Z_{m-2}(q)\right)\cdot|Q_1| \\ &\quad + \frac{q-3}{2}\cdot S_{m-2}(q)\cdot|Q_1| + S_{m-2}(q)\cdot|Q_0|, \\ N_m(q) &= q^m - S_m(q) - Z_m(q), \text{ and} \\ \Delta(\mathfrak{G}_{m,q}) &= S_m(q) + Z_m(q) - 1. \end{aligned}$$

PROOF. We rewrite the equation $\sum_{i=1}^m a_i^2 = y$ as $a_1^2 + a_2^2 = y - \sum_{i=3}^m a_i^2$ and apply Corollary 3.5. □

Theorem 4.3 *Let $m \geqslant 1$ be arbitrary. For $q \equiv 1 \mod 4$ we have*

$$Z_m(q) = \begin{cases} q^{m-1} & \text{for } m \text{ odd}, \\ q^{m-1} + q^{\frac{m}{2}} - q^{\frac{m-2}{2}} & \text{for } m \text{ even}, \end{cases}$$

$$S_m(q) = \begin{cases} \frac{1}{2}\left(q^m - q^{m-1} + q^{\frac{m+1}{2}} - q^{\frac{m-1}{2}}\right) & \text{for } m \text{ odd}, \\ \frac{1}{2}\left(q^m - q^{m-1} - q^{\frac{m}{2}} + q^{\frac{m-2}{2}}\right) & \text{for } m \text{ even}, \end{cases}$$

$$N_m(q) = \begin{cases} \frac{1}{2}\left(q^m - q^{m-1} - q^{\frac{m+1}{2}} + q^{\frac{m-1}{2}}\right) & \text{for } m \text{ odd}, \\ \frac{1}{2}\left(q^m - q^{m-1} - q^{\frac{m}{2}} + q^{\frac{m-2}{2}}\right) & \text{for } m \text{ even}, \end{cases}$$

and $\Delta(\mathfrak{G}_{m,q}) =$

$$\begin{cases} \frac{1}{2}\left(q^m + q^{m-1} + q^{\frac{m+1}{2}} - q^{\frac{m-1}{2}}\right) - 1 & \text{for } m \text{ odd}, \\ \frac{1}{2}\left(q^m + q^{m-1} + q^{\frac{m}{2}} - q^{\frac{m-2}{2}}\right) - 1 & \text{for } m \text{ even}. \end{cases}$$

For $q \equiv 3 \mod 4$ we have $Z_m(q) =$

$$\begin{cases} q^{m-1} & \text{for } m \text{ odd}, \\ q^{m-1} + (-q)^{\frac{m}{2}} + (-q)^{\frac{m-2}{2}} & \text{for } m \text{ even}, \end{cases}$$

$S_m(q) =$

$$\begin{cases} \frac{1}{2}\left(q^m - q^{m-1} - (-q)^{\frac{m+1}{2}} - (-q)^{\frac{m-1}{2}}\right) & \text{for } m \text{ odd}, \\ \frac{1}{2}\left(q^m - q^{m-1} - (-q)^{\frac{m}{2}} - (-q)^{\frac{m-2}{2}}\right) & \text{for } m \text{ even}, \end{cases}$$

$N_m(q) =$

$$\begin{cases} \frac{1}{2}\left(q^m - q^{m-1} + (-q)^{\frac{m+1}{2}} + (-q)^{\frac{m-1}{2}}\right) & \text{for } m \text{ odd}, \\ \frac{1}{2}\left(q^m - q^{m-1} - (-q)^{\frac{m}{2}} - (-q)^{\frac{m-2}{2}}\right) & \text{for } m \text{ even}, \end{cases}$$

and $\Delta(\mathfrak{G}_{m,q}) =$

$$\begin{cases} \frac{1}{2}\left(q^m + q^{m-1} - (-q)^{\frac{m+1}{2}} - (-q)^{\frac{m-1}{2}}\right) - 1 & \text{for } m \text{ odd}, \\ \frac{1}{2}\left(q^m + q^{m-1} + (-q)^{\frac{m}{2}} + (-q)^{\frac{m-2}{2}}\right) - 1 & \text{for } m \text{ even}. \end{cases}$$

PROOF. Induction on m utilizing Lemma 4.2. □

With strongly regular graphs in mind we consider the number of common neighbors.

Theorem 4.4 *If $N_c^+(m,q)$ denotes the number of common neighbors of 0 and e_1 in $\mathbb{F}_q^m \backslash \{0, e_1\}$, then for odd m we have* $N_c^+(m,q) =$

$$\frac{q^{m-2}(q+1)^2 + (-1)^{\frac{(m-1)(q-1)}{4}} q^{\frac{m-3}{2}}\left(3q^2-2q-1\right)}{4} - 2$$

and for even $m \geqslant 2$ we have $N_c^+(m,q) =$

$$\frac{q^{m-2}(q+1)^2 + 2(-1)^{\frac{m(q-1)}{4}} q^{\frac{m-2}{2}}(q-1)}{4} - 2.$$

PROOF. Clearly we have $N_c^+(1,q) = q-2$. For $m \geqslant 1$ we count the number of solutions $(v_1,\ldots,v_m)$ of the equation system

$$\begin{aligned} v_1^2 + \sum_{i=2}^{m} v_i^2 &= x^2, \\ (v_1-1)^2 + \sum_{i=2}^{m} v_i^2 &= y^2. \end{aligned}$$

There are $\left(\frac{q+1}{2}\right)^2$ different pairs (x^2,y^2) for $x,y \in \mathbb{F}_q$. For given x^2,y^2 we have $v_1 = \frac{x^2-y^2+1}{2}$ and $\sum_{i=2}^{m} v_i^2 = \frac{4x^2y^2-\left(x^2+y^2-1\right)^2}{4} = -\left(\frac{x^2-y^2-1}{2}\right)^2 + y^2 =: d$. Each of the $\left(\frac{q+1}{2}\right)^2$ cases leads to a specific $d \in \mathbb{F}_q$. Now let a_i be the number of pairs (x^2,y^2) which result in $d=i$. Then we obtain

$$\sum_{i\in\mathbb{F}_q} a_i = \left(\frac{q+1}{2}\right)^2. \tag{2}$$

By $b_{i,m}$ we denote the number of vectors $(v_2 \ \ldots \ v_m) \in \mathbb{F}_q^{m-1}$ with $\sum_{j=2}^{m} v_j^2 = i$. With this we have

$$N_c^+(m,q) = \sum_{i\in\mathbb{F}_q} a_i \cdot b_{i,m} - 2. \tag{3}$$

Due to $N_c^+(1,q) = q-2$ and $b_{i,1} = 0$ for $i \neq 0$ we have $a_0 = q$. If $i,j \in \square_q\backslash\{0\}$ or $i,j \notin \square_q$ then we have $b_{i,m} = b_{j,m}$. Next we show

$$a_s := \sum_{i\in\square_q\backslash\{0\}} a_i = \begin{cases} \frac{(q+1)(q-1)}{8} & \text{for } q \equiv 1 \mod 4 \\ \frac{(q-1)(q-3)}{8} & \text{for } q \equiv 3 \mod 4, \end{cases} \tag{4}$$

from which we can conclude

$$a_n := \sum_{i\notin\square_q} a_i = \begin{cases} \frac{(q-1)(q-3)}{8} & \text{for } q \equiv 1 \mod 4 \\ \frac{(q+1)(q-1)}{8} & \text{for } q \equiv 3 \mod 4, \end{cases}$$

due to Equation (2). We use the information for dimension $m=2$. For $i \notin \square_q$ we have $b_{i,2} = 0$ and for $i \in \square_q\backslash\{0\}$ we have $b_{i,2} = 2$. For $q \equiv 3 \mod 4$ we have $b_{0,2} = 1$ and for $q \equiv 1 \mod 4$ we have $b_{0,2} = 1$. Inserting this and the formula for $N_c^+(2,q)$ in Equation (3) yields Equation (4).

Using $b_{0,m} = Z_{m-1}(q)$, $b_{i,m} = \frac{2}{q-1} \cdot S_{m-1}(q)$ for $i \in \square_q\backslash\{0\}$, and $b_{i,m} = \frac{2}{q-1} \cdot N_{m-1}(q)$ for $i \notin \square_q$ we get

$$\begin{aligned} N_c^+(m,q) &= q \cdot Z_{m-1}(q) + a_s \cdot \frac{2}{q-1} \cdot S_{m-1}(q) \\ &\quad + a_n \cdot \frac{2}{q-1} \cdot N_{m-1}(q) - 2 \end{aligned}$$

and we obtain the stated formula by using Theorem 4.3. □

So for dimension $m = 3$ we have $N_c^+(3,q) = \frac{q^3+5q^2-q-9}{4}$ for $q \equiv 1 \mod 4$ and $N_c^+(3,q) = \frac{q^3-q^2+3q-7}{4}$ for $q \equiv 3 \mod 4$.

Lemma 4.5 *For odd dimension $m \geqslant 3$ the graph of integral distances $\mathfrak{G}_{m,q}$ is not a strongly regular graph.*

PROOF. Let us assume that $\mathfrak{G}_{m,q}$ is strongly regular. Then there exist corresponding parameters (v,k,λ,μ) with

$$\begin{aligned} v &= q^m, \\ k &= \Delta(\mathfrak{G}_{m,q}), \text{ and} \\ \lambda &= N_c^+(m,q). \end{aligned}$$

For a strongly connected graph we have the identity $(v-k-1)\mu = k(k-\lambda-1)$, see e. g. [19]. Using Theorem 4.3 and Theorem 4.4 we can utilize this identity to determine μ. For $q \equiv 1 \mod 4$ and m odd we have

$$\begin{aligned} k(k-\lambda-1) &= \frac{q^{\frac{m-3}{2}}(q-1)(q+1)\left(q^{\frac{m-1}{2}}-1\right)}{8} \\ &\quad \cdot \left(q^m + q^{m-1} + q^{\frac{m+1}{2}} - q^{\frac{m-1}{2}} - 2\right), \end{aligned}$$

$$v-k-1 = \frac{q^{\frac{m-1}{2}} \cdot (q-1) \cdot \left(q^{\frac{m-1}{2}}-1\right)}{2}, \quad \text{and}$$

$$\mu = \frac{(q+1)\left(q^m + q^{m-1} + q^{\frac{m+1}{2}} - q^{\frac{m-1}{2}} - 2\right)}{4q}.$$

For $q \equiv 3 \mod 4$ and m odd we obtain $k(k-\lambda-1)$

$$\begin{aligned} &= \frac{q^{\frac{m-3}{2}}(q-1)(q+1)\left(q^{\frac{m-1}{2}} - (-1)^{\frac{m-1}{2}}\right)}{8} \\ &\quad \cdot \left(q^m + q^{m-1} - (-q)^{\frac{m+1}{2}} - (-q)^{\frac{m-1}{2}} - 2\right), \end{aligned}$$

$$v-k-1 = \frac{q^{\frac{m-1}{2}}(q-1)\left(q^{\frac{m-1}{2}} - (-1)^{\frac{m-1}{2}}\right)}{2}, \quad \text{and}$$

$$\mu = \frac{(q+1)\left(q^m + q^{m-1} - (-q)^{\frac{m+1}{2}} - (-q)^{\frac{m-1}{2}} - 2\right)}{4q}.$$

Since for odd $m \geqslant 3$ the denominator of μ is divisible by q and the numerator is not divisible by q, the graph of integral distances $\mathfrak{G}_{m,q}$ is not a strongly regular graph in these cases. □

If we accomplish the same computation for even m then for $q \equiv 1 \mod 4$ we get $k(k-\lambda-1) =$

$$\frac{q^{m-2}(q-1)(q+1)\left(q^{\frac{m}{2}}-1\right)\left(q^{\frac{m}{2}} + q^{\frac{m-2}{2}} + 2\right)}{8},$$

$$\begin{aligned} v-k-1 &= \frac{q^{\frac{m-2}{2}}(q-1)\left(q^{\frac{m}{2}}-1\right)}{2}, \quad \text{and} \\ \mu &= \frac{q^{\frac{m-2}{2}}(q+1)\left(q^{\frac{m}{2}} + q^{\frac{m-2}{2}} + 2\right)}{4}, \end{aligned}$$

and for $q \equiv 3 \mod 4$ we obtain

$$k(k-\lambda-1) = \frac{q^{m-2}(q-1)(q+1)\left(q^{\frac{m}{2}}-(-1)^{\frac{m}{2}}\right)}{8} \cdot \left(q^{\frac{m}{2}}+q^{\frac{m-2}{2}}+2(-1)^{\frac{m}{2}}\right),$$

$$v-k-1 = \frac{q^{\frac{m-2}{2}}(q-1)\left(q^{\frac{m}{2}}-(-1)^{\frac{m}{2}}\right)}{2}, \quad \text{and}$$

$$\mu\frac{q^{\frac{m-2}{2}}(q+1)\left(q^{\frac{m}{2}}+q^{\frac{m-2}{2}}+2(-1)^{\frac{m}{2}}\right)}{4}.$$

So in both cases we have $\mu \in \mathbb{N}$ for even dimension m. Therefore the graph $\mathfrak{G}_{m,q}$ could be strongly regular for even dimension m, and indeed this is our conjecture:

Conjecture 4.6 *If $N_c^0(m,q)$ denotes the number of common neighbors of 0 and an element v with $\langle v,v\rangle = 0$ in $\mathbb{F}_q^m \backslash \{0,v\}$, then for even $m \geqslant 2$ we have $N_c^0(m,q) = N_c^+(m,q)$ and for odd m we have $N_c^0(m,q) =$*

$$N_c^+(m,q) - (-1)^{\frac{(q-1)(m-1)}{4}} \cdot q^{\frac{m-3}{2}} \cdot \frac{q^2-1}{4}.$$

For even dimension m the graph of integral distances $\mathfrak{G}_{m,q}$ is a strongly regular graph.

We remark that due to Theorem 4.8 $N_c^0(m,q)$ is well defined. For $q = p^1$ being a prime we have verified Conjecture 4.6 for small values using computer calculations. More explicitly, Conjecture 4.6 is valid for $(m = 3, p \leqslant 2029)$, $(m = 4, p \leqslant 283)$, $(m = 5, p \leqslant 97)$, $(m = 6, p \leqslant 59)$, $(m = 7, p \leqslant 31)$, and $(m = 8, p \leqslant 23)$. For all dimensions $m \geqslant 3$ the graph of integral distances $\mathfrak{G}_{m,q}$ is at least a slightly generalization of a strongly regular graph.

Definition 4.7 *An **association scheme** with s associate classes on a finite set Ω is a coloring of the edges of the complete undirected graph with vertex-set Ω by s colors such that*

(i) for all i,j,k in $\{1,\ldots,s\}$ there is an integer p_{ij}^k such that, whenever $\{\alpha,\beta\}$ is an edge of color k then $p_{ij}^k =$

$$\left|\{\gamma \in \Omega \mid \{\alpha,\gamma\} \text{ has color } i \text{ and } \{\gamma,\beta\} \text{ has color } j\}\right|;$$

(ii) every color is used at least once;

(iii) there are integers a_i for $i \in \{1,\ldots,s\}$ such that each vertex is contained in exactly a_i edges of color i.

For an introduction to association schemes we refer the interested reader to e. g. [2, 20].

Theorem 4.8 *Let $\Omega = \mathbb{F}_q^m$. In the complete graph with vertex set Ω we color edges with squared distance 0 by 1, edges with squared distance in $\square_q \backslash \{0\}$ by 2, and edges with squared distance not in $\square_q$ by 3. For dimension $m \geqslant 3$ this coloring forms a three-class association scheme on Ω.*

PROOF. From Lemma 3.21 we know that the automorphism group $\mathrm{Aut}\left(\mathbb{F}_q^m, \Delta\right)$ acts transitively on the edges of each of the three color classes. Thus the integers p_{ij}^k and a_i exist. Due to Theorem 4.3 every color is used at least once. □

For the special class of three-class association schemes and their properties we refer to [18]. We remark that we have $a_1 = Z_m(q) - 1$, $a_2 = S_m(q)$, and $a_3 = N_m(q)$ for $\mathbb{F}_q^m$, so that we can apply Theorem 4.3 to obtain explicit formulae. The values p_{ij}^k are harder to compute. For $m = 3$ we only state a conjecture:

Conjecture 4.9 *The intersection numbers p_{ij}^1, p_{ij}^2, and p_{ij}^3 for the three class association scheme corresponding to the integral distances in $\mathbb{F}_q^3$ are given by*

$$\begin{pmatrix} q-2 & \frac{q(q-1)}{2} & \frac{q(q-1)}{2} \\ \frac{q(q-1)}{2} & \frac{q(q-1)(q-3)}{4} & \frac{q(q-1)^2}{4} \\ \frac{q(q-1)}{2} & \frac{q(q-1)^2}{4} & \frac{q(q-1)(q+1)}{4} \end{pmatrix},$$

$$\begin{pmatrix} q+1 & \frac{(q+1)(q-3)}{2} & \frac{(q+1)(q-1)}{2} \\ \frac{(q+1)(q-3)}{2} & 1+\frac{(q-3)(q-1)^2}{4} & \frac{(q+1)(q-1)^2}{4} \\ \frac{(q+1)(q-1)}{2} & \frac{(q+1)(q-1)^2}{4} & \frac{(q+1)(q-1)^2}{4} \end{pmatrix},$$

$$\begin{pmatrix} q-1 & \frac{(q-1)^2}{2} & \frac{(q+1)(q-1)}{2} \\ \frac{(q-1)^2}{2} & \frac{(q-1)^3}{4} & \frac{(q-1)^3}{4} \\ \frac{(q+1)(q-1)}{2} & \frac{(q-1)^3}{4} & \frac{q^3+q^2-5q-1}{4} \end{pmatrix}$$

for $q \equiv 3 \mod 4$ and by

$$\begin{pmatrix} q-2 & \frac{q(q-1)}{2} & \frac{q(q-1)}{2} \\ \frac{q(q-1)}{2} & \frac{q(q-1)(q+1)}{4} & \frac{q(q-1)^2}{4} \\ \frac{q(q-1)}{2} & \frac{q(q-1)^2}{4} & \frac{q(q-1)(q-3)}{4} \end{pmatrix},$$

$$\begin{pmatrix} q-1 & \frac{(q-1)(q+1)}{2} & \frac{(q-1)^2}{2} \\ \frac{(q-1)(q+1)}{2} & \frac{q^3+q^2-5q-1}{4} & \frac{(q-1)^3}{4} \\ \frac{(q-1)^2}{2} & \frac{(q-1)^3}{4} & \frac{(q-1)^3}{4} \end{pmatrix},$$

$$\begin{pmatrix} q+1 & \frac{(q-1)(q+1)}{2} & \frac{(q-3)(q+1)}{2} \\ \frac{(q-1)(q+1)}{2} & \frac{(q-1)^2(q+1)}{4} & \frac{(q-1)^2(q+1)}{4} \\ \frac{(q-3)(q+1)}{2} & \frac{(q-1)^2(q+1)}{4} & \frac{q^3-5q^2+7q+1}{4} \end{pmatrix}$$

for $q \equiv 1 \mod 4$.

5 Maximum cardinality of integral point sets in $\mathbb{F}_q^m$

By Definition 2.2 an integral point set $\mathcal{P}$ over $\mathbb{F}_q^m$ is a subset of $\mathbb{F}_q^m$, where all pairs of points are at integral distance. In Definition 2.3 we have introduced the notion $\mathcal{I}(\mathbb{F}_q, m)$ for the maximum cardinality of an integral point set over $\mathbb{F}_q^m$. Since for $m = 1$ or $2|q$ all distances in $\mathbb{F}_q^m$ are integral, we have $\mathcal{I}(\mathbb{F}_q, m) = q^m$ in these cases. We have already stated $\mathcal{I}(\mathbb{F}_q, 2) = q$ for $2 \nmid q$ in Theorem 2.6. Combining this with the obvious bound $\mathcal{I}(\mathbb{F}_q, m) \leqslant q \cdot \mathcal{I}(\mathbb{F}_q, m-1)$ we obtain

$$\mathcal{I}(\mathbb{F}_q, 3) \leqslant q^2 \tag{5}$$

for $2 \nmid q$.

Theorem 5.1 *If* $q \equiv 1 \mod 4$ *then we have* $\mathcal{I}(\mathbb{F}_q, 3) = q^2$.

PROOF. Consider the point set

$$\mathcal{P} := \{(a, \omega_q a, b) \mid a, b \in \mathbb{F}_q\}.$$

This point set is an integral point set of cardinality q^2. Geometrically it is a hyperplane. □

Using the graph of integral distances $\mathfrak{G}_{m,q}$ from Section 4 the problem of determining $\mathcal{I}(\mathbb{F}_q, m)$ is transfered to the well known problem of the determination of the maximum cardinality of cliques, these are complete subgraphs, in $\mathfrak{G}_{m,q}$. For the latter problem there are software packages as e. g. CLIQUER [16], available.

Thus for small values m and q the maximum cardinality $\mathcal{I}(\mathbb{F}_q, m)$ can be determined exactly using computer calculations. In the remaining part of this section we will deal with the cases $q \nmid 2$, $m \geqslant 3$. Since $\mathfrak{G}_{m,q}$ consists of q^m vertices one should reduce the problem whenever possible. One possibility is to predescribe points which must be contained in the clique. Due to Lemma 3.21 it suffices to investigate the two cases where we predescribe $0 \in \mathbb{F}_q^m$ and an arbitrary element of $\mathcal{P}^+$ or $\mathcal{P}_0$.

Now let us consider the special case of dimension $m = 3$ and $q \equiv 3 \mod 4$. Due to $\mathcal{I}(\mathbb{F}_q, m) \geqslant q$ we can restrict our search on cliques C with cardinality at least $q + 1$. Thus there exists $z \in \mathbb{F}_q$ such that the hyperplane $\{(x, y, z) \mid x, y \in \mathbb{F}_q\}$ contains at least two points of a clique C with cardinality at least $q + 1$. We assume $z = 0$ and since for $q \equiv 3 \mod 4$ the equation $x^2 + y^2$ has the unique solution $(0, 0)$ w.l.o.g. we predescribe the points $(0, 0, 0)$ and $(1, 0, 0)$. Additionally we know the following: Either in such a clique C there exists a third point in the hyperplane with third coordinate being equal to zero, or there exist two points in a hyperplane with third coordinate being equal to an element of $\mathbb{F}_q^*$, or every hyperplane with fix third coordinate contains at least one element of the clique C. Using theses properties we were able to determine the following values of $\mathcal{I}(\mathbb{F}_q, 3)$ for small $q \equiv 3 \mod 4$:

q	3	7	11	19	23	27	31
$\mathcal{I}(\mathbb{F}_q, 3)$	4	8	11	19	23	28	31

These first few values and some theoretical considerations lead us to:

Conjecture 5.2 *We have* $\mathcal{I}(\mathbb{F}_q, 3) \in \{q, q + 1\}$ *for* $q \equiv 3 \mod 4$.

For any given point $(a, b, c) \neq (0, 0, 0)$ the point set $(a, b, c) \cdot \mathbb{F}_q$ is an integral point set over $\mathbb{F}_q^3$ of cardinality q. For $q \equiv 3 \mod 4$ there is another nice construction of an integral point set in $\mathbb{F}_q^3$ with cardinality q. At first we construct an integral point set *on a circle*, see [8]. Therefore we consider the field $\mathbb{F}_q' := \mathbb{F}_q[x]/\left(x^2 + 1\right)$. For $z = a + bx \in \mathbb{F}_q'$ with $a, b \in \mathbb{F}_q$ we set $\bar{z} := a - bx \in \mathbb{F}_q'$, which mimics the complex conjugation. Now let z be a generator of the cyclic group $\mathbb{F}_q'\backslash\{0\}$. We define $\mathcal{C}_q' := \left\{z \in \mathbb{F}_q' \mid z\bar{z} = 1\right\}$. It is not difficult to check that $\mathcal{C}_q'$ corresponds to an integral point set over $\mathbb{F}_q^2$ of cardinality $\frac{q+1}{2}$, see [8]. By $\mathcal{C}_q$ we denote the corresponding integral point set over $\mathbb{F}_q^3$, where the third coordinates of the points are equal to zero. Now we define the set $L := \left\{x \mid x^2 + 1 \in \square_q\right\}$ which has cardinality $\frac{q-1}{2}$ for $q \equiv 3 \mod 4$. With this notation we can state:

Lemma 5.3 *For* $q \equiv 3 \mod 4$ *the set* $\mathcal{C}_q \cup (0, 0, 1) \cdot L$ *is an integral point set over* $\mathbb{F}_q^3$ *with cardinality* q.

Lemma 5.4 *For* $q \equiv 3 \mod 4$ *there exists a hyperplane* H *with squared distances being either* 0 *or non-squares.*

PROOF. Due to Corollary 3.5 there exist $a, b \in \mathbb{F}_q$ with $a^2 + b^2 = -1$. We set $v_1 := (a, b, 1)$ and $v_2 := (-b, a, 0)$. This gives $\langle v_1, v_1\rangle = 0$, $\langle v_2, v_2\rangle = -1 \notin \square_q$, and $\langle v_1, v_2\rangle = 0$. Now let $H := \{xv_1 + yv_2 \mid x, y \in \mathbb{F}_q\}$. The squared distance of two elements $x_i v_1 + y_i v_2 \in H$, $i = 1, 2$, is given by $(y_1 - y_2)^2 \cdot \langle v_2, v_2\rangle \notin \square_q\backslash\{0\}$. □

Corollary 5.5 *Let* $\mathcal{P}$ *be an integral point set in* $\mathbb{F}_q^3$ *for* $q \equiv 3 \mod 4$. *Either* $|\mathcal{P}| \leqslant q$ *or some squared distances equal zero.*

PROOF. We consider a covering of $\mathbb{F}_q^3$ by q translations of the plane of Lemma 5.4. □

We remark that our two examples of integral point sets of cardinality q for $q \equiv 3 \mod 4$ do not contain a squared distance being equal to zero.

Since for $q \equiv 3 \mod 4$ integral point sets over $\mathbb{F}_q^3$ of cardinality $q + 1$ seem to be something special we want to list the examples that we have found by our clique search. For $\mathcal{I}(\mathbb{F}_{27}, 3) = 28$ an example is given by $\big\{(2 + 2w + 2w^2, 2 + w^2, w^2)$, $(0, 2w + 2w^2, 1 + 2w)$, $(1, 1 + w + w^2, w)$, $(2, 0, 0)$, $(2, w^2, 2 + w)$, $(2, 2w^2, 1 + 2w)$, $(2, 2w + 2w^2, 1 + 2w)$, $(w, 2 + 2w, 2 + 2w + 2w^2)$, $(2w, 2w^2, 2 + 2w + w^2)$, $(2 + 2w, w^2, 2 + w + w^2)$, $(2 + 2w, w + 2w^2, w + 2w^2)$, $(2 + 2w, 2w + 2w^2, 2w)$, $(w^2, 2 + w + w^2, 1 + 2w^2)$, $(1 + w^2, 2w + 2w^2, 2w)$, $(0, 0, 0)$, $(2 + w^2, 1 + 2w, 2w^2)$, $(1 + w + w^2, w^2, 2 + w^2)$, $(2 + w + w^2, w^2, 0)$, $(1, 0, 0)$, $(2w + w^2, 1 + 2w + 2w^2, 2 + 2w + 2w^2)$, $(2 + 2w + w^2, 2 + 2w^2, 1)$, $(1, 0, 1 + w^2)$, $(1 + 2w^2, w + w^2, 2w)$, $(w + 2w^2, 1 + w^2, 1 + w + 2w^2)$, $(2 + w + 2w^2, 2 + w, 2 + w + 2w^2)$, $(2 + w + 2w^2, 2 + 2w, 2w + w^2)$, $(1 + 2w + 2w^2, 2 + w + w^2, 2w + w^2)$, $(1 + 2w + 2w^2, 2w + 2w^2, 2w)\big\}$, where we use $\mathbb{F}_{27} \simeq \mathbb{F}_3[w]/\left(w^3 + w^2 + w + 2\right)$. For $q = 7$ we have $\big\{$ $(0, 0, 0)$, $(1, 0, 0)$, $(0, 0, 1)$, $(1, 5, 5)$, $(2, 1, 3)$, $(3, 1, 2)$, $(5, 5, 1)$, $(6, 3, 6)$ $\big\}$ and for $q = 3$ we have

$$\Big\{(0, 0, 0), (1, 0, 0), (2, 1, 1), (2, 2, 1)\Big\}$$

as examples.

For higher dimensions we know some more exact numbers, see [9, 10]: $\mathcal{I}(\mathbb{F}_3, 4) = 9$, $\mathcal{I}(\mathbb{F}_3, 5) = 27$, $\mathcal{I}(\mathbb{F}_3, 6) =$

33, $\mathcal{I}(\mathbb{F}_5,4) = 25$, $\mathcal{I}(\mathbb{F}_5,5) = 125$, $\mathcal{I}(\mathbb{F}_7,4) = 49$, $\mathcal{I}(\mathbb{F}_7,5) = 343$, and $\mathcal{I}(\mathbb{F}_{11},4) = 121$.

To obtain lower bounds we can consider pairs of integral point sets $\mathcal{P}_1 \subset \mathbb{F}_q^{m_1}$ and $\mathcal{P}_2 \subset \mathbb{F}_q^{m_2}$, where all squared distances in $\mathcal{P}_2$ are equal to zero. With $\mathcal{P} := \{(p_1|p_2) \mid p_1 \in \mathcal{P}_1, p_2 \in \mathcal{P}_2\}$ we obtain an integral point set in $\mathbb{F}_q^{m_1+m_2}$ with cardinality $|\mathcal{P}_1| \cdot |\mathcal{P}_2|$.

Theorem 5.6 *For $q \equiv 1 \mod 4$, $m \geqslant 1$, and $2n \leqslant m$ we have*

$$\mathcal{I}(\mathbb{F}_q, m) \geqslant q^n \cdot \mathcal{I}(\mathbb{F}_q, m-2n) \geqslant q^{\lceil \frac{m}{2} \rceil},$$

where we set $\mathcal{I}(\mathbb{F}_q, 0) = 1$.

PROOF. We set

$$\mathcal{P}_2 := \{(a_1, a_1\omega_q, \ldots, a_n, a_n\omega_q) \mid a_1, \ldots, a_n \in \mathbb{F}_q\}$$

in the above described construction. Thus we have $\mathcal{I}(\mathbb{F}_q, m) \geqslant q^n \cdot \mathcal{I}(\mathbb{F}_q, m-2n)$ for all $2n \leqslant m$. The remaining inequality can by proven by induction on m using $\mathcal{I}(\mathbb{F}_q, m) \geqslant q$. □

Lemma 5.7 *There exists an integral point set $\mathcal{P}_2$ in $\mathbb{F}_q^4$ of cardinality q^2, where all squared distances are equal to zero.*

PROOF. Let (x, y) be a solution of $x^2 + y^2 = -2$ in $\mathbb{F}_q$. By Corollary 3.5 there are at least $q - 1 \geqslant 1$ such solutions. We consider the vectors $u = (x, y, 1, 1)$ and $v = (-y, x, -1, 1)$. Clearly u and v are linearly independent and fulfill $\langle u, v\rangle = 0$, $\langle u, u\rangle = 0$, and $\langle v, v\rangle = 0$. We set $\mathcal{P}_2 := \{au + bv \mid a, b \in \mathbb{F}_q\}$. It suffices to check $d^2(0, au + bv) = 0$ for all $a, b \in \mathbb{F}_q$. Indeed we have

$$\begin{aligned} d^2(0, au + bv) &= \langle au + bv, au + bv\rangle \\ &= a^2\langle u, u\rangle + 2ab\langle u, v\rangle + b^2\langle v, v\rangle \\ &= 0. \end{aligned}$$

□

Theorem 5.8 *For $q \equiv 3 \mod 4$, $m \geqslant 1$, and $4n \leqslant m$ we have*

$$\begin{aligned} \mathcal{I}(\mathbb{F}_q, m) &\geqslant q^{2n} \cdot \mathcal{I}(\mathbb{F}_q, m-4n) \\ &\geqslant q^{2\cdot\lfloor \frac{m}{4} \rfloor + \left\lceil \frac{m}{4} - \lfloor \frac{m}{4} \rfloor \right\rceil} \\ &\geqslant q^{\lfloor \frac{m}{2} \rfloor}, \end{aligned}$$

where we set $\mathcal{I}(\mathbb{F}_q, 0) = 1$.

PROOF. We choose $\mathcal{P}_2$ as the n-fold cartesian product from the integral point set of Lemma 5.7 in the construction described above Lemma 5.6. Thus we have $\mathcal{I}(\mathbb{F}_q, m) \geqslant q^{2n} \cdot \mathcal{I}(\mathbb{F}_q, m-4n)$ for all $2n \leqslant m$. The remaining inequality can by proven by induction on m using $\mathcal{I}(\mathbb{F}_q, m) \geqslant q$. □

We would like to remark that there is a nice connection between integral distances over $\mathbb{F}_q^m$ and coding theory, at least for $q = 3$. Therefore let us denote by $d_H : \mathbb{F}_q^m \times \mathbb{F}_q^m \to \mathbb{N}_0$ the Hamming distance, which maps two vectors $u = (u_1, \ldots, u_m)$ and $v = (v_1, \ldots, v_m)$ to $d_H(u, v) = |\{i \mid u_i \neq v_i,\ 1 \leqslant i \leqslant m\}|$. For $q = 3$ and $u, v \in \mathbb{F}_3^m$ we have $d^2(u, v) \in \square_3$ if and only if $d_H(u, v) \not\equiv 2 \mod 3$.

If we denote by $\mathbb{H}^m$ the subsets of $\mathbb{F}_2^m$, where all Hamming distances are congruent to 0 or 1 modulo 4, then we have $\mathcal{I}(\mathbb{Z}_4, m) = 2^m \cdot \max_{S \in \mathbb{H}^m} |S|$. For the proofs and a more detailed description we refer the reader to [10].

6 Conclusion and outlook

For the study of discrete structures the knowledge of their automorphism group is very important. In Section 3 we have completed the determination of the automorphism group of $\mathbb{F}_q^m$ with respect to integral distances.

The graphs $\mathfrak{G}_{q,m}$ of integral distances are interesting combinatorial objects. We were able to determine a few parameters and properties, but the large part remains unsettled. It would be nice to have a proof of Conjecture 4.6, which maybe is not too difficult.

Section 5 gives a first glimpse on the maximum cardinalities $\mathcal{I}(\mathbb{F}_q, m)$ of integral point sets over $\mathbb{F}_q^m$. It remains a task for the future to determine some more exact numbers or lower and upper bounds. E.g. for small q we have no idea for a general construction of integral point sets with maximum cardinality. A detailed analysis of the parameters of the 3-class association schemes including the eigenvalues of the corresponding graphs could be very usefull to utilize some general upper bounds on clique sizes. A geometrical description of the point sets achieving $\mathcal{I}(\mathbb{F}_q, 3) = q + 1$ for $q \equiv 3 \mod 4$ would be interesting.

There are some similarities between integral point sets over $\mathbb{F}_q^m$ and integral point sets over Euclidean spaces $\mathbb{E}^m$. E.g. the constructions which lead to the maximum cardinality $\mathcal{I}(\mathbb{F}_q, m)$ in $\mathbb{F}_q^m$ often coincide with the constructions which lead to integral point sets over $\mathbb{E}^m$ with minimum diameter, see [12, 14, 15].

Bibliography

[1] A. Antonov and M. Brancheva, *Algorithm for finding maximal Diophantine figures*, Spring Conference 2007 of the Union of Bulgarian Mathematicians, 2007.

[2] R. A. Bailey, *Association schemes. Designed experiments, algebra and combinatorics*, Cambridge Studies in Advanced Mathematics 84. Cambridge: Cambridge University Press. 387 p., 2004.

[3] A. Blokhuis, *On subsets of $GF(q^2)$ with square differences*, Indag. Math. **46** (1984), 369–372.

[4] S. Dimiev, *A setting for a Diophantine distance geometry*, Tensor (N.S.) **66** (2005), no. 3, 275–283. MR MR2189847

[5] R. E. Fullerton, *Integral distances in banach spaces*, Bull. Amer. Math. Soc. **55** (1949), 901–905.

[6] H. Harborth, *Integral distances in point sets*, Butzer, P. L. (ed.) et al., Karl der Grosse und sein Nachwirken. 1200 Jahre Kultur und Wissenschaft in Europa. Band 2: Mathematisches Wissen. Turnhout: Brepols, 1998, pp. 213–224.

[7] B. Huppert, *Endliche Gruppen. I*, Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen. 134. Berlin-Heidelberg-New York: Springer-Verlag. 793 p., 1967.

[8] M. Kiermaier and S. Kurz, *Inclusion-maximal integral point sets in affine planes over finite fields*, (submitted).

[9] A. Kohnert and S. Kurz, *Integral point sets over* $\mathbb{Z}_n^m$, Electron. Notes Discrete Math. **27** (2006), 65–66.

[10] A. Kohnert and S. Kurz, *Integral point sets over* $\mathbb{Z}_n^m$, Discrete Appl. Math., (to appear).

[11] T. Kreisel and S. Kurz, *There are integral heptagons, no three points on a line, no four on a circle*, Discrete Comput. Geom., (to appear).

[12] S. Kurz, *Konstruktion und Eigenschaften ganzzahliger Punktmengen*, Ph.D. thesis, Bayreuth. Math. Schr. 76. Universität Bayreuth, 2006.

[13] S. Kurz, *Integral point sets over finite fields*, (submitted).

[14] S. Kurz and R. Laue, *Upper bounds for integral point sets*, Australas. J. Comb. **39** (2007), 233–240.

[15] S. Kurz and A. Wassermann, *On the minimum diameter of plane integral point sets*, Ars Combin., (to appear).

[16] S. Niskanen and P. R. J. Östergård, *Cliquer user's guide, version 1.0*, Tech. Report T48, Communications Laboratory, Helsinki University of Technology, Espoo, Finland, 2003.

[17] D. E. Taylor, *The geometry of the classical groups*, Sigma Series in Pure Mathematics Volume 9. Heldermann Verlag Berlin, 1992.

[18] E. R. van Dam, *Graphs with few eigenvalues*, Ph.D. thesis, University of Gent, 1996.

[19] D. B. West, *Introduction to graph theory*, 2nd ed., New Delhi: Prentice-Hall of India. 608 p., 2005.

[20] P.-H. Zieschang, *Theory of association schemes*, Springer Monographs in Mathematics. Berlin: Springer. 283 p., 2005.

Chapter 9

Inclusion-maximal integral point sets over finite fields

MICHAEL KIERMAIER[1] AND SASCHA KURZ[2]

ABSTRACT. We consider integral point sets in affine planes over finite fields. Here an integral point set is a set of points in $\mathbb{F}_q^2$ where the formally defined Euclidean distance of every pair of points is an element of $\mathbb{F}_q$. From another point of view we consider point sets over $\mathbb{F}_q^2$ with few and prescribed directions. So this is related to Rédei's work. Another motivation comes from the field of ordinary integral point sets in Euclidean spaces $\mathbb{E}^m$.

In this article we study the spectrum of integral point sets over $\mathbb{F}_q$ which are maximal with respect to inclusion. We give some theoretical results, constructions, conjectures, and some numerical data.

2000 MSC: 51E20; 05B25.

Key words and phrases: integral distances, exhaustive search, finite geometry, Paley graphs.

1 Introduction

The study of geometrical objects with integral edge lengths has been attractive for mathematicians for ages. The first result may be obtained be the Pythagoreans considering rectangles with integral side and diagonal lengths. A slight generalization of this problem is even unsolved up to our times. Is there a perfect box? This is a rectangular parallelepiped with all edges, face diagonals and space diagonals of integer lengths [11, 15]. In a more general context one is interested in the study of integral point sets, see [12, 21, 22] for an overview. As originally introduced integral point sets are sets of n points in the m-dimensional Euclidean space $\mathbb{E}^m$ with pairwise integral distances. Here the most results are known for dimension $m = 2$, see e. g. [12, 13, 18, 21, 22, 27]. Although integral point sets were studied for a long time our knowledge is still very limited.

[1]Michael Kiermaier, University of Bayreuth, Department of Mathematics, D-95440 Bayreuth, Germany.
E-mail adress: michael.kiermaier@uni-bayreuth.de

[2]Sascha Kurz, University of Bayreuth, Department of Mathematics, D-95440 Bayreuth, Germany.
E-mail adress: sascha.kurz@uni-bayreuth.de

So Stancho Dimiev [10] came up with the idea of studying integral point sets over finite rings with the hope that the situation in the finite case is easier and that some structure of the problem may be preserved. So for a commutative ring $\mathcal{R}$ with 1 we consider point sets $\mathcal{P}$ over $\mathcal{R}^2$. For two points $u = (u_1, u_2), v = (v_1, v_2)$ in $\mathcal{R}^2$ we define the squared distance as $d^2(u, v) := N(u - v) := (u_1 - v_1)^2 + (u_2 - v_2)^2 \in \mathcal{R}$. We say that two points u, v are at integral distance if there is an element $r \in \mathcal{R}$ with $d^2(u, v) = r^2$, meaning that the distance is an element of $\mathcal{R}$. Here an integral point set is a set of points in $\mathcal{R}^2$ with pairwise integral distances. For residue rings $\mathcal{R} = \mathbb{Z}_n$ first results were obtained in [10, 16].

If the ring $\mathcal{R}$ is a finite field we clearly have a bit more algebraic tools at hand to attack the problem in this special case. So in [19] one of the authors studied integral point sets over $\mathbb{F}_q^2$ and classified those integral point sets with maximum cardinality up to isomorphism almost completely, see Section 3 for the definition of isomorphic integral point sets. To state the classification result we need some notation. For an odd prime power q there are exactly $\frac{q+1}{2}$ squares in $\mathbb{F}_q$. The set of squares will be denoted by $\square_q$. We have $-1 \in \square_q$ if and only if $q \equiv 1 \pmod 4$. In this case ω_q will denote a fixed element with $\omega_q^2 = -1$. With this we can state:

Theorem 1.1 *(Kurz, 2007 [19])*
Let $q = p^r$ be a prime power. If $2|q$ then $\mathbb{F}_q^2$ is an integral point set otherwise the maximum cardinality of an integral point set $\mathcal{P}$ over $\mathbb{F}_q^2$ is given by q. If $q \equiv 3 \pmod 4$ then each integral point set of this maximum cardinality is isomorphic to $(1, 0) \cdot \mathbb{F}_q$. If $q = p \equiv 1 \pmod 4$ then each integral point set of this maximum cardinality is isomorphic $(1, 0) \cdot \mathbb{F}_q$, $(1, \omega_q) \cdot \mathbb{F}_q$, or $(1, \omega_q) \cdot \square_q \cup (1, -\omega_q) \cdot \square_q$.

The key ingredient for this result was a theorem on point sets over $\mathbb{F}_q^2$ with few directions. Here two points (x_1, y_1), (x_2, y_2) have the direction $\frac{y_1 - y_2}{x_1 - x_2} \in \mathbb{F}_q \cup \{\infty\}$.

Theorem 1.2 *(Ball, Blokhuis, Brouwer, Storme, Szőnyi, 1999 [7]; Ball 2003 [5])*
Let $f : \mathbb{F}_q \to \mathbb{F}_q$, where $q = p^n$, p prime, $f(0) = 0$. Let $N = |D_f|$, where D_f is the set of directions determined by the function f. Let e (with $0 \leqslant e \leqslant n$) be the largest integer

such that each line with slope in D_f meets the graph of f in a multiple of p^e points. Then we have the following:

(1) $e = 0$ *and* $\frac{q+3}{2} \leqslant N \leqslant q+1$,

(2) $p^e > 2$, $e|n$, *and* $\frac{q}{p^e} + 1 \leqslant N \leqslant \frac{q-1}{p^e-1}$,

(3) $e = n$ *and* $N = 1$.

Moreover, if $p^e > 2$, *then* f *is a linear map on* $\mathbb{F}_q$ *viewed as a vector space over* $\mathbb{F}_{p^e}$. *If* $e = 0$ *and* $N = \frac{q+3}{2}$ *then* f *is affinely equivalent to* $f(x) = x^{\frac{q-1}{2}}$. *(All possibilities for* N *can be determined in principle.)*

In [3] the Bulgarian group around Dimiev considered integral point sets over $\mathbb{F}_p^2$ for $p \equiv 3 \mod 4$ which are maximal with respect to inclusion. They classified the maximal integral point sets up to isomorphism for $p = 7, 11$ and conjectured that the maximal integral point sets have either cardinality $\frac{p+3}{2}$ or p. In the latter case all p points are on a line. Theorem 1.1 clears the situation for cardinality p. In this article we disprove their conjecture about the spectrum of possible cardinalities of maximal integral point sets and classify them for $q \leqslant 47$.

2 The graph of integral distances

It turns out that it is usefull to model integral points sets as cliques of certain graphs.

Definition 2.1 *For a fixed prime power* $q = p^r$ *we define the graph* $\mathfrak{G}$ *with vertex set* $\mathbb{F}_q^2$, *where two vertices* v *and* w *are adjacent if* $d(v, w) \in \square_q$. *So two different vertices are connected by an edge exactly if they are at integral distance. The graph* $\mathfrak{G}$ *will be called* graph of integral distances.

Furthermore, we recall that for $q \equiv 1 \pmod 4$ *the Paley-graph* Paley(q) *is defined as the graph with vertex set* $\mathbb{F}_q$ *where two vertices* v *and* w *are adjacent if* $v - w \in \square_q \backslash \{0\}$.

2.1 The case $q \equiv 3 \pmod 4$

Theorem 2.2 *For* $q \equiv 3 \pmod 4$ *it holds:* $\mathfrak{G} \cong$ Paley$\left(q^2\right)$.

PROOF. We define the two sets

$$M := \left\{(x, y) \in \mathbb{F}_q^2 \mid x^2 + y^2 \in \square_q\right\}$$

and

$$N := \left\{(x, y) \in \mathbb{F}_q^2 \mid x + yi \in \square_{q^2}\right\}$$

Obviously, $|N| = |\square_{q^2}| = \frac{q^2+1}{2}$, and by Lemma 3.3: $|M| = |P_0| + \frac{q-1}{2}|P_1| = |N|$. Let $(x, y) \in N$. Then there exist $a, b \in \mathbb{F}_q$ with $(a + bi)^2 = x + yi$. That implies $x = a^2 - b^2$ and $y = 2ab$, and we get $x^2 + y^2 = \left(a^2 + b^2\right)^2 \in \square_q$. Hence $(x, y) \in M$. Because of the finiteness of M and N we get $M = N$ and the proof is complete. □

Now we can apply the existing theory for the Paley-graphs on our situation. For example, $\mathfrak{G}$ is a strongly regular graph with parameters $(v, k, \lambda, \mu) =$

$$\left(q^2, \frac{q^2-1}{2}, \frac{q^2-5}{4}, \frac{q^2-1}{4}\right)$$

In [6] Aart Blokhuis determined the structure of cliques of maximal size in Paley graphs of square order: A clique of maximal size of $\mathfrak{G}$ is an affine line in $\mathbb{F}_q^2$. This implies that the size of a maximal integral point set in $\mathbb{F}_q$ is q, and—anticipating the definitions of the next section—these point sets are unique up to isomorphism.

Maximal cliques of size $\frac{q+1}{2}$ and $\frac{q-1}{2}$ in Paley graphs of square order can be found in [4].

2.2 The case $q \equiv 1 \pmod 4$

Theorem 2.3 *For* $q \equiv 1 \pmod 4$, $\mathfrak{G}$ *is a strongly regular graph with parameters* $(v, k, \lambda, \mu) =$

$$\left(q^2, \frac{(q-1)(q+3)}{2}, \frac{(q+1)(q+3)}{4} - 3, \frac{(q+1)(q+3)}{4}\right)$$

PROOF. The graph consists of q^2 vertices of degree $\frac{(q-1)(q+3)}{2}$ (there are $\frac{q+3}{2}$ integral directions and $q-1$ further points of one direction). We consider two points u and v which are at a non-integral distance. From each point there are $\frac{q+3}{2}$ integral directions. Since the direction from u to v is non-integral and non-parallel lines intersect in exactly one point we have $\mu = \frac{q+3}{2} \cdot \frac{q+1}{2}$. For the determination of λ we consider two points u and v at integral distance. Thus the direction from u to v is integral and all points on this line have integral distances to u and v. There are $\frac{q+1}{2}$ further integral directions from u and from v respectively Each pair intersects, if not parallel, in exactly one point. By a short calculation we can verify the stated value for λ. □

We remark that the parameters of the complementary graph of $\mathfrak{G}$ are $(v, k, \lambda, \mu) =$

$$\left(q^2, \frac{(q-1)^2}{2}, \frac{(q-1)(q-3)}{4} + 1, \frac{(q-1)(q-3)}{4}\right).$$

In both cases $\mathfrak{G}$ corresponds to an orthogonal array. We have $\mathfrak{G} \in \mathrm{OA}\left(q, \frac{q+1}{2}\right)$ for $q \equiv 3 \mod 4$ and $\mathfrak{G} \in \mathrm{OA}\left(q, \frac{q+3}{2}\right)$.

3 Automorphism group

It will be convenient to identify the affine plane $\mathbb{F}_q^2$ with the ring $\mathbb{F}_q[i]$ where i is a root of the polynomial $X^2 + 1 \in \mathbb{F}_q[X]$. With this identification, the map $N : (\mathbb{F}_q[i], \cdot) \to (\mathbb{F}_q, \cdot)$ is a monoid homomorphism. In the case $q \equiv 3 \pmod 4$ we have $-1 \notin \square_q$, so $X^2 + 1$ is irreducible and $\mathbb{F}_q[i] \cong \mathbb{F}_{q^2}$. For $q \equiv 1 \pmod 4$, $\mathbb{F}_q$ is a finite ring with two non-trivial ideals, namely $\mathbb{F}_q(\omega_q + i)$ and $\mathbb{F}_q(\omega_q - i)$. These ideals are of order q and contain the

zero-divisors of $\mathbb{F}_q[i]$. An element $z = x + iy \in \mathbb{F}_q[i]$ is a zero-divisor iff $N(z) = 0$.

In the introduction we announced that we want to classify maximal integral point sets up to isomorphism. So we have to specify what we consider as an automorphism. Now we say that an automorphism of $\mathbb{F}_q^2$ is a bijection $\tau : \mathbb{F}_q^2 \to \mathbb{F}_q$ with $d(x,y) \in \square_q \Leftrightarrow d(\tau(x),\tau(y)) \in \square_q$ for every $x, y \in \mathbb{F}_q^2$. This is exactly the automorphism group G of the corresponding graph $\mathfrak{G}$ of integral distances. For Paley graphs of square order the automorphism group was known since a while, see [9, 14, 28]. If we also request that automorphisms τ map lines to lines, then the automorphism group of $\mathbb{F}_q^2$ for $2 \nmid q$ was determined in [19].

Theorem 3.1 *(Kurz, 2007 [19])*
Let $q = p^r \neq 5, 9$ an odd prime power, G the automorphism group of $\mathbb{F}_q^2$ and $H := G \cap A\Gamma L(2, \mathbb{F}_q)$. Then H is generated by

1. $x \mapsto x + v$ *for all* $v \in \mathbb{F}_q^2$,
2. $x \mapsto \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \cdot x$,
3. $x \mapsto \begin{pmatrix} a & b \\ b & -a \end{pmatrix} \cdot x$ *for all* $\{a, b\} \subseteq \mathbb{F}_q^2$ *such that* $a^2 + b^2$ *is a square, and*
4. $(a, b) \mapsto \left(a^{p^i}, b^{p^i}\right)$ *for* $i \in \mathbb{N}$.

In the next section we will describe an algorithm which calculates maximal integral point sets up to isomorphism. In order to make this algorithm really fast we want to demand weaker conditions on automorphisms as in [19] and not use the multiplication in $\mathbb{F}_q[i]$ or that lines must be mapped onto lines. Strictly speaking we choose the automorphism group G of the corresponding graph $\mathfrak{G}$ of integral distances instead of H. It will turn out that a distinction between these two slightly different definitions of automorphisms is not necessary since we have $G \simeq H$ in many cases.

Definition 3.2 *A triple (a, b, c) is called* Pythagorean triple over $\mathbb{F}_q$ *if* $a^2 + b^2 = c^2$.

In the following it will be useful to have a parametric representation of the Pythagorean triples over $\mathbb{F}_q$.

Lemma 3.3 *For $2 \nmid q$ let $c \in \mathbb{F}_q$ and P_c the set of Pythagorean triples (a, b, c) over $\mathbb{F}_q$.*

(a) Case $c = 0$

$$P_0 = \begin{cases} \{(t, \pm t\omega_q, 0) \mid t \in \square_q\} & \text{if } q \equiv 1 \pmod 4, \\ \{(0,0,0)\} & \text{if } q \equiv 3 \pmod 4, \end{cases}$$

$$|P_0| = \begin{cases} 2q - 1 & \text{if } q \equiv 1 \pmod 4, \\ 1 & \text{if } q \equiv 3 \pmod 4. \end{cases}$$

(b) Case $c \neq 0$

$$P_c = \left\{(\pm c, 0, c)\right\} \cup \left\{ \left(\frac{t^2-1}{t^2+1} \cdot c, \frac{2t}{t^2+1} \cdot c, c \right) \;\middle|\; t \in \mathbb{F}_q^*, t^2 \neq 1 \right\}$$

$$|P_c| = \begin{cases} q - 1 & \text{if } q \equiv 1 \pmod 4, \\ q + 1 & \text{if } q \equiv 3 \pmod 4. \end{cases}$$

(c) There are exactly q^2 Pythagorean triples over $\mathbb{F}_q$.

PROOF. Part (a) is easy to verify. For part (b) there are 4 solutions with $ab = 0$, these are $\{(0, \pm c, c), (\pm c, 0, c)\}$. For $ab \neq 0$ we get:

$$a^2 + b^2 = c^2 \quad \Leftrightarrow \quad \frac{c-a}{b} \cdot \frac{c+a}{b} = 1.$$

Setting $t := \frac{c+a}{b} \in \mathbb{F}_q^\star$, we obtain

$$\frac{a}{b} = \frac{t - t^{-1}}{2} \quad \text{and} \quad \frac{c}{b} = \frac{t + t^{-1}}{2}$$

Because of $a \neq 0$, $c \neq 0$ we have $t^2 \notin \{-1, 1\}$. It follows

$$a = \frac{t - t^{-1}}{t + t^{-1}} \cdot c \quad \text{and} \quad b = \frac{2}{t + t^{-1}} \cdot c$$

It is easily checked that for all admissible values of t, the resulting triples (a, b, c) are pairwise different Pythagorean triples.

The expression for the number of solutions follows because -1 is a square in $\mathbb{F}_q$ exactly if $q \equiv 1 \pmod 4$.

With part (a) and part (b) we get the number of Pythagorean triples over $\mathbb{F}_q$ as

$$\sum_{c \in \mathbb{F}_q} |P_c| = |P_0| + (q-1)|P_1| = q^2$$

So also part (c) is shown. □

With the help of Lemma 3.3 we can easily deduce for $q \neq 5, 9$,

$$|H| = \begin{cases} q^2(q-1)^2 r & \text{if } q \equiv 1 \pmod 4, \\ q^2(q-1)(q+1) r & \text{if } q \equiv 3 \pmod 4. \end{cases}$$

For $q = 5$ we have $|H| = 800$ and for $q = 9$ we have $|H| = 31104$. It is not difficult to prove $G \simeq H$ for $q = p \neq 5, 9$ being a prime. But since we need the automorphism groups only for small q we simply have utilized `nauty` [23] for $q \leqslant 167$. We have obtained $|G| = 28800$ for $q = 5$, $|G| = 186624$ for $q = 9$, and

$$|G| = \begin{cases} q^2(q-1)^2 r^2 & \text{if } q \equiv 1 \pmod 4, \\ q^2(q-1)(q+1) r^2 & \text{if } q \equiv 3 \pmod 4. \end{cases}$$

for the remaining cases with $q \leqslant 167$.

From Theorem 3.1 we can deduce:

Corollary 3.4 *For two points $p_1 \neq p_2 \in \mathbb{F}_q'$ at integral distance there exists an automorphism φ with either $\varphi(p_1) = 0$, $\varphi(p_2) = 1$ or $\varphi(p_1) = 0$, $\varphi(p_2) = 1 + \omega_q x$.*

4 Inclusion-maximal integral point sets over $\mathbb{F}_q^2$

For the classification of inclusion-maximal integral point sets over $\mathbb{F}_q^2$ we can use Corollary 3.4 to prescribe the point $(0,0)$. Now we build up a graph $\mathcal{G}_q$ which consists of the elements of $\mathbb{F}_q^2$ which are at integral distance to $(0,0)$. Between two nodes $x, y \in \mathbb{F}_q^2$ there is an edge if and only if $d(x,y) \in \square_q$. For practical purposes the generation of all $\frac{q^2-1}{2}$ points which are at integral distance to $(0,0)$ can be easily done by a for loop, as in [3]. For theoretic applications one can deduce a parametric solution from Lemma 3.3. The cliques of $\mathcal{G}_q$ are in bijection to the inclusion-maximal integral point sets over $\mathbb{F}_q^2$. Thus we may use a clique search program, f.e. `cliquer` [24], to search for inclusion-maximal integral point sets.

For the classification up to isomorphism we use an orderly algorithm in combination with `nauty` [23] as described in [26] on the graph $\mathcal{G}_q$. To guarantee that this approach yields the correct classification we have to ensure that the automorphism group of $\mathcal{G}_q$ equals the automorphism group of the original problem. In our case we have simply checked this condition using `nauty`. We remark that we consider G as the automorphism group of the original problem and not $H \leqslant G$.

If we denote by $\mathcal{A}_{q,s}$ the number of non-isomorphic inclusion-maximal integral point sets over $\mathbb{F}_q^2$ we have obtained the following results with the above described algorithm. For $q \equiv 3 \mod 4$ we have:

q	Σ	3	5	7	8	9	10	11	12	13	14
3	1	1									
7	2		1	1							
11	4			3				1			
19	54			25	7	19		4			
23	294			85	108	80	7	9		4	
27	645			27	411	142	50	12			
31	6005			60	2004	2734	933	199	26	46	
43	231890			15	1748	54700	109127	54759	9785	1490	156
47	805783			12	1097	125545	434029	210725	28533	4904	628

q	Σ	15	16	17	19	23	25	27	31	43	47
3	1										
7	2										
11	4										
19	54				1						
23	294					1					
27	645	2						1			
31	6005			2					1		
43	231890	87		20		2				1	
47	805783	230	27	50			2				1

Conjecture 4.1 *For each* $q \equiv 3 \mod 4$ *there exist* $l_q, r_q \in \mathbb{N}$ *such that* $r_q \leqslant \frac{q-1}{2}$, $\mathcal{A}_{q,l_q} > 0$, $\mathcal{A}_{q,r_q} > 0$, $\mathcal{A}_{q,\frac{q+3}{2}} > 0$, $\mathcal{A}_{q,q} > 0$, *and* $\mathcal{A}_{q,s} = 0$ *for* $s \notin \left\{l_q, \ldots, r_q, \frac{q+3}{2}, q\right\}$.

For $q \equiv 1 \mod 4$ we have:

q	Σ	5	6	7	8	9	10	11	12	13	14	15
5	1	1										
9	4		2			2						
13	30		2	11	8	5	1			3		
17	107			8	57	24	12	2		1		
25	488			9	122	148	108	41	23	17	8	4
29	9693			6	893	4264	2864	1230	284	116	22	6
37	103604			1	314	17485	44952	24067	10645	4835	906	234
41	347761			1	1169	61940	149839	86159	33941	10854	2891	646

q	Σ	15	16	17	18	19	20	21	22	23	25	29	37	41
5	1													
9	4													
13	30													
17	107		3											
25	488	1	2		1					4				
29	9693	3	2								3			
37	103604	89	55	11	2	3	1			1		3		
41	347761	136	131	27	16		4	3	1	1			3	

Conjecture 4.2 *For each* $q \equiv 1 \mod 4$ *there exist* $l_q, r_q, \in \mathbb{N}$ *such that* $\mathcal{A}_{q,l_q} > 0$, $\mathcal{A}_{q,r_q} > 0$, $\mathcal{A}_{q,q} > 0$, *and* $\mathcal{A}_{q,s} = 0$ *for* $s \notin \{l_q, \ldots, r_q, q\}$.

So clearly the spectrum of possible cardinalities of inclusion-maximal integral point sets of $\mathbb{F}_q^2$ is a bit more complicated as conjectured in [3]. We would like to remark that for $q \in \{59, 67, 71, 79, 83, 103, 107, 127, 131, 139, 151, 163\}$ the second largest inclusion-maximal integral point set has size $\frac{q+3}{2}$, which was verified using an clique search approach. Besides the maximum and the second largest cardinality of an integral point set also the minimum cardinality of a maximal integral point set is of interest. Here we remark that we have $l_q = 7$ for $q \neq 13$ and $11 \leqslant q \leqslant 47$. For $q \in \{49, 53, 59, 61, 67, 73\}$ we have $l_q = 8$, $l_{71} = 9$, and $l_{79} \in \{8, 9\}$.

Later on we will prove $l_q \geqslant 5$ for $q \geqslant 5$, see Lemma 5.1.

Conjecture 4.3 *For each* $w \in \mathbb{N}$ *there exists a* $q_w \in \mathbb{N}$ *such that we have* $l_q \geqslant w$ *for* $q \geqslant q_w$, *meaning* $\mathcal{A}_{q,s} = 0$ *for* $s < w$ *and* $q \geqslant q_w$.

From Theorem 1.1 we can conclude the following corollary:

Corollary 4.4 *For* $2 \mid q$ *we have* $\mathcal{A}_{q,q^2} = 1$ *and all other numbers equal* 0. *For* $2 \nmid q$ *we have* $\mathcal{A}_{q,s} = 0$ *if* $s > q$. *Additionally we have* $\mathcal{A}_{q,q} \geqslant 1$ *if* $q \equiv 3 \mod 4$ *and* $\mathcal{A}_{q,q} \geqslant 3$ *if* $q \equiv 1 \mod 4$.

Conjecture 4.5 *For* $q \equiv 3 \mod 4$ *and* $q \geqslant 7$ *the second largest cardinality of an inclusion-maximal integral point set over* $\mathbb{F}_q^2$ *is* $\frac{q+3}{2}$.

To have a deeper look at the second largest inclusion-maximal integral point sets we need some lemmas from [19].

Lemma 4.6 *In* $\mathbb{F}_q[i]$ *the set* $N^{-1}(1) = \{z \in \mathbb{F}_q[i] \mid z\bar{z} = 1\}$ *is a cyclic multiplicative group.*

PROOF. If $-1 \notin \square_q$ then $\mathbb{F}_q[i]$ is a field and thus C must be cyclic. For the case $-1 \in \square_q$ we utilize the bijection

$$\rho_q : \mathbb{F}_q^* \to N^{-1}(1), \quad t \mapsto \frac{1+t^2}{2t} + \omega_q \frac{1-t^2}{2t} x.$$

Now we only have to check that the mapping is a group isomorphism, namely

$$\rho_q(i \cdot j) = \rho_q(i) \cdot \rho_q(j).$$

□

Lemma 4.7 *For $z \in \mathcal{R}'$ with $z\bar{z} = 1$ the set $\mathcal{P} = \{z^{2i} \mid i \in \mathbb{N}\}$ is an integral point set.*

PROOF. With $c := a - b$ we have

$$\begin{aligned} d\left(z^{2a}, z^{2b}\right) &= \left(z^{2a} - z^{2b}\right) \cdot \overline{(z^{2a} - z^{2b})} \\ &= \left(z^{2c} - 1\right) \cdot \overline{z^{2c} - 1} \\ &= 2 - z^{2c} 2 - \overline{z^{2c}} \\ &= \left(\underbrace{z^c i - \overline{z^c} i}_{\in \mathcal{R}}\right)^2 \end{aligned}$$

□

These two lemmas allow us to do a circle construction. We choose a generator z of the cyclic group $N^{-1}(1)$ and set $\mathcal{P}_W := \{z^{2i} \mid i \in \mathbb{N}\} \cup \{0\}$. With this we have

$$|\mathcal{P}_W| = \begin{cases} \frac{q+1}{2} & \text{if } q \equiv 1 \pmod 4, \\ \frac{q+3}{2} & \text{if } q \equiv 3 \pmod 4. \end{cases}$$

It is easy to check that $\mathcal{P}_W$ is an integral point set. For the order of the automorphism group we would like to mention

$$|\mathrm{Aut}(\mathcal{P}_W)| = \begin{cases} (q-1)r & \text{if } p^r = q \equiv 1 \pmod 4, \\ (q+1)r & \text{if } p^r = q \equiv 3 \pmod 4. \end{cases}$$

Theorem 4.8 *For $q \notin \{5, 9\}$ $\mathcal{P}_W$ is a maximal integral point set.*

PROOF. We identify the affine plane $\mathbb{F}_q^2$ with the field $\mathbb{F}_q[i]$. Let ζ a generator of the cyclic group $N^{-1}(1)$. Assume that there is a point $a \in \mathbb{F}_q[i] \setminus \mathcal{P}_W$ such that $\mathcal{P}_W \cup \{a\}$ is an integral point set. Then $N(a) \in \square_q$. The map $\rho : z \mapsto \zeta z^2$ is an integral automorphism. Let A the orbit of a with respect to $\langle\rho\rangle$. Then $\mathcal{P}' := \mathcal{P}_W \cup A$ is an integral point set, because $N\left(\zeta^{2n} a - \zeta^{2m} a\right) = N(a) N\left(\zeta^{2n} - \zeta^{2m}\right) \in \square_q$ and $N\left(\zeta^{2n} a - \zeta^{2m}\right) = N\left(a - \zeta^{2m-2n}\right) \in \square_q$.

Furthermore, let $\xi = u + vi \neq 1$ with $N(\xi) = 1$ and $u, v \in \mathbb{F}_q$. Because of $N(\xi) = u^2 + v^2 = 1$, we have $u \neq 1$. So $N(\xi - 1) = 2(1-u) \neq 0$ and $\xi - 1$ is invertible. Thus $\zeta^n a = a$ is equivalent to $(\zeta^n - 1) = 0$, and we get that $|A|$ equals the multiplicative order of ζ. For $q \equiv 3 \pmod 4$ we get $|\mathcal{P}_W \cup A| = q + 2$, a contradiction to the maximum cardinality of an integral point set, see Theorem 1.1. In the case $q \equiv 1 \pmod 4$ we have $|\mathcal{P}_W \cup A| = q$. We can easily check that in $\mathcal{P}_W \setminus \{0\}$ no three points are collinear. Thus with $\mathcal{P}_W \setminus \{0\}$ also $\mathcal{P}' = \mathcal{P}_W \cup A$ determines at least $\frac{q-3}{2}$ different integral directions. So for $p > 3$ we are either in case (1) or case (3) of Theorem 1.2. In case (1) there is a subset of $\frac{q+1}{2}$ collinear points. This is possible only for $q = 9$. For case (3) all q points are situated on a line. This is possible only for $q = 5$. For the other values $q < 11$ we have checked the stated result via a computer calculation.

So only the case $p = 3$ is left. Here we only have to consider case (2) of Theorem 1.2 and $e = 1$. So every line through two points of $\mathcal{P}'$ meets the point set $\mathcal{P}'$ in a multiple of $p^e = 3$ points. Let us fix a point $P \in \mathcal{P}_W \setminus \{0\}$. There are $\frac{q-3}{2}$ different lines trough P and a further point of $\mathcal{P}_W \setminus \{0\}$. For each of these lines l we have $|A \cap l| \leqslant 2$. One such line l' meets 0. Thus $A \cap l = \emptyset$. For all other lines l ($\#= \frac{q-5}{2}$) we have $|A \cap l| = 1$. Let $B := \{b \in \mathbb{F}_q[i] \mid N(b) = N(a)\}$. We have $|B| = q - 1$. So all points of $B \setminus A$ lie on these lines l and l'. There are two points $P_1, P_2 \in A$ left which are not on these lines l or l'. Since we have that the point 0 is met by the line l'' through P_1 and P_2, we have that no further point of $\mathcal{P}'$ is situated on l''. Thus we have two additional integral directions $\overline{PP_1}$ and $\overline{PP_2}$. Thus there are in total at least $\frac{q+1}{2}$ integral directions determined by $\mathcal{P}'$, which is too much for case (2). □

Remark 4.9 *For $q \in \{5, 9\}$ the set $\mathcal{P}_W$ can be extended to an integral point set of size q.*

To describe another construction we need some further lemmas. (Most of them are already stated and proven in [19].)

Lemma 4.10 *An integral point set over $\mathbb{F}_q^2$ determines at most $\frac{q+3}{2}$ different directions if $-1 \in \square_q$ and at most $\frac{q+1}{2}$ different directions if $-1 \notin \square_q$.*

PROOF. We consider the points $p = a + bi$ at integral distance to 0. Thus there exists an element $c' \in \mathbb{F}_q$ with $a^2 + b^2 = c'^2$. In the case $a = 0$ we obtain the direction ∞. Otherwise we set $d := \frac{b}{a}$ and $c := \frac{c'}{a}$, yielding $1 = c^2 - d^2 = (c - d)(c + d)$, where d is the direction of the point. Now we set $c + d =: t \in \mathbb{F}_q^*$ yielding $c = \frac{t + t^{-1}}{2}$, $d = \frac{t - t^{-1}}{2}$. The two values t and $-t^{-1}$ produce an equal direction. Since $t = -t^{-1} \Leftrightarrow t^2 = -1$ we get the desired bounds. □

Definition 4.11 *A line with slope $d = \frac{y}{x}$ is called vanishing line if $x^2 + y^2 = 0$. We call the direction d a vanishing direction. In all other cases d is called an integral direction if $1 + d^2 \in \square_q$ or non-integral direction if $1 + d^2 \notin \square_q$. The slope $d = \frac{1}{0} = \infty$ is integral.*

We remark that a vanishing line can only occur for $-1 \in \square_q$ and in this case there are exactly two different corresponding slopes, $d = \omega_q$ and $d = -\omega_q$. A line with an integral direction forms an integral point set. Similar a line with a non-integral directions forms a non-integral point set. The vanishing lines form both integral and non-integral point sets.

It is well known that $\mathrm{PGL}(2, q)$ acts transitively one the pairs of a line l and a point p not on l. For the automorphism group of integral point sets we have a similar result.

Lemma 4.12 *If L_i is the set of integral lines, L_n the set of non-integral lines, and L_v the set of vanishing lines in $\mathbb{F}_q^2$, then the automorphism group Aut of integral point sets acts transitively on the pairs (l, p) where $l \in L$, $p \in \mathbb{F}_q^2$, $p \notin l$ for $L \in \{L_i, L_n, L_v\}$.*

PROOF. We can easily check that the automorphism group Aut acts transitively on L_i, L_n, and L_v. Also after applying an automorphism l and p are not incident. Let $d = \frac{y}{x}$ be the slope of l. Now the multiplication by an invertible element

$r \in \mathbb{F}_q^*$ or the addition of a vector $r \cdot (x,y)^T$ let l fix. These two types of automorphisms suffice to map each two points $p, p' \notin l$ onto each other. □

Lemma 4.13 *If $-1 \notin \square_q$, $2 \nmid q$ and $\mathcal{P}$ is a non-collinear integral point set over $\mathbb{F}_q^2$, then each line l contains at most $\frac{q-1}{2}$ points.*

PROOF. If l is a line with an integral pair of points on it, then its slope is an integral direction. Now we consider the intersections of lines with integral directions containing a point $p \notin l$, with the line l. □

We remark that this lemma was already proved in [4, 8].

Lemma 4.14 *If $\mathcal{P}$ is a non-collinear integral point set over $\mathbb{F}_q^2$ then every line l contains at most $\frac{q+1}{2}$ points.*

PROOF. Analog to the proof of Lemma 4.13. □

Now we construct another maximal integral point set $\mathcal{P}_L$. Therefore let us choose a non-vanishing integral line l and an arbitrary point p not on l. Let p' be the mirror point of p on l. If we draw the lines of integral directions from p we receive some intersections with l. These points together with p and p' form an integral point set $\mathcal{P}_l$ (*orthogonal* directions are either both integral, both non-integral, or both vanishing). For $q \equiv 3 \mod 4$ we have $|\mathcal{P}_L| = \frac{q+3}{2}$ and for $q \equiv 1 \mod 4$ we have $|\mathcal{P}_L| = \frac{q+5}{2}$.

Theorem 4.15 *The integral point set $\mathcal{P}_L$ is maximal for $q \equiv 3 \mod 4$.*

PROOF. We identify the affine plane $\mathbb{F}_q^2$ with the field $\mathbb{F}_q[i]$. Without loss of generality we choose the line $\mathbb{F}_q$ and the point i not on $\mathbb{F}_q$, that is $\mathcal{P} = \mathbb{F}_q \cup \{\pm i\}$. For a point $P = x + iy \in \mathbb{F}_q[i] \setminus \mathbb{F}_q$, let σ_P be the map $\mathbb{F}_q[i] \to \mathbb{F}_q[i]$, $z \mapsto x + yz$ and $S(P)$ the set of the $\frac{q-1}{2}$ points in $\mathbb{F}_q$ which have integral distance to P. For all automorphisms ϕ it holds $S(\phi(P)) = \phi(S(P))$. Our strategy is to prove that $S(i) = S(P)$ and $d(i,P) \in \square_q$ only holds for $P = \pm i$.

It is easily checked that for all $P \in \mathbb{F}_q[i]$ we have $\sigma_P(\mathbb{F}_q) = \mathbb{F}_q$, $\sigma_P(i) = P$ and σ_P is an automorphism.

Now we define the set of automorphisms $A = \{\sigma_P : P \in \mathbb{F}_q[i] \setminus \mathbb{F}_q\}$ and the subset $B = \{\sigma_{(x,y)} : x \in \mathbb{F}_q, y \in \mathbb{F}_q \setminus \{0,1\}\}$. Clearly, A is a subgroup of G and acts regularly on $\mathbb{F}_q[i] \setminus \mathbb{F}_q$, so $\sigma_P^k = \sigma_{\sigma_P^k(i)}$. For $\sigma_{(x,y)} \in B$ it holds:

- $\sigma_{(x,y)}$ has exactly one fixed point Q on $\mathbb{F}_q[i]$, namely $Q = \frac{x}{1-y}$. Furthermore, $d(i, x+yi) \in \square_q \Leftrightarrow Q \in \mathbb{F}_q$.
- For each $k \in \mathbb{N}$: $\sigma_{(x,y)}^k(z) = x\frac{y^k-1}{y-1} + y^k z$ and in particular: $\sigma_{(x,y)}^{q-1} = \mathrm{id}$.
- For all $z \in \mathbb{F}_q[i]$: $\sigma^k(z) - z = \left(y^k - 1\right)\left(\frac{x}{y-1} + z\right)$, so the point set $\left\{\sigma_{(x,y)}^k(z) : k \in \mathbb{N}\right\} \subseteq z + \mathbb{F}_q \cdot \left(\frac{x}{y-1} + z\right)$ is collinear.

It follows that for $\sigma \in B$ we have $\sigma^k \in B \cup \{\mathrm{id}\}$ and that the order of each element in B divides $q-1$.

Now we assume that $P = x + yi \neq \pm i$ is a point not in $\mathcal{P}$ such that $S(P) = S(i)$ and $d(P,i) \in \square_q$.

(1) The case $\sigma_P \notin B$:
In this case σ_P is a translation and has no fixed point. Since $\gcd(q, q-1) = 1$ we clearly have $S(i) \neq \sigma_P(S(i)) = S(\sigma_P(i)) = S(P)$, a contradiction.

(2) The case $\sigma_P \in B$, where the order p of σ_P is prime:
As seen above, p divides $q-1$. The group action of $\langle\sigma_P\rangle$ on $S(i)$ partitions $S(i)$ into orbits of size p and one fixed point. Hence $p \mid -1 + |S(i)| = q-3$, which yields $p = 2$. In B there is only one automorphism of order 2, it is $z \mapsto -z$. So $P = \sigma_P(i) = -i$, a contradiction.

(3) The case $\sigma_P \in B$, where the order k of σ_P is not prime:
Because of $\sigma_P(i) = P \neq i$ we have $k \neq 1$. Since $k \mid q-1$ and $4 \nmid q-1$, k has a prime factor $p \neq 2$. We set $\tau := \sigma^{p^{-1}k}$, which is an element of B of order p. With $Q = \tau(i)$ we have $\tau = \sigma_Q$. The points i, $P = \sigma(i)$ and $Q = \tau(i)$ are collinear, so $d(i,Q) \in \square_q$. One easily verifies $Q \notin \mathcal{P}$ and $S(Q) = S(P) = S(i)$. Now the previous case applied to $\tau = \sigma_Q$ gives a contradiction.

□

For $q \equiv 1 \mod 4$ the situation is a bit harder and we need the following result of Weil, see e. g. [25]:

Theorem 4.16 *Let $f(x)$ be a polynomial over $\mathbb{F}_q$ of degree d without repeated factors and $N := \left|\left\{(x,y) \in \mathbb{F}_q^2 \mid y^2 = f(x)\right\}\right|$ then for $q \geqslant 5$ we have*

$$|N - q| \leqslant (d-1)\sqrt{q}.$$

Theorem 4.17 *The integral point set $\mathcal{P}_L$ is maximal for $9 < q = p^1 \equiv 1 \mod 4$.*

PROOF. We apply the same strategy as in the proof of Theorem 4.15 and adopt the notation. Nevertheless $\mathbb{F}_q[i]$ is not a field for $q \equiv 1 \mod 4$ we can define P, σ_P, A, B, and $\sigma_{(x,y)}$ in the same way. The three statements for $\sigma_{(x,y)} \in B$ remain valid. Also the order of each element in B divides $q-1$. Let us again assume that $P = x + yi \neq \pm i$ is a point not in $\mathcal{P}$ such that $S(P) = S(i)$ and $d(P,i) \in \square_q$. Since $\gcd(q, q-1) = 1$ we conclude $\sigma_P \in B$, see the proof of Theorem 4.15. We have $S(i) = \left\{(u,0) \mid u^2 + 1 \in \square_q\right\} = \left\{(u,0) \mid (u-x)^2 + y^2 \in \square_q\right\}$ and $(0,0), (x,0) \in S(i) = S(P)$. Thus we have the implications $(u,0) \in S(i) \Rightarrow (-u,0) \in S(i)$ and $(u,0) \in S(i) \Rightarrow (2x-u,0) \in S(i)$. We conclude $\{j \cdot (u,0) \mid j \in \mathbb{N}\} \subseteq S(i)$. For q being a prime this is only possible for $x = 0$.

So in the remaining cases we have $x = 0$. Thus we have $S(i) = \left\{(u,0) \mid u^2 + 1 \in \square_q\right\} = \left\{(uy,0) \mid (uy)^2 + 1 \in \square_q\right\}$. We remark that the equation $1 + u^2 = s^2$ has the parameter solution $s = \frac{t+t^{-1}}{2}$, $u = \frac{t^{-1}-t}{2}$ for $t \in \mathbb{F}_q^*$ since $0 \neq s - u = t$. So for all $t \in \mathbb{F}_q^*$

the term $1+y^2\left(\frac{t^{-1}-t}{2}\right)^2$ is a square. By multiplying with $4t^2$ we can conclude that $f(t) := 4t^2+y^2\left(1-t^2\right)^2$ is a square for all $t \in \mathbb{F}_q$. Thus for the N in Theorem 4.16 we have $N \geqslant 2q-4$. So for $q \geqslant 25$ we have that $f(t)$ contains a repeated factor. We simply check the cases $q \leqslant 23$ by computer and now assume $q \geqslant 25$. So there exists an t with $f(t) = f'(t) = 0$ or there exist $a, b, c \in \mathbb{F}_q$ with $f(t) = b\left(a+t^2\right)^2$. We have

$$f'(t) = 8t - 4ty^2 + 4y^2t^3 = t \cdot \left(8 - 4y^2 + 4y^2t^2\right) = 0$$

in the first case. Since $f(0) = y^2 \neq 0$ we have $t^2 = 1 - \frac{2}{y^2}$. Inserting yields $f(t) = 4 - \frac{4}{y^2} = 0$ which is equivalent to $y^2 = 1$ or $y = \pm 1$. In the second case we get $b = y^2$, $a^2 = 1$, and $2(a+1)y^2 = 4$. We conclude $a = 1$ and $y^2 = 1$. Thus $P = \pm 1$. □

We remark that if we would choose l as a vanishing line in the construction of $\mathcal{P}_L$ for $q \equiv 1$ then resulting integral point set could be completed to $(1, \pm\omega_q) \cdot \square_q$.

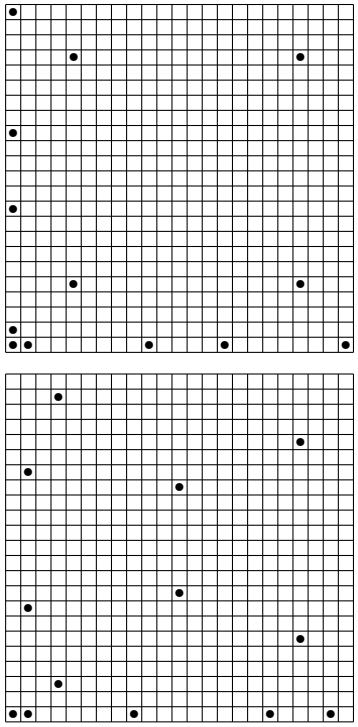


Figure 1: The integral point sets $\mathcal{P}_1$ and $\mathcal{P}_2$.

We summarize that for $q \equiv 3 \bmod 4$ by Theorem 4.8 and Theorem 4.15 we have two constructions showing $\mathcal{A}_{\frac{q+3}{2},q} \geqslant 2$ for $q \geqslant 11$. One may conjecture $\mathcal{A}_{\frac{q+3}{2},q} = 2$ for $q \geqslant 27$ and speak of *sporadic* solutions in the cases $q = 11, 19, 23$. The *sporadic* solutions also have a nice geometric pattern. By z_{11}, z_{19}, and z_{23} we denote an arbitrary generator of the multiplicative group $N^{-1}(1)$ in $\mathbb{F}_{11}[i]$, $\mathbb{F}_{19}[i]$, and $\mathbb{F}_{23}[i]$, respectively. For $q = 23$ the examples are given by

$$\mathcal{P}_1 = \{0\} \cup 1 \cdot \left\langle z_{23}^6 \right\rangle \cup 3 \cdot \left\langle z_{23}^6 \right\rangle \cup 9 \cdot \left\langle z_{23}^6 \right\rangle$$

and

$$\mathcal{P}_2 = \{0\} \cup 1 \cdot \left\langle z_{23}^8 \right\rangle \cup 2 \cdot z_{23}^4 \cdot \left\langle z_{23}^8 \right\rangle \cup 6 \cdot z_{23}^4 \cdot \left\langle z_{23}^8 \right\rangle \cup 8 \cdot \left\langle z_{23}^8 \right\rangle,$$

see Figure 1. For $q = 19$ one of the two examples has a similar shape and is given by

$$\mathcal{P}_3 = \{0\} \cup 1 \cdot \left\langle z_{19}^4 \right\rangle \cup 3 \cdot \left\langle z_{19}^4 \right\rangle,$$

see Figure 2.

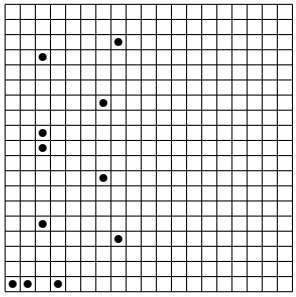


Figure 2: The integral point set $\mathcal{P}_3$.

The second sporadic example $\mathcal{P}_4$ for $q = 19$ and the sporadic example $\mathcal{P}_5$ for $q = 11$ have a different geometric pattern. They are subsets of $N^{-1}(1) \cup \mathbb{F}_q \subset \mathbb{F}_q[i]$, see Figure 3.

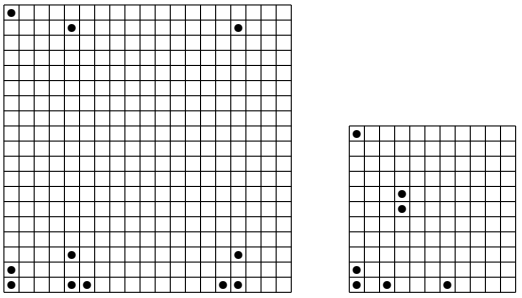


Figure 3: The integral point sets $\mathcal{P}_4$ and $\mathcal{P}_5$.

5 Remarks on integral point sets over $\mathbb{E}^2$

It is interesting to mention that the situation for integral point sets in $\mathbb{E}^2$ is somewhat similar. Since we have an infinite number of points there must not be an integral point set of maximum cardinality. So here we ask for the minimum possible diameter $d(2, n)$ of an integral point set in the Euclidean plane $\mathbb{E}^2$ with pairwise integral distances, where the diameter is the largest occurring distance. Without any extra condition n points on a line would yield an integral point set with small diameter. To make it more interesting one forces integral point sets in $\mathbb{E}^2$ to be two dimensional. Here all known non-collinear examples of integral point sets with minimum diameter consist of a line with $n-1$ points and

one point apart, see [22, 27]. If we forbid 3 points to be collinear integral point sets on circles seem to be the examples with minimum diameter. The situation stays more or less the same if we consider integral point sets over $\mathbb{Z}^2$. These results on the structure of integral point sets over $\mathbb{E}^2$ or $\mathbb{Z}^2$ are up to now only conjectures which are verified for the first few numbers n of points. So this is one motivation to study integral point sets over $\mathbb{F}_q^2$ in the hope that here the situation is easier to handle.

Besides the characterization of the inclusion-maximal integral point sets with largest or second largest cardinality another interesting question is the characterization of those inclusion-maximal integral point sets with minimum cardinality. From our data we may conjecture that for $q \geqslant 11$ we have $\mathcal{A}_{q,s} = 0$ for $s \leqslant 6$. Again we can compare this situation to the situation in $\mathbb{E}^2$. A result due to Almering [1, 2] is the following. Given any integral triangle Δ in the plane, the set of all points x with rational distances to the three corners of Δ is dense in the plane. Later Berry generalized this results to to triangles where the squared side lengths and at least one side length are rational. In $\mathbb{Z}^2$ the situation is a bit different. In [17] the authors search for inclusion-maximal integral triangles over $\mathbb{Z}^2$. They exist but seem to be somewhat rare. There are only seven inclusion-maximal integral triangles with diameter at most 5000. The smallest possible diameter is 2066. In a forthcoming paper [20] one of the authors has extended this list, as a by-product, up to diameter 15000 with in total 126 inclusion-maximal integral triangles. So is very interesting that we have the following lemma:

Lemma 5.1 *If $\mathcal{P}$ is an inclusion-maximal integral point set over $\mathbb{F}_q^2$ for $q \geqslant 5$ then we have $|\mathcal{P}| \geqslant 5$.*

PROOF. For small q we use our classification of maximal integral point sets over $\mathbb{F}_q^2$. If $2|q$ then the only inclusion maximal integral point set over $\mathbb{F}_q$ has size q^2. So we assume w.l.o.g. that q is odd. Clearly an integral point set of cardinality 1 is not inclusion maximal. An integral point set $\mathcal{P}$ of cardinality two can be completed by all other points on the line defined by $\mathcal{P}$. The similar statement holds for three collinear points. So let us assume that we have an inclusion maximal integral triangle $\Delta = \{p_1, p_2, p_3\}$ over $\mathbb{F}_q^2$. Let l be the line through p_2 and p_3. Starting from point p_1 there are at least $\frac{q+1}{2}$ integral directions. Lets draw lines through p_1 for these integral directions. Two of them meet p_2 and p_3, respectively. Since at most of the remaining directions can be parallel to l we can expand Δ by least $\frac{q-5}{2} > 1$ points if $q \geqslant 7$. We remark that for suitable large q the cardinality $|\mathcal{P}| = 4$ may be only possible if $\mathcal{P} = \{P_1, P_2, P_3, P_4\}$ is a point set without a collinear triple. There are six configurations of lines defined by (P_a, P_b), (P_c, P_d) with $\{a, b, c, d\} = \{1, 2, 3, 4\}$. In at least one case the line through P_a, P_b and the line through P_c, P_d are not parallel. W.l.o.g. we assume $a = 1$, $b = 2$, $c = 3$, $d = 4$, $P_1, P_2 \in \mathbb{F}_q$, and $P_3, P_4 \notin \mathbb{F}_q$. The line through P_3 and P_4 intersects the line $\mathbb{F}_q$ in a point $P_5 \in \mathbb{F}_q$. Since $\mathcal{P} \cup \{P_5\}$ is an integral point set and $P_5 \notin \mathcal{P}$ we have the stated result. □

Bibliography

[1] J. H. J. Almering, *Rational quadrilaterals*, Indag. Math. **25** (1963), 192–199.

[2] J. H. J. Almering, *Rational quadrilaterals II*, Indag. Math. **27** (1965), 290–304.

[3] A. Antonov and M. Brancheva, *Algorithm for finding maximal Diophantine figures*, Spring Conference 2007 of the Union of Bulgarian Mathematicians, 2007.

[4] R. D. Baker, G. L. Ebert, J. Hemmeter, and A. Woldar, *Maximal cliques in the Paley graph of square order*, J. Statist. Plann. Inference **56** (1996), no. 1, 33–38.

[5] S. Ball, *The number of directions determined by a function over a finite field*, J. Combin. Theory Ser. A **104** (2003), no. 2, 341–350.

[6] A. Blokhuis, *On subsets of $GF(q^2)$ with square differences*, Indag. Math. **46** (1984), 369–372.

[7] A. Blokhuis, S. Ball, A. E. Brouwer, L. Storme, and T. Szőnyi, *On the number of slopes of the graph of a function defined on a finite field*, J. Combin. Theory Ser. A **86** (1999), no. 1, 187–196.

[8] R. H. Bruck, *Finite nets. II: Uniqueness and imbedding*, Pacific J. Math. **13** (1963), 421–457.

[9] L. Carlitz, *A theorem on permutations in a finite field*, Proc. Amer. Math. Soc. **11** (1960), 456–459. MR MR0117223 (22 #8005)

[10] S. Dimiev, *A setting for a Diophantine distance geometry*, Tensor (N.S.) **66** (2005), no. 3, 275–283. MR MR2189847

[11] R. K. Guy, *Unsolved problems in number theory. 2nd ed.*, Unsolved Problems in Intuitive Mathematics. 1. New York, NY: Springer- Verlag. xvi, 285 p., 1994.

[12] H. Harborth, *Integral distances in point sets*, Butzer, P. L. (ed.) et al., Karl der Grosse und sein Nachwirken. 1200 Jahre Kultur und Wissenschaft in Europa. Band 2: Mathematisches Wissen. Turnhout: Brepols, 1998, 213–224.

[13] H. Harborth, A. Kemnitz, and M. Möller, *An upper bound for the minimum diameter of integral point sets*, Discrete Comput. Geom. **9** (1993), no. 4, 427–432.

[14] T. Khoon Lim and C. E. Praeger, *On generalised Paley graphs and their automorphism groups*, Michigan Math. J. (to appear).

[15] V. Klee and S. Wagon, *Old and new unsolved problems in plane geometry and number theory*, The Dolciani Mathematical Expositions. 11. Washington, DC: Mathematical Association of America. xv, 333 p., 1991.

[16] A. Kohnert and S. Kurz, *Integral point sets over $\mathbb{Z}_n^m$*, Electron. Notes Discrete Math. **27** (2006), 65–66.

[17] A. Kohnert and S. Kurz, *A note on Erdös-Diophantine graphs and Diophantine carpets*, Math. Balkanica (N.S.) **21** (2007), no. 1-2, 1–5.

[18] T. Kreisel and S. Kurz, *There are integral heptagons, no three points on a line, no four on a circle*, Discrete Comput. Geom., (to appear).

[19] S. Kurz, *Integral point sets over finite fields*, (submitted).

[20] S. Kurz, *On generating integer heronian triangles*, (submitted).

[21] S. Kurz, *Konstruktion und Eigenschaften ganzzahliger Punktmengen*, Ph.D. thesis, Bayreuth. Math. Schr. 76. Universität Bayreuth, 2006.

[22] S. Kurz and A. Wassermann, *On the minimum diameter of plane integral point sets*, Ars Combin. (to appear).

[23] B. D. McKay, *Practical graph isomorphism*, Congr. Numer. **30** (1981), 45–87.

[24] S. Niskanen and P. R. J. Östergård, *Cliquer user's guide, version 1.0*, Tech. Report T48, Communications Laboratory, Helsinki University of Technology, Espoo, Finland, 2003.

[25] T. K. Petersen, *Polynomials over finite fields whose values are squares*, Rose-Hulman Undergraduate Mathematics Journal **2** (2001), no. 1, 12 p.

[26] G. F. Royle, *An orderly algorithm and some applications in finite geometry*, Discrete Math. **185** (1998), no. 1-3, 105–115.

[27] J. Solymosi, *Note on integral distances*, Discrete Comput. Geom. **30** (2003), no. 2, 337–342.

[28] D. B. Surowski, *Automorphism groups of certain unstable graphs*, Math. Slovaca **53** (2003), no. 3, 215–232. MR MR2025019 (2004k:05110)

Chapter 10

Maximal integral point sets over $\mathbb{Z}^2$

ANDREY R. ANTONOV[1] AND SASCHA KURZ[2]

ABSTRACT. Geometrical objects with integral side lengths have fascinated mathematicians through the ages. We call a set $P = \{p_1, \ldots, p_n\} \subset \mathbb{Z}^2$ a maximal integral point set over $\mathbb{Z}^2$ if all pairwise distances are integral and every additional point p_{n+1} destroys this property. Here we consider such sets for a given cardinality and with minimum possible diameter. We determine some exact values via exhaustive search and give several constructions for arbitrary cardinalities. Since we cannot guarantee the maximality in these cases we describe an algorithm to prove or disprove the maximality of a given integral point set. We additionally consider restrictions as no three points on a line and no four points on a circle.

2000 MSC: 52C10; 52C45, 05D99, 11D99, 52-04.
Key words and phrases: integral distances, diameter, exhaustive search, maximality.

1 Introduction

Geometrical objects with integral side lengths have fascinated mathematicians through the ages. A very early example is the Pythagorean triangle with side lengths 3, 4, and 5. A universal framework for most of these objects are integral point sets. By an integral point set we understand a set of n points in an m dimensional Euclidean vector space $\mathbb{E}^m$, where the pairwise distances between the points are integral. Those integral point sets were studied by many authors, see [9] for an overview. From a combinatorial point of view for a given cardinality n and a given dimension m the question on the minimum possible diameter $d(n, m)$, this is the largest distance between any two points, arises, see [16, 19, 20] for an overview.

To obtain some interesting discrete structures one could also require some additional properties. One possibility is to request that besides the distances also the coordinates must be integral. Another classical possibility is to forbid subsets of three points on a line or four points on a circle. The question of P. Erdős whether there exists a set of seven points in the plane with no three points on a line, no four points on a circle, and pairwise integral distances, has recently been answered positively, see [14]. If all three mentioned additional properties are required simultaneously, one speaks of n_m-clusters, see [22]. In this article we request that besides the distances also the coordinates of the points are integral and restrict ourselves to dimension 2. Additionally we consider the cases where no three points are on a line or no four points are on a circle.

In finite geometry one is sometimes interested in point configurations which are maximal with respect to some property. This means that it is not possible to add a point without destroying the requested property. Here we consider integral point sets which are maximal, meaning that there does not exist an additional point x with integral distances to the other points of the point set.

1.1 Related work

There have been extensive studies on integral point sets in Euclidean spaces. Some authors also consider other spaces, e. g. Banach spaces [6], integral point sets over rings [13], or integral point sets over finite fields [2, 11, 15]. In [3] the authors consider integral point sets over $\mathbb{Z}^2$ and conjecture some examples to be maximal. As an answer to their open problems in [12], the authors describe an algorithm to prove the maximality of a given integral point set and prove the conjectures of [3].

1.2 Our contribution

In this paper we describe algorithms to efficiently test integral point sets for maximality and to determine possible extension points. To deal with the isomorphism problem we describe an algorithm which transforms a given plane integral point set into a normal form in $O(n^2)$ time, where n is the cardinality of the point set. We give several constructions of integral point sets over $\mathbb{Z}^2$ which have a given cardinality and fulfill additional conditions, such as that there are "no three points on a line" or "no four points on a circle". Although we cannot prove the maximality of the point sets obtained with the proposed constructions in general, we conjecture this property for many of our constructions. By

[1]Andrey Radoslavov Antonov, Department of Mathematics, University of Chemical Technology and Metallurgy - Sofia, Bulgaria.
E-mail adress: andrio@uctm.edu
[2]Sascha Kurz, Fakultät für Mathematik, Physik und Informatik, Universität Bayreuth, Germany.
E-mail adress: sascha.kurz@uni-bayreuth.de

exhaustive search we have determined some exact minimum diameters of integral point sets over $\mathbb{Z}^2$ with given cardinality and with or without additional conditions. We give constructive upper bounds in most cases and conjecture them to be the exact values.

1.3 Outline of the paper

In Section 3 we state the basic definitions and in Section 2 we describe the basic algorithms to deal with maximal integral point sets over $\mathbb{Z}^2$. These include an algorithm to exhaustively generate Heronian triangles up to isomorphism, an algorithm to determine all possible embeddings of an Heronian triangle on the integer grid $\mathbb{Z}^2$, and an algorithm that determines all points of $\mathbb{Z}^2$ which have integral distances to three given points in $\mathbb{Z}^2$ with pairwise integral distances. The last mentioned algorithm enables us to algorithmically prove or disprove the maximality of a given integral point set. Since we intend to consider integral point sets up to isomorphism, we introduce normal forms of integral point sets and algorithms to obtain them in Section 4. We deal with the key question of maximal integral point sets over $\mathbb{Z}^2$ with given cardinality and minimum diameter in Section 5. Several constructions for maximal integral point sets, where the maximality is not guaranteed but very likely, are described in Section 6. In Section 7 we deal with additional properties as "no three points on a line" and "no four points on a circle". We finish with a short conclusion and an outlook in Section 8.

2 Basics

Definition 2.1 *An integral point set over $\mathbb{Z}^2$ is a non-collinear set $\mathcal{P}$ of n points in the integer grid $\mathbb{Z}^2$, where the points have pairwise integral distances.*

For brevity we only speak of integral point sets and assume that the coordinates of the points are integral numbers, too.

Definition 2.2 *We call an integral point set $\mathcal{P}$ over $\mathbb{Z}^2$ maximal if for every $x \in \mathbb{Z}^2 \backslash \mathcal{P}$ the point set $\mathcal{P} \cup \{x\}$ is not an integral point set.*

The existence of maximal integral point sets in the plane is guaranteed by a famous theorem of N. H.. Anning and P. Erdős, respectively its proof.

Theorem 2.3 *An infinite set $\mathcal{P}$ of points in the Euclidean space $\mathbb{E}^m$ with pairwise integral distances is situated on a line. [1, 4]*

PROOF. We only prove the statement for dimension $m = 2$, as the generalization is obvious. If A, B, and C are three points not on a line, we set $k = \max\{\overline{AC}, \overline{BC}\}$ and consider points P such that $\left|\overline{PA} - \overline{PC}\right|$ and $\left|\overline{PB} - \overline{PC}\right|$ are integral. Due to the triangle inequalities the attained values are in $\{0, 1, \ldots, k\}$. Thus the point P lies on the intersection of two distinct hyperbolas, where we have at most $k+1$ choices for each hyperbola. Thus there are at most $4(k+1)^2$ possible locations for the point P. □

This proof can clearly be converted into a constructive algorithm. Given three points $A = (x_1, y_1)$, $B = (x_2, y_2)$, and $C = (x_3, y_3)$ in $\mathcal{P} \subset \mathbb{Z}^2$, which are not on a line, the problem of determining points $P = (x_4, y_4)$ at integral distance to A, B, and C is reduced to the problem of solving the equation system

$$\left| \begin{array}{rcl} \sqrt{\Delta x_{1,4}^2 + \Delta y_{1,4}^2} - \sqrt{\Delta x_{3,4}^2 + \Delta y_{3,4}^2} & = & d_1 \\ \sqrt{\Delta x_{2,4}^2 + \Delta y_{2,4}^2} - \sqrt{\Delta x_{3,4}^2 + \Delta y_{3,4}^2} & = & d_2 \end{array} \right|, \quad (1)$$

where $\Delta x_{i,j} := x_i - x_j$, $\Delta y_{i,j} := y_i - y_j$, $d_1 \in \{-\overline{AC}, \ldots, \overline{AC}\} \subset \mathbb{Z}$, and $d_2 \in \{-\overline{BC}, \ldots, \overline{BC}\} \subset \mathbb{Z}$. If there exists no integral solution in $\mathbb{Z}^2 \backslash \mathcal{P}$, then the point set $\mathcal{P}$ is maximal. This algorithm was already used in [12] to prove the maximality of the two integral point sets of Figure 1.

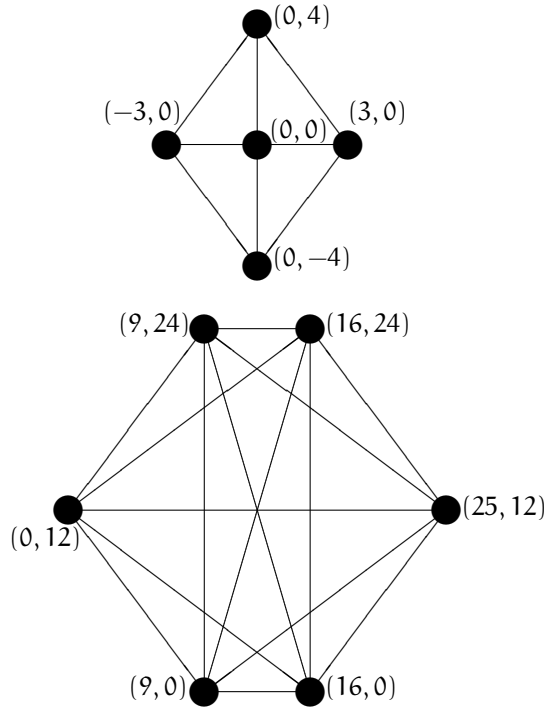


Figure 1: Examples of maximal integral point sets.

Since this algorithm is essential for our article we will go into the details how to solve Equation System (1). To get rid of some of the square roots we add $w := \sqrt{\Delta x_{3,4}^2 + \Delta y_{3,4}^2}$ on both sides and square the expressions afterwards:

$$\begin{array}{rcl} \Delta x_{1,4}^2 + \Delta y_{1,4}^2 & = & d_1^2 + 2d_1 \cdot w + \Delta x_{3,4}^2 + \Delta y_{3,4}^2 \\ \Delta x_{2,4}^2 + \Delta y_{2,4}^2 & = & d_2^2 + 2d_2 \cdot w + \Delta x_{3,4}^2 + \Delta y_{3,4}^2 \end{array}$$

Rearranging yields

$$\left| \begin{array}{l} x_1^2 + y_1^2 - x_3^2 - y_3^2 - d_1^2 + 2x_4 \Delta x_{3,1} + 2y_4 \Delta y_{3,1} = 2d_1 w \\ x_2^2 + y_2^2 - x_3^2 - y_3^3 - d_2^2 + 2x_4 \Delta x_{3,2} + 2y_4 \Delta y_{3,2} = 2d_2 w \end{array} \right|. \quad (2)$$

If $d_1 = 0$ then the first equation corresponds to a linear equation

$$c_1 x_4 + c_2 y_4 + c_3 = 0, \tag{3}$$

where not both c_1 and c_2 are equal to zero, since $A \neq C$. If we square the second equation of (2) we can substitute one variable using Equation (3) and obtain a quadratic equation in one variable, which can be easily solved. The case where $d_2 = 0$ is similar. Here we use the second equation of (2) to obtain Equation (3) (we have $c_1 \neq 0$ or $c_2 \neq 0$ due to $B \neq C$), and substitute it into the squared version of the first equation to obtain the quadratic equation in one variable. In the remaining case we have $d_1, d_2 \neq 0$. Here we subtract d_1 times the second equation of (2) from d_2 times the first equation of (2) to obtain Equation (3) (we have $c_1 \neq 0$ or $c_2 \neq 0$ since the points A, B, and C are not located on a line). Now we can square one of the two equations of (2) and subsitute one variables using Equation (3). Again we end up with a quadratic equation in one variable. At the end we have to check if the obtained values (x_4, y_4) are solutions of the original Equation System (1).

Definition 2.4 *For an integral point set $\mathcal{P}$ its diameter $diam(\mathcal{P})$ is given by the largest distance between a pair of its points.*

We remark that the upper integral point set of Figure 1 has diameter 8 and the lower integral point set of Figure 1 has diameter 25.

3 Exhaustive generation of maximal integral point sets

To obtain interesting examples of maximal integral point sets we utilize computers to exhaustively generate maximal integral point sets. In the following we will describe the algorithm used. For a given diameter d we loop over all non-isomorphic Heronian triangles (having integral side lengths and integral area) $\Delta = (a, b, c)$ with diameter $d = \max\{a, b, c\}$. Utilizing the Heron formula

$$A = \frac{\sqrt{(a+b+c)(a+b-c)(a-b+c)(-a+b+c)}}{4}$$

for the area of a triangle we can generate this list e. g. by the following short algorithm:

Algorithm 3.1 *(Generation of Heronian triangles)*
input: diameter d
output: complete list of Heronian triangles with diameter d *up to isomorphism*
begin
  $a = d$
  for $b = \left\lfloor \frac{a+2}{2} \right\rfloor, \ldots, a$ **do**
    for $c = a + 1 - b, \ldots, b$ **do**
      if $\frac{\sqrt{(a+b+c)(a+b-c)(a-b+c)(-a+b+c)}}{4} \in \mathbb{Z}$ **then**
        output (a, b, c)
end

For a more sophisticated and efficient algorithm we refer to [18]. The next step is to embed a given Heronian triangle $\Delta = (a, b, c)$ in the plane integer grid $\mathbb{Z}^2$. Here we can utilize two conjectures, which are theorems for dimension $m = 2$, see e. g. [5].

Conjecture 3.2 *Let $\mathcal{P} \subset \mathbb{Q}^m$ be a finite set of points such that the distances between any two points of $\mathcal{P}$ are integers. In this case one can find an Euclidean motion T such that $T(\mathcal{P}) \subset \mathcal{P}^m$.*

Conjecture 3.3 *Let $\mathcal{P} \subset \mathbb{Z}^m$ be a finite set of points such that the distances between any two points of $\mathcal{P}$ are integers and divisible by an integer k. In this case one can find a set $\mathcal{P}' \subset \mathbb{Z}^m$ such that $\mathcal{P}' \cdot k$ (the set $\mathcal{P}'$ scaled by a factor k) is congruent to $\mathcal{P}$.*

Since Conjecture 3.2 is a well known theorem for dimension $m = 2$, see e. g. [5], for every Heronian triangle $\Delta(a, b, c)$ there exists an embedding in the plane integer grid $\mathbb{Z}^2$. We remark that there may be several embeddings for the same triangle $\Delta = (a, b, c)$, which lead to different results. If we consider the number of points $(x_4, y_4) \in \mathbb{Z}^2 \backslash \mathcal{E}$ which are at integral distance to an embedded triangle $\mathcal{E} = \{(x_1, y_1), (x_2, y_2), (x_3, y_3)\}$, we can distinguish three different embeddings of the Heronian triangle $\Delta_1 = (25, 20, 15)$. The embedding $\mathcal{E}_1 = \{(0, 0), (0, 25), (12, 16)\}$ of Δ_1 yields 12 points (x_4, y_4) at integral distance to the corners of Δ_1 given by $\mathcal{E}_1$. For the embedding $\mathcal{E}_2 = \{(0, 0), (15, 20), (0, 20)\}$ we obtain 16 such points, and for the embedding $\mathcal{E}_3 = \{(0, 0), (7, 24), (16, 12)\}$ we obtain only 5 such points. Determining the possible embeddings of a given Heronian triangle $\Delta = (a, b, c)$ is a rather easy task. W.l.o.g. we assume $a = \max\{a, b, c\}$ and $x_2 = 0 = y_2$. Since the point (x_3, y_3) is at distance a to the point (x_2, y_2), we have to solve the Diophantine equation

$$x_3^2 + y_3^2 = a^2$$

in integers. This is a well known problem. One might even store for each small number (e. g. $a \leqslant 10\,000$) $a \in \mathbb{N}$ a list of the corresponding solutions (x_3, y_3). Now the coordinates of the remaining point A are given as solutions of the equation system

$$\left| \begin{array}{rcl} (x_2 - x_1)^2 + (y_2 - y_1)^2 & = & c^2 \\ (x_3 - x_1)^2 + (y_2 - y_1)^2 & = & b^2 \end{array} \right|, \tag{4}$$

which can be easily solved. As an algorithm for the embedding of an Heronian triangle in $\mathbb{Z}^2$ we obtain:

Algorithm 3.4 *(Embedding of an Heronian Triangle)*
input: Heronian Triangle $\Delta = (a, b, c)$
output: complete list of different embeddings of Δ *in* $\mathbb{Z}^2$
begin
  $x_2 = 0, y_2 = 0$
  loop over *the integer solutions* (x_3, y_3) *of* $x_3^2 + y_3^2 = a^2$ **do**
    loop over *the integer solutions* (x_1, y_1) *of Equation System (4)* **do**
      output $\left\{ (x_1, y_1), (x_2, y_2), (x_3, y_3) \right\}$
end

The next step is to determine the points $(x_4, y_4) \in \mathbb{Z}^2$ which are at integral distance to a given embedded triangle $\left\{ (x_1, y_1), (x_2, y_2), (x_3, y_3) \right\}$:

Algorithm 3.5 *(Enlargement of an embedded triangle)*
input: Embedded triangle
$\mathcal{E} = \left\{ (x_1, y_1), (x_2, y_2), (x_3, y_3) \right\} \subset \mathbb{Z}^2$
output: complete list of points $(x_4, y_4) \in \mathbb{Z}^2 \backslash \mathcal{E}$ *which are at integral distance to* $\mathcal{E}$
begin
  loop over *the integer solutions* (x_4, y_4) *of Equation System (1)* **do**
    if $(x_4, y_4) \notin \mathcal{E}$ **then**
      output (x_4, y_4)
end

We remark that the previous algorithms have to be implemented using an arithmetic which is able to do integer calculations with unlimited precision, since the occurring numbers can increase very quickly. We have utilized the software package CLN [8] for this purpose.

Now we utilize the set of points given by Algorithm 3.5 to build up a graph $\mathcal{G}(\mathcal{E})$. The vertices are given by the possible points (x_4, y_4). Two points (x_4, y_4) and (x_4', y_4') are connected by an edge if and only if $\sqrt{(x_4 - x_4')^2 + (y_4 - y_4')^2}$ is an integer. A complete subgraph of $\mathcal{G}(\mathcal{E})$ is called a clique. A clique $\mathcal{C}_1$ is called maximal if it is not properly contained in another clique $\mathcal{C}_2$ of $\mathcal{G}(\mathcal{E})$. Clearly the cliques of $\mathcal{G}(\mathcal{E})$ are in bijection to integral point sets $\mathcal{P} \subset \mathbb{Z}^2$ containing $\mathcal{E}$ as a subset. The same statement holds for maximal cliques of $\mathcal{G}(\mathcal{E})$ and maximal integral point sets $\mathcal{P} \subset \mathbb{Z}^2$ containing $\mathcal{E}$ as a subset. Thus we can use a clique-search package as CLIQUER [21] to exhaustively generate maximal integral point sets $\mathcal{M}$ over $\mathbb{Z}^2$.

Let us consider an example. If we apply our algorithm on the embedded triangle

$$\mathcal{E}_2 = \left\{ (0,0), (15,20), (0,20) \right\}$$

with diameter 25, we obtain a set

$$\begin{aligned} \Big\{ &(0,28), (0,40), (0,56), (0,132), (0,-92), (0,-16), \\ &(0,12), (-15,20), (15,0), (-21,20), (105,-36), \\ &(21,20), (-48,20), (48,20), (-99,20) \Big\} \end{aligned}$$

of 16 possible points to enlarge the integral point set $\mathcal{E}_2$. The clique-search program CLIQUER determines five maximal cliques which correspond to the following five maximal integral point sets:

$$\begin{aligned} \mathcal{M}_1 &= \left\{ (0,0), (15,20), (0,20), (15,0) \right\}, \\ \mathcal{M}_2 &= \left\{ (0,0), (15,20), (0,20), (0,-92), (105,-36) \right\}, \\ \mathcal{M}_3 &= \Big\{ (0,0), (15,20), (0,20), (0,40), (0,56), \\ &\quad (0,-16), (-15,20), (-48,20), (48,20) \Big\}, \\ \mathcal{M}_4 &= \Big\{ (0,0), (15,20), (0,20), (0,40), (-15,20), \\ &\quad (-21,20), (21,20), (-48,20), (48,20), \\ &\quad (-99,20), (99,20) \Big\}, \text{ and} \\ \mathcal{M}_5 &= \Big\{ (0,0), (15,20), (0,20), (0,28), (0,40), (0,56), \\ &\quad (0,132), (0,-92), (0,-16), (0,12), (-15,20) \Big\}. \end{aligned}$$

It is interesting to have a look at the cardinalities and diameters of these maximal integral point sets. We have $|\mathcal{M}_1| = 4$, $\operatorname{diam}(\mathcal{M}_1) = 25$, $|\mathcal{M}_2| = 5$, $\operatorname{diam}(\mathcal{M}_2) = 119$, $|\mathcal{M}_3| = 9$, $\operatorname{diam}(\mathcal{M}_3) = 96$, $|\mathcal{M}_4| = 11$, $\operatorname{diam}(\mathcal{M}_1) = 198$, $|\mathcal{M}_5| = 11$, and $\operatorname{diam}(\mathcal{M}_5) = 224$. Although we start with a point set $\mathcal{E}_2$ of small diameter, the resulting maximal integral point sets $\mathcal{M}_i$ may have a large diameter. We are not aware of a formula to bound $\operatorname{diam}(\mathcal{M})$ with respect to $\operatorname{diam}(\mathcal{E})$. A second somewhat disappointing fact of our algorithm is, that each subset $\mathcal{E}'$ of three non-collinear points of an maximal integral point set $\mathcal{M}$ produces $\mathcal{M}$. Thus our algorithm produces many identical copies of maximal integral point sets with large cardinality. We will deal with this fact and the isomorphism problem in the next section.

The algorithms described in this section focus on the maximality of the integral point sets. They should not be used to exhaustively generate all maximal integral point sets up to a given diameter. To perform this task the algorithms to exhaustively generate integral point sets with or without additional properties are better suited, see [16, 20], and ignore the maximality condition in the first run. All integral point sets with required cardinalities and small diameters can then be tested if they are maximal.

4 Normal forms and automorphisms for integral point sets over $\mathbb{Z}^2$

In this section we aim to consider isomorphisms which preserve certain properties of maximal integral point sets. Since a main property of an integral point set is the set of distances between its points, we only consider distance-preserving isomorphisms, so called isometries. In the Euclidean plane the isometries are given by compositions of translations $T_{u,v} : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} u \\ v \end{pmatrix}$, rotations $R_\theta : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix}$, and reflections at one of the two axes. Each isometry can be written as $I_{t,O} : x \mapsto t + O \cdot x$, where $t \in \mathbb{R}^2$ is a translation vector and $O \in \mathbb{R}^{2\times 2}$ an orthogonal matrix. Next we restrict ourselves to mappings which map integral coordinates onto integral coordinates. Thus we have $t \in \mathbb{Z}^2$ and $O \in \mathbb{Z}^{2\times 2}$. Each such isometry $I_{t,O}$ maps integral point sets onto integral point sets. It is easy to figure out that there are only 8 orthogonal matrices

in $\mathbb{Z}^{2\times 2}$. So we define

$$\mathrm{Aut} := \left\{ I_{t,O} : t \in \mathbb{Z}^2,\ O \in \left\{ \begin{pmatrix} \pm 1 & 0 \\ 0 & \pm 1 \end{pmatrix}, \begin{pmatrix} \pm 1 & 0 \\ 0 & \mp 1 \end{pmatrix}, \begin{pmatrix} 0 & \pm 1 \\ \pm 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & \pm 1 \\ \mp 1 & 0 \end{pmatrix} \right\} \right\}$$

as the automorphism group of plane integral point sets.

We call two integral point sets $\mathcal{P}$ and $\mathcal{P}'$ isomorphic, if there exists a mapping $I_{t,O} \in \mathrm{Aut}$ such that $I_{t,O}(\mathcal{P}) = \mathcal{P}'$. So our aim is to develop an algorithm which can check whether two given integral point sets are isomorphic. For this purpose we want to use the technique of normal forms of discrete objects. This means that we have a function τ which fulfills the following: If $\mathcal{O}$ is the orbit of an integral point set $\mathcal{P}$ under the group Aut then $\tau(\mathcal{P}) = \tau(\mathcal{P}')$ for each $\mathcal{P}' \in \mathcal{O}$. Additionally for two integral point sets of different orbits the function τ should have different images. Having such a function τ at hand we can easily decide whether two integral point sets $\mathcal{P}$ and $\mathcal{P}'$ are isomorphic, by checking whether $\tau(\mathcal{P}) = \tau(\mathcal{P}')$ or not.

In order to describe such a function τ we need to define a total ordering $\preceq$ on $\mathbb{Z}^2$:

(1) if $|a| < |c|$, then we set $\begin{pmatrix} a \\ b \end{pmatrix} \prec \begin{pmatrix} c \\ d \end{pmatrix}$,

(2) if $a > 0$, then we set $\begin{pmatrix} -a \\ b \end{pmatrix} \prec \begin{pmatrix} a \\ d \end{pmatrix}$,

(3) if $|b| < |d|$, then we set $\begin{pmatrix} a \\ b \end{pmatrix} \prec \begin{pmatrix} a \\ d \end{pmatrix}$, and

(4) if $b > 0$, then we set $\begin{pmatrix} a \\ -b \end{pmatrix} \prec \begin{pmatrix} a \\ b \end{pmatrix}$

for all $a, b, c, d \in \mathbb{Z}$. We set $\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} c \\ d \end{pmatrix}$ if and only if we have $a = c$ and $b = d$. By $x_1 \preceq x_2$ we mean $x_1 \prec x_2$ or $x_1 = x_2$. One of the properties of this total ordering $\preceq$ is, that we have $\begin{pmatrix} 0 \\ 0 \end{pmatrix} \preceq x$ for all $x \in \mathbb{Z}^2$, so $\begin{pmatrix} 0 \\ 0 \end{pmatrix} \preceq x$ is the smallest element in $\mathbb{Z}^2$. Using $\prec$ we can bijectively identify an integral point set $\mathcal{P}$ with a list $\mathcal{L}(\mathcal{P})$ of its points, which is sorted in ascending order with respect to $\preceq$. Now we extend our total ordering $\preceq$ onto such lists by utilizing the lexicographic ordering. This allows us to define our normalization function by

$$\tau(\mathcal{P}) = \min_{\preceq} \left\{ \mathcal{L}(\sigma(\mathcal{P})) : \sigma \in \mathrm{Aut} \right\}.$$

To obtain a finite algorithm for the determination of $\tau(\mathcal{P})$ we use the fact, that for every point set $\mathcal{P} \neq \emptyset$ the *minimum* list-representation $\mathcal{L}(\sigma(\mathcal{P}))$ starts with $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$:

Algorithm 4.1 *(Normalization of an integral point set)*
input: integral point set $\mathcal{P} = \{p_1, \ldots, p_n\}$
output: minimum list representation $\tau(\mathcal{P})$
begin
$\mathtt{champion} = \mathcal{L}(\mathcal{P})$
$M_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $M_2 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$,
$M_3 = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$, $M_4 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$,
$M_5 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $M_6 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$,
$M_7 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, $M_8 = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}$
for $i = 1, \ldots, n$ **do**
  for $j = 1, \ldots, 8$ **do**
    $\mathtt{tmp} = \mathcal{L}\left(M_j \cdot \{p_1 - p_i, \ldots, p_n - p_i\}\right)$
    if $\mathtt{tmp} \prec \mathtt{champion}$ **then**
      $\mathtt{champion} = \mathtt{tmp}$
return champion
end

We remark that Algorithm 4.1 runs in $O(n^2)$ time. As an example we consider the two integral point sets from Figure 1. Their normal forms or minimum list representations are given by

$$\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ -3 \end{pmatrix}, \begin{pmatrix} 0 \\ 3 \end{pmatrix}, \begin{pmatrix} -4 \\ 0 \end{pmatrix}, \begin{pmatrix} 4 \\ 0 \end{pmatrix} \right]$$

and

$$\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ -7 \end{pmatrix}, \begin{pmatrix} -12 \\ 9 \end{pmatrix}, \begin{pmatrix} -12 \\ -16 \end{pmatrix}, \begin{pmatrix} -24 \\ 0 \end{pmatrix}, \begin{pmatrix} -24 \\ -7 \end{pmatrix} \right],$$

respectively.

For a given integral point set $\mathcal{P}$ there may exist rotation matrices $M \in \mathbb{R}^{2\times 2}$, such that $M(\mathcal{P})$ has integral coordinates, which are different from the eight orthogonal matrices in $\mathbb{Z}^{2\times 2}$. But for these matrices there is no guarantee for a proper extension $\mathcal{E} \supset \mathcal{P}$, which is also an integral point set over $\mathbb{Z}^2$, such that $M(\mathcal{E})$ has integral coordinates. Examples are given by the sets $\mathcal{E}_1$, $\mathcal{E}_2$, $\mathcal{E}_3$ in Section 3. This means that for a given maximal integral point set $\mathcal{M}$ over $\mathbb{Z}^2$ there can exist an orthogonal matrix $M \in \mathbb{R}^{2\times 2}$, such that $M(\mathcal{M})$ is also an integral point set over $\mathbb{Z}^2$, but which is not maximal.

We may call a maximal integral point set $\mathcal{M}$ over $\mathbb{Z}^2$ strongly maximal, if such a matrix M does not exist. To check whether a given integral point set $\mathcal{P}$ is strongly maximal, we only have to consider all possible embeddings of $\mathcal{P}$ in $\mathbb{Z}^2$, which are finitely many. Another possibility is to slightly alter Algorithm 3.5 by looping over the rational (instead of integral) solutions (x_4, y_4) of Equation System (1). Now the algorithm leads to point sets with integral distances and rational coordinates. But due to Conjecture 3.2 (which is a theorem for dimension $m = 2$), there exist embeddings with integral coordinates.

To clear the situation with integral and rational coordinates we will have to give some facts from the general theory of integral point sets (without integral coordinates). So, let $\mathcal{P}$ be a set of points in the m-dimensional Euclidean space $\mathbb{E}^m$ with pairwise integral distances. By $\mathcal{S} \subseteq \mathcal{P}$ we denote an integral simplex, which is a set of $m+1$ points, and by $\mathrm{vol}_m(\mathcal{S})$ we denote the m-dimensional volume spanned by

the $m+1$ points. Since the pairwise distances are integral we can write $\mathrm{vol}_m(\mathcal{S}) = q \cdot k$ with $q \in \mathbb{Q}$ and k being a square free integer. If $\mathrm{vol}_m(\mathcal{S}) \neq 0$ the square free integer k is unique and we set $\mathrm{char}(\mathcal{S}) = k$, which we call the characteristic of $\mathcal{S}$. Using this notation we can cite two results from [17]:

Theorem 4.2 *In an m-dimensional integral point set $\mathcal{P}$ all simplices $\mathcal{S} = \{v_0, v_1, \ldots, v_m\}$ with $vol_m(\mathcal{S}) \neq 0$ have the same characteristic $\mathrm{char}(\mathcal{S}) = k$.*

So we can speak of *the* characteristic $\mathrm{char}(\mathcal{P})$ of an integral point set $\mathcal{P}$.

Lemma 4.3 *An integral m-dimensional simplex $\mathcal{S} = \{v'_0, v'_1, \ldots, v'_m\}$ with distance matrix $D = (d_{i,j}) \in \mathbb{N}$ for $0 \leqslant i, j \leqslant m$ and $vol_m(\mathcal{S}) \neq 0$ can be transformed via an isometry into the coordinates*

$$\begin{aligned} v_0 &= (0,0,\ldots,0), \\ v_1 &= \left(q_{1,1}\sqrt{k_1}, 0, 0\ldots, 0\right), \\ v_2 &= \left(q_{2,1}\sqrt{k_1}, q_{2,2}\sqrt{k_2}, 0, \ldots, 0\right), \\ &\vdots \\ v_m &= \left(q_{m,1}\sqrt{k_1}, q_{m,2}\sqrt{k_2}, \ldots, q_{m,m}\sqrt{k_m}\right), \end{aligned}$$

where k_i is the squarefree part of $\frac{vol_i(v'_0,v'_1,\ldots,v'_i)^2}{vol_{i-1}(v'_0,v'_1,\ldots,v'_{i-1})^2}$, $q_{i,j} \in \mathbb{Q}$, and $q_{j,j}, k_j \neq 0$.

We remark that we always have $k_1 = 1$. The connection between the k_i and the characteristic $\mathrm{char}(\mathcal{P}) = k$ is given by

$$\mathrm{char}(\mathcal{P}) = \mathrm{char}(\mathcal{S}) = k = \text{square free part of } \prod_{i=1}^{m} k_i.$$

Thus plane integral point sets $\mathcal{P}$ with rational coordinates are exactly those with characteristic $\mathrm{char}(\mathcal{P}) = 1$. Due to Conjecture 3.2 plane integral point sets over $\mathbb{Z}^2$ correspond to plane integral point sets with characteristic 1. So in principle there is no need to care about the coordinates – this can still be done afterwards.

There is one further transformation that maps integral point sets over $\mathbb{Z}^2$ onto integral point sets over $\mathbb{Z}^2$: scaling by an integral factor λ. One handicap of this mapping is that the inverse mapping may lead to non-integral point sets. Another shortcoming is that maximal integral point sets may be mapped onto non-maximal integral point sets. An example is given by the maximal integral point set $\mathcal{P} = \left\{\binom{0}{0}, \binom{3}{0}, \binom{0}{4}, \binom{3}{4}\right\}$. If we scale it by a factor of 2 we obtain $2 \cdot \mathcal{P} = \left\{\binom{0}{0}, \binom{6}{0}, \binom{0}{8}, \binom{6}{8}\right\}$ an integral point set over $\mathbb{Z}^2$ which can be extended by the point $\binom{3}{4}$. In contrast to this example the integral point set $3 \cdot \mathcal{P} = \left\{\binom{0}{0}, \binom{9}{0}, \binom{0}{12}, \binom{9}{12}\right\}$ is maximal. One might conjecture that for every maximal integral point set $\mathcal{M}$ there exists an integer $\lambda > 1$ such that $\lambda \cdot \mathcal{M}$ is also maximal.

5 Maximal integral point sets with given cardinality and minimum diameter

From the combinatorial point of view a natural question is to ask for the minimum possible diameter $d_M(k, m)$ of a maximal integral point set $\mathcal{M} \subset \mathbb{Z}^m$ of cardinality k. If such a point set does not exist we set $d_M(k, m) = \infty$. Utilizing the exhaustive algorithm described in Section 3 we have obtained the results given in Table 1.

k	$d_M(k,2)$	corresponding point set
4	5	$\{(0,0),(3,4),(0,4),(3,0)\}$
5	8	$\{(0,0),(3,4),(0,4),(0,8),(-3,4)\}$
6	25	$\{(0,0),(12,16),(12,9),(-12,9),(-12,16),(0,25)\}$
7	30	$\{(0,0),(6,8),(0,8),(0,16),(-6,8),(-15,8),(15,8)\}$
8	65	$\{(0,0),(15,36),(0,16),(15,-20),(48,-20),(48,36),(63,0),(63,16)\}$
9	96	$\{(0,0),(15,20),(0,20),(0,40),(0,56),(0,-16),(-15,20),(-48,20),(48,20)\}$
10	$\leqslant 600$	$\{(0,0),(22,120),(0,120),(-27,120),(160,120),(182,0),(182,120),(-209,120),(209,120),(391,120)\}$
11	70	$\{(0,0),(5,12),(0,12),(0,24),(-5,12),(-9,12),(9,12),(-16,12),(16,12),(-35,12),(35,12)\}$

Table 1: Minimum possible diameters of maximal plane integral point sets with given cardinality.

Clearly we have $d_M(1,2) = d_M(2,2) = \infty$ since a line l through two different points P_1 and P_2 with integral coordinates and integral distance $\overline{P_1P_2}$ contains an infinite integral point set $\mathcal{P} = \{P_1 + \lambda \cdot (P_2 - P_1) : \lambda \in \mathbb{Z}\}$ as a subset. So the next value to determine is $d_M(3,2)$. Whether $d_M(3,2)$ is finite had been an open question of [3], which was answered in [12] by determining $d_M(3,2) = 2066$, – a diameter out of reach for our general exhaustive algorithm described in Section 3. But it can be easily adapted for this purpose. We alter Algorithm 3.1 by omitting right-angled triangles, since these obviously are not maximal. Then we skip Algorithm 3.4 and directly run the version of Algorithm 3.5 where we search for rational instead of integral solutions (x_4, y_4) of Equation System (1). If we have found the first solution (x_4, y_4) for a given triangle Δ we can immediately stop our investigations on Δ since it cannot be a maximal integral triangle. Using these reductions and skipping the time consuming clique search we were able to exhaustively search for (strongly) maximal integral triangles over $\mathbb{Z}^2$ with

diameter at most 15000 [12, 18]. There are exactly 126 such examples. Here we list the first, with respect to their diameter, ten examples, where we give the edge lengths and the coordinates in minimal list representation, which is unique in these cases:

$$\{2066,1803,505\}\left[(0,0)^T,(-336,-377)^T,(384,-2030)^T\right]$$
$$\{2549,2307,1492\}\left[(0,0)^T,(-700,-2451)^T,(1100,-1008)^T\right]$$
$$\{3796,2787,2165\}\left[(0,0)^T,(-387,-2760)^T,(1680,-3404)^T\right]$$
$$\{4083,2425,1706\}\left[(0,0)^T,(-410,-1656)^T,(1273,2064)^T\right]$$
$$\{4426,2807,1745\}\left[(0,0)^T,(-280,-2793)^T,(376,-4410)^T\right]$$
$$\{4801,2593,2210\}\left[(0,0)^T,(-1488,-1634)^T,(1632,2015)^T\right]$$
$$\{4920,4177,985\}\left[(0,0)^T,(-473,-864)^T,(4015,1152)^T\right]$$
$$\{5044,4443,2045\}\left[(0,0)^T,(-1204,-1653)^T,(2156,-4560)^T\right]$$
$$\{5045,4803,244\}\left[(0,0)^T,(-44,-240)^T,(240,4797)^T\right]$$
$$\{5186,5163,745\}\left[(0,0)^T,(-407,-624)^T,(4030,-3264)^T\right]$$

6 Constructions for maximal integral point sets over $\mathbb{Z}^2$

In this section we want to describe constructions for maximal integral point sets $\mathcal{M}$ of a given cardinality or a given shape. In most cases our constructions do not lead to integral point sets which are maximal in every case, but which yield candidates, which are very likely to be maximal (from an empiric point of view). W.l.o.g. we can assume that the origin $(0,0)^T$ is always contained in $\mathcal{M}$. Every further point $(a,b)^T$ meets $a^2+b^2=c^2$. In this case we call (a,b) a Pythagorean pair or (a,b,c) a Pythagorean triple. If additionally $\gcd(a,b)=\gcd(a,b,c)=1$ we speak of primitive pairs or triples. Given only one Pythagorean pair (a,b) we can perform the following two constructions for integral point sets over $\mathbb{Z}^2$:

Construction 6.1 *If (a,b) is a Pythagorean pair, then $\mathcal{P}_1(a,b) := \left\{\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}a\\0\end{pmatrix},\begin{pmatrix}0\\b\end{pmatrix},\begin{pmatrix}a\\b\end{pmatrix}\right\}$ is an integral point set of cardinality 4.*

Construction 6.2 *If (a,b) is a Pythagorean pair, then $\mathcal{P}_2(a,b) := \left\{\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}a\\0\end{pmatrix},\begin{pmatrix}-a\\0\end{pmatrix},\begin{pmatrix}0\\b\end{pmatrix},\begin{pmatrix}0\\-b\end{pmatrix}\right\}$ is an integral point set of cardinality 5.*

We call Construction 6.1 the *rectangle construction* of (a,b) and Construction 6.2 the *rhombus construction* of (a,b). If we choose (a,b) with $2|a$, $2|b$ then clearly $\mathcal{P}_1(a,b)$ cannot be maximal. On the other side $\mathcal{P}_1(9,12)$ is a maximal integral point set although $\gcd(9,12)=3$. Empirically, we have observed that for primitive pairs (a,b) the point set $\mathcal{P}_1(a,b)$ is maximal in many, but not all cases, see e. g. the non maximal integral point set $\mathcal{P}_1(7,24)$, which can be extended to the maximal integral point set $\left\{\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}7\\0\end{pmatrix},\begin{pmatrix}0\\24\end{pmatrix},\begin{pmatrix}7\\24\end{pmatrix},\begin{pmatrix}-9\\12\end{pmatrix},\begin{pmatrix}16\\12\end{pmatrix}\right\}$. For $(a,b)=(3,4)$ both constructions $\mathcal{P}_1(a,b)$ and $\mathcal{P}_2(a,b)$ yield maximal integral point sets. Empirically Construction 6.2 is a bit weaker, since it often happens that $\mathcal{P}_1(a,b)$ is maximal but $\mathcal{P}_2(a,b)$ is not, as for example for $(a,b)=(5,12)$. For the other direction we have no example. We would like to mention that $\mathcal{P}_2(5,12)$ can be extended to the very interesting maximal integral point set $\mathcal{M}=\left\{\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}5\\0\end{pmatrix},\begin{pmatrix}0\\12\end{pmatrix},\begin{pmatrix}0\\-12\end{pmatrix},\begin{pmatrix}-5\\0\end{pmatrix},\begin{pmatrix}9\\0\end{pmatrix},\begin{pmatrix}-9\\0\end{pmatrix},\begin{pmatrix}16\\0\end{pmatrix},\begin{pmatrix}-16\\0\end{pmatrix},\begin{pmatrix}35\\0\end{pmatrix},\begin{pmatrix}-35\\0\end{pmatrix}\right\}$, which has an intriguing geometrical structure, see Figure 2.

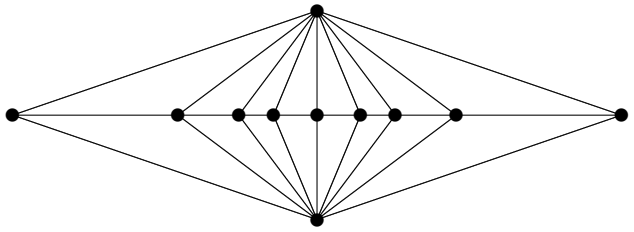


Figure 2: Extension of $\mathcal{P}_2(5,12)$ to a crab of cardinality 11.

Definition 6.3 *For positive integers $a,b_1,\ldots,b_k$ we call the point set*

$$\mathit{crab}\,(a,b_1,\ldots,b_k) := \left\{\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}0\\\pm a\end{pmatrix},\begin{pmatrix}\pm b_1\\0\end{pmatrix},\ldots,\begin{pmatrix}\pm b_k\\0\end{pmatrix}\right\}$$

a crab of order k.

We remark that the cardinality of $\mathrm{crab}\,(a,b_1,\ldots,b_k)$ is given by $2k+3$ and that the point set is symmetric w.r.t. the two coordinate axes. This point set is indeed integral if the pairs $(a,b_1),\ldots,(a,b_k)$ are Pythagorean pairs. So it is very easy to construct crabs, either directly or by extending $\mathcal{P}_2(a,b)$, see Subsection 6.1. Empirically the extension points of $\mathcal{P}_2(a,b)$ very often lie on one of the two axes. An example that this must not be the case in general is given by the primitive pair $(1480,969)$, where $\mathcal{P}_2(1480,969)$ can be extended to $\left\{\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}1480\\0\end{pmatrix},\begin{pmatrix}-1480\\0\end{pmatrix},\begin{pmatrix}0\\969\end{pmatrix},\begin{pmatrix}0\\-969\end{pmatrix},\begin{pmatrix}1040\\462\end{pmatrix},\begin{pmatrix}1040\\-462\end{pmatrix},\begin{pmatrix}-1040\\462\end{pmatrix},\begin{pmatrix}-1040\\-462\end{pmatrix}\right\}$.

6.1 Construction of crabs

Since many maximal integral point sets over $\mathbb{Z}^2$ are crabs we are interested in a method to construct them directly. From the general theory of integral point sets we know that integral point sets $\mathcal{P}$ over $\mathbb{R}^2$ with minimum diameter consist

of point sets with $n-1$ collinear points, see Figure 3, for $9 \leqslant n \leqslant 122$ points, see [16, 20]. For these point sets there is an interesting connection between the points of the point set $\mathcal{P}$ and divisors of a certain number D, see [16, 20].

Definition 6.4 *The decomposition number* D *of an integral triangle with side lengths* a, b, *and* c *is given by*

$$D = \frac{(a+b+c)(a+b-c)(a-b+c)(-a+b+c)}{\gcd\left(b^2-c^2+a^2, 2a\right)^2}.$$

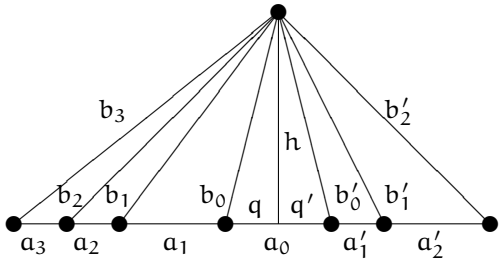


Figure 3: Plane integral point set $\mathcal{P}$ with $n-1$ points on a line.

Lemma 6.5 (Decomposition lemma)
The distances of a plane integral point set $\mathcal{P}$ *consisting of* n *points where a subset of* $n-1$ *points is collinear correspond to decompositions of the decomposition number* D *of the largest triangle of* $\mathcal{P}$ *into two factors.*

PROOF. We use the notation of Figure 3 and set

$$\begin{aligned} c_i &= q + \sum_{j=1}^{i} a_j \quad \text{for} \quad 0 \leqslant i \leqslant s, \\ c_i' &= q' + \sum_{j=1}^{i} a_j' \quad \text{for} \quad 0 \leqslant i \leqslant t. \end{aligned}$$

Pythagoras' Theorem yields $c_{i+1}^2 + h^2 = b_{i+1}^2$ and $c_i^2 + h^2 = b_i^2$ for $0 \leqslant i < s$. We subtract these equations from each other and get

$$a_{i+1}^2 + 2a_{i+1}\sum_{j=1}^{i} a_j + 2a_{i+1}q = b_{i+1}^2 - b_i^2.$$

Because the a_i and the b_i are positive integers we have $2a_{i+1}q \in \mathbb{N}$ for $0 \leqslant i < s$ and therefore $2\gcd(a_1, a_2, \ldots, a_s)\, q \in \mathbb{N}$. From $q + q' = a_0 \in \mathbb{N}$ we conclude $2\gcd(a_1, a_2, \ldots, a_s)\, q' \in \mathbb{N}$. With an analogous conclusion for the c_i' and $g = 2\gcd(a_1, \ldots, a_s, a_1', \ldots, a_t')$ we get $gq \in \mathbb{N}$ and $gq' \in \mathbb{N}$. A last use of Pythagoras' Theorem yields for $1 \leqslant i \leqslant s$ and for $1 \leqslant j \leqslant t$ the factorization of g^2h^2 into a product of two positive integers,

$$\begin{aligned} g^2h^2 &= (gb_i + gc_i)(gb_i - gc_i) \\ &= \left(gb_j' + gc_j'\right)\left(gb_j' - gc_j'\right). \end{aligned}$$

So we can obtain the possible values for c_i and c_i' by decomposing g^2h^2 into two factors.

If we are given the three side lengths a, b, and c of an integral triangle and want to determine the points on the side of length a so that the resulting point set is integral, then we can associate b with b_s, c with b_t', and a with $\sum_{i=1}^{s} a_i + a_0 + \sum_{i=1}^{t} a_i'$. With this we have

$$c_s = q + \sum_{j=1}^{s} a_j = \frac{b^2 - c^2 + a^2}{2a}.$$

Because g can also be defined as the smallest integer with $gc_s \in \mathbb{N}$ we receive

$$g = \frac{2a}{\gcd\left(b^2-c^2+a^2, 2a\right)}.$$

Due to the Heron formula $16A_\Delta^2 = (a+b+c)(a+b-c)(a-b+c)(-a+b+c)$ and the formula for the area of a triangle $2A_\Delta = ah$ we finally get

$$\begin{aligned} g^2h^2 &= \frac{g^2(a+b+c)(a+b-c)(a-b+c)(-a+b+c)}{4a^2} = \\ &= \frac{(a+b+c)(a+b-c)(a-b+c)(-a+b+c)}{\gcd\left(b^2-c^2+a^2, 2a\right)^2} = D. \end{aligned}$$

□

If we choose $g = 1$ and $h \in \mathbb{N}$ we can directly apply Lemma 6.5 to construct crabs. Let us look at an example. We choose $g = 1$ and $h = 2 \cdot 3 \cdot 5 = 30$. The divisors of $D = g^2h^2 = 900$ are given by $\{1, 2, 3, 4, 5, 6, 9, 10, 12, 15, 18, 20, 25, 30, 36, 45, 50, 60, 75, 90, 100, 150, 180, 225, 300, 450, 900\}$. If we have $D = f_1 \cdot f_2$, then $b = \frac{f_1+f_2}{2}$ and $c = \frac{f_1-f_2}{2}$. Thus we must have $f_1 > f_2$ and $f_1 \equiv f_2 \pmod 2$ to determine the values b_i of a corresponding crab. Here we have $b_1 = \frac{50-18}{2} = 16$, $b_2 = \frac{90-10}{2} = 40$, $b_3 = \frac{150-6}{2} = 72$, $b_4 = \frac{450-2}{2} = 224$, and $a = h = 30$. This yields the integral point set crab$(30, 16, 40, 72, 224)$ of cardinality $2 \cdot 4 + 3 = 11$ and diameter $\max\left\{2b_i, 2a, \sqrt{b_i^2 + a^2}\right\} = 2 \cdot \max\{b_i, a\} = 448$. Given the prime factorization $h = \prod_{i=1}^{r} p_i^{\alpha_i}$ it is not difficult to determine the k-value of the resulting crab. Let us fix $p_1 = 2$ and set $\tilde{\alpha}_1 = \max(\alpha_1 - 1, 0)$. With this we can state

$$k = \frac{(2\tilde{\alpha}_1 + 1) \cdot \prod_{i=2}^{r} (2\alpha_i + 1) - 1}{2}. \tag{5}$$

Using $h = p^k$, where p is an arbitrary odd prime, we are able to produce a crab of order k for each $k \geqslant 1$. Thus we have constructions for integral point sets of cardinality $2k+3$ for each $k \in \mathbb{N}$. To obtain small point sets with many points we should clearly choose integers with many divisors for h instead. As for all of our constructions the maximality of the resulting integral point set is not guaranteed, but very likely.

Construction 6.6 *For a given integer* h *there exists an integral point set decompose*(h) *which is a crab of order* k, *where* k *is given by Equation (5).*

If $h > 4$ then the diameter of decompose(h) is given by $h^2 - 1$ if h is odd and given by $\frac{h^2}{2} - 2$ if h is even.

Conjecture 6.7 *For each integer h the plane integral point set $\mathcal{P} = decompose(h)$ is maximal if $|\mathcal{P}| \geqslant 7$.*

Also, the recognition of a crab is a very easy task. Given an integral point set $\mathcal{P}$ over $\mathbb{Z}^2$ one can easily check whether a subset $\mathcal{L} \subset \mathcal{P}$ of $n-2$ points is collinear by using:

Lemma 6.8 *Three points (x_1, y_1), (x_2, y_2), and (x_3, y_3) in $\mathbb{R}^2$ are collinear if and only if we have*

$$\begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix} = 0.$$

Additionally the lines through $\mathcal{L}$ and $\mathcal{P}\backslash\mathcal{L}$ are perpendicular. If the point set is symmetric to these two lines then $\mathcal{P}$ is a crab.

Crabs are very dominating examples of maximal integral point sets over $\mathbb{Z}^2$. For the study of maximal integral point sets over $\mathbb{Z}^3$ one might try to generalize the construction of a crab. Let us remark in this context, that the existence of an integral point set with coordinates

$$\left\{ \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} x \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ y \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ z \end{pmatrix} \right\},$$

where $x, y, z \in \mathbb{Z}$ is equivalent to a famous open problem, the existence of a perfect box, see [7, Problem D18].

So far we have only used $g = 1$ in Lemma 6.5. Now we want to have a look at the case $g > 1$. So given integers $g \cdot h$ and g we can apply Lemma 6.5. For two factors $f_1 > f_2$ with $f_1 \cdot f_2 = g^2h^2$ we have

$$gb_i = \frac{f_1 + f_2}{2} \text{ and } gc_i = \frac{f_1 - f_2}{2}.$$

The values gb_i and gc_i are integers if and only if we have $f_1 \equiv f_2 \pmod 2$. Since not only the gb_i's but also the b_i's must be integers we have to require $f_1 + f_2 \equiv 0 \pmod g$. Let us have an example. We choose $gh = 672 = 2^5 \cdot 3 \cdot 7$ and $g = 5$. Now we look at the divisors of $g^2h^2 = 451584 = 2^{10} \cdot 3^3 \cdot 7^2$ and determine the suitable pairs (f_1, f_2) fulfilling

$$\begin{aligned} &f_1 \cdot f_2 = g^2h^2, \quad f_1 > f_2, \quad f_1 \equiv f_2 \pmod 2, \\ &\text{and } f_1 + f_2 \equiv 0 \pmod 5, \end{aligned}$$

$$\begin{aligned} &\Big\{(784, 576), (896, 504), (1176, 384), (1344, 336), \\ &(1536, 294), (1764, 256), (2016, 224), (2304, 196), \\ &(3136, 144), (3584, 126), (4704, 96), (5376, 84), \\ &(7056, 64), (8064, 56), (12544, 36), (18816, 24), \\ &(28224, 16), (32256, 14), (75264, 6), (112896, 4)\Big\}. \end{aligned}$$

The corresponding values b_i are given by

$$\begin{aligned} &\Big\{136, 140, 156, 168, 183, 202, 224, 250, 328, 371, 480, \\ &546, 712, 812, 1258, 1884, 2824, 3227, 7527, 11290\Big\} \end{aligned}$$

and the corresponding values gc_i are given by

$$\begin{aligned} \mathcal{C} \;=\; &\Big\{104, 196, 396, 504, 621, 754, 896, 1054, 1496, \\ &1729, 2304, 2646, 3496, 4004, 6254, 9396, 14104, \\ &16121, 37629, 56446\Big\}. \end{aligned}$$

Clearly the c_i cannot be integers unless h is an integer. So let us consider the points on the left of the base point F of the height h. They correspond to values gc_i which all fulfill $gc_i \equiv m \pmod g$, for a fixed $m \in \{1, \ldots, g-1\}$. The points on the right hand side of F correspond to the values gc_i fulfilling $gc_i \equiv -m \pmod g$. So let us choose $m = 1$. Since all elements of our candidate set $\mathcal{C}$ are congruent to ± 1 modulo 5 we obtain an integral point set of cardinality $|\mathcal{C}| + 1 = 21$:

$$\begin{aligned} \mathcal{P} \;=\; &\left\{ \begin{pmatrix} 0 \\ \frac{672}{5} \end{pmatrix}, \begin{pmatrix} \frac{-196}{5} \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{-396}{5} \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{-621}{5} \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{-896}{5} \\ 0 \end{pmatrix}, \right. \\ &\begin{pmatrix} \frac{-1496}{5} \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{-2646}{5} \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{-3496}{5} \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{-9396}{5} \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{-16121}{5} \\ 0 \end{pmatrix}, \\ &\begin{pmatrix} \frac{-56446}{5} \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{104}{5} \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{504}{5} \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{754}{5} \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{1054}{5} \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{1729}{5} \\ 0 \end{pmatrix}, \\ &\left. \begin{pmatrix} \frac{2304}{5} \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{4004}{5} \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{6254}{5} \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{14104}{5} \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{37629}{5} \\ 0 \end{pmatrix} \right\} \end{aligned}$$

After a suitable transformation and applying Algorithm 4.1 we obtain the minimum coordinate representation

$$\begin{aligned} &\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ -168 \end{pmatrix}, \begin{pmatrix} -40 \\ 30 \end{pmatrix}, \begin{pmatrix} 64 \\ -48 \end{pmatrix}, \begin{pmatrix} -88 \\ 66 \end{pmatrix}, \begin{pmatrix} 112 \\ -84 \end{pmatrix}, \right. \\ &\begin{pmatrix} 144 \\ -108 \end{pmatrix}, \begin{pmatrix} 180 \\ -135 \end{pmatrix}, \begin{pmatrix} -196 \\ 147 \end{pmatrix}, \begin{pmatrix} 224 \\ -168 \end{pmatrix}, \begin{pmatrix} -288 \\ 216 \end{pmatrix}, \\ &\begin{pmatrix} 320 \\ -240 \end{pmatrix}, \begin{pmatrix} 504 \\ -378 \end{pmatrix}, \begin{pmatrix} -560 \\ 420 \end{pmatrix}, \begin{pmatrix} 640 \\ -480 \end{pmatrix}, \begin{pmatrix} -920 \\ 690 \end{pmatrix}, \\ &\left. \begin{pmatrix} 1584 \\ -1188 \end{pmatrix}, \begin{pmatrix} -2176 \\ 1632 \end{pmatrix}, \begin{pmatrix} 2660 \\ -1995 \end{pmatrix}, \begin{pmatrix} -5940 \\ 4455 \end{pmatrix}, \begin{pmatrix} 9112 \\ -6834 \end{pmatrix} \right]. \end{aligned}$$

We call point sets arising from Lemma 6.5, where $g > 1$ and $h \notin \mathbb{N}$ semi-crabs, see Figure 4 for a drawing of our example.

Definition 6.9 *For positive integers $g, \tilde{h}$ and non-zero integers $\tilde{b}_1, \ldots, \tilde{b}_k$ we call the point set*

$$\textit{semi-crab}(g, \tilde{h}, \tilde{b}_1, \ldots, \tilde{b}_k) := \left\{ \begin{pmatrix} 0 \\ \frac{h}{g} \end{pmatrix}, \begin{pmatrix} \frac{\tilde{b}_1}{g} \\ 0 \end{pmatrix}, \ldots, \begin{pmatrix} \frac{\tilde{b}_k}{g} \\ 0 \end{pmatrix} \right\}$$

a semi-crab of order k.

Construction 6.10 *For given positive integers g and gh, where $h \notin \mathbb{N}$, there exists an integral point set decompose(g, gh) which is isomorphic to a semi-crab.*

Conjecture 6.11 *For each pair of integers gh, g the plane integral point set $\mathcal{P} = decompose(gh, g)$ is maximal if $|\mathcal{P}| \geqslant 7$.*

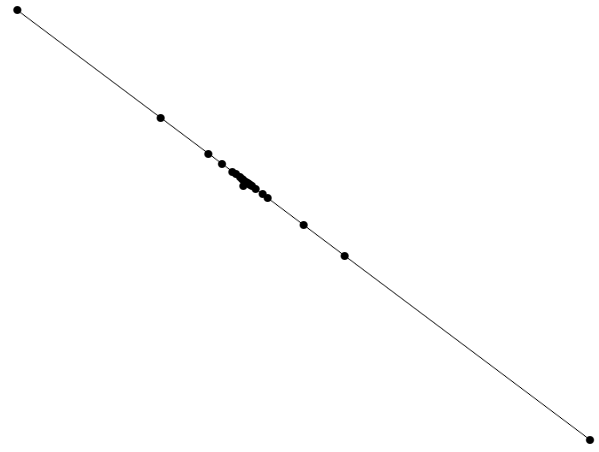


Figure 4: A semi-crab of cardinality 21 and diameter 18815.

Since Construction 6.6 can only produce integral point sets of odd cardinality, Construction 6.10 is a convenient completion. It is not difficult to implement an algorithm that exhaustively generates crabs and semi-crabs up to a given diameter by utilizing Lemma 6.5.

Let us have a look at the possible values for $g > 1$. If we would choose $g = 2$ then due to $2 \nmid g^2h^2$ all divisors of g^2h^2 would be odd and we would have $m \equiv 1 \pmod 2$, which is not possible. Thus $2 \nmid g$. For $g = 3$ the only possibility for $f_1 + f_2 \equiv 0 \pmod 3$ is $f_1 \equiv 1 \pmod 3$ and $f_2 \equiv 2 \pmod 3$, which is not possible since $1 \cdot 2 \equiv 2 \pmod 3$ is not a square in $\mathbb{Z}_3$. Thus $g = 5$ is the first valid possibility. More generally we can state that if g is a prime then we have $g \equiv 1 \pmod 4$, since -1 has to be a square in $\mathbb{Z}_g$.

6.2 Construction of integral point sets on circles

In Addition to the construction of crabs there is another useful construction of integral point sets of $\mathbb{Z}^2$ with large cardinality, see [10] for a similar construction over the ring $\mathbb{Z}\left[\frac{-1+\sqrt{-3}}{2}\right]$. Let $p_j \equiv 1 \pmod 4$ be distinct primes over $\mathbb{N}$. We consider the ring $\mathbb{Z}[i]$, where every integer p_j has a unique prime factorization $p_j = \omega_j \cdot \overline{\omega_j}$, where $\overline{c}$ denotes the complex conjugate of c. We may write $\omega_j = a_j + b_j i$, with integers a_j, b_j. With multiplicities $\nu_j \in \mathbb{N}$ we set

$$R = \prod_{j=1}^{r} p_j^{\nu_j}$$

and for each of the $\tau(R)$ divisors of R,

$$\prod_{j=1}^{r} p_j^{u_j} = \prod_{j=1}^{r} \omega_j^{u_j} \overline{\omega}_j^{u_j}, \quad 0 \leqslant u_j \leqslant \nu_j$$

we define $\eta_{2h} = \prod_{j=1}^{r} \omega_j^{\nu_j+u_j} \overline{\omega}_j^{\nu_j-u_j}$, $\eta_{2h-1} = i \cdot \eta_{2h}$ for $1 \leqslant h \leqslant \tau(R)$. With this we define vertices ξ_s for $1 \leqslant s \leqslant 2\tau(R)$ by

$$\xi_{2h-k} = \frac{\eta_{2h-k}^2}{R}, \quad 1 \leqslant h \leqslant \tau(R),\ k \in \{0,1\}.$$

We set $\eta_s = x_s + y_s i$ with $x_s, y_s \in \mathbb{Z}$ for $1 \leqslant s \leqslant 2\tau(R)$. We have

$$|\eta_s|^2 = \eta_s \overline{\eta}_s = x_s^2 + y_s^2 = \prod_{j=1}^{r} \omega_j^{2\nu_j} \overline{\omega}_j^{2\nu_j} = \prod_{j=1}^{r} p_j^{2\nu_j} = R^2.$$

This yields $x_s^2 = R^2 - y_s^2$, which we use to calculate

$$\begin{aligned} R^2 \cdot |\xi_s - \xi_t|^2 &= \left|\eta_s^2 - \eta_t^2\right|^2 \\ &= \left|x_s^2 - y_s^2 - x_t^2 + y_t^2 + i \cdot (2x_s y_s - 2x_t y_t)\right|^2 \\ &= \left|\left(2y_t^2 - 2y_s^2\right) + i \cdot (2x_s y_s - 2x_t y_t)\right|^2 \\ &= 4\left(y_t^2 - y_s^2\right)^2 + 4\left(x_s y_s - x_t y_t\right)^2 \\ &= 4\left(y_t^2 - y_s^2\right)\left(x_s^2 - x_t^2\right) + 4\left(x_s y_s - x_t y_t\right)^2 \\ &= 2^2\left(x_s y_t - x_t y_s\right)^2. \end{aligned}$$

Thus the distance between ξ_s and ξ_t is given by $|\xi_s - \xi_t| = \frac{1}{R}|x_s y_t - x_t y_s|$. Since $\eta_s \overline{\eta}_t$ $= (x_s + y_s i)(x_t - y_t i) = x_s x_t + y_s y_t + i(x_t y_s - x_s y_t)$ and

$$\begin{aligned} \eta_s \overline{\eta}_t &= i^{k_s} \bar{i}^{k_t} \prod_{j=1}^{r} \omega_j^{\nu_j+u_j} \overline{\omega}_j^{\nu_j-u_j} \prod_{j=1}^{r} \overline{\omega}_j^{\nu_j+w_j} \omega_j^{\nu_j-w_j} \\ &= i^{k_s-k_t} \prod_{j=1}^{r} \omega_j^{2\nu_j+u_j-w_j} \overline{\omega}_j^{2\nu_j-u_j+w_j} \\ &= R \cdot i^{k_s-k_t} \prod_{j=1}^{r} \omega_j^{\nu_j+u_j-w_j} \overline{\omega}_j^{\nu_j-u_j+w_j} \in \mathbb{Z}[i] \end{aligned}$$

we have that the distance between ξ_s and ξ_t is integral for every $1 \leqslant s, t \leqslant 2\tau(R)$. Additionally we can add the center of the circle to this point set to obtain an integral point set of cardinality $2 \cdot \tau(R) + 1$ having rational coordinates. After a suitable rotation we can achieve integral coordinates.

So let us have an example. We choose $R = 5 \cdot 13 = 65$ and successively obtain

$$\begin{aligned} &\omega_1 = 2 + i, \quad \omega_2 = 3 + 2i, \\ &\eta_1 = 65i, \quad \eta_2 = 65, \\ &\eta_3 = -52 + 39i, \quad \eta_4 = 39 + 52i, \\ &\eta_5 = -60 + 25i, \quad \eta_6 = 25 + 60i, \\ &\eta_7 = -56 - 33i, \quad \eta_8 = -33 + 56i, \\ &\xi_1 = -65, \quad \xi_2 = 65, \\ &\xi_3 = \frac{91}{5} - \frac{312}{5}i, \quad \xi_4 = -\frac{91}{5} + \frac{312}{5}i, \\ &\xi_5 = \frac{595}{13} - \frac{600}{13}i, \quad \xi_6 = -\frac{595}{13} + \frac{600}{13}i, \\ &\xi_7 = \frac{2047}{65} + \frac{3696}{65}i, \quad \xi_8 = -\frac{2047}{65} - \frac{3696}{65}i. \end{aligned}$$

After adding the origin $(0,0)^T$ and applying a suitable rotation and translation we obtain the maximal integral point set

$$\begin{aligned} \mathcal{P} = \Bigg[&\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}0\\-32\end{pmatrix}, \begin{pmatrix}-30\\40\end{pmatrix}, \begin{pmatrix}-30\\-72\end{pmatrix}, \begin{pmatrix}-63\\-16\end{pmatrix}, \\ &\begin{pmatrix}-96\\40\end{pmatrix}, \begin{pmatrix}-96\\-72\end{pmatrix}, \begin{pmatrix}-126\\0\end{pmatrix}, \begin{pmatrix}-126\\-32\end{pmatrix}\Bigg] \end{aligned}$$

in minimum coordinate representation.

Construction 6.12 *For a given* R *which has only prime factors* p *fulfilling* $p \equiv 1 \pmod 4$ *there exists an integral point set circle*(R) *consisting of* $2\cdot\tau(R)$ *points on a circle of radius* R *together with its center, where* $\tau(R)$ *denotes the number of divisors of* R.

From the above it is easy to deduce that the $2\tau(R)$ points on the circle all have pairwise even distances and that the diameter of this point set is given by $2R$. Using this we can give another construction.

Construction 6.13 *For a given* R *which has only prime factors* p *fulfilling* $p \equiv 1 \pmod 4$ *there exists an integral point set* $\widetilde{\text{circle}}(R)$ *consisting of* $2\cdot\tau(R)$ *points on a circle of radius* $\frac{R}{2}$*, where* $\tau(R)$ *denotes the number of divisors of* R.

Conjecture 6.14 *The plane integral point sets given by Construction 6.12 and Construction 6.13 are maximal.*

We can generalize the idea of Construction 6.13 in some way. Let t be an arbitrary integer, R be a integer having only prime factors fulfilling $p \equiv 1 \pmod 4$, and $\mathcal{P}(R)$ be the integral point set given by Construction 6.12 with radius R. By $\mathcal{P}(R,t)$ we denote the point set which arises from $\mathcal{P}(R)$ by scaling the point set with a factor $\frac{1}{t}$, this means dividing all distances by t. Thus $\mathcal{P}(R,t)$ is a point set with pairwise rational distances and rational coordinates. With this we can construct a graph $\mathcal{G}$ containing the points of $\mathcal{P}(R,t)$ as its vertices. Two vertices of $\mathcal{G}$ are connected by an edge, if and only if the corresponding points have an integral distance in $\mathcal{P}(R,t)$. The maximal cliques $\mathcal{C}$ of $\mathcal{G}$ correspond to integral point sets $\mathcal{P}(R,t,\mathcal{C})$.

Construction 6.15 *For a given* R *which has only prime factors* p *fulfilling* $p \equiv 1 \pmod 4$ *and a given integer* t *there exist integral point sets circle*$(R,t,\mathcal{C})$ *consisting of points on a circle of radius* $\frac{R}{t}$*, where* $\mathcal{C}$ *is a maximal clique of the above described graph. As an abbreviation we use circle*(R,t) *instead of circle*$(R,t,\mathcal{C})$.

Conjecture 6.16 *For* $t = 8$ *Construction 6.15 gives maximal integral point sets of cardinality* $\tau(R)$.

7 Maximal integral point sets over $\mathbb{Z}^2$ with further conditions

In Table 2 we have summarized the constructions yielding the smallest diameter of a maximal integral point set over $\mathbb{Z}^2$. Some of the values $d_M(k,2)$ could be determined exactly by an exhaustive search, but for most values of k we only have upper bounds (and 301 as lower bound). In some cases, denoted by $\overset{\star}{\lesssim}$, we were not able to check the maximality of the constructed point sets, since their diameter was too large.

Looking at Table 2 we observe, that the constructions of crabs (Construction 6.6 and Construction 6.10) are very

k	$d_M(k,2)$	construction
3	$= 2066$	$\Delta(2066, 1803, 505)$
4	$= 5$	$\mathcal{P}_1(3,4) = \widetilde{\text{circle}}(5)$
5	$= 8$	$\mathcal{P}_2(3,4) = \text{crab}(3,4)$
6	$= 25$	$\widetilde{\text{circle}}\left(5^2\right)$
7	$= 30$	$\text{crab}\,(8,6,15)$
8	$= 65$	$\widetilde{\text{circle}}\,(5\cdot 13)$
9	$= 130$	$\text{circle}\,(5\cdot 13)$
10	$\leqslant 625$	$\widetilde{\text{circle}}\left(5^4\right)$
11	$= 70$	$\text{decompose}\left(2^2\cdot 3\right)$
12	$= 325$	$\widetilde{\text{circle}}\left(5^2\cdot 13\right)$
13	$\leqslant 650$	$\text{circle}\left(5^2\cdot 13\right)$
14	$\leqslant 15625$	$\widetilde{\text{circle}}\left(5^6\right)$
15	$\leqslant 8190$	$\text{decompose}\left(2^7\right)$
16	$\leqslant 1105$	$\widetilde{\text{circle}}\,(5\cdot 13\cdot 17)$
17	$= 286$	$\text{decompose}\left(2^3\cdot 3\right)$
18	$\leqslant 4225$	$\widetilde{\text{circle}}\left(5^2\cdot 13^2\right)$
19	$\leqslant 8450$	$\text{circle}\left(5^2\cdot 13^2\right)$
20	$\leqslant 8125$	$\widetilde{\text{circle}}\left(5^4\cdot 13\right)$
21	$\leqslant 16250$	$\text{circle}\left(5^4\cdot 13\right)$
22	$\leqslant 53360$	$\text{decompose}\left(2^2\cdot 3\cdot 7\cdot 11, 5\right)$
23	$\leqslant 1150$	$\text{decompose}\left(2^4\cdot 3\right)$
24	$\leqslant 5525$	$\widetilde{\text{circle}}\left(5^2\cdot 13\cdot 17\right)$
25	$\leqslant 11050$	$\text{circle}\left(5^2\cdot 13\cdot 17\right)$
26	$\leqslant 112895$	$\text{decompose}\left(2^6\cdot 3\cdot 7, 5\right)$
27	$\leqslant 2590$	$\text{decompose}\left(2^3\cdot 3^2\right)$
28	$\overset{\star}{\lesssim} 203125$	$\widetilde{\text{circle}}\left(5^6\cdot 13\right)$
29	$\leqslant 1798$	$\text{decompose}\left(2^2\cdot 3\cdot 5\right)$
30	$\leqslant 105625$	$\widetilde{\text{circle}}\left(5^4\cdot 13^2\right)$
31	$\overset{\star}{\lesssim} 211250$	$\text{circle}\left(5^4\cdot 13^2\right)$
32	$\leqslant 27625$	$\widetilde{\text{circle}}\left(5^3\cdot 13\cdot 17\right)$
33	$\leqslant 55250$	$\text{circle}\left(5^3\cdot 13\cdot 17\right)$
34	$\leqslant 142295$	$\text{decompose}\left(2^3\cdot 3\cdot 7\cdot 11, 5\right)$
35	$\leqslant 18430$	$\text{decompose}\left(2^6\cdot 3\right)$
36	$\leqslant 40625$	$\widetilde{\text{circle}}\left(5^5\cdot 13\right)$
37	$\leqslant 10366$	$\text{decompose}\left(2^4\cdot 3^2\right)$
38	$\overset{\star}{\lesssim} 571535$	$\text{decompose}\left(2^4\cdot 3^3\cdot 7, 5\right)$
39	$\overset{\star}{\lesssim} 4816895$	$\text{decompose}\left(2^9\cdot 3\cdot 7, 5\right)$
40	$\leqslant 138125$	$\widetilde{\text{circle}}\left(5^4\cdot 13\cdot 17\right)$
41	$\leqslant 73726$	$\text{decompose}\left(2^7\cdot 3\right)$
42	$\overset{\star}{\lesssim} 677375$	$\text{decompose}\left(2^6\cdot 3^2\cdot 7, 5\right)$
43	$\overset{\star}{\lesssim} 4573799$	$\text{decompose}\left(2^3\cdot 3^2\cdot 5\cdot 7\cdot 11, 17\right)$
44	$\overset{\star}{\lesssim} 6614998$	$\text{decompose}\left(2^4\cdot 3^2\cdot 5^2\cdot 7, 13\right)$
45	$\overset{\star}{\lesssim} 7001315$	$\text{decompose}\left(2^3\cdot 3^2\cdot 7^2, 5\right)$
46	$\overset{\star}{\lesssim} 64833614$	$\text{decompose}\left(2^2\cdot 3^4\cdot 5\cdot 7\cdot 11, 17\right)$
47	$\leqslant 7198$	$\text{decompose}\left(2^3\cdot 3\cdot 5\right)$
48	$\overset{\star}{\lesssim} 160225$	$\widetilde{\text{circle}}\left(5^2\cdot 13\cdot 17\cdot 29\right)$
49	$\overset{\star}{\lesssim} 320450$	$\text{circle}\left(5^2\cdot 13\cdot 17\cdot 29\right)$
50	$\overset{\star}{\lesssim} 4064255$	$\text{decompose}\left(2^7\cdot 3^2\cdot 7, 5\right)$

Table 2: Best known constructions for maximal integral point sets over $\mathbb{Z}^2$ in arbitrary position.

dominating. The resulting point sets contain $n-2$ and $n-1$ collinear points out of n points, respectively. So it may be interesting to study maximal integral point sets over $\mathbb{Z}^2$, where no three points are collinear. We also say, that a point set is in semi-general position, if no three points are collinear. By $\overline{d_M}(k,2)$ we denote the minimum possible diameter of these point sets. We can check for this further condition, that no three points are collinear, by applying Lemma 6.8.

k	$\overline{d_M}(k,2)$	construction
3	$= 2066$	$\Delta(2066,1803,505)$
4	$= 5$	$\mathcal{P}_1(3,4)$=$\widetilde{\text{circle}}(5)$
5	$= 120$	see Figure 5
6	$= 25$	$\widetilde{\text{circle}}(5^2)$
7	$= 925$	see Figure 6
8	$= 65$	$\widetilde{\text{circle}}(5\cdot 13)$
9	$= 1045$	see Figure 7
10	$= 625$	$\widetilde{\text{circle}}(5^4)$
11	$\overset{\star}{\lesssim} 2434375$	$\text{circle}(5^{10},8)$
12	$= 325$	$\widetilde{\text{circle}}(5^2\cdot 13)$
13	$\overset{\star}{\lesssim} 60859375$	$\text{circle}(5^{12},8)$
14	$\leqslant 15625$	$\widetilde{\text{circle}}(5^6)$
15	$\leqslant 26390$	$\text{circle}(5^4\cdot 13^2,8)$
16	$= 1105$	$\widetilde{\text{circle}}(5\cdot 13\cdot 17)$
17	$\overset{\star}{\lesssim} 38037109375$	$\text{circle}(5^{16},8)$
18	$= 4225$	$\widetilde{\text{circle}}(5^2\cdot 13^2)$
19	$\overset{\star}{\lesssim} 950927734375$	$\text{circle}(5^{18},8)$
20	$= 8125$	$\widetilde{\text{circle}}(5^4\cdot 13)$
21	$\overset{\star}{\lesssim} 659750$	$\text{circle}(5^6\cdot 13^2,8)$
22	$\overset{\star}{\lesssim} 9765625$	$\widetilde{\text{circle}}(5^{10})$
23	$\overset{\star}{\lesssim} 595928935571106$	$\text{circle}(5^{22},8)$
24	$= 5525$	$\widetilde{\text{circle}}(5^2\cdot 13\cdot 17)$
25	$\overset{\star}{\lesssim} 4462500$	$\text{circle}(5^4\cdot 13^4,8)$
26	$\overset{\star}{\lesssim} 244140625$	$\widetilde{\text{circle}}(5^{12})$
27	$\overset{\star}{\lesssim} 305218$	$\text{circle}(5^2\cdot 13^2\cdot 17^2,8)$
28	$\overset{\star}{\lesssim} 203125$	$\widetilde{\text{circle}}(5^6\cdot 13)$
29	$\overset{\star}{\lesssim} 9311389618298531250$	$\text{circle}(5^{28},8)$
30	$\leqslant 105625$	$\widetilde{\text{circle}}(5^4\cdot 13^2)$
31	$\overset{\star}{\lesssim} 232784740457463281250$	$\text{circle}(5^{30},8)$
32	$\leqslant 27625$	$\widetilde{\text{circle}}(5^3\cdot 13\cdot 17)$
33	$\overset{\star}{\lesssim} 412343750$	$\text{circle}(5^{10}\cdot 13^2,8)$
34	$\overset{\star}{\lesssim} 152587890625$	$\widetilde{\text{circle}}(5^{16})$
35	$\overset{\star}{\lesssim} 111562500$	$\text{circle}(5^6\cdot 13^4,8)$

Table 3: Best known constructions for maximal integral point sets over $\mathbb{Z}^2$ in semi-general position - part 1.

Using the methods and algorithms described in this article, we were able to obtain some exact values and some upper bounds for $\overline{d_M}(k,2)$. The results are summarized in Table 3 and Table 4. We would like to remark that we additionally have the lower bounds $\overline{d_M}(k,2) \geqslant 5525$ for $k \in \{11,13,14,15,17\}$ and $\overline{d_M}(k,2) \geqslant 10001$ for $k \geqslant 19$, $k \neq 20,24$.

k	$\overline{d_M}(k,2)$	construction
36	$\leqslant 71825$	$\widetilde{\text{circle}}(5^2\cdot 13^2\cdot 17)$
37	$\overset{\star}{\lesssim} 3637261569647863769531250$	$\text{circle}(5^{36},8)$
38	$\overset{\star}{\lesssim} 3814697265625$	$\widetilde{\text{circle}}(5^{18})$
39	$\overset{\star}{\lesssim} 10314771205$	$\text{circle}(5^{12}\cdot 13^2,8)$
40	$\leqslant 138125$	$\widetilde{\text{circle}}(5^4\cdot 13\cdot 17)$
41	$\overset{\star}{\lesssim} 22732884810299148559570312 50$	$\text{circle}(5^{40},8)$
42	$\overset{\star}{\lesssim} 2640625$	$\widetilde{\text{circle}}(5^6\cdot 13^2)$
43	$\overset{\star}{\lesssim} 56832212025747871398925781250$	$\text{circle}(5^{42},8)$
44	$\overset{\star}{\lesssim} 126953125$	$\widetilde{\text{circle}}(5^{10}\cdot 13)$
45	$\overset{\star}{\lesssim} 7630450$	$\text{circle}(5^4\cdot 13^2\cdot 17^2,8)$
46	$\overset{\star}{\lesssim} 2384185791015625$	$\widetilde{\text{circle}}(5^{22})$
47	$\overset{\star}{\lesssim} 3552013251609241962432861328 1250$	$\text{circle}(5^{46},8)$
48	$\overset{\star}{\lesssim} 160225$	$\widetilde{\text{circle}}(5^2\cdot 13\cdot 17\cdot 29)$
49	$\overset{\star}{\lesssim} 18854062500$	$\text{circle}(5^6\cdot 13^6,8)$
50	$\overset{\star}{\lesssim} 17850625$	$\widetilde{\text{circle}}(5^4\cdot 13^4)$

Table 4: Best known constructions for maximal integral point sets over $\mathbb{Z}^2$ in semi-general position - part 2.

We would like to have a closer look on the smallest known examples of maximal integral point sets in semi-general position consisting of an odd number of points. For cardinality 5 the two smallest point sets with respect to the diameter are given in minimum coordinate representation by

$$\left[\binom{0}{0}, \binom{0}{-78}, \binom{-20}{21}, \binom{-20}{-99}, (-52,-39)\right]$$

and

$$\left[\binom{0}{0}, \binom{0}{-80}, \binom{-45}{28}, \binom{-45}{-108}, \binom{-96}{-40}\right],$$

see Figure 5 for a drawing of the first point set. Both point sets consist of four point on a circle $\mathcal{C}$ of radii $\frac{29\cdot 101}{40}$ and $\frac{13\cdot 53}{10}$, respectively. In each case the fifth point does not lie on this circle $\mathcal{C}$, but the line through this point and the center of $\mathcal{C}$ is a symmetry axis of the point set.

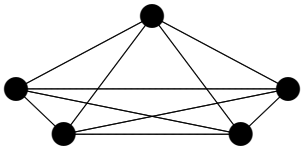


Figure 5: The smallest maximal integral point set of cardinality 5 in semi-general position.

For cardinality 7 the two smallest examples are given by

$$\left[\binom{0}{0}, \binom{0}{-285}, \binom{-180}{240}, \binom{-440}{-384}, \binom{-700}{240}, \binom{-880}{0}, \binom{-880}{-285}\right]$$

and

$$\left[\binom{0}{0}, \binom{0}{-855}, \binom{-540}{720}, \binom{-1320}{-1152}, \binom{-2100}{720}, \binom{-2640}{0}, \binom{-2640}{-855}\right],$$

see Figure 6 for a graphical representation of the first example. The geometric shape of the corresponding two point sets is similar to the case of cardinality 5. In each case 6 points are situated on a circle $\mathcal{C}$ of radii $\frac{5^2 \cdot 37}{2}$ and $\frac{3 \cdot 5^2 \cdot 37}{2}$, respectively. Again we have the symmetry axis through the seventh point and the center of $\mathcal{C}$.

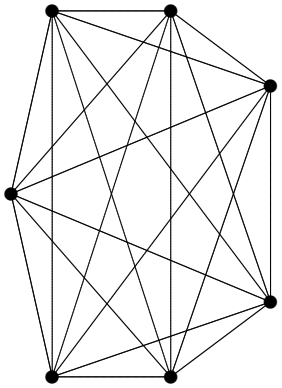


Figure 6: The smallest maximal integral point set of cardinality 7 in semi-general position.

For cardinality 9 the two smallest examples are given by

$$\left[\binom{0}{0}, \binom{0}{-504}, \binom{-64}{-252}, \binom{612}{255}, \binom{612}{-759}, \binom{720}{210}, \binom{720}{-714}, \binom{836}{123}, \binom{836}{-627}\right]$$

and

$$\left[\binom{0}{0}, \binom{0}{-672}, \binom{-123}{164}, \binom{-123}{-836}, \binom{-816}{340}, \binom{-816}{-1012}, \binom{-960}{280}, \binom{-960}{-952}, \binom{-1323}{-336}\right],$$

see Figure 7 for a graphical representation of the first example. Here in both examples all nine points are situated of circles of radii $\frac{5^2 \cdot 13^2}{8}$ and $\frac{5^2 \cdot 13^2}{6}$, respectively. They both can be obtained using Construction 6.15.

Now we observe that the constructions based on circles, Construction 6.12, Construction 6.13, and Construction 6.15, are very dominating in this context. The next natural step is to also forbid four points on a circle. If no three

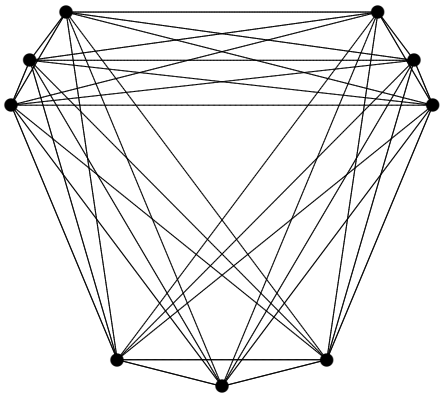


Figure 7: The smallest maximal integral point set of cardinality 9 in semi-general position.

points are on a line and no four points on a circle we speak of general position. By $\dot{d}_M(k,2)$ we denote the minimum possible diameter of a maximal plane integral point set in general position over $\mathbb{Z}^2$. Without the maximality condition these point sets are also known as k_2-cluster [22]. As we cannot apply our most successful constructions based on crabs and circles in this case, examples are scarce.

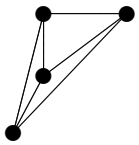


Figure 8: The smallest maximal integral point set of cardinality 4 in general position.

For the check whether four points are situated on a circle we have a well known criterion similar to Lemma 6.8:

Lemma 7.1 *Four points* (x_1,y_1), (x_2,y_2), (x_3,y_3), (x_4,y_4) *in* $\mathbb{R}^2$ *are situated on a circle if and only if*

$$\begin{vmatrix} x_1 & y_1 & x_1^2+y_1^2 & 1 \\ x_2 & y_2 & x_2^2+y_2^2 & 1 \\ x_3 & y_3 & x_3^2+y_3^2 & 1 \\ x_4 & y_4 & x_4^2+y_4^2 & 1 \end{vmatrix} = 0$$

holds.

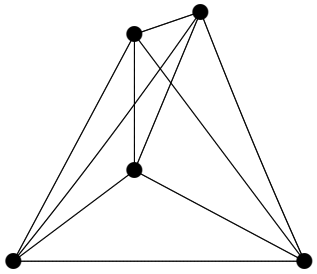


Figure 9: The smallest maximal integral point set of cardinality 5 in general position.

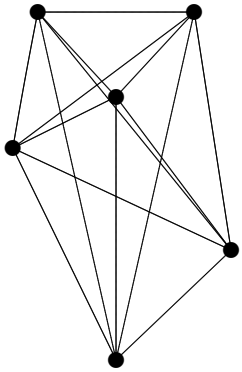


Figure 10: The smallest maximal integral point set of cardinality 6 in general position.

In Table 5 we have summarized our knowledge on $\dot{d}_M(k,2)$. For the lower bound $\dot{d}_M(7,2) > 599000$ we refer to [18]. Whether $\dot{d}_M(7,2)$ is finite (even if we drop the maximality condition) is an open problem, see [7, 22]. If we drop the maximality condition and the condition on the integrality of the coordinates (in other words characteristic one), then very recently two such examples were found, see [14]. The smallest example for $k = 6$ is indeed the smallest integral point set of characteristic one in general position with cardinality 6.

k	$\dot{d}_M(k,2)$	construction
3	$= 2066$	$\Delta(2066, 1803, 505)$
4	$= 87$	$\left[\binom{0}{0}, \binom{0}{-33}, \binom{-16}{30}, \binom{44}{-33}\right]$, see Figure 8
5	$= 165$	$\left[\binom{0}{0}, \binom{0}{-72}, \binom{-35}{12}, \binom{64}{-120}, \binom{-90}{-120}\right]$, see Figure 9
6	$= 1886$	$\left[\binom{0}{0}, \binom{0}{-828}, \binom{-448}{-414}, \binom{-720}{132}, \binom{-1260}{-1023}, \binom{-1840}{-414}\right]$, see Figure 10
7	> 599000	

Table 5: Best known constructions for maximal integral point sets over $\mathbb{Z}^2$ in general position.

We would also like to give the coordinates for the second smallest examples. For cardinality 4 we have

$$\left[\binom{0}{0}, \binom{0}{-69}, \binom{-20}{-21}, \binom{-92}{0}\right],$$

for cardinality 5 we have

$$\left[\binom{0}{0}, \binom{0}{-153}, \binom{-60}{144}, \binom{-140}{-48}, \binom{-176}{57}\right],$$

and for cardinality 6 we have

$$\left[\binom{0}{0}, \binom{-135}{-1008}, \binom{420}{1008}, \binom{735}{-392}, \binom{1155}{616}, \binom{1290}{1624}\right].$$

8 Conclusion and outlook

We have described several constructions for integral point sets over $\mathbb{Z}^2$ with given cardinality that fulfill some further properties. Although the maximality of the resulting integral point sets cannot be guaranteed so far, we conjecture them to be in many cases. We have described efficient algorithms for exhaustive generation of maximal integral point sets over $\mathbb{Z}^2$ and for testing the maximality of a given integral point set. Some exact values of minimum diameters for given cardinalities could be obtained and several values are constructed as upper bounds and conjectured to be the exact values.

It remains a task to prove the maximality of point sets resulting from some of our constructions in general. Clearly similar problems could be considered in higher dimensions.

Bibliography

[1] N. H. Anning and P. Erdős, *Integral distances*, Bull. Amer. Math. Soc. **51** (1945), 598–600.

[2] A. Antonov and M. Brancheva, *Algorithm for finding maximal Diophantine figures*, Spring Conference 2007 of the Union of Bulgarian Mathematicians, 2007.

[3] S. Dimiev and K. Markov, *Gauss Integers and Diophantine Figures*, Mathematics and Mathematical Education **31** (2002), 88–95.

[4] P. Erdős, *Integral distances*, Bull. Amer. Math. Soc. **51** (1945), 996.

[5] J. Fricke, *On heron simplices and integer embedding*, preprint (2001).

[6] R. E. Fullerton, *Integral distances in banach spaces*, Bull. Amer. Math. Soc. **55** (1949), 901–905.

[7] R. K. Guy, *Unsolved problems in number theory. 2nd ed.*, Unsolved Problems in Intuitive Mathematics. 1. New York, NY: Springer- Verlag. xvi, 285 p., 1994.

[8] B. Haible and R. Kreckel, *CLN, a class library for numbers*, 2005, http://www.ginac.de/CLN/.

[9] H. Harborth, *Integral distances in point sets*, Karl der Grosse und sein Nachwirken. 1200 Jahre Kultur und Wissenschaft in Europa. Band 2: Mathematisches Wissen. Turnhout: Brepols (P. L. Butzer et al., eds.), 1998, 213–224.

[10] H. Harborth, A. Kemnitz, and M. Möller, *An upper bound for the minimum diameter of integral point sets*, Discrete Comput. Geom. **9** (1993), no. 4, 427–432.

[11] M. Kiermaier and S. Kurz, *Inclusion-maximal integral point sets in affine planes over finite fields*, (submitted).

[12] A. Kohnert and S. Kurz, *A note on Erdös-Diophantine graphs and Diophantine carpets*, Math. Balkanica (N.S.) **21** (2007), no. 1-2, 1–5.

[13] A. Kohnert and S. Kurz, *Integral point sets over* $\mathbb{Z}_n^m$, Discrete Appl. Math. (to appear).

[14] T. Kreisel and S. Kurz, *There are integral heptagons, no three points on a line, no four on a circle*, Discrete. Comput. Geom. (to appear).

[15] S. Kurz, *Integral point sets over finite fields*, (submitted).

[16] S. Kurz, *Konstruktion und Eigenschaften ganzzahliger Punktmengen*, Ph.D. thesis, Bayreuth. Math. Schr. 76. Universität Bayreuth, 2006.

[17] S. Kurz, *On the characteristic of integral point sets in* $\mathbb{E}^m$, Australas. J. Combin. **36** (2006), 241–248.

[18] S. Kurz, *On the generation of heronian triangles*, (submitted).

[19] S. Kurz and R. Laue, *Upper bounds for integral point sets*, Australas. J. Combin. **39** (2007), 233–240.

[20] S. Kurz and A. Wassermann, *On the minimum diameter of plane integral point sets*, Ars Combin. (to appear).

[21] S. Niskanen and P. R. J. Östergård, *Cliquer user's guide, version 1.0*, Tech. Report T48, Communications Laboratory, Helsinki University of Technology, Espoo, Finland, 2003.

[22] L. C. Noll and D. I. Bell, n-*clusters for* $1 < n < 7$, Math. Comp. **53** (1989), no. 187, 439–444.

Chapter 11

Bounds for the minimum oriented diameter

SASCHA KURZ[1] AND MARTIN LÄTSCH[2]

ABSTRACT. We consider the problem of finding an orientation with minimum diameter of a connected and bridgeless graph. Fomin et. al. [7] discovered a relation between the minimum oriented diameter and the size of a minimal dominating set. We improve their upper bound.

2000 MSC: 05C12;05C20,05C69.
Key words and phrases: diameter, orientation, domination.

1 Introduction

An orientation of an undirected graph G is a directed graph whose arcs correspond to assignments of directions to the edges of G. An orientation H of G is strongly connected if every two vertices in H are mutually reachable in H. An edge e in an undirected connected graph G is called a bridge if $G-e$ is not connected. A connected graph G is bridgeless if $G-e$ is connected for every edge e, i. e. there is no bridge in G.

Conditions when an undirected graph G admits a strongly connected orientation were determined by Robbins in 1939 [25]. Necessary and sufficient conditions are that G is connected and bridgeless. Chung et. al provided a linear-time algorithm for testing whether a graph has a strong orientation and finding one if it does [1].

Definition 1.1 *Let $\vec{G}$ be a strongly connected directed graph. By $\mathrm{diam}\left(\vec{G}\right)$ we denote the diameter of $\vec{G}$. For a simple connected graph G without bridges we define* $\overrightarrow{\mathrm{diam}}_{\min}(G) :=$

$$\min\left\{\mathrm{diam}\left(\vec{G}\right) : \vec{G} \text{ is an orientation of } G\right\},$$

which we call the minimum oriented diameter of a simple graph G. By $\gamma(G)$ we denote the smallest cardinality of a dominating set of G.

[1]Sascha Kurz, Fakultät für Mathematik, Physik und Informatik, Universität Bayreuth, Germany.
E-mail adress: sascha.kurz@uni-bayreuth.de
[2]Martin Lätsch, Zentrum für Angewandte Informatik, Universität zu Köln, Germany.
E-mail adress: laetsch@zpr.uni-koeln.de

We are interested in graphs G which have a large minimum oriented diameter $\overrightarrow{\mathrm{diam}}_{\min}(G)$ in dependence of its domination number $\gamma(G)$. Therefore we set

$$\Xi(\gamma) := \max\left\{\overrightarrow{\mathrm{diam}}_{\min}(G) : \gamma(G) \leqslant \gamma \text{ for } G \text{ being a connected and bridgeless graph}\right\}.$$

The aim of this note is to prove a better upper bound on $\Xi(\gamma)$. The previously best known result [7] was:

Theorem 1.2

$$\Xi(\gamma) \leqslant 5\gamma - 1.$$

Our main results are

Theorem 1.3

$$\Xi(\gamma) \leqslant 4\gamma$$

and

Conjecture 1.4

$$\Xi(\gamma) = \left\lceil \frac{7\gamma(G)+1}{2} \right\rceil.$$

Clearly we have that $\Xi(\gamma)$ is weak monotone increasing. At first we observe that we have $\Xi(\gamma) \geqslant \left\lceil \frac{7\gamma(G)+1}{2} \right\rceil$. Therefore we consider the following set of examples, where we have depicted the vertices of a possible minimal dominating set by a filled black circle, see Figure 1.

If we formalize this construction of graphs G, which is depicted for $\gamma(G) = \gamma = 1,2,3,4$ we obtain examples which attain the proposed upper bound $\left\lceil \frac{7\gamma(G)+1}{2} \right\rceil$ for all $\gamma \in \mathbb{N}$. In the following we always depict vertices in a given dominating set by a filled circle.

1.1 Related results

Instead of an upper bound of $\overrightarrow{\mathrm{diam}}_{\min}(G)$ in dependence of $\gamma(G)$ one is also interested in an upper bound in dependence of the diameter $\mathrm{diam}(G)$. Here the best known result is given by [2]:

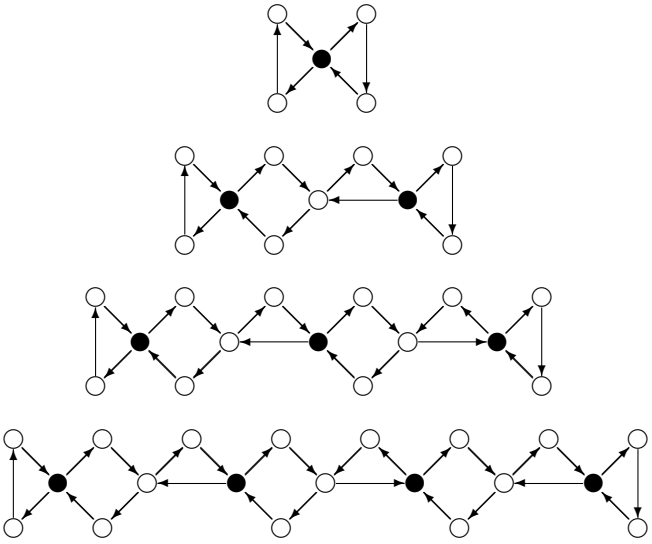


Figure 1: Examples with large minimum oriented diameter in dependence of the domination number $\gamma(G)$.

Theorem 1.5 *(Chvátal and Thomassen, 1978) Let* $f(d)$ *denote the best upper bound on* $\overrightarrow{diam}_{min}(G)$ *where* $d = diam(G)$ *and* G *is connected and bridgeless.*
If G *is a connected and bridgeless graph then we have*

$$f(d) \geqslant \frac{1}{2} \cdot diam(G)^2 + diam(G) \text{ and}$$
$$f(d) \leqslant 2 \cdot diam(G) \cdot (diam(G) + 1).$$

In [2] it was also shown that we have $f(2) = 6$. Examples achieving this upper bound are given by the Petersen graph and by the graph obtained from K_4 by subdividing the three edges incident to one vertex. Recently in [21] $9 \leqslant f(3) \leqslant 11$ was shown.

The oriented diameter is trivially at least the diameter. Graphs where equality holds are said to be *tight*. In [15] some Cartesian products of graphs are shown to be tight. For $n \geqslant 4$ the n-cubes are tight [22]. The discrete tori $C_n \times C_m$ which are tight are completely determined in [20].

The origin of this problem goes back to 1938, where Robbins [25] proves that a graph G has a strongly connected orientation if and only if G has no cut-edge. As an application one might think of making streets of a city one-way or building a communication network with links that are reliable only in one direction.

There is a huge literature on the minimum oriented diameter for special graph classes, see e. g. [11, 12, 13, 14, 16, 17, 18, 19, 23].

From the algorithmic point of view the following result is known [2]:

Theorem 1.6 *The problem whether* $\overrightarrow{diam}_{min}(G) \leqslant 2$ *is* $\mathcal{NP}$*-hard for a given graph* G.

We remark that the proof is based on a transformation to the problem whether a hypergraph of rank 3 is two-colorable.

2 Preliminaries

A vertex set $D \subseteq V(G)$ of a graph G is said to be a dominating set of G if for every vertex $u \in V(G)\backslash D$ there is a vertex $v \in D$ such that $\{u, v\} \in E(G)$. The minimum cardinality of a dominating set of a graph G is denoted by $\gamma(G)$. If P is a path we denote by $|P|$ its length which equals the number of its edges. An elementary cycle C of a graph $G = (V, E)$ is a list $[v_0, \ldots, v_k]$ of vertices in V, where $v_0 = v_k$, $|\{v_0, \ldots, v_{k-1}\}| = k$ and $\{v_i, v_{i+1}\} \in E$ for $0 \leqslant i < k$. Similarly $|C|$ denotes the length of C which equals the number of its edges and vertices. For other not explicitly mention graph-theoretic terminology we refer the reader to [6] for the basic definitions.

Our strategy to prove bounds on $\Xi(\gamma)$ is to apply some transformations on connected and bridgeless graphs attaining $\Xi(\gamma)$ to obtain some structural results. Instead of considering graphs G from now on we will always consider pairs (G, D), where D is a dominating set of G.

Definition 2.1 *For a graph* G *and a dominating set* D *of* G *we call* $\{u, v\} \subseteq V(G)\backslash D$ *an isolated triangle if there exists an* $w \in D$ *such that all neighbors of* u *and* v *are contained in* $\{u, v, w\}$ *and* $\{u, v\} \in E(G)$. *We say that the isolated triangle is associated with* $w \in D$.

Definition 2.2 *A pair* (G, D) *is in first standard form if*

(1) $G = (V, E)$ *is a simple connected graph without a bridge,*

(2) D *is a dominating set of* G *with* $|D| = \gamma(G)$,

(3) for $u, v \in D$ *we have* $\{u, v\} \notin E$,

(4) for each $u \in V\backslash D$ *there exists exactly one* $v \in D$ *with* $\{u, v\} \in E$,

(5) G *is edge-minimal, meaning one can not delete an edge in* G *without creating a bridge, destroying the connectivity or destroying the property of* D *being a dominating set, and*

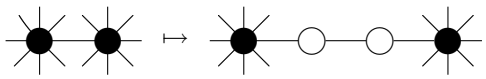
(6) for $|D| = \gamma(G) \geqslant 2$ *every vertex in* D *is associated with exactly one isolated triangle and for* $|D| = \gamma(G) = 1$ *the vertex in* D *is associated with exactly two isolated triangles.*

Lemma 2.3

$$\Xi(\gamma) = \max\Big\{ \overrightarrow{diam}_{min}(G) : |D| \leqslant \gamma, \\ (G, D) \text{ is in first standard form} \Big\}.$$

PROOF. For a given $\gamma \in \mathbb{N}$ we start with a connected and bridgeless graph G' attaining $\Xi(\gamma) = \overrightarrow{diam}_{min}(G')$ and minimum domination number $\gamma(G')$. Let D' be an arbitrary dominating set of G' fulfilling $|D'| = \gamma(G')$. Our aim is to apply some graph transformations onto (G', D') to obtain a pair (G, D) in first standard form fulfilling $\overrightarrow{diam}_{min}(G) \geqslant \overrightarrow{diam}_{min}(G')$ and $|D| \leqslant |D'|$.

At the start conditions (1) and (2) are fulfilled. If there is an edge e between two nodes of D then we recursively apply the following graph transformation until there exists no such edge:



If there exists a node $v \in V\backslash D$ with at least $r \geqslant 2$ neighbors $d_1, \ldots, d_r$ in D then we replace the edge (v, d_i) $i = 2, \ldots, r$ with a path of length 2. We iterate this until case (4) is fulfilled. In Figure 2 we have depicted the graph transformation for $r = 2, 3$.

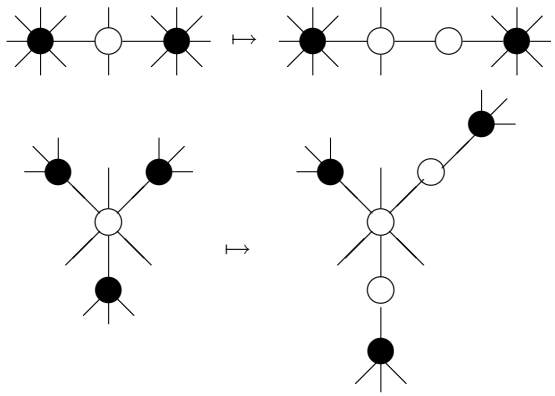


Figure 2: Graph transformation to fulfill condition (4) of Definition 2.2

So after a finite number of transformation we have constructed a pair (G, D) which fulfills conditions (1), (3), (4) of the first standard form where D is a dominating set of G and (G, D) also fulfills

$$\gamma(G) \leqslant |D| \leqslant |D'| = \gamma(G')$$

and

$$\infty > \overrightarrow{diam}_{min}(G) \geqslant \overrightarrow{diam}_{min}(G').$$

To additionally fulfill condition (5) of the first standard form we only need to delete the controversial edges. If $\gamma(G) < |D| \leqslant \gamma(G')$ we would have a contradiction to the minimality of $\gamma(G')$. Since adding isolated triangles does not contradict with the other properties and also does not decrease the minimum oriented property we can assume that every vertex of D is associated with enough isolated triangles. For two vertices x and y in two different isolated triangles being associated with the same vertex $w \in D$ we have $d(x, y) \leqslant 4$ in every strongly connected orientation. Thus we can delete some isolated triangles to achieve the stated number of isolated triangles for every vertex in the dominating set D. Finally we have a pair (G, D) in first standard form. □

So in order to prove bounds on $\Xi(\gamma)$ we can restrict ourselves on pairs (G, D) in first standard form. Due to Theorem 1.2 we can assume $\gamma(G) = |D| \geqslant 2$ both for the proof of Theorem 1.3 and also for Conjecture 1.4.

Corollary 2.4 *If (G, D) is a pair in first standard form then we have*

(i) for all $u, v \in D$ the distance fulfills $d(u, v) \geqslant 3$ and

(ii) for all $u \in V(G)\backslash D$ there exists exactly one $f(u) \in D$ with $\{u, f(u)\} \in E(G)$.

Let G be a connected and bridgeless undirected graph, D be a dominating set of G and H be a strongly connected orientation of G. By $diam_i(H, D)$ we denote

$$\max\left\{d_H(u, v) : \left|\{u, v\} \cap (V(H)\backslash D)\right| = i\right\}.$$

Clearly we have $diam(H) =$
$\max\left\{diam_0(H, D), diam_1(H, D), diam_2(H, D)\right\}$.
Now we refine a lemma from [7]:

Lemma 2.5 *Let G' and G be connected and bridgeless graphs such that G is a subgraph of G' and D is a dominating set of both G' and G. Then for every strongly connected orientation H of G there is an orientation H' of G' such that* $diam(H') \leqslant$

$$\max\left\{diam_0(H, D)+4, diam_1(H, D)+2, diam_2(H, D)\right\}.$$

PROOF. (We rephrase most of the proof from [7].) We adopt the direction of the edges from H to H'. For the remaining edges we consider connected components Q of $G'\backslash V(G)$ and direct some edges having ends in Q as follows.

If Q consists of one vertex x then x is adjacent to at least one vertex u in D and to another vertex $v \neq u$ (the graph G is bridgeless and D is a dominating set). If also v is an element of D then we direct one edge from x and the second edge towards x. Otherwise v is in $V\backslash D$. In this case we direct the edges $[x, u]$ and $[v, x]$ in the same direction as the edge $[f(v), v]$. If there are more edges incident with x (in both cases) we direct them arbitrarily. Then, we have assured the existence of vertices $u', v' \in D$ such that $d_{H'}(x, v') \leqslant 2$ and $d_{H'}(u', x) \leqslant 2$.

Suppose that there are at least two vertices in the connected component Q. Choose a spanning tree T in this component rooted in a vertex v. We orient edges of this tree as follows: If a vertex x of the tree has odd distance from v, then we orient all the tree edges adjacent to x from x outwards. Also, for every such vertex x we orient the edges between x and $V(G)$ towards x if the distance from v on the tree is even, and towards $V(G)$ otherwise, see Figure 1 in [7]. The rest of the edges in the connected component Q are oriented arbitrarily.

In such an orientation H', for every vertex $x \in Q$ there are vertices $u, v \in D$ such that $d_{H'}(x, v) \leqslant 2$ and $d_{H'}(u, x) \leqslant 2$. Therefore, for every $x, y \in V(G')$ the distance between x and y in H' is at most $\max\left\{diam_{2-i}(H, D) + 2i \mid 0 \leqslant i \leqslant 2\right\}$. □

Due to the isolated triangles being associated with the vertices of the dominating set D, for every pair (G, D) in first standard form, there exists an orientation H of G such that

$$\overrightarrow{diam}_{min}(G) = diam(H) = \max\left\{diam_0(H,D)+4, diam_1(H,D)+2, diam_2(H,D)\right\}. \quad (1)$$

If we say that H is an optimal or a minimal orientation of (G, D) we mean an orientation that fulfills Equation (1).

In [7] the authors have described a nice construction to obtain such a subgraph G for a given connected and bridgeless graph G' fulfilling $|V(G)| \leqslant 5 \cdot \gamma(G') - 4$:

For $\gamma(G') = 1$ we may simply choose the single vertex in D as our subgraph D. Now we assume $|D| = \gamma(G') \geqslant 2$. Iteratively, we construct a tree T_k for $k = 1, \ldots, |D|$. The tree T_1 is composed by one vertex x_1 in D. To construct T_{k+1} from T_k we find a vertex x_{k+1} in $D \backslash V(T_k)$ with minimum distance to T_k. The tree T_{k+1} is the union of T_k with a shortest path from x_{k+1} to T_k. Since D is a dominating set this path has length at most 3. We say that the edges of this path are *associated* with x_{k+1}. At the last step we obtain a dominating tree T with $D \subseteq T$ and with $|V(T)| \leqslant 2(|D| - 1) + |D|$.

In order to transform T in a connected and bridgeless graph we construct a sequence of subgraphs G_k for $k = 1, \ldots, |D|$. We say that $x_j \in D$ is *fixed* in G_k if no edge associated with x_j is a bridge in G_k. Notice that x_1 is fixed in T because it does not have any associated edge.

We set $G_1 = T$. Assume we have constructed the subgraph G_k. If x_{k+1} is already fixed in G_k we set $G_{k+1} = G_k$. If x_{k+1} is not fixed in G_k we add a subgraph M to G_k to obtain G_{k+1}.

Let P_k be the path added to T_k to obtain T_{k+1}. We only consider the case where P_k has length three. The other cases can be done similarly. Let us assume that P_k is given by $P_k = (x_{k+1}, u, v, x_j)$ with $u, v \notin D$, and $x_j \in D$, $j \leqslant k$. Moreover let us denote the edges of P_k by e, e' and e''. If we remove all edges e, e', e'' of P_k from T we obtain four subtrees T^1, T^2, T^3 and T^4 containing x_{k+1}, u, v and x_j, respectively.

Among all shortest paths in $G' \backslash e$ connecting T^1 with $T^2 \cup T^3 \cup T^4$ we select P as one whose last vertex belongs to T^i with i maximum. Among all shortest paths in $G' \backslash e''$ connecting T^4 with $T^1 \cup T^2 \cup T^3$ we select Q as one whose first vertex belongs to T^i with i minimum. Let R be any shortest path in $G' \backslash e'$ connecting $T^3 \cup T^4$ with $T^1 \cup T^2$.

Since G' is a connected and bridgeless graph the paths P, Q, R exist. Since $D \subseteq V(T)$ and the set D is a dominating set, the length of paths P, Q and R is at most 3. Moreover, if the length of P is three its end vertices belong to D. The same holds for the paths Q and R.

The definition of M is given according to the following cases. If the last vertex of P belongs to T^4 we define $M = P$. If the last vertex of P belongs to T^3 or it belongs to T^2 and the first vertex of Q belongs to T^2 we define $M = P \cup Q$. If none of the previous cases hold the first vertex of R belongs to T^2 and the last one belongs to T^3. We define $M = P \cup Q \cup R$.

For the analysis that $\left|V\left(G_{|D|}\right)\right| \leqslant 5 \cdot \gamma(G') - 4$ we refer to [7].

Since a shortest path does contain every vertex at most once, we can combine the above described construction of a subgraph with Lemma 2.5 to obtain the bound $\Xi(\gamma) \leqslant 5\gamma - 1$.

Lemma 2.6

$$\Xi(1) = 4 \text{ and } \Xi(2) = 8.$$

PROOF. At first we observe that the examples from Figure 1 give $\Xi(1) \geqslant 4$ and $\Xi(2) \geqslant 8$. For the other direction let (G, D) be a pair in first standard form attaining $\overrightarrow{diam}_{min}(G) = \Xi(\gamma(G))$. For $\gamma = \gamma(G) = 1$ we have $|D| = 1$, choose the single vertex of D as a subgraph and apply Lemma 2.5. Going through the cases of the above described subgraph construction for $\gamma = \gamma(G) = 2$ we obtain up to symmetry the two possibilities given in Figure 3. By H we denote the depicted corresponding orientation of the edges. Since in both cases we have $diam_0(H, D) \leqslant 4$ and $diam_1(H, D), diam_2(H, D) \leqslant 5$ we can apply Lemma 2.5 to obtain the stated result. □

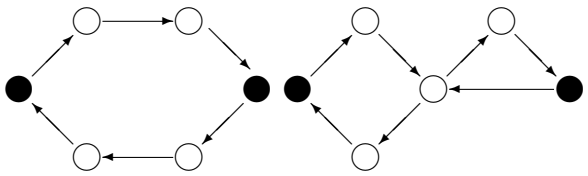


Figure 3: The two possible subgraphs for $\gamma(G) = 2$.

With Lemma 2.5 in mind we would like to restrict our investigations on connected and bridgeless subgraphs containing the dominating set.

Definition 2.7 *For a pair (G', D) in first standard form we call G a minimal subgraph of (G', D), if*

(1) G is a subgraph of G' containing the vertex set D,

(2) G is connected and bridgeless,

(3) for every vertex $v \in V(G) \backslash D$ we have $\{v, f(v)\} \in E(G)$, where $f : V(G') \backslash D \to D$ is the function from the first standard form of (G', D), and

(4) G is vertex and edge-minimal with respect to properties (1), (2), and (3).

Corollary 2.8 *If G is a minimal subgraph of (G', D) in first standard form, we have*

(1) $|V(G)| \leqslant 5 \cdot |D| - 4$ and

(2) there exists no chord $\{u, v\} \in E(G)$, where $\{u, v\} \cap D$ is an empty set.

Definition 2.9 *Let* G *be a minimal subgraph of* (G', D) *in first standard form. We construct a graph* $\widetilde{G}$ *from* G *by adding isolated triangles at vertices of* D *such that* $(\widetilde{G}, D)$ *is in first standard form. We call* $\widetilde{G}$ *a minimal completion and we say that* H *is a minimal or an optimal orientation of* G*, if* H *is strongly connected and we have*

$$\overrightarrow{diam}_{min}(\widetilde{G}) \geqslant \max\left\{diam_0(H,D)+4, diam_1(H,D)+2, diam_2(H,D)\right\}.$$

By considering the isolated triangles being associated to the vertices of the dominating set D we can easily check, that every minimal subgraph G of a pair (G', D) in first standard form admits a minimal orientation H and that we have $\overrightarrow{diam}_{min}(G') \leqslant \overrightarrow{diam}_{min}\left(\widetilde{G}\right)$. If G does only fulfill conditions (1)-(3) of Definition 2.7 then we may consider a minimal subgraph G'' of (G', D), which contains G as a subgraph. With this we can call an orientation H of G minimal or optimal if it is minimal or optimal for G''.

Definition 2.10 *We call a pair* (G', D) *in first standard form critical, if* $\Xi(\gamma(G')) = \overrightarrow{diam}_{min}(G')$.

Definition 2.11 *We call a minimal subgraph* G *of* (G', D) *in first standard form critical if for a minimal orientation* H *of* G *we have*

$$\Xi(\gamma(G')) = \max\left\{diam_0(H,D)+4, diam_1(H,D)+2, diam_2(H,D)\right\}.$$

Together with Lemma 2.5 we obtain:

Lemma 2.12

$$\Xi(\gamma) = \max\Big\{\min\Big\{\max\{diam_0(H,D)+4, diam_1(H,D)+2, diam_2(H,D)\} : H \text{ is strongly connected orientation of } G\Big\} : G \text{ is critical minimal subgraph of } (G', D) \text{ in first standard form with } |D| = \gamma\Big\}.$$

Sometimes it is useful to know some basic facts about strongly connected orientations of graphs.

Lemma 2.13 *(1) If* H *is a strongly connected orientation of an undirected graph* G *and* C *is a directed cycle without repeated edges in* H*, then inverting of the edges of* C *yields another strongly connected orientation of* G.

(2) If H *is a strongly connected orientation of an undirected graph* G *and* P_1 *and* P_2 *are two edge-disjoint directed paths from* x *to* y*, then inverting* P_2 *yields another strongly connected orientation of* G.

(3) If H *is a strongly connected orientation of an undirected graph* G *then inverting all edges yields another strongly connected orientation with equal diameter.*

3 Reductions

In this section we will propose some reductions for critical minimal subgraphs G of pairs (G', D) in first standard form, in order to provide some tools for an inductive proof of a better upper bound on $\Xi(\gamma)$.

Lemma 3.1 *Let* G *be a critical minimal subgraph of* (G', D) *in first standard form with* $\gamma = \gamma(G') = |D| \geqslant 3$*. If* G *contains vertices* $x, y \in D$*,* $l_1, l_2, r_1, r_2 \in V(G)\backslash D$*, two edge disjoint paths* $P_1 = [x, l_1, r_1, y]$*,* $P_2 = [x, l_2, r_2, y]$*, all neighbors of* l_1, r_1 *are in* $\{x, l_1, r_1, y\}$*, and all neighbors of* l_2, r_2 *are in* $\{x, l_2, r_2, y\}$*, then we have* $\Xi(\gamma) \leqslant \Xi(\gamma - 1) + 3$.

PROOF. Let $\widetilde{G}$ be the graph which arises from G by deleting l_1, l_2, r_1, r_2 and identifying x with y. Now let $\widetilde{D} := D\backslash\{y\}$ and $\widetilde{H}$ be an arbitrary minimal orientation of $\widetilde{G}$. Thus we have $diam_0\left(\widetilde{H}, \widetilde{D}\right) \leqslant \Xi(\gamma - 1) + 4$, $diam_1\left(\widetilde{H}, \widetilde{D}\right) \leqslant \Xi(\gamma-1)+2$, and $diam_2\left(\widetilde{H}, \widetilde{D}\right) \leqslant \Xi(\gamma-1)$. We construct an orientation H of G by directing the two paths P_1 and P_2 in opposing directions, and by taking the directions from $\widetilde{H}$. Now we analyze the distance $d_H(u, v)$ in H for all pairs $u, v \in V(G)$. If both u and v are in $\{l_1, l_2, r_1, r_2\}$, then we have $d_H(u, v) \leqslant 5 \leqslant \Xi(\gamma - 1) + 3$. If none of u and v is in $\{l_1, l_2, r_1, r_2\}$, then we have $d_H(u, v) \leqslant d_{\widetilde{H}}(u, v) + 3$. In the remaining case we have $d_H(u, v) \leqslant d_{\widetilde{H}}(u, v)+5$. Thus we have

$$\begin{aligned} diam_2(H, D) &\leqslant \max\left\{diam_2\left(\widetilde{H},\widetilde{D}\right)+3, diam_1\left(\widetilde{H},\widetilde{D}\right)+5, 5\right\} \\ &\leqslant \Xi(\gamma - 1) + 3, \\ diam_1(H, D) &\leqslant \max\left\{diam_1\left(\widetilde{H},\widetilde{D}\right)+3, diam_0\left(\widetilde{H},\widetilde{D}\right)+5, 5\right\} \\ &\leqslant \Xi(\gamma - 1) + 1, \text{ and} \\ diam_0(H, D) &\leqslant diam_0\left(\widetilde{H},\widetilde{D}\right)+3 \\ &\leqslant \Xi(\gamma - 1) - 1, \end{aligned}$$

which yields $\Xi(\gamma) \leqslant \Xi(\gamma - 1) + 3$. □

We remark that Lemma 3.1 corresponds to a graph containing the left graph of Figure 3 as an induced subgraph, where the vertices corresponding to the empty circles have no further neighbors in the whole graph.

Lemma 3.2 *Let* G *be a critical minimal subgraph of* (G', D) *in first standard form with* $\gamma = \gamma(G') = |D| \geqslant 3$*. If* G *contains vertices* $x, y, z \in D$*, four edge disjoint paths* $P_1 = [x, v_1, v_2, v_3, y]$*,* $P_2 = [y, v_4, v_5, v_6, z]$*,* $P_3 = [x, u_1, u_2, y]$*,* $P_4 = [y, u_3, u_4, z]$*, and all edges being adjacent to vertices in* $I := \{v_1, v_2, v_3, v_4, v_5, v_6, u_1, u_2, u_3, u_4\}$ *are contained in* $P := P_1 \cup P_2 \cup P_3 \cup P_4$*, then we have* $\Xi(\gamma) \leqslant \Xi(\gamma - 2) + 7$.

PROOF. At first we want to determine some structure information on the vertices v_i, u_j and the adjacent edges. We have $f(v_1) = f(u_1) = x$, $f(v_3) = f(v_4) = f(u_2) = f(u_3) = y$, and $f(v_6) = f(u_4) = z$. Since all edges

being adjacent to vertices in I are contained in P we have $f(v_2), f(v_5) \in \{x, y, z\}$. Some vertices may have several labels. By $v_i \sim$ we denote the set of labels which correspond to the same vertex as v_i. Similarly we define $u_i \sim$.

Let us at first assume $|I| = 10$, meaning, that each vertex has a unique label. In this case we may consider the edge $\{v_2, f(v_2)\}$ which is not contained in P to see that G would not be a minimal subgraph of (G', D) in first standard form.

Due to the 14 pairwise different edges of P and the information on the values of f we have

(a) $v_1 \sim \subseteq \{v_1, v_5\}$, $v_3 \sim \subseteq \{v_3, v_5\}$, $v_4 \sim \subseteq \{v_2, v_4\}$, $v_6 \sim \subseteq \{v_2, v_6\}$,

(b) $u_1 \sim \subseteq \{u_1, v_2, v_5\}$, $u_2 \sim \subseteq \{u_2, v_2, v_5\}$, $u_3 \sim \subseteq \{u_3, v_2, v_5\}$, $u_4 \sim \subseteq \{u_4, v_2, v_5\}$,

(c) $v_2 \sim \subseteq \{v_2, v_4, v_5, v_6, u_1, u_2, u_3, u_4\}$, $v_5 \sim \subseteq \{v_1, v_2, v_3, v_5, u_1, u_2, u_3, u_4\}$.

Next we assume $|I| = 9$ which means that exactly one vertex in I has two different labels and all other vertices have unique labels.

(1) If $v_1 = v_5$ then v_2, v_3, and v_4 could be deleted.

(2) If $v_3 = v_5$ then v_4 could be deleted.

(3) If $u_1 = v_2$ then by considering the edge $\{v_5, f(v_5)\} \notin P$ we could conclude that either v_4 or v_6 could be deleted.

(4) If $u_1 = v_5$ then u_2 could be deleted.

(5) If $u_2 = v_2$ then by considering the edge $\{v_5, f(v_5)\} \notin P$ we could conclude that either v_4 or v_6 could be deleted.

(6) If $u_2 = v_5$ then v_4 could be deleted.

(7) If $v_2 = v_5$ then v_3 and v_4 could be deleted.

Thus the vertices v_1, v_3, u_1, u_2 are unique. Using symmetry we conclude that also the vertices v_4, v_6, u_3, and u_4 are unique. Since we have also dealt with the only left possibility $v_2 = v_5$ we can conclude $|I| \leqslant 8$.

We proceed similar as in the proof of Lemma 3.1 and let $\widetilde{G}$ be the graph arising from G by deleting the vertices u_i, v_i, y and by identifying x and z. Obviously $\widetilde{G}$ is connected and bridgeless. Now let $\widetilde{D} := D \backslash \{y, z\}$ and $\widetilde{H}$ be an arbitrary minimal orientation of $\widetilde{G}$. Thus we have $\mathrm{diam}_0\left(\widetilde{H}, \widetilde{D}\right) \leqslant \Xi(\gamma - 2) - 4$, $\mathrm{diam}_1\left(\widetilde{H}, \widetilde{D}\right) \leqslant \Xi(\gamma - 2) - 2$, and $\mathrm{diam}_2\left(\widetilde{H}, \widetilde{D}\right) \leqslant \Xi(\gamma - 2)$.

We construct an orientation H of G by directing the two pairs of paths (P_1, P_3), (P_2, P_4) in opposing directions such that the arcs $[v_3, y]$, $[y, v_4]$ are directed different, by taking the directions from $\widetilde{H}$ and by directing remaining edges arbitrarily.

Now we analyze the distance $d_H(u, v)$ in H for all pairs $u, v \in V(G)$. Due to $d_H(x, z), d_H(z, x) \leqslant 7$, $d_H(y, x), d_H(y, z), d_H(x, y), d_H(z, y) \leqslant 4$ we have $d_H(u, v) \leqslant d_{\widetilde{H}}(u, v) + 7$ for $u, v \notin I$. Now we consider $d_H(u, v)$ for $u, v \in I \cup \{x, y, z\}$. Due to $L := |I \cup \{x, y, z\}| \leqslant 11$ we clearly have $d_H(u, v) \leqslant 10$. We assume $L = 11$ since otherwise we would have $d_H(u, v) \leqslant 9$. Now we have a closer look at the directed cycle $C := P_1 \circ P_4 \circ P_2 \circ P_3$ of length 14 consisting of 11 vertices. It is not possible to visit all 11 vertices going along edges of the cycle C without visiting a vertex twice. Thus we have $d_H(u, v) \leqslant 9$ for $u, v \in I \cup \{x, y, z\}$. Summarizing our results gives

$$\begin{aligned}
\mathrm{diam}_2(H, D) &\leqslant \max\left\{\mathrm{diam}_2\left(\widetilde{H}, \widetilde{D}\right) + 7, \mathrm{diam}_1\left(\widetilde{H}, \widetilde{D}\right) + 9, 9\right\} \\
&\leqslant \Xi(\gamma - 2) + 7, \\
\mathrm{diam}_1(H, D) &\leqslant \max\left\{\mathrm{diam}_1\left(\widetilde{H}, \widetilde{D}\right) + 7, \mathrm{diam}_0\left(\widetilde{H}, \widetilde{D}\right) + 9, 9\right\} \\
&\leqslant \Xi(\gamma - 2) + 5, \text{ and} \\
\mathrm{diam}_0(H, D) &\leqslant \mathrm{diam}_0\left(\widetilde{H}, \widetilde{D}\right) + 7 \\
&\leqslant \Xi(\gamma - 2) + 3,
\end{aligned}$$

which yields $\Xi(\gamma) \leqslant \Xi(\gamma - 2) + 7$. □

We remark that Lemma 3.2 corresponds to a graph containing the right graph of Figure 3 two times as an induced subgraph for $x, y, z \in D$ corresponding to the black circle, where the vertices corresponding to the empty circles have no further neighbors in the whole graph.

Lemma 3.3 *Let G be a critical minimal subgraph of (G', D) in first standard form with $\gamma = \gamma(G') = |D| \geqslant 3$ and x a vertex contained in the dominating set D. If removing x produces at least three connectivity components C_1, C_2, C_3, ..., then we have*

$$\Xi(\gamma) \leqslant \max\left\{\Xi(\gamma - i) + \Xi(i) - 4 : 1 \leqslant i \leqslant \gamma - 1\right\}.$$

PROOF. Let $\widetilde{C}_i$ be the induced subgraphs of $V(C_i) \cup \{x\}$ in G. We set $D_i = \{x\} \cup (V(C_i) \cap D)$ and $\gamma_i := |D_i| - 1$ so that we have $1 + \sum_i \gamma_i = \gamma$. Since G is a minimal subgraph we have $\gamma_i \geqslant 1$ for all i. Now we choose arbitrary minimal orientations $\widetilde{H}_i$ of the $\widetilde{C}_i$. Thus we have $\mathrm{diam}_0\left(\widetilde{H}_i, D_i\right) \leqslant \Xi(\gamma_i + 1) - 4$, $\mathrm{diam}_1\left(\widetilde{H}_i, D_i\right) \leqslant \Xi(\gamma_i + 1) - 2$, and $\mathrm{diam}_2\left(\widetilde{H}_i, D_i\right) \leqslant \Xi(\gamma_i + 1)$ for all i. Since $\widetilde{C}_i$ and $\widetilde{C}_j$ are edge-disjoint for $i \neq j$ we can construct an orientation H of G by taking the directions of the $\widetilde{H}_i$. Now we analyze the distance $d_H(u, v)$ in H for all pairs $u, v \in V(G)$. If u and v are contained in the same component $\widetilde{C}_i$ we have $d_H(u, v) = d_{\widetilde{H}_i}(u, v)$. If u is contained in $\widetilde{C}_i$ and v is contained in $\widetilde{C}_j$, then we have $d_H(u, v) \leqslant d_{\widetilde{H}_i}(u, x) + d_{\widetilde{H}_j}(x, v)$. Thus we have

$$\begin{aligned}
&\mathrm{diam}_2(H, D) \\
&\leqslant \max_{i \neq j}\left\{\mathrm{diam}_2\left(\widetilde{H}_i, D_i\right), \mathrm{diam}_1\left(\widetilde{H}_i, D_i\right) + \mathrm{diam}_1\left(\widetilde{H}_j, D_j\right)\right\} \\
&\leqslant \max\left\{\Xi(\gamma_i + 1), \Xi(\gamma_i + 1) + \Xi(\gamma_j + 1) - 4 : i \neq j\right\}, \\
&\mathrm{diam}_1(H, D) \\
&\leqslant \max_{i \neq j}\left\{\mathrm{diam}_1\left(\widetilde{H}_i, D_i\right), \mathrm{diam}_1\left(\widetilde{H}_i, D_i\right) + \mathrm{diam}_0\left(\widetilde{H}_j, D_j\right)\right\} \\
&\leqslant \max_{i \neq j}\left\{\Xi(\gamma_i + 1) - 2, \Xi(\gamma_i + 1) + \Xi(\gamma_j + 1) - 6\right\},
\end{aligned}$$

and $\mathrm{diam}_0(H,D)$

$$\leqslant \max\left\{\mathrm{diam}_0\left(\widetilde{H}_i,D_i\right)+\mathrm{diam}_0\left(\widetilde{H}_j,D_j\right) : i\neq j\right\}$$
$$\leqslant \max\left\{\Xi(\gamma_i+1)+\Xi(\gamma_j+1)-8 : i\neq j\right\}.$$

Since we have at least three connectivity components it holds $\gamma_i+\gamma_j\leqslant\gamma-2$ for all $i\neq j$. Using this and $\Xi(n-1)\leqslant\Xi(n)$ we conclude $\Xi(\gamma)\leqslant\max\left\{\Xi(\gamma-i)+\Xi(i)-4 : 1\leqslant i\leqslant\gamma-1\right\}$. □

Lemma 3.4 *Let G be a critical minimal subgraph of (G',D) in first standard form with $\gamma=\gamma(G')=|D|\geqslant 3$ and x a vertex not contained in the dominating set D. If removing x produces at least three connectivity components $C_1, C_2, C_3, \ldots,$ then $\Xi(\gamma)$ is at most*

$$\max_{2\leqslant i\leqslant\gamma-1}\left\{\Xi(i)+\Xi(\gamma+1-i)-7,\Xi(i-1)+\Xi(\gamma+1-i)-4\right\}.$$

PROOF. W.l.o.g. let $f(x)$ be contained in C_1. Let $\widetilde{C}_1$ be the induced subgraph of $V(C_1)\cup\{x\}$ in G and $D_1=D\cap V(C_1)$. For $i\geqslant 2$ let $\widetilde{C}_i$ be the induced subgraph of $V(C_i)\cup\{x\}$ in G with additional vertices y_i, z_i, additional edges $\{x,y_i\},\{x,z_i\},\{y_i,z_i\}$, and $D_i=(V(C_i)\cap D)\cup\{z_i\}$. We set $\gamma_1=|D_1|\geqslant 1$ and $\gamma_i=|D_i|-1\geqslant 1$ for $i\geqslant 2$ so that we have $\sum_i\gamma_i=\gamma$. By $\widetilde{H}_i$ we denote an optimal orientation of C_i. W.l.o.g. we assume that in $\widetilde{H}_1$ the edge $\{f(x),x\}$ is directed from $f(x)$ to x and that for $i\geqslant 2$ in $\widetilde{H}_i$ the edges $\{x,y_i\}$, $\{x,z_i\}$, $\{y_i,z_i\}$ are directed from x to y_i, from y_i to z_i and from z_i to x. Due to the minimality of the orientations $\widetilde{H}_i$ we have $\mathrm{diam}_0\left(\widetilde{H}_1,D_1\right)\leqslant\Xi(\gamma_1)-4$, $\mathrm{diam}_1\left(\widetilde{H}_1,D_1\right)\leqslant\Xi(\gamma_1)-2$, $\mathrm{diam}_2\left(\widetilde{H}_1,D_1\right)\leqslant\Xi(\gamma_1)$, and for $i\geqslant 2$ we have $\mathrm{diam}_0\left(\widetilde{H}_i,D_i\right)\leqslant\Xi(\gamma_i+1)-4$, $\mathrm{diam}_1\left(\widetilde{H}_i,D_i\right)\leqslant\Xi(\gamma_i+1)-2$, $\mathrm{diam}_2\left(\widetilde{H}_i,D_i\right)\leqslant\Xi(\gamma_i+1)$.

We construct an orientation H of G by taking the directions of the common edges with the $\widetilde{H}_i$. Now we analyze the distance $d_H(u,v)$ in H for all pairs $u,v\in V(G)$. We only have to consider the cases where u and v are in different connectivity components. Let us first assume $u\in\widetilde{C}_i$, $v\in\widetilde{C}_j$ with $i,j\geqslant 2$. We have

$$\begin{aligned} d_H(u,v) &\leqslant d_{\widetilde{H}_i}(u,x)+d_{\widetilde{H}_j}(x,v)\\ &\leqslant d_{\widetilde{H}_i}(u,z_i)-2+d_{\widetilde{H}_j}(z_j,v)-1,\end{aligned}$$

since every directed path from a vertex $u\in V(G)$ to z_i in $\widetilde{H}_i$ uses the arcs $[x,y_i]$, $[y_i,z_i]$, and every directed path from z_j to a vertex $v\in V(G)$ in $\widetilde{H}_j$ uses the arc $[z_j,x]$. Now let u be in $\widetilde{C}_1$ and v be in $\widetilde{C}_i$ with $i\geqslant 2$. Since the edge $\{f(x),x\}$ is directed from $f(x)$ to x, both in H and in $\widetilde{H}_1$, we can conclude

$$\begin{aligned} d_H(u,v) &\leqslant d_{\widetilde{H}_1}(u,x)+d_{\widetilde{H}_i}(x,v)\\ &\leqslant d_{\widetilde{H}_1}(u,f(x))+1+d_{\widetilde{H}_i}(z_i,v)-1.\end{aligned}$$

If $u\in\widetilde{C}_i$ with $i\geqslant 2$ and $v\in\widetilde{C}_1$, then we similarly conclude

$$\begin{aligned} d_H(u,v) &\leqslant d_{\widetilde{H}_i}(u,x)+d_{\widetilde{H}_1}(x,v)\\ &\leqslant d_{\widetilde{H}_i}(u,z_i)-2+d_{\widetilde{H}_1}(x,v).\end{aligned}$$

Thus using $\Xi(i-1)\leqslant\Xi(i)$ for $i\in\mathbb{N}$ and $\gamma_i+\gamma_j\leqslant\gamma-1$ for all $i\neq j$ in total we have $\mathrm{diam}_2(H,D)$

$$\begin{aligned}
&\leqslant \max\Big\{\mathrm{diam}_2\Big(\widetilde{H}_1,D_1\Big),\mathrm{diam}_2\Big(\widetilde{H}_i,D_i\Big),\mathrm{diam}_1\Big(\widetilde{H}_i,D_i\Big)\\
&\quad+\mathrm{diam}_1\Big(\widetilde{H}_j,D_j\Big)-3,\mathrm{diam}_1\Big(\widetilde{H}_1,D_1\Big)+\mathrm{diam}_1\Big(\widetilde{H}_i,D_i\Big),\\
&\quad\mathrm{diam}_2\Big(\widetilde{H}_1,D_1\Big)+\mathrm{diam}_1\Big(\widetilde{H}_i,D_i\Big)-2\Big\}\\
&\leqslant \max\Big\{\Xi(\gamma-1),\Xi(\gamma_i+1)+\Xi(\gamma_j+1)-7,\\
&\quad\Xi(\gamma_1)+\Xi(\gamma_i+1)-4 : 2\leqslant i<j\Big\}\\
&\leqslant \max\Big\{\Xi(i)+\Xi(\gamma+1-i)-7,\\
&\quad\Xi(i-1)+\Xi(\gamma+1-i)-4 : 2\leqslant i\leqslant\gamma-1\Big\},
\end{aligned}$$

$\mathrm{diam}_1(H,D)$

$$\begin{aligned}
&\leqslant \max\Big\{\mathrm{diam}_1\Big(\widetilde{H}_1,D_1\Big),\mathrm{diam}_1\Big(\widetilde{H}_i,D_i\Big),\mathrm{diam}_0\Big(\widetilde{H}_i,D_i\Big)\\
&\quad+\mathrm{diam}_1\Big(\widetilde{H}_j,D_j\Big)-3,\mathrm{diam}_0\Big(\widetilde{H}_1,D_1\Big)+\mathrm{diam}_1\Big(\widetilde{H}_i,D_i\Big),\\
&\quad\mathrm{diam}_1\Big(\widetilde{H}_1,D_1\Big)+\mathrm{diam}_0\Big(\widetilde{H}_i,D_i\Big),\mathrm{diam}_2\Big(\widetilde{H}_1,D_1\Big)+\\
&\quad\mathrm{diam}_0\Big(\widetilde{H}_i,D_i\Big)-2,\mathrm{diam}_1\Big(\widetilde{H}_1,D_1\Big)+\mathrm{diam}_1\Big(\widetilde{H}_i,D_i\Big)-2\Big\}\\
&\leqslant \max\Big\{\Xi(\gamma-1)-2,\Xi(\gamma_i+1)+\Xi(\gamma_j+1)-9,\\
&\quad\Xi(\gamma_1)+\Xi(\gamma_i+1)-6 : 2\leqslant i<j\Big\}\\
&\leqslant \max\Big\{\Xi(i)+\Xi(\gamma+1-i)-9,\Xi(i-1)+\\
&\quad\Xi(\gamma+1-i)-6 : 2\leqslant i\leqslant\gamma-1\Big\},
\end{aligned}$$

and $\mathrm{diam}_0(H,D)$

$$\begin{aligned}
&\leqslant \max\Big\{\mathrm{diam}_0\Big(\widetilde{H}_1,D_1\Big),\mathrm{diam}_0\Big(\widetilde{H}_i,D_i\Big),\mathrm{diam}_0\Big(\widetilde{H}_i,D_i\Big)\\
&\quad+\mathrm{diam}_0\Big(\widetilde{H}_j,D_j\Big)-3,\mathrm{diam}_0\Big(\widetilde{H}_1,D_1\Big)+\mathrm{diam}_0\Big(\widetilde{H}_i,D_i\Big),\\
&\quad\mathrm{diam}_1\Big(\widetilde{H}_1,D_1\Big)+\mathrm{diam}_0\Big(\widetilde{H}_i,D_i\Big)-2\Big\}\\
&\leqslant \max\Big\{\Xi(\gamma-1)-4,\Xi(\gamma_i+1)+\Xi(\gamma_j+1)-11,\\
&\quad\Xi(\gamma_1)+\Xi(\gamma_i+1)-8 : 2\leqslant i<j\Big\}\\
&\leqslant \max\Big\{\Xi(i)+\Xi(\gamma+1-i)-11,\Xi(i-1)+\\
&\quad\Xi(\gamma+1-i)-8 : 2\leqslant i\leqslant\gamma-1\Big\}.
\end{aligned}$$

Thus we can conclude that $\Xi(\gamma)$ is at most $\max_{2\leqslant i\leqslant\gamma-1}\Big\{\Xi(i)+\Xi(\gamma+1-i)-7,\Xi(i-1)+\Xi(\gamma+1-i)-4\Big\}$. □

Now we are ready to determine the next exact value of $\Xi(\gamma)$:

Lemma 3.5

$$\Xi(3)=11.$$

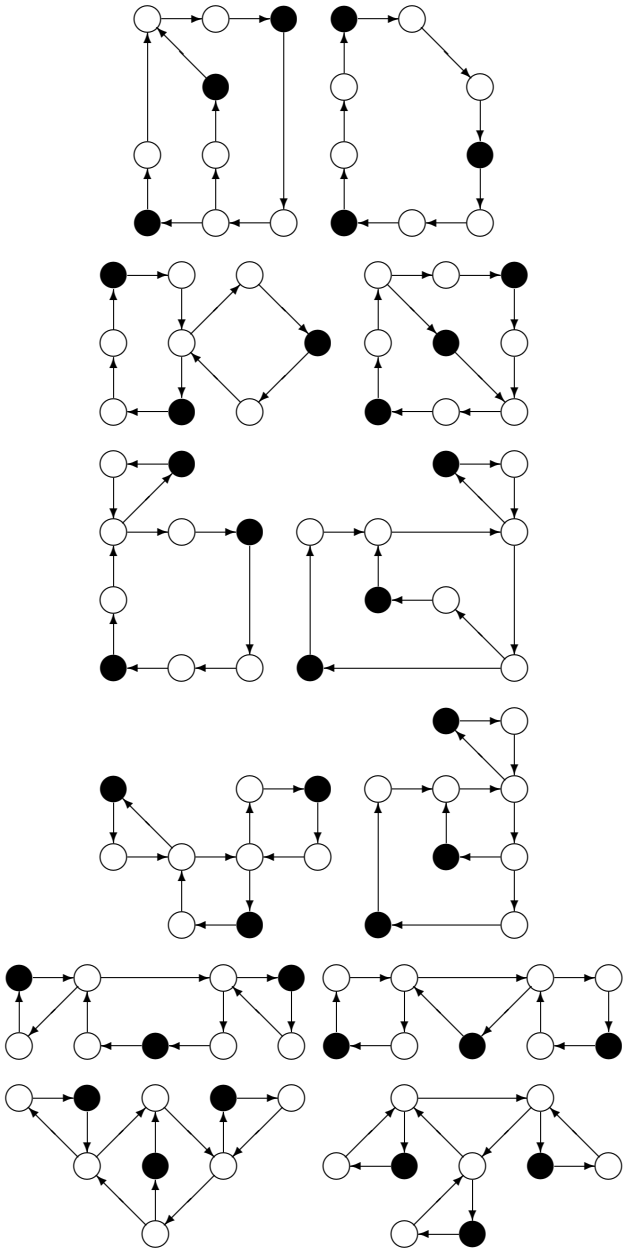


Figure 4: Orientations for the proof of Lemma 3.5 - part 1.

PROOF. The third example from Figure 1 gives $\Xi(3) \geqslant 11$. Going through the cases of the subgraph construction being described in front of Lemma 2.6 we are able to explicitly construct a finite list of possible subgraphs for $\gamma = 3$. This fall differentiation is a bit laborious but not difficult. We can assume that these graphs G are minimal subgraphs of a suitable pair (G', D) in first standard form. During our construction we can drop all graphs which are not minimal, e. g. graphs containing a chord where no end vertex lies in the dominating set D. Doing this we obtain a list of 25 non-isomorphic minimal subgraphs. In Figure 4 and Figure 5 we give suitable orientations for the cases, where we can not apply Lemma 3.1, Lemma 3.2, or Lemma 3.4. □

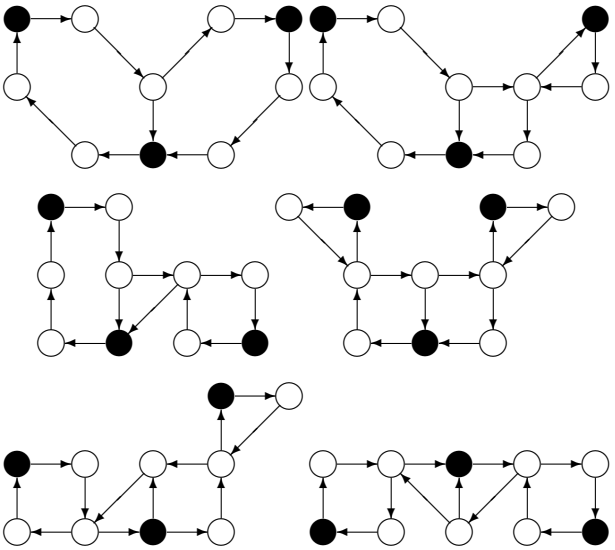


Figure 5: Orientations for the proof of Lemma 3.5 - part 2.

Going over the proofs of the previous lemmas again, we can conclude some further, in some sense weaker, reduction results. Similarly as in Lemma 3.2 we can prove:

Lemma 3.6 *Let G be a critical minimal subgraph of (G', D) in first standard form with $\gamma = \gamma(G') = |D| \geqslant 3$. If G contains vertices $x, y, \in D$, two edge disjoint paths $P_1 = [x, u_1, u_2, u_3, y]$, $P_2 = [x, v_1, v_2, y]$, and all edges being adjacent to vertices in $I := \{u_1, u_2, u_3, v_1, v_2\}$ are contained in $P_1 \cup P_2$, then we have $\Xi(\gamma) \leqslant \Xi(\gamma - 1) + 4$.*

Lemma 3.7 *Let G be a critical minimal subgraph of (G', D) in first standard form with $\gamma = \gamma(G') = |D| \geqslant 3$ and x a vertex contained in the dominating set D. If removing x produces two connectivity components C_1 and C_2 then we have*

$$\Xi(\gamma) \leqslant \max\left\{\Xi(\gamma + 1 - i) + \Xi(i) - 4 \,:\, 2 \leqslant i \leqslant \gamma - 1\right\}.$$

PROOF. We can rephrase most of the proof of Lemma 3.3. Our estimations on $\mathtt{diam}_i(H, D)$ remain valid. Since we only have two connectivity components we do not have $\gamma_i + \gamma_j \leqslant \gamma - 2$ for $i \neq j$. Instead we have $\gamma_1 + \gamma_2 = \gamma - 1$ and $\gamma_1, \gamma_2 \leqslant \gamma - 2$. Combining this with $\Xi(n-1) \leqslant \Xi(n)$ we obtain the stated upper bound. □

Lemma 3.8 *Let G be a critical minimal subgraph of (G', D) in first standard form with $\gamma = \gamma(G') = |D| \geqslant 3$ and x a vertex not contained in the dominating set D. If removing x produces at least two connectivity components C_1, C_2 where $f(x) \in C_1$ and $|V(C_1) \cap D| \geqslant 2$ then we have*

$$\Xi(\gamma) \leqslant \max\left\{\Xi(i) + \Xi(\gamma + 1 - i) - 4 \,:\, 2 \leqslant i \leqslant \gamma - 1\right\}.$$

PROOF. We can rephrase most of the proof of Lemma 3.4. Using $\Xi(i-1) \leqslant \Xi(i)$ for all $i \in \mathbb{N}$ and the fact that we have

exactly two connectivity components C_1 and C_2 yields

$$diam_2(H,D) \leqslant \max\left\{\Xi(\gamma-1), \Xi(\gamma_1)+\Xi(\gamma_2+1)-4\right\}$$
$$diam_1(H,D) \leqslant \max\left\{\Xi(\gamma-1)-2, \Xi(\gamma_1)+\Xi(\gamma_2+1)-6\right\}$$
$$diam_0(H,D) \leqslant \max\left\{\Xi(\gamma-1)-4, \Xi(\gamma_1)+\Xi(\gamma_2+1)-8\right\}.$$

Due to $\Xi(i-1) \leqslant \Xi(i)$, $2 \leqslant y_1 \leqslant \gamma-1$, and $1 \leqslant \gamma_2 \leqslant \gamma-1$ we have

$$\Xi(\gamma) \leqslant \max\left\{\Xi(i)+\Xi(\gamma+1-i)-4 \,:\, 2 \leqslant i \leqslant \gamma-1\right\}.$$

□

We would like to remark that Lemmas 3.1, 3.2, 3.3, 3.4 can be used in an induction proof of Conjecture 1.4, whereas Lemmas 3.6, 3.7, 3.8 can only be used in an induction proof of Theorem 1.3.

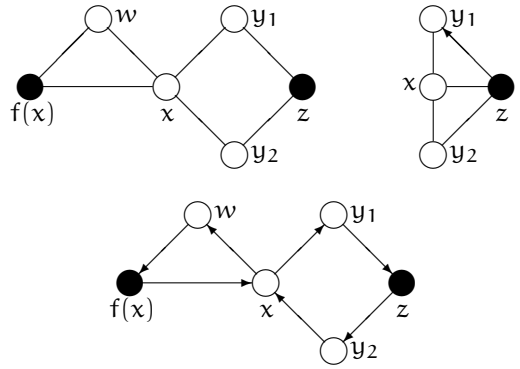


Figure 6: The situation of Lemma 3.9 if we can not apply Lemma 3.8.

In order to prove Theorem 1.3 we need some further reduction lemmas.

Lemma 3.9 *Let* G *be a critical minimal subgraph of* (G', D) *in first standard form with* $\gamma = \gamma(G') = |D| \geqslant 3$ *and* x *a vertex not contained in the dominating set* D. *If removing* x *produces at least two connectivity components* C_1, C_2, *where* $f(x) \in C_1$ *and there exist* $y_1 \neq y_2 \in V(G)\backslash D$ *fulfilling* $f(y_1) = f(y_2)$ *and* $\{x, y_1\}, \{x, y_2\} \in E(G)$ *then we either can apply Lemma 3.8 or we have* $\Xi(\gamma) \leqslant \Xi(\gamma-1)+4$.

PROOF. If $|V(C_1) \cap D| \geqslant 2$ we can apply Lemma 3.8 to obtain a contradiction to the minimality of (G', D). Thus we may assume $|V(C_1) \cap D| = 1$. Since G is a minimal subgraph, we have $V(C_1) = \{f(x), w\}$ and the neighbors of $f(x)$ and w in G are contained in $\{f(x), w, x\}$. As an abbreviation we set $f(y_1) = f(y_2) = z \in D$. See the upper left drawing in Figure 6 for a graphical representation of the situation. Now we consider the subgraph $\widetilde{C}_2$ consisting of the induced subgraph of $V(C_2) \cup \{x\}$ with the additional edge $\{x, f(y_1)\}$. Let H_2 be an optimal orientation of $\widetilde{C}_2$, where we assume that the arc $[z, y_1]$ is directed from z to y_1, see the upper right graph of Figure 6. Now we construct an orientation H of G by taking the directions from H_2 and redirecting some edges. We direct x to w, w to $f(x)$, $f(x)$ to x to y_1, y_1 to z, z to y_2, and y_2 to x, see the lower drawing of Figure 6.

Now we analyze the distance $d_H(a, b)$ between two vertices in $V(G)$. If a and b are both in $\widetilde{C}_2$, then we can consider a shortest path P in H_2. It may happen that P uses some of the redirected edges. In this case P contains at least two vertices from $\{x, y_1, y_2, z\}$. If P uses more than two vertices from $\{x, y_1, y_2, z\}$ then we only consider those two vertices which have the largest distance on P. Looking at our redirected edges in H we see, the distance between two such vertices is at most three, so that we have $d_H(a, b) \leqslant d_{H_2}(a, b) + 3$ in this case.

Now let b be in $\widetilde{C}_2$. We consider a shortest path P in H_2 from z to b. In H we have $d_H(f(x), z) \leqslant 3$ by considering the path $[f(x), x, y_1, z]$. Since $d_H(z, y_2) = 1$ we have $d_H(f(x), b) \leqslant d_{H_2}(z, b) + 4$. Similarly we obtain $d_H(w, b) \leqslant d_{H_2}(z, b) + 5$. With $D_2 = D\backslash\{f(x)\}$ the set D_2 is a dominating set of $\widetilde{C}_2$ and we can check that $|D_2| = \gamma\left(\widetilde{C}_2\right)$ holds. Since $z \in D_2$ and H_2 is an optimal orientation, for $b_1 \in D_2$, $b_2 \notin D_2$ we have $d_{H_2}(z, b_1) \leqslant \Xi(\gamma-1) - 4$ and $d_{H_2}(z, b_2) \leqslant \Xi(\gamma-1) - 2$ yielding $d_H(f(x), b_1) \leqslant \Xi(\gamma-1)$, $d_H(f(x), b_2) \leqslant \Xi(\gamma-1)+2$, $d_H(w, b_1) \leqslant \Xi(\gamma-1)+1$, and $d_H(w, b_2) \leqslant \Xi(\gamma-1)+3$. This is compatible with $\Xi(\gamma) \leqslant \Xi(\gamma-1)+4$ due to $f(x), b_1 \in D$ and $w, b_2 \notin D$.

Now let a be in $\widetilde{C}_2$. we consider a shortest path P in H_2 from a to z. In H we have $d_H(z, f(x)) \leqslant 4$ by considering the path $[z, y_2, x, w, f(x)]$. Since P can not use an arc from y_1 to z (this arc is directed in the opposite direction in H_2) either P contains a vertex in $\{x, y_2\}$ or P also exists in H, so that we have $d_H(a, f(x)) \leqslant d_{H_2}(a, z) + 4$. Similarly we obtain $d_H(a, w) \leqslant d_{H_2}(a, z) + 3$. Since H_2 we conclude similarly as in the above paragraph that all distances are compatible with $\Xi(\gamma) \leqslant \Xi(\gamma-1)+4$. □

Lemma 3.10 *Let* G *be a minimal subgraph of a pair* (G', D) *in first standard form. If there exist* $z_1, z_2 \in V(G)\backslash D$ *with* $f(z_1) = f(z_2)$ *and* $\{z_1, z_2\} \in E(G)$, *then either* z_1 *or* z_2 *is a cut vertex.*

PROOF. If z_1 has no other neighbors besides z_2 and $x := f(z_1)$ then either z_2 is a cut vertex or z_1 can be deleted from G without destroying the properties of Definition 2.7. We assume that neither z_1 nor z_2 is a cut vertex. Thus both z_1 and z_2 have further neighbors y_1 and y_2, respectively. Since $\{z_1, z_2\}$ can not be deleted we have $y_1 \neq y_2$. Let P_1 be a shortest path from y_1 to z_2 in $G\backslash\{z_1\}$. Since $\{z_1, z_2\}$ can not be deleted P_1 contains the edge $\{x, z_2\}$. Similarly there exists a shortest path from y_2 to z_1 containing the edge $\{x, z_1\}$. Thus in the end the existence of P_1 and P_2 shows that $\{z_1, z_2\}$ could be deleted, which is a contradiction to the minimality of G. □

Lemma 3.11 *Let* G *be a minimal subgraph of a pair* (G', D) *in first standard form. Let* x, y_1, y_2 *be three vertices not in the dominating set* D *with* $\{x, y_1\}, \{x, y_2\} \in E(G)$ *and*

$f(y_1) \neq f(x) \neq f(y_2)$ *either one vertex of* x, y_1, y_2 *is a cut vertex, or* $f(y_1) \neq f(y_2)$.

PROOF. We assume as contrary that none of x, y_1, y_2 is a cut vertex and $f(y_1) = f(y_2)$. Now we consider $G\backslash\{x\}$, which must be connected. Thus there must exist a path P connecting $f(x)$ to $f(y_1) = f(y_2)$ and either one of the edges $\{x, y_1\}$, $\{x, y_2\}$ is a chord or one of the vertices y_1, y_2 could be deleted from G, which is a contradiction to the minimality of G. □

4 Proof of the main theorem

In this section we want to prove Theorem 1.3. We use the techniques of induction on $\gamma(G)$ and minimal counter examples with respect to $\gamma(G)$.

Definition 4.1 *We call a minimal subgraph* G *of* (G', D) *in first standard form a minimal counter example to Theorem 1.3 if we have* $\max\left\{\mathrm{diam}_0(H, D) + 4, \mathrm{diam}_1(H, D) + 2, \mathrm{diam}_2(H, D)\right\} > 4\gamma$ *for a minimal orientation* H *and* $\gamma = |D|$ *is minimal with this property.*

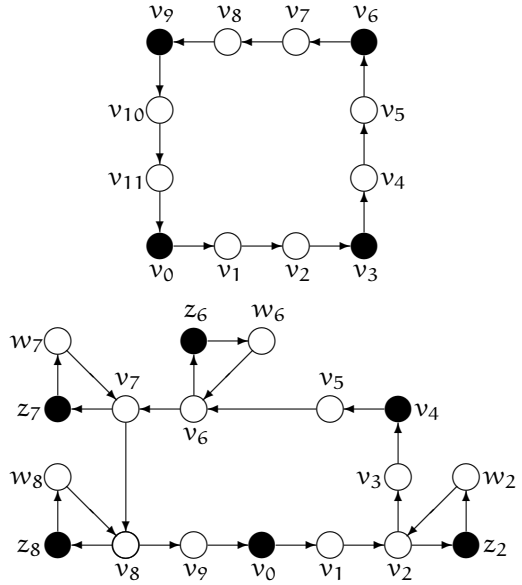


Figure 7: The situation of Lemma 4.2 and the situation of Lemma 4.3.

Lemma 4.2 *Let* G *be a minimal subgraph of* (G', D) *in first standard form which is a minimal counter example to Theorem 1.3, then there can not exist an elementary cycle* $C = [v_0, \ldots, v_{3k} = v_0]$ *in* G *with* $k \geqslant 2$ *and the* $v_{3j} \in D$ *for all* $0 \leqslant j < k$.

PROOF. We assume the existence of such a cycle C, see the upper graph in Figure 7 for an example, and consider another graph $\widetilde{G}$ arising from G by:

(1) deleting the edges of C,

(2) deleting the vertices v_{3j} for $0 < j < k$,

(3) inserting vertices u_j and edges $\{v_0, v_j\}$, $\{v_0, u_j\}$, $\{u_j, v_j\}$ for all $0 < j < 3k$ with $3 \nmid j$, and by

(4) identifying all vertices $v_{3j} \in G$ with the vertex $v_0 \in \widetilde{G}$, meaning that we replace edges $\{v_{3j}, x\}$ in G by edges $\{v_0, x\}$ in $\widetilde{G}$.

We remark that this construction does not produce multiple edges since (G', D) is in first standard form. The set $\widetilde{D} := D\backslash\left\{v_3, v_6, \ldots, v_{3k-3}\right\}$ is a dominating set of $\widetilde{G}$ with $\left|\widetilde{D}\right| = |D|-k+1$. Let $\widetilde{H}$ be a minimal orientation of $(\widetilde{G}, \widetilde{D})$. We construct an orientation H of G by taking over the directions of all common edges with $\widetilde{H}$ and by orienting the edges of C from v_j to v_{j+1}, see the upper graph in Figure 7.

Now we analyze the distances in H. For brevity we set $I := \left\{v_{3j} : 0 \leqslant j < k\right\}$ (these are the vertices in G which are associated with v_0 in $\widetilde{G}$). The distance of two vertices in I in the orientation H is at most $3k-3$ and the distance of two vertices in $V(C)$ is at most $3k-1$. Thus we may assume $|D| > k$. Let a, b be vertices in $V(G)$.

(1) If a and b are elements of $\{v_j : 0 \leqslant j < 3k\}$ then we have $d_H(a, b) \leqslant 3k-1 < 4|D|-4$.

(2) If a and b are not in I then we consider a shortest path $\widetilde{P}$ in $\widetilde{H}$ connecting a and b.

(3) If $a \in I$ and $b \notin I$ then we consider a shortest path $\widetilde{P}$ in $\widetilde{H}$ connecting v_0 and b.

(4) The case $a \notin I$ and $b \in I$ then we consider a shortest path $\widetilde{P}$ in $\widetilde{H}$ connecting a and v_0.

Let $\widetilde{P}$ be an arbitrary shortest path in $\widetilde{H}$ connecting a and b. It may happen that in H this path $\widetilde{P}$ does not exist since it may contain the vertex v_0 corresponding to two different vertices v_{3i} and v_{3j} in G or may contain one of the edges $\{v_0, v_j\}$, $\{v_0, u_j\}$, or $\{u_j, v_j\}$ with $3 \nmid j$.

Now we want to construct a path P which does connect a and b in H. The path $\widetilde{P}$ may use one of the edges $\{v_0, v_j\}$, $\{v_0, u_j\}$, or $\{u_j, v_j\}$ with $3 \nmid j$. Deleting all these edges decomposes $\widetilde{P}$ in at least two parts $\widetilde{P}_1, \ldots, \widetilde{P}_m$ with $\left|\widetilde{P}_1\right| + \left|\widetilde{P}_m\right| \leqslant \left|\widetilde{P}\right| - 1$. Using a suitable segment $\widetilde{C}$ of the cycle C we obtain a path $P = \widetilde{P}_1 \cup \widetilde{C} \cup \widetilde{P}_m$ of length at most $\left|\widetilde{P}_1\right| + \left|\widetilde{P}_m\right| + \left|\widetilde{C}\right| \leqslant \left|\widetilde{P}\right| + 3k - 2$. If $\widetilde{P}$ does not use one of these edges then it can only happen that v_0 is used in $\widetilde{P}$ corresponding to two different vertices v_{3i} and v_{3j} in G. In this case we can use a suitable segment $\widetilde{C}$ of the cycle C, which starts and ends in a vertex of I, to obtain a path P connecting a and b in H of length at most $\left|\widetilde{P}\right| + 3k - 3$.

Now we are ready to prove that G is not a counter example. If $\gamma(\widetilde{G}) < \left|\widetilde{D}\right|$ then we have $\mathrm{diam}\left(\widetilde{H}\right) \leqslant 4\cdot\left|\widetilde{D}\right|-4 = 4 \cdot |D| - 4k$ due to the minimality of G. In each of the cases (1)-(4) we have $d_H(a, b) \leqslant 4 \cdot |D| - k - 2 \leqslant 4 \cdot |D| - 4$

for all $a, b \in G$. Otherwise we have $\gamma(\widetilde{G}) = \left|\widetilde{D}\right|$ and $\widetilde{D}$ is a minimal dominating set of $\widetilde{G}$. In this case we have

$$\begin{aligned}
\mathrm{diam}_2(H,D) &\leqslant \max\Big\{\mathrm{diam}_2\Big(\widetilde{H},\widetilde{D}\Big)+3k-2,\\
&\qquad \mathrm{diam}_1\Big(\widetilde{H},\widetilde{D}\Big)+3k-1,3k-1\Big\}\\
&\leqslant 4\cdot|D|-k+2\\
&\leqslant 4\cdot|D|\\
\mathrm{diam}_1(H,D) &\leqslant \max\Big\{\mathrm{diam}_1\Big(\widetilde{H},\widetilde{D}\Big)+3k-2,\\
&\qquad \mathrm{diam}_0\Big(\widetilde{H},\widetilde{D}\Big)+3k-1,3k-1\Big\}\\
&\leqslant 4\cdot|D|-k\\
&\leqslant 4\cdot|D|-2\\
\mathrm{diam}_0(H,D) &\leqslant \max\Big\{\mathrm{diam}_0\Big(\widetilde{H},\widetilde{D}\Big)+3k-2,3k-3\Big\}\\
&\leqslant 4\cdot|D|-k-2\\
&\leqslant 4\cdot|D|-4
\end{aligned}$$

□

Lemma 4.3 *Let G be a minimal subgraph of (G',D) in first standard form which is a minimal counter example to Theorem 1.3, then there can not exist an elementary cycle $C=[v_0,\dots,v_l=v_0]$ in G with the following properties:*

(1) $v_0 \in D$,

(2) $|V(C)\cap D| \geqslant 2$,

(3) $l \geqslant 6$, *and*

(4) if $v_j \notin D$ then either $f(v_j) \in \{v_{j-1}, v_{j+1}\}$ or v_j is a cut vertex in G where the component containing $f(v_j)$ contains exactly one vertex of D.

PROOF. We assume the existence of such a cycle C. By y we denote the number of cut vertices v_j in C and by Y the corresponding set. For all $v \in Y$ we have $f(v) \notin C$ since otherwise we could apply Lemma 3.8. If $e=\{v',v''\}$ would be a chord of C then $|\{v',v''\}\cap D| = 1$ since (G',D) is in first standard form and G is a minimal subgraph, which especially means that we can not delete the edge e. We assume w.l.o.g. $v' \in D$ and conclude $f(v'')=v'$. Thus v'' is not a cut vertex and due to property (4) the edge e is not a chord. Finally we conclude that C is chordless. For $y=0$ we would have $v_{3j}\in D$ due to $l \geqslant 6$ and the property $f(v_j)\in\{v_{j-1},v_{j+1}\}$ for vertices $v_j \notin D$. Thus we may assume $y \geqslant 1$ since otherwise we could apply Lemma 4.2. For each $v_j \in Y$ we set $z_j = f(v_j) \notin V(C)$ and denote by $w_j \in V(G)\backslash(V(C)\cup D)$ the vertex which is adjacent to v_j and z_j. By k we denote the number of vertices v_j in $V(C)$ which are also contained in D. Due to condition (2) we have $k \geqslant 2$. The two neighbors on the cycle C of a vertex in Y both are not contained in D. For a vertex $v \in V(C)\backslash(D\cup Y)$ one neighbor on C is $f(v)$ and the other neighbor lies in $V(C)\backslash D$. Thus the length $|C|$ of the cycle is given by $3k+y \geqslant 7$. On the lower side of Figure 7 we have depicted an example with $k=2$ and $y=4$.

Now we consider another graph $\widetilde{G}$ arising from G by:

(1) deleting the edges of C,

(2) deleting the vertices $\Big(\{z_j,w_j : 0<j<l\}\cup(V(C)\cap D)\Big)\backslash\{v_0\}$,

(3) inserting vertices u_j and edges $\{v_0,v_j\}$, $\{v_0,u_j\}$, $\{u_j,v_j\}$ for all $0<j<l$ with $v_j\notin D$, and by

(4) identifying all vertices $v_j \in D$ with the vertex $v_0 \in \widetilde{G}$, meaning that we replace edges $\{v_j,x\}$ in G by edges $\{v_0,x\}$ in $\widetilde{G}$.

We remark that this construction does not produce multiple edges since (G',D) is in first standard form. The set $\widetilde{D} := D\backslash\Big\{v_1,\dots,v_{l-1},z_1,\dots,z_{l-1}\Big\}$ is a dominating set of $\widetilde{G}$ with $\left|\widetilde{D}\right| = |D|-k-y+1$. Let $\widetilde{H}$ be a minimal orientation of $(\widetilde{G},\widetilde{D})$. We construct an orientation H of G by taking over the directions of all common edges with $\widetilde{H}$ and by orienting the edges of C from v_j to v_{j+1}. The missing edges corresponding to z_j and w_j are oriented from v_j to z_j, from z_j to w_j, and from w_j to v_j, see the graph on the lower side of Figure 7. For brevity we set $A = V(C)\cup\Big\{w_j,z_j : 0<j<l\Big\}$.

Now we analyze the distances in H. For $a_1,b_1\in A$ we have $d_H(a_1,b_1)\leqslant 3k+y+3$, for $a_2,b_2\in V(C)$ we have $d_H(a_2,b_2)\leqslant 3k+y-1$, and for $a_3,b_3\in V(C)\cap D$ we have $d_H(a_3,b_3)\leqslant 3k+y-3$. Thus we may assume $|D|>k+y$. Let a,b be vertices in $V(G)$.

(1) If a and b are elements of A then we have $d_H(a,b)\leqslant 3k+y+3<4|D|-4$.

(2) If a and b are not in A then we consider a shortest path $\widetilde{P}$ in $\widetilde{H}$ connecting a and b.

(3) If $a\in A$ and $b\notin A$ then we consider a shortest path $\widetilde{P}$ in $\widetilde{H}$ connecting v_0 and b.

(4) The case $a\notin A$ and $b\in A$ then we consider a shortest path $\widetilde{P}$ in $\widetilde{H}$ connecting a and v_0.

Let $\widetilde{P}$ be a shortest path in $\widetilde{H}$ connecting two vertices a and b. Similarly as in the proof of Lemma 4.2 we construct a path P in H connecting a and b. Doing the same analysis we obtain $|P|\leqslant\left|\widetilde{P}\right|+3k+y-2$. Starting or ending at a vertex z_i or w_i increases the length by at most 2.

If $\gamma\left(\widetilde{G}\right)<\left|\widetilde{D}\right| = |D|-k-y+1$ then we would have $d_H(a,b)\leqslant 4|D|-k-3y+1\leqslant 4|D|-4$. Thus we may assume $\gamma\left(\widetilde{G}\right)=\left|\widetilde{D}\right|=|D|-k-y+1$, meaning that $\widetilde{D}$ is a minimal dominating set. With this clearly we have $d_{\widetilde{H}}(v_0,b), d_{\widetilde{H}}(a,v_0)\leqslant 4\cdot\left|\widetilde{D}\right|-2$ for all $a,b\in\widetilde{G}$ and $d_{\widetilde{H}}(v_0,b'), d_{\widetilde{H}}(a',v_0)\leqslant 4\cdot\left|\widetilde{D}\right|-4$ for all $a',b'\in\widetilde{D}$.

For $k+y \geqslant 3$ and $|D| \geqslant k+y+1$ we have

$$\begin{aligned}
diam_2(H,D) &\leqslant \max\left\{diam_2\left(\widetilde{H},\widetilde{D}\right)+3k+y-2,\right.\\
&\qquad \left. diam_1\left(\widetilde{H},\widetilde{D}\right)+3k+y,3k+y+3\right\}\\
&\leqslant 4\cdot|D|-k-3y+2\\
&\leqslant 4\cdot|D|\\
diam_1(H,D) &\leqslant \max\left\{diam_1\left(\widetilde{H},\widetilde{D}\right)+3k+y,\right.\\
&\qquad \left. diam_0\left(\widetilde{H},\widetilde{D}\right)+3k+y,3k+y+3\right\}\\
&\leqslant 4\cdot|D|-k-3y+2\\
&\leqslant 4\cdot|D|-2\\
diam_0(H,D) &\leqslant \max\left\{diam_0\left(\widetilde{H},\widetilde{D}\right)+3k+y,3k+y+3\right\}\\
&\leqslant 4\cdot|D|-k-3y\\
&\leqslant 4\cdot|D|-4
\end{aligned}$$

□

Now we are ready to prove Theorem 1.3:

PROOF.(of Theorem 1.3)
Let G be a minimal subgraph of (G',D) in first standard form which is a minimal counter example to Theorem 1.3. Due to Lemma 2.6 and Lemma 3.5 we can assume $|D| \geqslant 4$. We show that we have $|V(G)| \leqslant 4\cdot(|D|-1)+1$. In this case we can utilize an arbitrary orientation H of G. Since a shortest path uses every vertex at most once we would have $diam(H) \leqslant 4\cdot(|D|-1)$. Applying Lemma 2.5 we conclude $diam_{min}(G') \leqslant 4\cdot|D| = 4\cdot\gamma(G')$, which is a contradiction to G being a minimal counter example to Theorem 1.3 and instead proves this theorem.

At first we summarize some structure results for minimal counter examples to Theorem 1.3.

(1) We can not apply one of the Lemmas 3.3, 3.4, 3.7, 3.8, or 3.10. So if $v \in V(G)$ is a cut vertex we have $v \notin D$ and there exists a unique vertex $t(V) \notin D$ such that we have $\{v,f(v)\},\{f(v),t(v)\},\{t(v),v\} \in E(G)$ and all neighbors of $f(v),t(v)$ are contained in $\{f(v),t(v),v\}$.

(2) Due to Lemma 3.9, Lemma 3.11, and (1) there do not exist pairwise different vertices $x,y_1,y_2 \in V(G)\backslash D$ with $\{x,y_1\},\{x,y_2\} \in E(G)$ and $f(y_1)=f(y_2)$.

(3) We can not apply Lemma 4.2 or Lemma 4.3 on G.

In order to bound $|V(G)|$ from above we perform a technical trick and count the number of vertices of a different graph $\widetilde{G}$. Therefore we label the cut vertices of G by $v_1,\ldots,v_m$. With this we set

$$\widetilde{D} = \Big(D \cup \{v_i : 1 \leqslant i \leqslant m\}\Big)\backslash\{f(v_i) : 1 \leqslant i \leqslant m\}.$$

The graph $\widetilde{G}$ arises from G by deleting the $f(v_i),t(v_i)$ for $1 \leqslant i \leqslant m$ and by replacing the remaining edges $\{v_i,x\}$ by a pair of two edges $\{v_i,y_{x,i}\},\{y_{x,i},x\}$, where the $y_{x,i}$ are new vertices. We have $\left|\widetilde{D}\right| = |D|$, $\left|V\left(\widetilde{G}\right)\right| \geqslant |V(G)|$, the set $\widetilde{D}$ is a dominating set of $\widetilde{G}$, and $\widetilde{G}$ is a subgraph of a suitable pair in first standard form. If $\widetilde{G}$ would not be a minimal subgraph than also G would not be a minimal subgraph. We have the following structure results for $\widetilde{G}$:

(a) There do not exist two vertices $u,v \in V(\widetilde{G})\backslash\widetilde{D}$ with $\{u,v\} \in E(\widetilde{G})$ and $f(u)=f(v)$.

(b) There do not exist pairwise different vertices $x,y_1,y_2 \in V(\widetilde{G})\backslash\widetilde{D}$ with $\{x,y_1\},\{x,y_2\} \in E\left(\widetilde{G}\right)$ and $f(y_1)=f(y_2)$.

(c) We can not apply Lemma 4.2 or Lemma 4.3 on $\widetilde{G}$.

Since our construction of $\widetilde{G}$ has removed all such configurations (a) holds. If in (b) $f(y_1)=f(y_2)$ is an element of D then such a configuration also exists in G, which is a contradiction to (2). If $f(y_1)$ corresponds to a v_i in G, then y_1 and y_2 would correspond to two new vertices $y_{i,e}$ and $y_{i,e'}$. In this case we would have a double edge from x to v_i in G, which is not true. Thus (b) holds. Since all vertices in $\widetilde{D}\backslash D$ correspond to cut vertices in G also (c) holds.

In order to prove $\left|V\left(\widetilde{G}\right)\right| \leqslant 4\cdot\left(\left|\widetilde{D}\right|-1\right)+1$ we construct a tree T fulfilling

(i) $\widetilde{D} \subseteq V(T)$ and

(ii) if $v_1 \in V(T)\backslash\widetilde{D}$ then we have $\{f(v_1),v_1\} \in E(T)$.

Therefore we iteratively construct trees T_k for $1 \leqslant k \leqslant \left|\widetilde{D}\right|$. The tree T_1 is composed of a single vertex $x_1 \in \widetilde{D}$. The tree T_1 clearly fulfills condition (ii). To construct T_{k+1} from T_k we find a vertex x_{k+1} in $\widetilde{D}\backslash V(T_k)$ with the minimum distance to T_k. The tree T_{k+1} is the union of T_k with a shortest path P_{k+1} from x_{k+1} to T_k. Since $\widetilde{D}$ is a dominating set this path P_{k+1} has length at most three. Since $\widetilde{G}$ is a subgraph of a suitable pair in first standard form P_{k+1} has length at least two. For $P_{k+1}=[x_{k+1},v_1,v_2]$ we have $v_1,v_2 \notin \widetilde{D}$ due to the first standard form and $f(v_1)=x_{k+1}$, $v_2 \in V(T_k)$. Since condition (ii) is fulfilled for T_k it is also fulfilled for T_{k+1} in this case. In the remaining case we have $P_{k+1}=[x_{k+1},v_1,v_2,v_3]$ with $v_1,v_2 \notin V(T_k)$, $v_1,v_2 \notin \widetilde{D}$, and $v_3 \in V(T_k)$. If $f(v_2)$ would not be contained in $V(T_k)$ then $[f(v_2),v_2,v_3]$ would be a shorter path connecting $f(v_2)$ to T_k. Thus we have $f(v_2) \in V(T_k)$ and we may assume $v_3=f(v_2)$. (We may simply consider the path $[x_{k+1},v_1,v_2,f(v_2)]$ instead of P_{k+1}.) Due to $x_{k+1} \notin V(T_k)$ and T_k fulfilling condition (ii), these conditions are also fulfilled for T_{k+1}. In the end we obtain a tree $T_{|\widetilde{D}|}$ fulfilling condition (i) and condition (ii). By considering the paths P_k we conclude $|V(T)| \leqslant \left|\widetilde{D}\right|+2\left(\left|\widetilde{D}\right|-1\right)$.

Clearly we have some alternatives during the construction of $T_{|\widetilde{D}|}$. Now we assume that T is a subtree of $\widetilde{G}$ fulfilling conditions (i) and (ii), and having the maximal number of vertices. In the next step we want to prove some properties of the vertices in T.

Let $v \in \widetilde{D}$ and let $u \in V(\widetilde{G})\backslash V(T)$ be a neighbor of v in $\widetilde{G}$. We prove that every neighbor u' of u in $\widetilde{G}$ is contained in $V(T)$. Clearly we have $u' \notin \widetilde{D}$. Due to (a) we have $f(u') \neq v$. If $u' \notin V(T)$ then adding the edges $A := \left\{\{v,u\},\{u,u'\},\{u',f(u')\}\right\}$ gives an elementary cycle $C = [v_0,\ldots,v_l]$ in $\left(V(T)\cup\{u,u'\}, E(T)\cup A\right)$, where $v_0 = v_l$ and $l \geqslant 6$. Since we can not apply Lemma 4.2 there exists an index j (reading the indices modulo l) fulfilling

$$v_j \in \widetilde{D} \quad \text{and} \quad v_{j+1}, v_{j+2}, v_{j+3} \in V(T)\backslash\widetilde{D}.$$

Since the edge $\{v_{j+1}, v_{j+2}\}$ is contained in $E(T)$ also the edge $\{v_{j+2}, f(v_{j+2})\}$ is contained in $E(T)$. Similarly we conclude that the edge $\{v_{j+3}, f(v_{j+3})\}$ is contained in $E(T)$. If v_{j+1} has no further neighbors besides v_j and v_{j+2} in T then

$$\begin{aligned} T' \;:=\; & \Big((V(T)\cup\{u,u'\})\backslash\{v_{j+1}\}, \\ & (E(T)\cup A)\backslash\Big\{\{v_j,v_{j+1}\},\{v_{j+1},v_{j+2}\}\Big\}\Big) \end{aligned}$$

would be a subtree of $\widetilde{G}$ fulfilling the conditions (i) and (ii) with a larger number of vertices than T. Thus such an u' can not exist in this case. If v_{j+1} has further neighbors in T, then deleting the edge $\{v_{j+1}, v_{j+2}\}$ and adding the edges and vertices of A would also yield a subtree of $\widetilde{G}$ fulfilling the conditions (i) and (ii) with a larger number of vertices than T.

The same statement also holds for $v \in V(T)\backslash\widetilde{D}$ since we may consider $f(u)$ instead v. Thus in $\widetilde{G}$ we have $\{u,v\} \cap V(T) \neq \emptyset$ for every edge $\{u,v\} \in E(\widetilde{G})$.

For a graph K and a vertex $v \in V(K)$ we denote by $S(K,v)$ the uniquely defined maximal connected and bridgelesssubgraph of K containing v. If every edge being adjacent to v is a bridge or v does not have any edges, then S consists only of vertex v. We remark that $u \in S(K,v)$ is an equivalence relation $\sim_K$ for all vertices $u,v \in V(K)$. By F we denote the set of vertices in $V(T)$ which are either contained in $\widetilde{D}$ or have a degree in $V(T)$ of at least three. We have

$$|V(T)| + |F| \leqslant 4\cdot\left|\widetilde{D}\right| - 2,$$

which can be proved by induction on $|V(T_k)| + |F\cap V(T_k)| \leqslant 4\cdot k - 2$ for $1 \leqslant k \leqslant \left|\widetilde{D}\right|$. Clearly we have $|V(T_1)| + |F\cap V(T_1)| = 2 \leqslant 4\cdot 1 - 2$. The tree T_{k+1} arises from T_k by adding a path P_{k+1} of length at most three. If $|P_{k+1}| = 3$ then we have $F\cap V(T_{k+1}) = \left(F\cap V(T_k)\right)\cup\{x_{k+1}\}$ and $|V(T_{k+1})| \leqslant |V(T_k)| + 3$. For $|P_{k+1}| = 2$ we have $|V(T_{k+1})| \leqslant |V(T_k)| + 2$ and $|F\cap V(T_{k+1})| \leqslant |F\cap V(T_k)| + 2$.

For a graph K containing T as a subgraph we denote by $N(K)$ the number $\left|\{S(K,v) : v \in F\}\right|$ of equivalence classes of $\sim_K$. Since T is a tree we have $N(T) = |F|$. Now we recursively construct a sequence of graphs G_i for $1 \leqslant i \leqslant |F|$ fulfilling

$$\begin{aligned} & |V(G_i)| + N(G_i) \leqslant 4\cdot\left|\widetilde{D}\right| - 2,\ N(G_i) \leqslant i, \\ & \text{and } T \subseteq G_i \subseteq \widetilde{G}. \end{aligned} \tag{2}$$

This yields a graph G_1 containing at most $4\cdot\left|\widetilde{D}\right|-3$ vertices, where each two elements of $\widetilde{D}$ are connected by at least two edge disjoint paths. So either we have $\left|V(\widetilde{G})\right| \leqslant 4\cdot\left|\widetilde{D}\right| - 3$ or $\widetilde{G}$ and G are not minimal subgraphs.

During the following analysis we often delete a vertex v or an edge e from the tree T in such a way that it decomposes in exactly two subtrees T^1 and T^2. Since T contains no cut vertices there exists a path M in $\widetilde{G}$ without v or without e connecting T^1 and T^2. Since there does not exist an edge $\{u_1,u_2\} \in E\left(\widetilde{G}\right)$ with $\{u_1,u_2\}\cap V(T) = \emptyset$ we have $|M| \leqslant 2$ if M is a shortest path.

For $G_{|F|} = T$ condition (2) holds. Now for $i \geqslant 2$ let G_i be given. If there exists a vertex $u \in V(\widetilde{G})\backslash V(G_i)$ having neighbors $x,y \in V(G_i)$ with $S(G_i,x) \neq S(G_i,y)$ we define G_{i-1} by adding vertex u and adding all edges, being adjacent with u in $\widetilde{G}$, to G_i. With this we have $|V(G_{i-1})| = |V(G_i)| + 1$ and $N(G_{i-1}) = N(G_i) - 1$, so that condition (2) is fulfilled for G_{i-1}.

Now we deal with the cases where $i \geqslant 2$ and where such vertices u, x, y do not exist. We use the setwise defined distance

$$d_K(A,B) := \min\left\{d_K(a,b) \,:\, a \in A,\ b \in B\right\}.$$

Now we choose $f_1, f_2 \in F$ with $S(G_i,f_1) \neq S(G_i,f_2)$, where $d_{G_i}(S(G_i,f_1), S(G_i,f_2))$ is minimal. Clearly we have $1 \leqslant d_{G_i}(S(G_i,f_1), S(G_i,f_2)) \leqslant 3$. By P_{f_1,f_2} we denote the corresponding shortest path connecting $S(G_i,f_1)$ with $S(G_i,f_2)$.

If $|P_{f_1,f_2}| = [v_0,v_1]$ and the edge $\{v_0,v_1\}$ is not contained in $E(T)$, then we simply add this edge to G_i to obtain G_{i-1}. So we may assume that $\{v_0,v_1\} \in E(T)$. Deleting $\{v_0,v_1\}$ in T decomposes T into two subtrees T^1 and T^2, where we assume w.lo.g. that $f_1 \in V\left(T^1\right)$ and $f_2 \in V\left(T^2\right)$. Due to $d_{\widetilde{G}\backslash\{v_0,v_1\}}\left(T^1,T^2\right) \leqslant 2$ we can obtain a graph G_{i-1} adding add most one vertex, where $S(G_{i-1},f_1) = S(G_{i-1},f_2)$ holds.

If $P_{f_1,f_2} = [v_0,v_1,v_2]$ and $v_1 \notin V(T)$ the we can add v_1 and add all its edges to G_i to obtain G_{i-1}. So we may assume $v_1 \in V(T)$. If $\{v_0,v_1\}$ or $\{v_1,v_2\}$ would not be contained in $E(T)$, then we may simply add it to G_i, without increasing the number of vertices, and are in a case $|P_{f_1,f_2}| = 1$. So we may assume $\{v_0,v_1\},\{v_1,v_2\} \in E(T)$. Due to $S(G_i,v_0) \neq S(G_i,v_1) \neq S(G_i,v_2)$ and the minimality of P_{f_1,f_2} we have $v_1 \notin F$. Thus v_1 has degree two in T and removing v_1 decomposes T into two subtrees T^1 and T^2, where we assume w.lo.g. that $f_1 \in V\left(T^1\right)$ and $f_2 \in V\left(T^2\right)$. Since there does not exist a cut vertex in $\widetilde{G}$ we have $d_{\widetilde{G}\backslash\{v_1\}}\left(T^1,T^2\right) \leqslant 2$ and we can obtain a graph G_{i-1} adding add most one vertex, where $S(G_{i-1},f_1) = S(G_{i-1},f_2)$ holds.

The remaining case is $P_{f_1,f_2} = [v_0, v_1, v_2, v_3]$. Due to the minimality of P_{f_1,f_2} we have $f(v_2) \in V(S(G_i, f_2))$ and $f(v_1) \in V(S(G_i, f_1))$. Thus we may assume $v_0, v_3 \in \widetilde{D}$. Additionally we have $\{v_1, v_2\} \cap V(T) \neq \emptyset$. If $v_j \notin V(T)$ we may simply add v_j and its edges to G_i to obtain G_{i-1}. So we may assume $v_1, v_2 \in V(T)$. W.l.o.g. we assume $\{v_1, v_2\} \in E(T)$. Otherwise there exists an edge $\{v_1, v_4\} \in E(T)$ with $v_4 \neq v_0$ and we could choose $f_1 = f(v_1)$, $f_2 = f(v_4)$. The vertices v_1 and v_2 both have degree two in T. Deleting v_1 in T gives two subtrees T^1 and T^2, where we can assume $v_0 \in V(T^1)$ and $v_2 \in V(T^2)$. Since there does not exist a cut vertex in $\widetilde{G}$ we have $d_{\widetilde{G}\setminus\{v_1\}}(T^1, T^2) \leqslant 2$ and denote the corresponding shortest path by R_1. If $R_1 = [r_0, r_1, r_2]$ does not end in v_2 then we could obtain G_{i-1} by adding vertex r_1 and its edges to G_i. Similarly we may delete vertex v_2 to obtain a shortest path R_2 which ends in v_1. But in this case the edge $\{v_1, v_2\}$ could be deleted from $\widetilde{G}$, which is a contradiction to the minimality of $\widetilde{G}$. □

We conjecture that if G is a critical minimal subgraph of a pair (G', D) in first standard form then we always can apply one of the lemmas 3.1, 3.2, 3.3, 3.4, 3.6, 3.7, 3.8, 3.9, 3.10, 3.11, 4.2, or 4.3.

We would like to remark that our reduction technique is constructive in the following sense: If we have a graph G and a dominating set D, not necessarily a minimal dominating set of G, then we can construct an orientation H of G in polynomial time fulfilling $\mathrm{diam}(H) \leqslant 4 \cdot |D|$: At first we apply the transformations of the proof of Lemma 2.3 to obtain a graph $\widetilde{G}$, which fulfills conditions (1), (3)-(6) of Definition 2.2 and where D remains a dominating set. In the following we will demonstrate how to obtain an orientation $\widetilde{H}$ of $\widetilde{G}$ fulfilling $\mathrm{diam}\left(\widetilde{H}\right) \leqslant 4 \cdot |D|$. From such an orientation we can clearly reconstruct an orientation H of G. Since Lemma 2.5 does not use the minimality of the dominating set D we can restrict our consideration on a minimal subgraph $\hat{G}$ of $\widetilde{G}$. Since none of the lemmas in Section 3 uses the minimality of the domination set D, we can apply all these reduction steps on $\hat{G}$. These steps can easily be reversed afterwards. The proofs of Lemma 4.2 and Lemma 4.3 have to be altered very slightly to guarantee a suitable reduction also in the case where D is not minimal. (Here only the analysis is affected, not the construction.) We end up with a graph $\dot{G}$ with dominating set $\dot{D}$ (here $\dot{D}$ arises from D by applying the necessary reduction steps). Since in the proof of Theorem 1.3 we show $\left|V\left(\dot{G}\right)\right| \leqslant 4 \cdot \left|\dot{D}\right| - 3$ we can choose an arbitrary strong orientation and reverse all previous steps to obtain an orientation H of G with $\mathrm{diam}(H) \leqslant 4 \cdot |D|$. We remark that all steps can be performed in polynomial time.

5 Conclusion and outlook

In this article we have proven

$$\overrightarrow{\mathrm{diam}}_{\min}(G) \leqslant 4 \cdot \gamma(G)$$

for all connected and bridgeless graphs and conjecture

$$\overrightarrow{\mathrm{diam}}_{\min}(G) \leqslant \left\lceil \frac{7\gamma(G)+1}{2} \right\rceil$$

to be the true upper bound. Lemma 3.5 shows that Theorem 1.3 is not tight for $\gamma = 3$. Some of our reduction steps in Section 3 can also be used for a proof of Conjecture 1.4. Key ingredients might be the lemmas 4.2 and 4.3, which can be utilized as reductions for Conjecture 1.4 if $k + y$ is large enough. Figure 4 and Figure 5 indicate several cases which can not be reduced so far.

Besides a proof of Conjecture 1.4 one might consider special subclasses of general graphs to obtain stronger bounds on the minimum oriented diameter. E. g. for C_3-free graphs and C_4-free graphs we conjecture that the minimum oriented diameter is at most $3 \cdot \gamma + c$.

Bibliography

[1] F. R. K. Chung, M. R. Garey, and R. E. Tarjan, *Strongly connected orientations of mixed multigraphs*, Networks **15** (1985), 477–484.

[2] V. Chvátal and C. Thomassen, *Distances in orientations of graphs*, J. Combin. Theory Ser. B **24** (1978), 61–75.

[3] P. Dankelmann, *The diameter of directed graphs*, J. Combin. Theory Ser. B **94** (2005), no. 1, 183–186.

[4] P. Dankelmann, O. R. Oellermann, and J.-L. Wu, *Minimum average distance of strong orientations of graphs*, Discrete Appl. Math. **143** (2004), no. 1-3, 204–212.

[5] P. Dankelmann, H. C. Swart, and D. P. Day, *On strong distances in oriented graphs*, Discrete Math. **266** (2003), no. 1-3, 195–201.

[6] R. Diestel, *Graph theory*, second ed., Springer-Verlag, New York, 2000.

[7] F. V. Fomin, M. Matamala, E. Prisner, and I. Rapaport, *Bilateral orientations and domination*, Proceedings of the Brazilian Symposiun on Graphs, Algorithms and Combinatorics (GRACO 2001), Electron. Notes Discrete Math., vol. 7, Elesevier Science Publishers, 2001.

[8] F. V. Fomin, M. Matamala, E. Prisner, and I. Rapaport, *AT-free graphs: Linear bounds for the oriented diameter*, Discrete Appl. Math. **141** (2004), no. 1-3, 135–148.

[9] F. V. Fomin, M. Matamala, and I. Rapaport, *The complexity of approximating the oriented diameter of chordal graphs. (extended abstract)*, Kučera, Luděk (ed.), Graph-theoretic concepts in computer science. 28th international workshop, WG 2002, Český Krumlov, Czech Republic, June 13-15, 2002. Revised papers. Berlin: Springer. Lect. Notes Comput. Sci. 2573, 211-222 (2002), 2002.

[10] F. V. Fomin, M. Matamala, and I. Rapaport, *Complexity of approximating the oriented diameter of chordal graphs*, J. Graph Theory **45** (2004), no. 4, 255–269.

[11] K. M. Koh and K. L. Ng, *The orientation number of two complete graphs with linkages*, Discrete Math. **295** (2005), no. 1-3, 91–106.

[12] K. M. Koh and B. P. Tan, *The diameter of an orientation of a complete multipartite graph*, Discrete Math. **149** (1996), no. 1-3, 131–139.

[13] K. M. Koh and B. P. Tan, *The minimum diameter of orientations of complete multipartite graphs*, Graphs Combin. **12** (1996), no. 4, 333–339.

[14] K. M. Koh and E. G. Tay, *Optimal orientations of products of paths and cycles*, Discrete Appl. Math. **78** (1997), no. 1-3, 163–174.

[15] K. M. Koh and E. G. Tay, *On optimal orientations of cartesian products with a bipartite graph*, Discrete Appl. Math. **98** (1999), no. 1-2, 103–120.

[16] K. M. Koh and E. G. Tay, *On optimal orientations of cartesian products of graphs. II: Complete graphs and even cycles*, Discrete Math. **211** (2000), no. 1-3, 75–102.

[17] K. M. Koh and E. G. Tay, *On optimal orientations of* G *vertex-multiplications*, Discrete Math. **219** (2000), no. 1-3, 153–171.

[18] K. M. Koh and E. G. Tay, *On a conjecture concerning optimal orientations of the cartesian product of a triangle and an odd cycle*, Discrete Math. **232** (2001), no. 1-3, 153–161.

[19] K. M. Koh and E. G. Tay, *On optimal orientations of tree vertex-multiplications*, Australas. J. Combin. **34** (2006), 69–87.

[20] J.-C. König, D. W. Krumme, and E. Lazard, *Diameter-preserving orientations of the torus*, Networks **32** (1998), no. 1, 1–11.

[21] P. K. Kwok, Q. Liu, and A. B. West, *Oriented diameter of graphs with diameter 3*, (submitted).

[22] J. E. McCanna, *Orientations of the n-cube with minimum diameter*, Discrete Math. **68** (1988), no. 2-3, 309–310.

[23] J. Plesník, *Remarks on diameters of orientations of graphs*, Acta Math. Univ. Comenian **46/47** (1985), 225–236.

[24] J. Plesník, *On minimal graphs of diameter 2 with every edge in a 3-cycle*, Math. Slovaca **36** (1986), 145–149.

[25] H. E. Robbins, *A theorem on graphs, with an application to a problem in traffic control*, Amer. Math. Monthly **46** (1939), 281–283.

Chapter 12

Demand forecasting for companies with many branches, low sales numbers per product, and non-recurring orderings

SASCHA KURZ[1] AND JÖRG RAMBAU[2]

ABSTRACT. We propose the new Top-Dog-Index to quantify the historic deviation of the supply data of many small branches for a commodity group from sales data. On the one hand, the common parametric assumptions on the customer demand distribution in the literature could not at all be supported in our real-world data set. On the other hand, a reasonably-looking non-parametric approach to estimate the demand distribution for the different branches directly from the sales distribution could only provide us with statistically weak and unreliable estimates for the future demand. Based on real-world sales data from our industry partner we provide evidence that our Top-Dog-Index is statistically robust. Using the Top-Dog-Index, we propose a heuristics to improve the branch-dependent proportion between supply and demand. Our approach cannot estimate the branch-dependent demand directly. It can, however, classify the branches into a given number of clusters according to an historic oversupply or undersupply. This classification of branches can iteratively be used to adapt the branch distribution of supply and demand in the future.

2000 MSC: 90B90; 90B05.
Key words and phrases: revenue management, demand forecasting.

1 Introduction

Many retailers have to deal in their daily businesses with small profit margins. Their economic success lies mostly in the ability to forecast the customers' demand for individual products. More specifically: trade exactly what you can sell to your customers. This task has two aspects if your company has many branches in different regions: trade what your customers would like to buy because the product as such is attractive to them and provide a demand adjusted number of items for each branch or region.

In this paper we deal with the second aspect only: meet the branch distributed demand for products as closely as possible. The first aspect clearly also interferes with the total demand for a product over all branches. Therefore, we assume that we are given a fix total number of items per product which should be distributed over the set of branches to meet the branch-dependent demand distribution as closely as possible.

Our industry partner is a fashion discounter with more than 1 000 branches most of whose products are never replenished, except for the very few "never-out-of-stock"-products (NOS products): because of lead times of around three months, apparel replenishments would be too late anyway. In most cases the supplied items per product and apparel size lie in the range between 1 and 6.

The task can be formulated informally as follows: Given historic supply and sales data for a commodity group, find out some robust information on the demand distribution over branches in that commodity group that can be used to optimize or at least to improve the supply distribution over all branches.

We remark that trading fashion has the special feature that also the demand for different apparel size varies over the branches. In this article, however, we focus on the aspect of improving the supply distribution over all branches. The apparel size distribution problem is subject some other research in progress.

1.1 Related work

Demand forecasting for NOS items is an well-studied topic both in research and practice. The literature is overboarding, see, e. g., [1, 2, 3] for some surveys. For promotional items and other items with single, very short life cycles, however, we did not find any suitable demand forecasting methods.

The literature in revenue management (assortment optimization, inventory control, dynamic pricing) very often assumes the neglectability of out-of-stock substitution effects.

[1]Sascha Kurz, University of Bayreuth, Department of Mathematics, 95440 Bayreuth, Germany.
E-mail adress: sascha.kurz@uni-bayreuth.de
[2]Jörg Rambau, University of Bayreuth, Department of Mathematics, 95440 Bayreuth, Germany.
E-mail adress: joerg.rambau@uni-bayreuth.de

This out-of-stock substitution in the sales data of our partner, however, poses the biggest problem in our case. In our real-world application we have no replenishment, small volume deliveries per branch, lost sales with unknown or even no substitution, sales rates depending much more on the success of the individual product at the time it was offered than on the size. Therefore, estimating the absolute future demand distribution from historical sales data with no correction for out-of-stock substitution seems questionable.

Most demand forecasting tools used in practice are provided by specialized software companies. Quite a lot of software packages are available, see [6] for an overview. Our partner firm has checked several offers in the past and did – apart from the NOS segment — not find any optimization tools tailored to their needs.

1.2 Our contribution

We show that a reasonably-looking attempt to measure the demand distribution over all branches by measuring for each branch the sales over all products up to a certain day (to avoid out-of-stock substitution) does not work because of the high volatility in the sales rates of different products.

The key idea of this work is that estimating something weaker than the absolute fraction of total demand of a branch will result in stronger information that is still sufficient to improve on the demand consistency of the supply of branches.

More specifically, we propose the new Top-Dog-Index (TDI) that can measure the branch dependent deviation of demand from supply, even for very small sales amounts or short selling periods. This yields, in particular, an estimate for the direction in which the supply was different from demand in the past for each branch.

On the one hand, the TDI is a rather coarse measurement; on the other hand, we can show that on our real-world data set it is statistically robust in the sense that the TDIs of the branches relative to each other are surprisingly similar on several independent samples from the sales data and their complements.

To show the value of the information provided by the TDI, we propose a dynamic optimization procedure that shifts relative supply among branches until the deviation measured is as small as possible.

Of course, the impact of such an optimization procedure has to be evaluated in practice. This is subject of future research.

1.3 Outline of the paper

In Section 2 we state the real-world problem we are interested in. Moreover, we give an abstract problem formulation. An obvious approach of determining the demand distribution of the branches directly from historic sales data is shown to be inappropriate on our given set of sales data in Section 3. We propose our new Top-Dog-Index in Section 4. We analyze its statistical robustness and its distinctive character in clustering branches according to the deviation of the historic ratio between supply and demand. In Section 5 we describe an heuristic iterative procedure that uses the information from the Top-Dog-Indices to alter the supply distribution towards a suitable distribution that more or less matches the demand distribution over branches. An outlook and a conclusion will be given in Section 6.

2 The real-world problem and an abstract problem formulation

Our industry partner is a fashion discounter with over $1\,000$ branches. Products can not be replenished and the number of sold items per product and branch is rather small. There are no historic sales data for a specific product available since every product is sold only for one selling period. The challenge for our industry partner is to determine a suitable total amount of items of a specific product which should be bought. For this part the knowledge and experience of the buyers employed by a fashion discounter is used. We seriously doubt that a software package based on historic sales data can do better. But there is another task being more accessible for computer aided forecasting methods. Once the total amount of sellable items of a specific product is determined, one has to decide how to distribute this total amount to a set of branches B which differ in their demand. The remaining part of this paper addresses the latter task.

In the following, we formulate this problem in a more abstract way. Given a set of branches B, a set of products P, a function $S(b,p)$ which denotes the historic supply of product p for each branch b, and historic sales transactions from which one can determine how many items of a given product p are sold in a given branch b at a given day of sales d. The target is to estimate a demand $\eta(b,\tilde{p})$ for a future product $\tilde{p} \notin P$ in a given branch b, where we can use $\sum_{b\in B} \eta(b,\tilde{p}) = 1$ as normalization. This estimation $\eta(b,\tilde{p})$ should be useable as a good advice for a supply $S(b,\tilde{p})$. No further information, e. g., on a stochastic model for the purchaser behavior, is available.

3 Some real-data analysis evaluating an obvious approach

The most obvious approach to determine a demand distribution over branches is to count the sold items per branch and divide by the total number of sold items. Here we have some freedom to choose the day of the sale where we measure these magnitudes. We have to balance two competing influences. An early measurement may provide numbers of sale which are statistically too small for a good estimate. On the other hand on a late day of sales there might be too much unsatisfied demand to estimate the demand since no replenishment is possible in our application.

The business strategy of our partner implies to cut prices until all items are sold. So, a very late measurement would only estimate the supply instead of the demand. As there is no expert knowledge to decide which is the *optimal* day

of sales to measure the sales and estimate the branch dependent demand distribution we have adapted a statistical test to measure the significance of the demand distributions obtained for each possible day of counting the sold items. Given a data set D, a day of sales d let $\phi_{b,d}(D)$ be the estimated demand for branch b determined using the amounts of sold items up to day d as described above.

We normalize the values $\phi_{b,d}(D)$ so that we have $\sum_{b \in B} \phi_{b,d}(D) = 1$ for each day of sales d, where B is the set of branches. A common statistical method to analyze the reliability of a prediction based on some data universe D is to randomly partition D into two nearly equally sized disjoint samples D_1 and D_2 with $D_1 \dot{\cup} D_2 = D$ and to compare the prediction based on D_1 with the prediction based on D_2. If the two predictions differ substantially than the used prediction method is obviously not very trustworthy or statistically speaking not very robust.

In the following part of this section we analyze the robustness of the prediction $\phi_{b,d}(D)$ for every possible sales day, meaning that even an *optimal* sales day for the measurement does not provide a prediction being good enough for our purpose. To measure exactly by how much two predictions $\phi_{\cdot,d}(D_1)$ and $\phi_{\cdot,d}(D_2)$ differ we introduce the following:

Definition 3.1 *For a given sales day* d *and two samples* D_1 *and* D_2 *we define the discrepancy* δ_d *as*

$$\delta_d(D_1, D_2) := \sum_{b \in B} |\phi_{b,d}(D_1) - \phi_{b,d}(D_2)|. \quad (1)$$

Similarly we define a discrepancy between supply and demand. We compare both discrepancies in Figure 1. The result: there is no measuring day for which the discrepancy between two samples is smaller than the discrepancy between a sample and the supply. In other words, if we consider the discrepancy between supply and demand as a measure for the inconsistency of the supply distribution with the demand distribution, then either the supply is not significantly inconsistent with demand (i. e., we should better change nothing) or the measurements on the various samples are significantly different (i. e., nothing can be learned about how to correct the supply distribution).

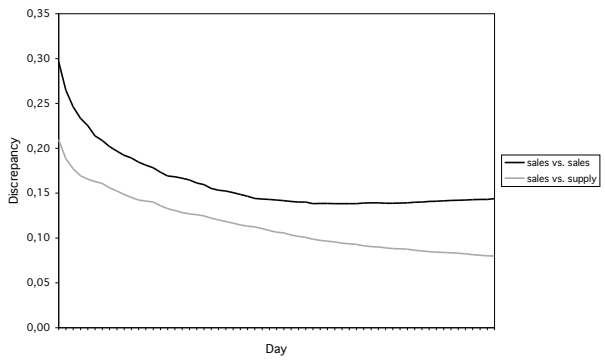


Figure 1: Discrepancy for the first 60 days.

An explanation why this obvious approach does not work well in our case is due to the small sale numbers and the interference of the demand of a branch with product attractivity and price cutting strategies. In Figure 2 we depict the change of prediction $\phi_{b,d}(D)$ over time for five characteristic but arbitrary branches

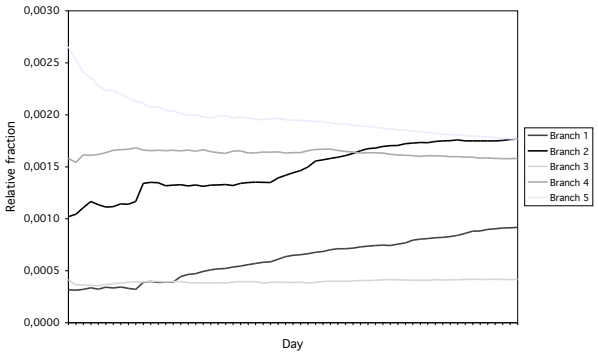


Figure 2: Prediction $\phi_{b,d}(D)$ over time.

We would like to remark that one of the authors currently advises two diploma theses which check some common parametric models for demand forecasting on historic sales data from literature. None of them gives significant information of the demand distribution over branches of our data set because the data does not exhibit any similarity to the parametric distributions coming from economic theory and the like. This may be due to the fact that the contaminating effects of promotion, mark-downs, openings/closings of competing stores prohibit a causal model for the demand. We do not claim that the assumptions of parametric demand models never hold, but in our application they are most certainly not met.

4 The Top-Dog-Index (TDI)

In the previous section we learned that in our application we cannot utilize the most obvious approach of looking at the sales distribution over the different branches on an arbitrary but fixed day of the selling period of each individual product. Since there is also no indication that any of the common parametric models for the demand estimation directly from sales data fit in our application we make no assumptions on a specific stochastic distribution of the purchaser behavior.

Our new idea dismisses the desire to estimate an absolute percental demand distribution for the branches. Instead we develop an index measuring the relative success of a branch in the competition of all branches that can be estimated from historic sales data in a stable way.

To motivate our distribution free measurement we consider the following thought experiment. For a given branch b and given product p let $\theta_b(p)$ denote the stock-out-day. Let us assume that we have $\theta_b(p) = \theta_{b'}(p)$ for all products p and all pairs of branches b, b'. In this situation one could certainly say that the branch-dependent demand is perfectly matched by the supply. Our measure tries to quantify the variation of the described ideal situation.

Therefore, we sort for each product p the stock-out-days $\theta_b(p)$ in increasing order. If for a fixed product p a branch b is among the best third according to this list it gets a *winning point* for p. If it is among the last third it is assigned a losing point for p. With B_p being the set of branches which trade product p and P being the set of the products traded by the company we can define more precisely:

Definition 4.1 *Let b be a branch. The Top-Dog-Count is defined as $W(b) :=$*

$$\left|\left\{p \in P \;\middle|\; \frac{1}{3}|B_p| \geqslant |\{b' \in B_p \mid \theta_{b'}(p) \leqslant \theta_b(p)\}|\right\}\right| \quad (2)$$

and the Flop-Dog-Count is defined as $L(b) :=$

$$\left|\left\{p \in P \;\middle|\; \frac{1}{3}|B_p| \geqslant |\{b' \in B_p \mid \theta_{b'}(p) \geqslant \theta_b(p)\}|\right\}\right|. \quad (3)$$

For a fix dampening parameter $C > 0$ let

$$TDI(b) := \frac{W(b) + C}{L(b) + C} \quad (4)$$

be the Top-Dog-Index (TDI) of branch b.

If the TDI of a branch b is significantly large compared to the TDIs of the other branches then we claim that branch b was undersupplied in the past. Similarly, if the TDI of branch b is significantly small compared to the TDIs of the other branches then we claim that branch b was oversupplied in the past. We give an heuristic optimization procedure past on this information in the section. The effect of the dampening parameter C is on the one hand that the TDI is well defined since division by zero is circumvented. On the other hand, and more important, the influence of small Top-Dog- or Flop-Dog-Counts, which are statistically unstable, is leveled to a decreased importance.

4.1 Statistical significance of the TDI

Similarly as in Section 3 we want to analyze the significance of the proposed Top-Dog-Index on some real sales data. Instead of two data sets D_1 and D_2 we use seven such samples D_i. Therefore we assign to each different product $p \in P$ a equi-distributed random number $r_p \in \{1,2,3,4\}$. The samples D_i are composed as summarized in Table 1.

$D_1 := \{p \in P$	$r_p \in \{1,2\}\}$
$D_2 := \{p \in P$	$r_p \in \{3,4\}\}$
$D_3 := \{p \in P$	$r_p \in \{1,3\}\}$
$D_4 := \{p \in P$	$r_p \in \{2,4\}\}$
$D_5 := \{p \in P$	$r_p \in \{3\}\}$
$D_6 := \{p \in P$	$r_p \in \{1,2,4\}\}$
$D_7 := \{p \in P$	$r_p \in \{1,2,3,4\}\}$

Table 1: Assignment of test sets.

For the interpretation we remark that the pairs (D_1, D_2), (D_3, D_4), and (D_5, D_6) are complementary. The whole data population is denoted by D_7 and equals P. We use $TDI(b, D_i)$ as an abbreviation of $TDI(b)$ where P is replaced by D_i.

Since the Top-Dog-Index is designed as a non-quantitative index we have to use another statistical test to assure ourselves that it gives some significant information. We find it convincing to regard the Top-Dog-Index as significant and robust whenever we have

$$\frac{TDI(b, D_i)}{TDI(b, D_j)} \approx \frac{TDI(b', D_i)}{TDI(b', D_j)} \quad (5)$$

for each pair of branches b, b' and each pair of samples D_i, D_j. In words we claim that the Top-Dog-Index is a relative index which is independent of the underlying sample if we consider a fixed universe D_7.

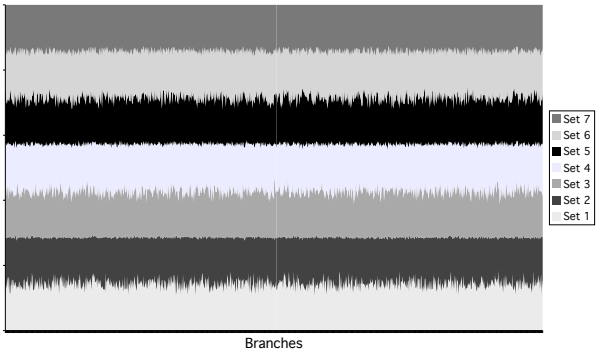


Figure 3: Relative distribution of the Top-Dog-Index on different data samples and branches.

Our first aim is to provide evidence that the $TDI(b)$ values are robust measurements. There is a nice way to look at equation (5) graphically. For each branch b let us plot a column of the relative values $\frac{TDI(b,D_i)}{\sum_j TDI(b,D_j)}$ for all i. The result for our data set is plotted in Figure 3.

To get the correct picture in the interpretation of the plot of Figure 3 we compare it to the extreme cases of deterministic numbers (i. e., $\frac{TDI(b,D_i)}{TDI(b,D_j)} = c_{ij} = c$ for all i and j), see Figure 4, and random numbers, see Figure 5.

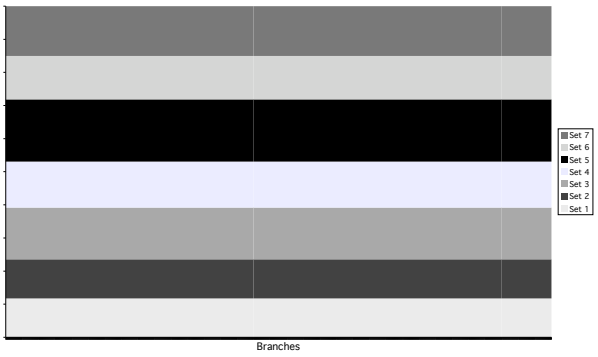


Figure 4: Relative distribution of deterministic numbers.

As a matter of fact, the regions of same color in the plot of the relative distribution of deterministic numbers in Figure 4 are formed by perfect rectangles, which are not forced in general to have equal height.

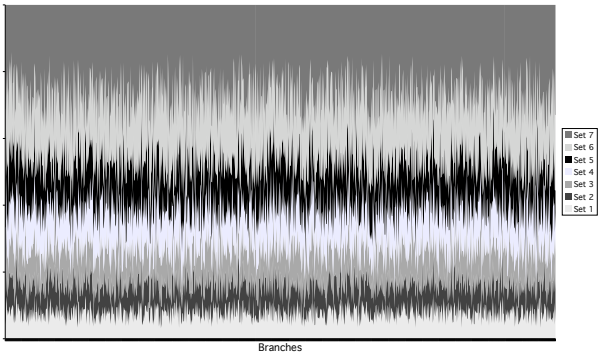


Figure 5: Relative distribution of in $[0.5, 1.5]$ equi-distributed random variables.

As an example for a *random plot* we depict in Figure 5 the relative distribution of random numbers being equi-distributed in the interval $[0.5, 1.5]$.

In the plots of Figure 3, 4, and 5 we can see that that the TDI on the given data set behaves more like a perfect deterministic estimation than a random number distribution. (Ideally, one should now quantify how large the probability is to obtain a TDI chart as in Figure 3 by a random measurement.) So there is empirical evidence that the TDI gives some stable information. As a comparison of the TDI and the method described in Section 3 we depict the corresponding relative distribution for measuring day 5 in Figure 6. Although a measurement on this day was the best we could find, it still produces more severe outliers than the TDI measurement.

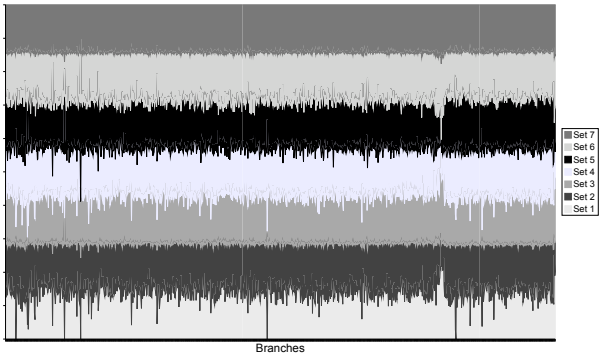


Figure 6: Relative distribution of $\phi_{\cdot,5}$.

Now the question remains whether this information is enough to cluster branches into oversupplied and undersupplied ones. More directly: is the distinctive character of the TDI strong enough? We consider this question in the next subsection. How the TDI information can be used to iteratively improve the branch dependent ratio between supply and demand will be the topic of Section 5.

4.2 The distinctive character of the TDI

If one forces the values of the TDIs to be contained in an interval of small length, then clearly a plot of the relative distributions would look like the plot of Figure 4. As an thought experiment just imagine how Figure 5 would look like, if we would use random numbers being equi-distributed in the interval $[0.9, 1.1]$ instead of being equi-distributed in the interval $[0.5, 1.5]$

Forcing the possible values of the TDIs in an interval of small length is feasible by choosing a sufficiently large dampening parameter C. So this parameter has to be chosen with care. We remind ourselves that we would like to use the TDIs to cluster branches. Therefore the TDIs should vary over a not to small range of values to have a good distinctive character. Clearly by using the TDI we can only detect possible improvements if the supply versus demand ratio actually inadequate in a certain level. In Figure 7 we have plotted the occurring TDIs of our data set to demonstrate there is indeed some variation of values in our data set, no matter which sample we consider (let alone the data universe). As one can see the TDIs vary widely enough to distinguish between historically under- and oversupplied branches.

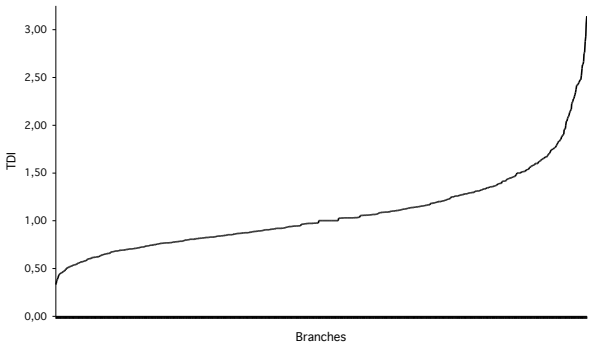


Figure 7: Occurring TDIs.

5 The heuristic supply optimization procedure based on the TDI

So far we have developed and statistically stable index capturing the deviation of supply from demand for each branch. Now we have to specify how we can use this information to improve the branch dependent ratio between supply and demand.

Let $S(b)$ be the historic supply of branch b being normalized so that we have $\sum_{b \in B} S(b) = 1$. Our aim is to estimate supplies $\tilde{S}(b)$, also fullfilling $\sum_{b \in B} \tilde{S}(b) = 1$, which are more appropriate concerning the satisfaction of demand by using the TDI information.

Therefore let us partition the interval $(0, \infty)$ of the positive real numbers into a given number of l appropriate chosen intervals $\mathcal{I}_j$. Further we need l appropriately chosen in-

crement numbers Δ_j. Our proposed update formula for the estimated branch dependent demand is given by

$$\tilde{S}(b) = \frac{S(b) + \Delta_{j(b)}}{\sum_{b' \in B} S(b') + \Delta_{j(b')}} \tag{6}$$

for all branches b, where $j(b)$ is the unique index with $TDI(b) \in \mathcal{I}_{j(b)}$.

We do not claim that the $\tilde{S}(b)$ are a good estimation for the demand of all branches. Our claim is that they approach a good estimation of the branch dependent demand if one iterates the described procedure over several rounds and carefully chooses the increment numbers Δ_j, which may vary over the time.

Once you have a new proposal $\tilde{S}(b)$ of the relative supply for each branch b, one only has to fit it into an integer valued supply for each new product p'. Given the problem of apparel size assortment and pre-packing, this is easier said than done and is subject of further studies.

In contrast to the other sections here we are somewhat imprecise and there is a lot of freedom, e. g., how to choose the intervals $\mathcal{I}_j$ and increment numbers Δ_j. That is for several reasons. On the one hand that is exactly the point where some expert from the business should calibrate the parameters to specific data of the company. One the other hand there are quite a lot of possibilities how to do it in detail. Their analysis will be a topic of future research. For the practical application we account rather simple than sophisticated variants in the first step.

6 Conclusion and outlook

We have introduced the new Top-Dog-Index which is capable to cluster branches of a retail company into oversupplied and undersupplied branches at a statistically robust level where more direct methods fail. The robustnest of this method is documented by some statistical tests based on real-world data.

We have also documented that the distinctive character of the proposed TDI is significant for our application: for the first time we can gain information about the demand distribution of branches from historic sales data on only few products with volatile success in sales rates and with unknown stock-out substitution effects, and this information does not depend too much on the sample of the sales data universe out of which the TDI is computed.

For the dynamic optimization of the supply distribution among branches, some fine tuning of parameters is needed; for a real-world implementation these details have to be fixed. This, together with a field study of the impacts of an improved supply distribution are research in progress.

Bibliography

[1] J. S. Armstrong (ed.), *Principles of forecasting: A handbook for researchers and practitioners*, Kluwer, 2001.

[2] A. Kok and M. Fisher, *Demand estimation and assortment optimization under substitution: Methodology and application*, Oper. Res. (to appear).

[3] S. Makridakis, S. Wheelwright, and R. Hyndman, *Forecasting:methods and applications*, Wiley, 2004.

[4] M. K. Mantrala and S. Rao, *A decision-support system that helps retailers decide order quantities and markdowns for fashion goods*, Interfaces **31** (2001), no. 3b, 146–165.

[5] B. P. Pashigian, *Demand uncertainty a sales: A study of fashion and markdown pricing*, The American Economic Review **78** (1988), no. 5, 936–953.

[6] J. Yurkiewicz, *Software survey: Forecasting 2000*, OR/MS Today **27** (2000), no. 1.

Chapter 13

The Top-Dog Index: A New Measurement for the Demand Consistency of the Size Distribution in Pre-Pack Orders for a Fashion Discounter with Many Small Branches

SASCHA KURZ[1], JÖRG RAMBAU[2], JÖRG SCHLÜCHTERMANN[3], AND RAINER WOLF[4] [5]

ABSTRACT. We propose the new Top-Dog-Index, a measure for the branch-dependent historic deviation of the supply data of apparel sizes from the sales data of a fashion discounter. A common approach is to estimate demand for sizes directly from the sales data. This approach may yield information for the demand for sizes if aggregated over all branches and products. However, as we will show in a real-world business case, this direct approach is in general not capable to provide information about each branch's individual demand for sizes: the supply per branch is so small that either the number of sales is statistically too small for a good estimate (early measurement) or there will be too much unsatisfied demand neglected in the sales data (late measurement). Moreover, in our real-world data we could not verify any of the demand distribution assumptions suggested in the literature. Our approach cannot estimate the demand for sizes directly. It can, however, individually measure for each branch the scarcest and the amplest sizes, aggregated over all products. This measurement can iteratively be used to adapt the size distributions in the pre-pack orders for the future. A real-world blind study shows the potential of this distribution free heuristic optimization approach: The gross yield measured in percent of gross value was almost one percentage point higher in the test-group branches than in the control-group branches.

[1]Sascha Kurz, Fakultät für Mathematik, Physik und Informatik, Universität Bayreuth, Germany.
E-mail adress: sascha.kurz@uni-bayreuth.de
[2]Jörg Rambau, Fakultät für Mathematik, Physik und Informatik, Universität Bayreuth, Germany.
E-mail adress: joerg.rambau@uni-bayreuth.de
[3]Jörg Schlüchtermann, Fakultät für Rechts- und Wirtschaftswissenschaften, Universität Bayreuth, Germany.
E-mail adress: j.schluechtermann@uni-bayreuth.de
[4]Rainer Wolf, Betriebswirtschaftliches Forschungszentrum für Fragen der mittelständischen Wirtschaft e. V. an der Universität Bayreuth, Germany.
E-mail adress: rainer.wolf@uni-bayreuth.de
[5]This research was supported by the *Bayerische Forschungsstiftung, Project* DISPO—A Decision Support for the Integrated Size and Price Optimization

2000 MSC: 90B05; 90B90.
Key words and phrases: revenue management, size optimization, demand forecasting, Top-Dog-Index, field study, parallel blind testing.

1 Introduction

The financial performance of a fashion discounter depends very much on its ability to predict the customers' demand for individual products. More specifically: trade exactly what you can sell to your customers. This task has two aspects: offer what your customers would like to wear because the product as such is attractive to them and offer what your customers can wear because it has the right size.

In this paper, we deal with the second aspect only: meet the demand for sizes as accurately as possible. The first aspect, demand for products, is a very delicate issue: Products in a fashion discounter are never replenished because of lead times of around three months. Therefore, there will never be historic sales data of an item at the time when the order has to be submitted (except for the very few "never-out-of-stock"-items, NOS items, for short).

When one considers the knowledge and experience of the professional buyers employed at a fashion discounter—acquired by visiting expositions, reading trade journals, and the like—it seems hard to imagine that a forecast for the demand for a product could be implemented in an automated decision support system at all. We seriously doubt that the success of fashion product can be assessed by looking at historic sales data only. In contrast to this, the demand for sizes may stay reasonably stable over time to extract useful information from historic sales data.

In the historic sales data the influences of demand for products and demand for sizes obviously interfere. Moreover, it was observed at our partners' branches that the de-

mand for sizes seems not to be constant over all around 1 200 branches.

The main question of this work is: how can we forecast the demand for sizes individually for each branch or for a class of branches?

1.1 Related Work

Interestingly enough, we have not found much work that exactly deals with our task. It seems that, at first glance, the problem of determining the size distribution in delivery prepacks can be considered as simple regression once you have historic sales data: Just estimate the historic size profile and fit your delivery to that. At least two trivial US-patents [6, 8] have recently been granted and published along these lines (which witnesses that the US patent system may not have employed the necessary expertise in their patent evaluations ...).

In our problem, however, the historic sales data is *not* necessarily equal to the historic demand data, and it is interesting how to find the demand data in the sales data in the presence of unsatisfied demand and very small delivery volumes per branch and per product.

The type of research closest to ours seems to be classified as assortment optimization. In a sense, we want to decide on the start inventory level of sizes in a pre-pack for an individual product in an individual branch. (Let us ignore for a moment that these unaggregated inventory levels are very small compared to other inventory levels, e. g., for grocery items.)

Mostly, the successful approaches deal rather with NOS items than with perishable and not replenishable fashion goods. For example, assortment optimization in the grocery sector [4, Section 4] — one of the very few papers documenting a field study — can usually neglect the effects of stockout substitution in sales data, which make demand estimation from sales data much more reliable. There is work on the specific influence of substitution on the *optimization of expected profit* (see, e. g., [7]), but the problem of how to *estimate demand parameters* from low-volume sales data in the presence of stock-out substitution remains.

Much more work has been published in the field of dynamic pricing, where in one line of research pricing and inventory decisions are linked. See [2, 3] for surveys.

A common aspect of all cited papers (and papers cited there) that separates their research from ours is the following: those papers, in some sense, postulate the possibility to estimate a product's demand in an individual branch directly from sales data, in particular from sales rates. In our real-world application we have no replenishment, small delivery volumes per branch, lost sales with unknown or even no substitution, and sales rates depending much more on the success of the individual product at the time it was offered than on the size. Therefore, estimating future absolute demand data from historic sales data directly seems to need extra ideas, except maybe for the data aggregated over many branches.

Size optimization can be found in a few offers of retail optimization systems, like 7thOnline (`http://www.7thonline.com`). It is not clear, on what kind of research these products are based and under which assumptions they work well. Our partner firm has checked several offers in the past and did not find any optimization tools that met their needs.

1.2 Our contribution

The main result of this work is: a useful forecast for the demand for sizes on the level of individual branches is too much to be asked for, but historic information about which sizes have been the scarcest and which sizes have been the amplest ones can be obtained by measuring the new *Top-Dog-Index (TDI)*. The TDI can be utilized in a dynamic heuristic optimization procedure, that adjusts the size distributions in the branches' corresponding pre-packs accordingly until the difference between the scarcest and the amplest sizes can not be improved anymore. The main benefit of the TDI: it measures the consistency of historic supply with sizes with the historic demand for sizes in a way that is not influenced by the attractivity of the product itself. This way, we can aggregate data over all products of a product group, thereby curing the problem with small delivery volumes per branch and product and size.

The potential of our TDI-approach is shown in a blind-study with 20 branches and one product group (womens' outer garments). Ten branches randomly chosen from the 20 branches (test-branches) received size pre-packs according to our heuristics' recommendations, ten received unchanged supply (control-branches). The result: One percentage point increase of gross yield per merchandise value for the test-branches against the control-branches. A conservative extrapolation of this result for our partner would already mean a significant increase of gross yield.

We have not seen any field study of this type documented in the literature so far. The only documented yield management studies in the apparel retail business (e. g., for dynamic pricing policies) try to prove the success of their methods by showing that they would have gained something on the set of data that was used to estimate the parameters of the model [1]. Such tests are very far from reality: a by-product of our study is that in our case it makes absolutely no sense to compare gross yield data of a fashion discounter across seasons, because the differences between the yields in different seasons are much larger than the differences caused by anything we are interested in. This was the reason for us to use the parallel blind testing instead.

Further tests are planned by our partner on a larger scale in the near future.

1.3 Outline of the paper

In Section 2, we briefly restate the real-world problem we are concerned with. In Section 3, we show our experience with straight-forward estimators for the demand distribution on sizes. In Section 4, we introduce the new Top Dog Index, which is utilized in Section 5 in a heuristic optimization

procedure. Section 6 is the documentation of a field study containing a blind testing procedure among two groups of branches: one supplied with and one supplied without the suggestions from the first step of the optimization heuristics. We summarize the findings in Section 7, including some ideas for further research.

2 The real-world problem

In this section, we state the problem we are concerned with. Before that we briefly provide the context in which our problem is embedded.

2.1 The supply chain of a fashion discounter

As in most other industries the overall philosophy of supply chain management in fashion retailing is to coordinate the material flow according to the market demand. The customer has to become the "conductor" of the "orchestra" of supply chain members. Forecasting the future demand is, therefore, crucial for all logistics activities. Special problems occur in cases like ours, when the majority of inventory items is not replenished, because the relationship between lead times and fashion cycles makes replenishments simply impossible. The resulting "textile pipeline" has strong interdependencies between marketing, procurement, and logistics.

The business model of our real world problem bases on a strict cost leadership strategy with sourcing in low cost countries, either East Asia or South East Europe. The transportation time is between one and six weeks, economies of scale are achieved via large orders.

2.2 Internal stock turnover of pre-packs

The material flow in our problem is determined by a central procurement for around 1 200 branches. All items are delivered from the suppliers to a central distribution center, where a so called "slow cross docking" is used to distribute the items to the branches. Some key figures may give an impression of the situation: 32 000 square meters, 80 workers, 30 000 tons of garment in 10 million lots per year. Each branch is delivered once a week with the help of a fixed routing system. This leads to a sound compromise between inventory costs and costs of stock turnover.

There are two extreme alternatives for the process of picking the items. The retailer can either work with one basic lot and deliver this lot or integer multiples of it for every article to the branches, or he develops individual lots for the shops. At the beginning of our analysis most of the articles were picked in one basic lot, but more than 40 other constellations were used additionally. The costs for picking the items are not relevant for the following analysis because only minor changes of internal processes are necessary. Only in later phases of our project a detailed analysis of the picking costs will be needed.

2.3 The problem under consideration

Recall that the stock turn-over is accelerated by ordering pre-packs of every product, i. e., a package containing a specified number of items of each size. We call the corresponding vector with a non-negative integral entry for every size a *lot-type*.

In this environment, we focus on the following problem: Given historic delivery data (in terms of pre-packed lots of some lot-type for each branch) and sales data for a group of products for each branch, determine for each branch a new lot-type with the same number of items that meets the relative demand for sizes more accurately.

In particular, find out from historic *sales* data some information about the relative *demands* for sizes. We stress the fact that stockout substitution in the data can not be neglected since replenishment does not take place (unsatisfied demand is lost and does not produce any sales data). We also stress the fact that we are not trying to improve the number of items delivered to each branch but the distribution of sizes for each branch only.

3 Some real-data analysis evaluating seemingly obvious approaches

To deal with the problem described in Section 2 our partner has provided us some historic sales data for approximately 1 200 branches over a time period of 12 months. The task was to forecast the size distribution of the future demand for each branch. A possible size distribution is depicted in Figure 1.

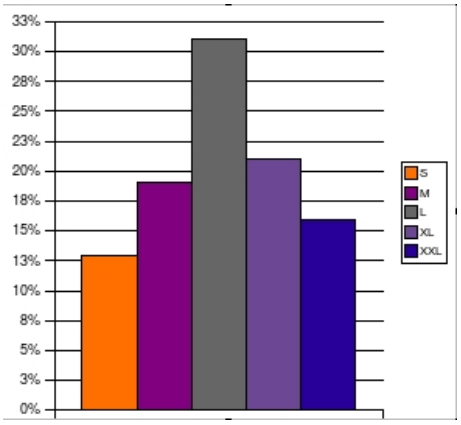


Table 1: A possible size distribution of the demand.

The most obvious way to determine a size distribution as in Figures 1 is to count the number of sold items per size and divide by the total number of sold items. Here we have some freedom to choose the day of the sale where we measure the amounts. We have to balance two competing facets. An early measurement may provide sales figures which are statistically too small for a good estimate while a late measurement may suffer from unsatisfied demand that is not present in the sales data.

The business strategy of our partner implies to cut prices until all items are sold. So, a very late measurement would only estimate the supply instead of the demand. As there is no expert knowledge to decide which is the *optimal* day of sales to count the amounts and estimate the size distribution we have applied a statistical test to measure the significance of the size distributions obtained for each possible day of counting the sold items.

Given a data set D, a day of sales d, and a size s let $\phi_{s,d}(D)$ be the estimated demand for size s measured on day d as described above. We normalize so that $\sum_s \phi_{s,d}(D) = 1$ for each day and each data set. Our statistical test partitions the original data set D randomly into two disjoint data sets D_1 and D_2. Naturally we would not trust a forecast ϕ whenever $\phi_{s,d}(D_1)$ and $\phi_{s,d}(D_2)$ are too far apart. Statistically speaking ϕ would not be robust in that case. To measure more precisely how far apart $\phi_{s,d}(D_1)$ and $\phi_{s,d}(D_2)$ are, we define the discrepancy δ_d as $\delta_d(D_1, D_2) := \sum_s |\phi_{s,d}(D_1) - \phi_{s,d}(D_2)|$. In Figure 2 we have depicted the average discrepancy over all branches for the first 60 days of sale.

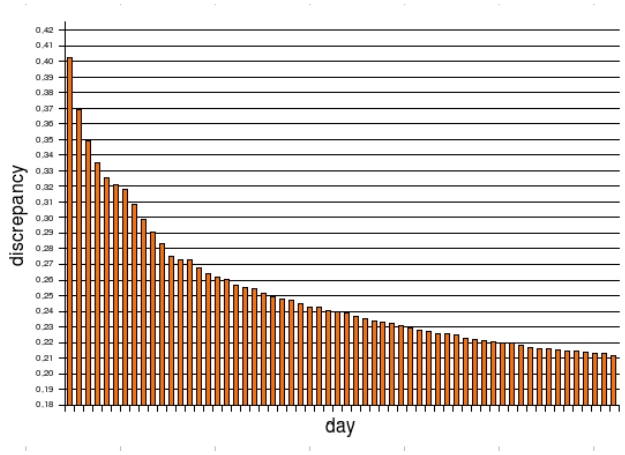


Table 2: Average discrepancy δ_d over all branches.

For our practical application the significance of ϕ is too low for all possible days of measurement. We remark that in our examined data set the discrepancy δ_d tends to 0.19 as d tends to the end of the sales period. An explanation why this obvious approach does not work in our case is due to the small sales numbers and the interference of product attractivity and price cutting strategies.

Even when we tried to use only use ordinal information generated from an estimated size distribution (some size is too scarce, some size too ample), we encountered different results from different samples (same size was scarce in one sample and ample in the other).

In the same manner we have checked some common parametric models (mostly based on an estimation of a constant sales rate for each individual product-size pair) for demand forecasting on historic sales data from literature. None of them gives significant information of the size distribution of the demand of our data set, since the sales rates vary drastically and depend more on the attractivity of the products than on the sizes. For the details we refer to two diploma theses [5, 9] from our group.

4 The Top-Dog-Index (TDI)

In the previous section we learned that in our application, first, we cannot trust the common parametric models for the demand distribution and, second, measuring a size profile directly from sales data may lead to more or less random decisions. The main reason for the latter is the former and the interference between attractiveness of offered products and compatibility of offered sizes. Because of this, the stockout saturation of sales data happens at almost all times during the sales period, and thus aggregating the sales data of different products yields no reasonable information.

Our new idea throws overboard the desire to estimate an absolute size profile of the demand in every branch. Instead, we try to define a measure for the scarcity of sizes during the sales process that can be estimated from historic sales data in a stable way.

The following thought experiment is the motivation for our distribution free measure: Consider a product, for which in a branch all sizes are sold out at the very same day. This can be regarded as the result of an ideal balance between sizes in the supply. Our measure tries to quantify the deviation of this ideal situation in historic sales data. How can this be done? In the following, we extract data of a new type from the sales process.

Fix a delivery period $\Delta := [0, T]$ from some day in the past normed to day 0 to day T. Let B be the set of all branches that are operating in time interval Δ, and let P be the set of all products in a group delivered in time interval Δ in sizes from a size set S. We assume that in each branch the product group can be expected to have homogeneous demand for sizes throughout the time period. Fix $b \in B$. For each $p \in P$ and for each $s \in S$ let $\theta_b(p, s)$ be the stockout-day of size s of product p, i. e., the day when the last item of p in size s was sold out in branch b.

Fix a size $s \in S$. Our idea is now to compare for how many products p size s has the earliest stockout-day $\theta(p, s)$ and for how many products p size s has latest stockout-day $\theta(p, s)$. These numbers have the following interpretation: If for many more products the stockout-day of the given size was first among all sizes, then the size was scarce. If for many more products the stockout-day of the given size was last among all sizes, then the size was ample.

In order to quantify this, we use the following approach. (In fact, it is not too important how we exactly quantify our idea, since we will never use the absolute quantities for decision making; we will only use the quantities relative to each other.)

Definition 4.1 (Top-Dog-Index) *Let $s \in S$ be a size and b be a branch.*

The Top-Dog-Count $W_b(s)$ *for s in b is defined as*

$$W_b(s) := \left| \left\{ p \in P \mid \theta_b(p, s) = \min_{s' \in S} \theta_b(p, s') \right\} \right|$$

and the Flop-Dog-Count $L_b(s)$ *in* b *is defined as*

$$L_b(s) := \left| \left\{ p \in P \mid \theta_b(p,s) = \max_{s' \in S} \theta_b(p,s') \right\} \right|.$$

Moreover, for a fixed dampening parameter $C > 0$ *let*

$$TDI_b(s) := \frac{W_b(s) + C}{L_b(s) + C}$$

be the Top-Dog-Index (TDI) *of Size* s *in Branch* b.

In the data of this work, we used $C = 15$.

4.1 Statistical significance of the Top-Dog-Index

In a similar way as in Section 3 we want to analyze the significance of the proposed Top-Dog-Index. Since this method is supposed to be applied to a real business case, we analyze the statistical significance in more detail. Instead of two data sets D_1 and D_2 as in Section 3 we utilize seven such sets D_i. Therefore, we assign a random number in $\{1,2,3,4\}$ to each different product. The sets are composed of the data of products where the corresponding random number lies in a characteristic subset of $\{1,2,3,4\}$. See Table 3 for the assignment. For the interpretation we remark that the pairs (D_1, D_2), (D_3, D_4), and (D_5, D_6) are complementary. The whole data set is denoted by D_7.

D_1	$\{1,2\}$
D_2	$\{3,4\}$
D_3	$\{1,3\}$
D_4	$\{2,4\}$
D_5	$\{3\}$
D_6	$\{1,2,4\}$
D_7	$\{1,2,3,4\}$

Table 3: Assignment of test sets.

Since the Top-Dog-Index is designed to provide mainly ordinal information, we have to use another statistical test to make sure that it yields some significant information. Let $TDI_b(s, D_i)$ denote the Top-Dog-Index in Branch b of Size s computed from the data in Data Set D_i. We find it convincing to regard the ordinal information generated by the Top-Dog-Index as robust whenever we have

$$\begin{aligned} & TDI_b(s, D_i) \gg TDI_b(s', D_i) \\ \iff \quad & TDI_b(s, D_j) \gg TDI_b(s', D_j) \end{aligned} \tag{1}$$

for each pair of sizes s, s' and each pair of data sets D_i, D_j. In words: the order of Top-Dog-Indices of various sizes does not change significantly when computed from a different sample. The following is a sufficient condition for this to happen:

$$\frac{TDI_b(s, D_i)}{\sum_j TDI_b(s, D_j)} \approx const_i \tag{2}$$

Our first aim is to provide evidence that the $TDI_b(s)$ values are robust measurements in this sense. There is a nice way to look at Equation (2) graphically. Let us fix a size s. For each branch b let us plot a column of the relative values $\frac{TDI_b(s,D_i)}{\sum_j TDI_b(s,D_j)}$ for all branch-size combinations and for all i. The columns corresponding to the same branch-size combination but different samples are stacked on top of each other. This way, each column stack has height one in total, and the size relations of the measurements based on the different samples can be assessed right away. To show the robustness of the mean and the median, we have added an additional column for each of them: The median of the seven estimates corresponds to the height of the top-most column, the mean to the height of the second-to-top-most column. The goal now is not to read off the exact values, but to get an intuitive impression how heavy an estimate depends on the data subset it was computed from.

The Top-Dog-Indices are plotted this way in Figure 1.

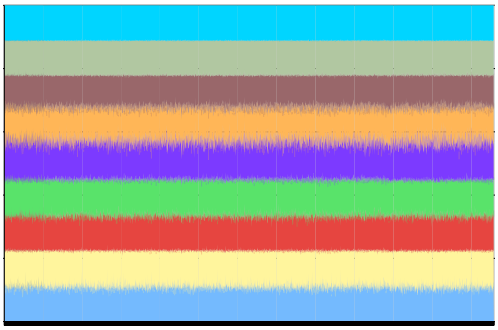


Figure 1: Relative distribution of the Top-Dog-Index on seven data subsets for all branch-size combinations; the two top-most columns are median and mean, resp.

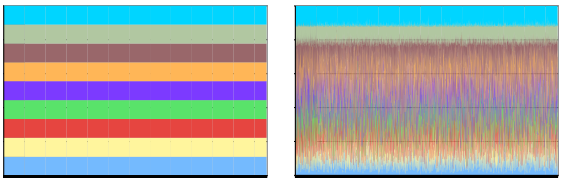


Figure 2: Relative distribution of deterministic and random numbers on seven data subsets and branches; the two top-most columns are median and mean, resp.

In order to provide some intuitively clear reference data to compare to the plot of Figure 1, we present the corresponding plots for the two extreme cases of deterministic numbers (i. e., $TDI_b(s, D_i) = TDI_b(s, D_j)$ for all i and j) and totally random numbers in Figure 2. In the complete deterministic case the areas of same color form perfect rectangles. In the random case the areas of same color corresponding to the data subsets form zig-zag lines; the median has fewer zig-zag than the mean, but both are quite stable because the random numbers are all from the same distribution.

It is immediately obvious that the plot of Figure 1 looks

more like the plot in the deterministic case as the plot in the random case. It is interesting to note that the dampening parameter in the computation of the TDI does indeed influence the amount of noise in the plot but had almost no influence on the order of TDI values, which is what we intend to use.

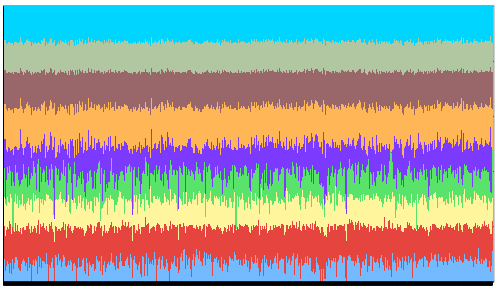


Figure 3: Relative distribution of sales up to Day 0 on on seven data subsets for all branch-size combinations; the two top-most columns are median and mean, resp.

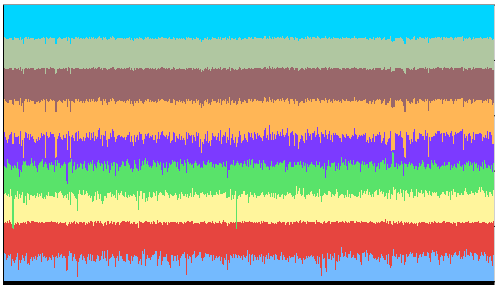


Figure 4: Relative distribution of sales up to Day 12 on on seven data subsets for all branch-size combinations; the two top-most columns are median and mean, resp.

Recall that we have analyzed the direct estimation of size distributions for each branch by looking at two complementary samples of the data set. In order to be able to directly compare the results of that attempt to Figure 1, we plot the size distributions for the same seven data subsets. This is shown in Figure 3 for an estimate from the sales up to Day 0 (the first day in the sales period) and in Figure 4 for an estimate from the sales up to Day 12.

It can be seen that Day 0-estimates are extremely dependent on the data subset even if only the relative values are taken into account, i. e., the estimates are not robust even in the weak sense measured in the plot. The Day-12 estimates are more robust, but not even close to what the TDI achieves in Figure 1. Moreover, as we said, those estimates are already quite close to the supply because of unsatisfied demand, and will therefore fail to measure the size distribution of the demand.

Generating a robust statistics for a set of data is, of course, easy: just assign the same deterministic number to each object. This is not what we want since this does not carry any useful information. We claim that the Top-Dog-Index exhibits differences between demands for sizes in each branch individually. That this is indeed the case, follows from the Top-Dog-Indices that we encountered in the field study documented below. But let us first discuss which actions we could take to improve the size distributions of the branches' supply.

So far we have argued that the Top-Dog-Index produces size-related information in a robust way, while other methods fail (at least for our given real-world data). In the next section we describe a procedure to harmonize demand and supply with respect to the size distribution. In Section 6 we provide evidence via a real-world blind study that this procedure helps to raise the gross yield in reality; thus, the Top-Dog-Index is correlated to the size distribution of the demand.

5 The heuristic size optimization procedure based on the TDI

Interesting for us is not the absolute TDI of a size but the TDIs of all sizes in a branch compared to each other, i. e., the ordinal information implied by the TDIs. The size with the maximal TDI among all sizes can be interpreted as the scarcest (the one that was sold the fastest) size in that branch; the size with the minimal TDI among all sizes can be interpreted as the amplest size in that branch. Of course, we have the problem of deciding whether or not a maximal TDI is significantly larger than the others. Since the absolute values of the TDIs have no real meaning we did not even try to assess this issue in a statistically profound way.

Our point of view is again that absolute forecasting is too much to be asked for. Therefore, we resort to a dynamic heuristic optimization procedure: sizes with "significantly" large TDIs (Top-Dog-Sizes) should receive larger volumes in future deliveries until their TDIs do not improve anymore, while sizes with "significantly" small TDIs (Flop-Dog-Sizes) need smaller supplies in the future. Whenever this leads to oversteering, the next TDI analysis will show this, and we go back one step. This is based on the assumption that the demand for sizes does not change too quickly over time. If it does then optimization methods based on historic sales data are useless anyway.

Let us describe our size distribution optimization approach in more detail.

We divide time into delivery periods (e. g., one quarter of a year). We assume that the sales period of any product in a delivery period ends at the end of the next delivery period (e. g., half a year after the beginning of the delivery period). Recall that a size distribution in the supply of a branch is given by a pre-pack configuration: a package that contains for each size a certain number of pieces of a product (compare Subsection 2.2).

We want to base our delivery decisions for an up-coming period on

- the pre-pack configuration of the previous period and
- the TDI information of the previous period giving us the deviation from the ideal balance

According to given restrictions from the distribution system, we assume that only one pre-pack configuration per branch is allowed. We may use distinct pre-pack configurations for different branches, though.

Since we are only dealing with the size distribution of the total supply but not with the total supply for a branch itself, the total number of pieces in a pre-pack has to stay constant. Since the TDI information only yields aggregated information over all products in the product group, all products of this group will receive identical pre-pack configurations in the next period, as desired.

In order to adjust the supply to the demand without changing the total number of pieces in a pre-pack, we will remove one piece of a Flop-Dog-Size from the pre-pack and add one piece of a Top-Dog-Size instead. At the end of the sales period (i. e., at the end of the next delivery period), we can do the TDI-analysis again and adjust accordingly.

Given the usual lead times of three months this leads to a heuristics that reacts to changes in the demand for sizes with a time lag of nine month to one year. Not exactly prompt, but we assume the demand for sizes to be more or less constant over longer periods of time.

The most interesting question for us was how much, in practice, could be gained by performing only one step of the heuristics explained above.

6 A real-world blind study

In this section we describe the set-up and the results of the blind study carried out by our partner. A summary of parameters can be found in Table 4.

Test period :	April through June 2006 (3 Months)
Data collection period:	April through September 2006 (6 Months)
Branches:	20 branches with unbalanced TDIs randomly classified into 10 test and 10 control branches
Pieces of merchandise:	approx. 4 000 pieces for all test and control branches
Merchandise value:	approx. 30 000 €

Table 4: Summary of parameters of the blind study.

6.1 Selection of branches

A reasonable selection of branches for a test and a control group had to meet essentially three requiremens: first, only those branches should be chosen whose TDI indicated that, in the past, the supply by sizes did not meet the demand for sizes; secondly, no branch should be chosen, where other tests were running during the test period; thirdly, the assignment of branches to test- and control group should be completely random. The reason for the third aspect was that this way all other influences on the gross yield than the selection of sizes would appear similarly in both the test group and the control group and, thus, would average out evenly.

test group	**special advertising**		**no advertising**	
Branch	remove	add	remove	add
1	L	XL	L	XL
2	M	XL	M	XL
3	L	XL	S	XL
4	L	XL	S	XL
5	M	XL	M	XL
6	M	XL	M	XL
7	M	XL	M	XL
8	L	XL	S	XL
9	M	XL	M	XL
10	M	XL	M	XL

Table 5: How the pre-packs were modified in the test group.

We suggested a set of 50 branches with interesting Top-Dog-Indices to our partner. Out of these 50 branches, our partner chose 20 branches where a potential re-packing of pre-packs would be possible. This set of 20 branches was fixed as the set of branches included in our blind study.

After that, a random number between 0 and 1 was assigned to each of the 20 branches. The 10 branches with the smallest random numbers were chosen to be the test group, the rest was taken as the control group.

6.2 Handling of pre-packs

Next, we had to specify the modifications to the size distribution in pre-packs on the basis of the TDI information. It turned out that additional side constraints had to be satisfied: Whenever a product would appear in an advertisement flyer, the pre-pack had contained at least one piece in each of the four main sizes S, M, L, and XL. That means, although sometimes the TDI suggested that S was the amplest size, we could not remove the only piece in S from the pre-pack. We removed a piece of the second amplest size (M or L, of which there were two in the unmodified pre-packs) instead. To all branch deliveries, one additional piece of XL was packed, because this was the scarcest size in every branch in the test group. This way, the total number of pieces was unchanged in every pre-pack, as suggested in Section 5.

Since all orders had been placed well before the decision to conduct a blind study, our partner re-packed all pre-packs for the test branches according to Table 5.

6.3 Time frame

The test included two relevant time periods: the first period from which the TDI data was extracted and the test period in which the recommendations based on the TDI data were implemented for the test group.

The TDI data was drawn from a delivery period of nine month (January through September 2005) and a sales period

of twelve month (January through December 2005).

The test data was drawn from a delivery period of three months (April through June 2006) and a sales period of six months (April through September 2006).

6.4 Data collection

In order to sort out contaminated data easily, our partner agreed to take stock to check inventory data for correctness every month. To receive a good estimate for the financial benefit of the supply modification proposed by the procedure described in the previous section we had defined some criteria how to detect contaminated data automatically via a computer program.

6.5 Data analysis

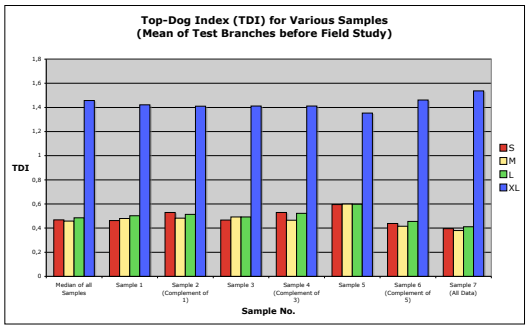


Figure 5: TDI in the test branches from historic sales data.

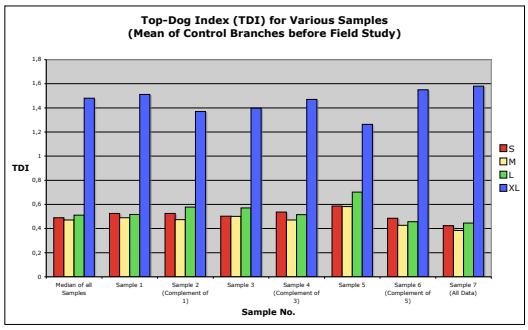


Figure 6: TDI in the control branches from historic sales data.

The initial situation was given by the data set described in Section 3. In Figure 5 we have depicted the initial Top-Dog-Indices for the test branches and in Figure 6 the initial Top-Dog-Indices for the control branches. Moreover, Figures 7 and 8 finally show the situations for the individual branches.

The analysis of our field study was intended to answer the following two main questions: are the Top-Dog-Indices better distributed in the test branches than in the control branches, and, if yes, does this have a significant monetary impact?

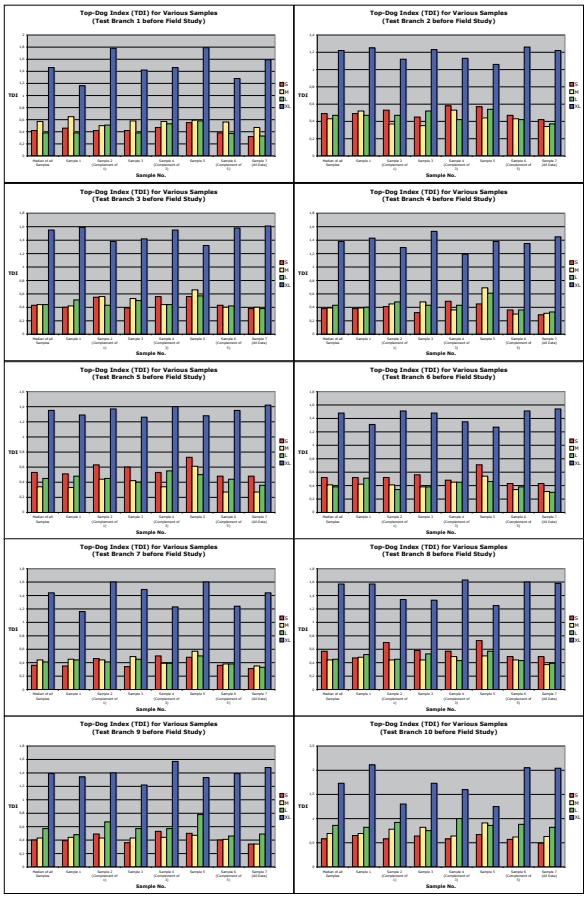


Figure 7: Individual TDIs in test branches from historic sales data.

To investigate the latter, we had to analyze some monetary variables. The most important monetary indices for our partner are the gross yield and the last price. The gross yield directly shows how much turnover was lost using a price cutting strategy to sell out all items. The last price tells us how far one was forced to dump items provoked by an inadequate size distribution of the supply. Since the values of different products vary widely we only consider relative values. So, the gross yield is given by the ratio of realized turnover and theoretic revenue without cutting the prices.

Since we had to deal with a large amount of lost or inconsistent data, we have applied two ways of evaluating gross yield and last price. Imagine that your data says that you have sold 10 items of a given article in a given branch but that the supply was only 8 items. Or the other way round that the initial supply was 10 items, during the time 8 items were sold, but all items are gone. Both of the described situations occurred significantly often in our data set.

Our first strategy to evaluate the given data was to ignore inconsistent data. In the first case, 8 sale transactions are

consistent. For the remaining two items the corresponding supply transaction is missing. So, we simply ignore these transactions. In the other case we would ignore the supply of two items.

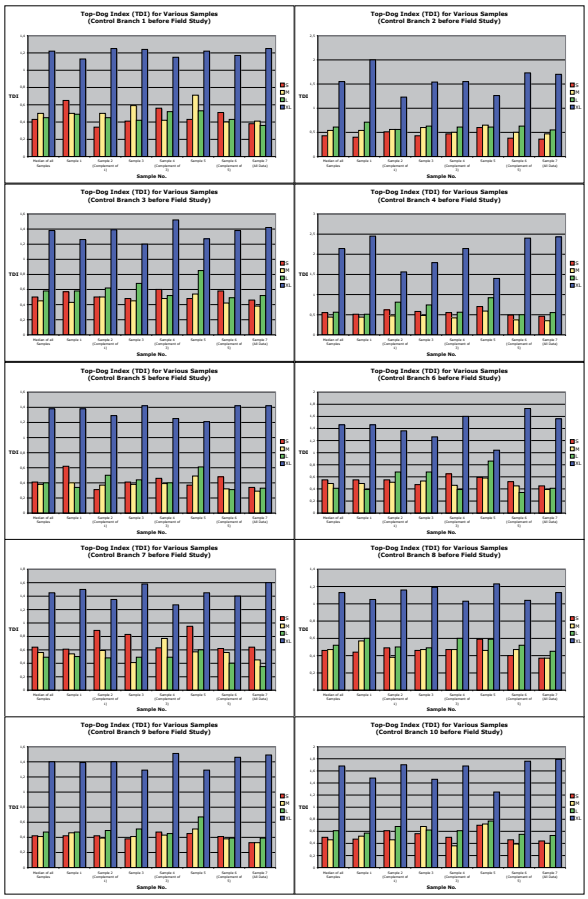


Figure 8: Individual TDIs in control branches from historic sales data.

The alternative to ignoring inconsistent data is to estimate or recover it from the rest of the data set. As an example, we would simply assume that there was a supply of 10 items instead of 8 items at the same price level in the first case. In the second case we would assume that the remaining 2 items were also sold. Maybe they were shoplifted, some sort of selling for a very cheap prize. So, we need an estimation for the selling prize of the two missing items. Here, we have used the last selling price over all sizes for this product in this branch as an estimate.

Neither evaluation method reflects reality exactly. Our hope was that both estimations encompass the true values. At least our partner accepts both values as a good approximation of reality. The truth may be somewhere in between both values.

6.6 Results

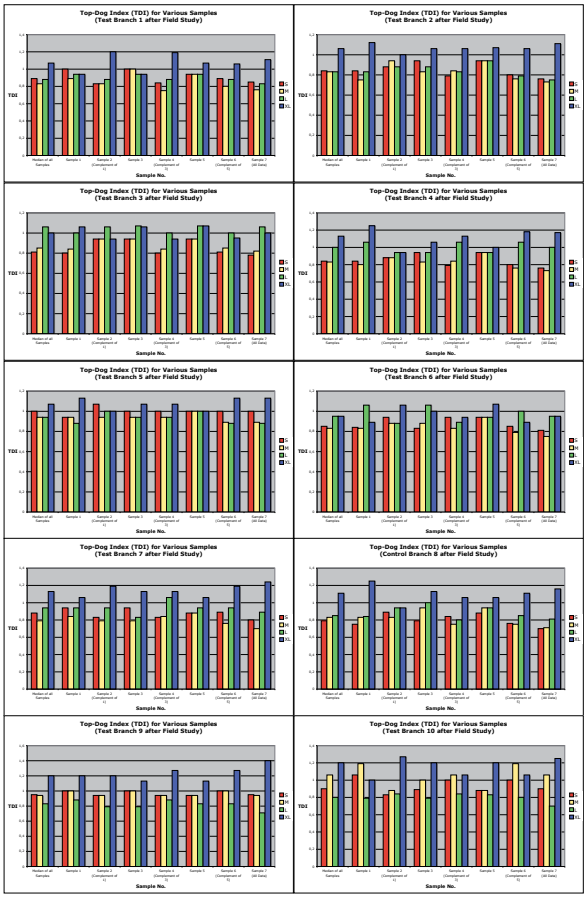


Figure 9: Individual TDIs in test branches in test period.

We have depicted the new Top-Dog-Indices after applying our proposed repacking in Figure 10 for the test branches and in Figure 12 for the control branches. Figures 9 and 11 finally show the results for the individual branches.

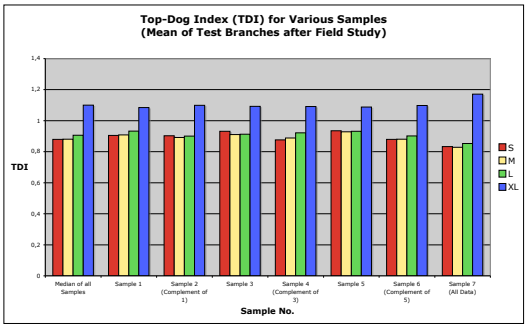


Figure 10: TDI in the test branches in test period.

We can see in Figure 8 and 11 that the absolute Top-Dog-Indices are not as constant over seasons as we would

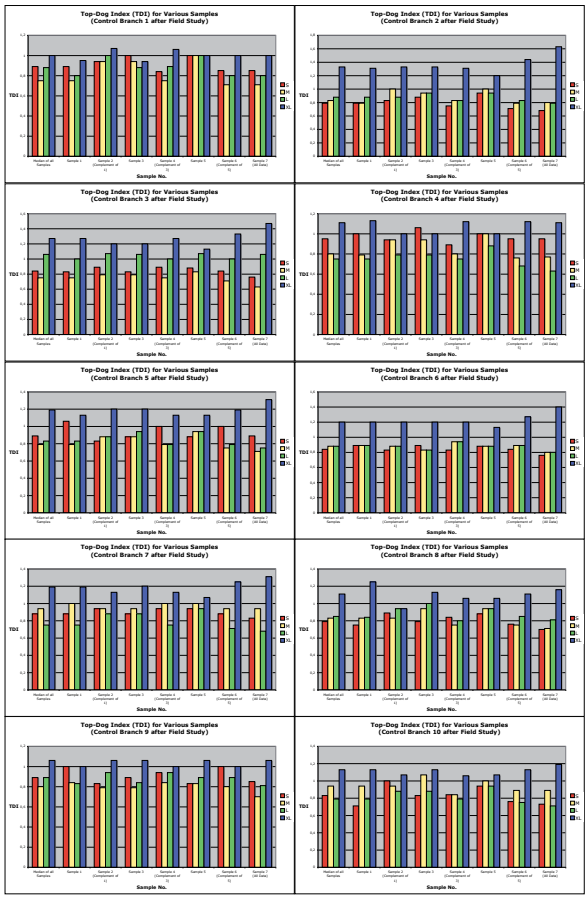


Figure 11: Individual TDIs in control branches in test period.

wish. However, the induced order on sizes seems quite stable across seasons, and this is all we try to exploit.

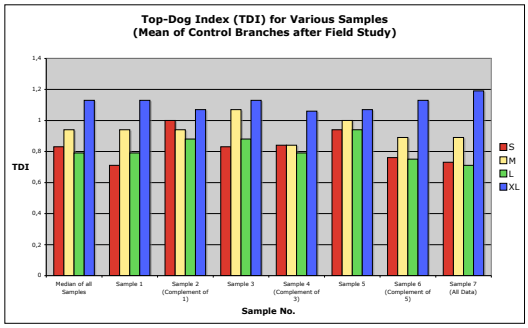


Figure 12: TDI in the control branches in test period.

We can see that it is rather hard to compare the Top-Dog-Indices of the same branches before and after the blind study. The situation on the real market almost never stays constant over time. There are so many influences not considered in our study that it would have been a bad idea to measure a possible raise of earnings directly. For this reason, the simultaneous observation of a test group and a control group makes all outer effects appear in both.

Comparing Figure 10 and Figure 12 based on the same time period, it appears that the Top-Dog-Indices of the test group have improved more.

More specifically, in Figures 9 and 11 we see that in some of the test branches (especially, Test Branches 5 and 6), Size XL is no longer too scarce, while it remains too scarce in other branches. Moreover, on average over all test branches, the Top-Dog-Indices of Sizes S, M, and L are better balanced, while in the control branches the corresponding Top-Dog-Indices differ.

While the former achievement might have been equally possible on the basis of a statistics aggregated over all branches, it seems that the latter result was made possible only by the branch dependent information from the Top-Dog-Index, since different sizes were removed from the prepacks in favor of XL. Moreover, some of the test branches still need fewer pieces in Size XL, some test branches don't. That is, in the next optimization step, the branch dependent information becomes vital also for the supply with Size XL.

But is an improved Top-Dog-Index really an improvement for the business? To answer this question, we have quantified the gross yields and the last prices in the test group and the control group, resp.

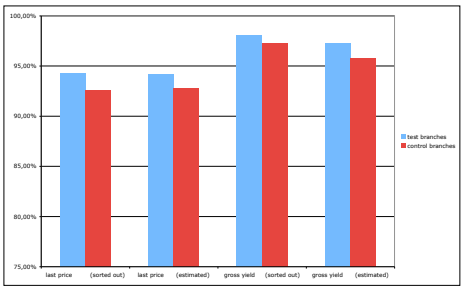


Figure 13: Average last prices/gross yields per merchandise value.

In Figure 13 we have compared the average values of the gross yield and the last price for the control and the test branches for both evaluation methods. The gross yield of the test branches is 98.0 % using the ignore and 97.2 % using the estimate method. For the control branches we have gross yields of 97.2 % (ignore) and 95.7 % (estimate). This corresponds to improvements of 0.85 and 1.5 percentage points, resp.

The improvement for the last price is even larger. The test branches show a last price level of 94.2 % (ignore) and 94.1 %; the control branches exhibit a last price level of 92.5 % (ignore) and 92.7 % (estimate). This corresponds to improvements of 1.7 and 1.4 percentage points, resp.

The drastically improved results for the respective "loser branches" (see Figure 14) and the reduced standard devia-

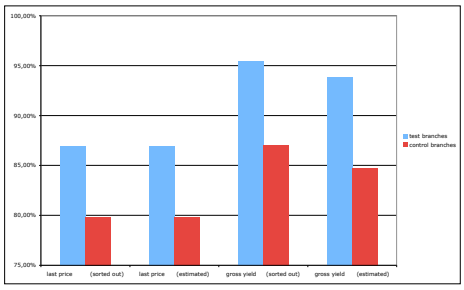


Figure 14: Minimal last prices/gross yields per merchandise value.

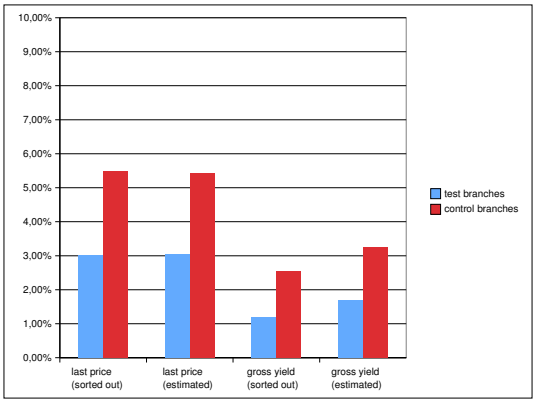


Figure 15: Standard deviation of last prices/gross yields per merchandise value.

tion (see Figure 15) in the data of the test branches provides evidence for the fact that our procedure was able to reduce the risk of a very low last price or a very low gross yield in an individual branch. This effect is desirable beyond the better earnings, as very low last prices undermine the image of the retailer.

6.7 Statistical evidence of the improvement of the gross yield

As in the previous sections we analyze our results concerning the gross yield from the statistical point of view. Faced with the fact of a widely varying gross yield over the branches with no appropriate theoretic sales model, we have to restrict ourselves to distribution-free statistics.

Therefore, we adapt the Wilcoxon rank sum test to our situation. This is a test method to find out whether or not two data sets are drawn from the same distribution. We sort in increasing order the gross yields of the 20 branches participating in our blind study. We associate the largest value with Rank 1, the second largest with Rank 2, and so on. Then we form the rank sums of the test branches and the control branches, resp. The more the rank sums differ, the less likely is the event that our method did not influence the gross yields/last prices at all.

For the Wilcoxon rank sum test, it is vital that we have partitioned the 20 branches for the blind study independently at random into test and control branches.

It is intuitively clear that a smaller rank sum for the gross yield/last price is more likely if the corresponding expected values are better. A lower rank sum can indeed occur by pure coincidence, but the probability decreases with the rank sum. As an example, the rank sum for the test branches regarding the gross yield measured by the ignore method is 89. If we had not changed anything, the chance to receive a rank sum of 89 or lower would have been 12.4 %. So, we have a certainty of 87.6 % that our proposed re-packing improved the situation. (More formally: the probability that the gross yields of the test branches and the gross yields of the control branches stem from the same distribution, i. e., nothing has changed systematically, is at most 12.4 %.)

Now we consider different scenarios. Let $y_i(b)$ be the gross yield measured with the method ignore of branch b and $y_e(b)$ the gross yield measured with the method estimate. By i_c we denote the scenario where we consider the values $y_i(b)$ for control branches and the values $y_i(b)+\frac{c}{100}$ for test branches. Similarly, we define the scenarios for e_c utilizing $y_e(b)$ instead of $y_i(b)$. In Table 6 we have given the rank sums and the certainties of some scenarios.

	$i_{-0.00}$	$e_{-0.00}$	$i_{-0.25}$	$e_{-0.25}$
control group	121	131	118	128
test group	89	79	92	82
certainty (%)	87.6	97.4	82.4	95.5
	$i_{-0.50}$	$e_{-0.50}$	$i_{-0.75}$	$e_{-0.75}$
control group	105	124	94	120
test group	105	86	116	90
certainty (%)	48.5	91.7	19.7	86.0
	$e_{-1.00}$	$e_{-1.50}$		
control group	112	107		
test group	98	103		
certainty (%)	68.5	54.4		

Table 6: Ranks sums and certainties of improvements of the gross yield.

How can we interpret these numbers? The first two columns of Table 6 show that with a certainty of 87.6 % (ignore) and 97.4 % (estimate) that our proposed modification increased the expected gross yield. In Scenario $i_{-0.25}$ we artificially decrease the gross yield (ignore) values by 0.25 percentage points. The monetary value associated with this specific decrease can be interpreted, e. g., as the implementation and consultancy costs of the modification. So, by a look at Table 6 we can say that with a certainty of 82.4 % our proposed modification yields an improvement of the gross yield (ignore) by at least 0.25 percentage points.

7 Conclusion and outlook

The distribution of fashion goods to the branches of a fashion discounter must meet the demand for sizes as accurately as possible. However, in our business case, an estimation of the relative demand for apparel sizes from historic sales data was not possible in a straight-forward way.

Our proposal is to use the *Top-Dog-Index (TDI)*, a measure that yields basically ordinal information about what were the scarcest and the amplest sizes in a product group in a historic sales period. This information was utilized to change the size distributions for future deliveries by replacing one piece of the amplest size by a piece of the scarcest size in every pre-pack (this can be seen as a subgradient improvement step in an iterative size distribution heuristics based on the TDI analysis).

Empirical evidence from a blind study with twenty branches (ten of them, randomly chosen, were supplied according to TDI-based recommendations; ten of them were supplied as before) showed a significant increase in gross yield: On average, the increase in the gross yield in our blind study was around one percentage point. The probability that gross yield improvements of at least 0.25 percentage points occurred is at least 87.6 (even 95.5 % if inconsistent data is repaired in a plausible way). And: this was the result of a single iteration of the optimization procedure, which did not result in perfectly balanced Top-Dog-Indices.

Given the large economies of scale of a fashion discounter, we consider the TDI a valuable contribution to revenue management tools in this business sector. Moreover, to the best of our knowledge, our blind study is the first published study that evaluates a revenue management method in the apparel retailer industry by comparing simultaneously obtained business results of test-branches (optimized) and control-branches (no action).

The draw-back of the TDI is its lack of information about the cardinal expected revenue for a given size distribution of the supply. This is partly due to the fact that the loss of a bad size distribution is closely related to the markdown policy of the discounter. This markdown policy, however, is itself subject of revenue management methods. Therefore, we regard the integration of size and price optimization, as is done in our project BFS-DISPO, as a valuable direction of further research .

Bibliography

[1] G. R. Bitran, R. Caldentey, and S. V. Mondscheinm, *Coordinating clearance markdown sales of seasonal products in retails chains*, Oper. Res. **46** (1998), 609–624.

[2] L. M. A. Chan, Z. J. M. Shen, D.Simchi-Levi, and J. L. Swann, *Coordination of pricing and inventory decisions: A survey and classification*, Handbook of Quantitative Supply Chain Analysis: Modeling in the E-Business Era (David Simchi-Levi, S. D. Wu, and Z. J. Max Shen, eds.), Kluwer Academic Publishers, 2006, 335–392.

[3] W. Elmaghraby and P. Keskinocak, *Dynamic pricing in the presence of inventory considerations: Research overview, current practices, and future directions*, Management Science **49** (2003), 1287–1309.

[4] A. G. Kök and M. L. Fisher, *Demand estimation and assortment optimization under substitution: methodology and application*, Preprint FRPS04-135, Duke University, 2006, Oper. Res. (to appear).

[5] I. Kouris, *Dynamic pricing strategies: Markdown procedure for a textile retailer*, Diplomarbeit, Universität Bayreuth, Fakultät für Mathematik, Physik und Informatik, 2007.

[6] M. Levanoni, Y. T. Leung, and S. E. Ramaswamy, *Method and apparatus suitable for demand forecasting*, US Patent 6976001, 2006.

[7] S. Mahajan and G. van Ryzin, *Stocking retail assortments under dynamic consumer substitution*, Oper. Res. **49** (2001), 334–351.

[8] D. M. Rose, S. M. Leven, and J. W. Woo, *Assortment decisions*, US Patent 7006981, 2006.

[9] C. Wopperer, *Optimierung der Preisreduzierungsstrategie eines Textildiscounters*, Diplomarbeit, Universität Bayreuth, Fakultät für Mathematik, Physik und Informatik, 2007.

Chapter 14

Lotsize optimization leading to a p-median problem with cardinalities

CONSTANTIN GAUL[1], SASCHA KURZ[2], AND JÖRG RAMBAU[3]

ABSTRACT. We consider the problem of approximating the branch and size dependent demand of a fashion discounter with many branches by a distributing process being based on the branch delivery restricted to integral multiples of lots from a small set of available lot-types. We propose a formalized model which arises from a practical cooperation with an industry partner. Besides an integer linear programming formulation and a primal heuristic for this problem we also consider a more abstract version which we relate to several other classical optimization problems like the p-median problem, the facility location problem or the matching problem.

2000 MSC: 90B80; 90C59, 90C10.

Key words and phrases: lotsize optimization, p-median problem, facility location problem, integer linear program formulation, primal heuristic, real world data, location-allocation.

1 Introduction

Usually, fashion discounters can only achieve small profit margins. Their economic success depends mostly in the ability to meet the customers' demands for individual products. More specifically: offer exactly what you can sell to your customers. This task has two aspects: offer what the customers would like to wear (attractive products) and offer the right volumes in the right places and the right sizes (demand consistent branch and size distribution).

In this paper we deal with the second aspect only: meet the branch and size specific demand for products as closely as possible. Our industry partner is a fashion discounter with more than 1 000 branches most of whose products are never replenished, except for the very few "never-out-of-stock"-products (NOS products): because of lead times of around three months, apparel replenishments would be too late anyway. In most cases the supplied items per product and apparel size lie in the range between 1 and 6. Clearly there are some difficulties to determine a good estimate for the branch and size dependent demand, but besides a few practical comments on this problem we will blind out this aspect of the problem completely.

The problem we deal with in this article comes from another direction. Our business partner is a discounter who has a lot of pressure to reduce its costs. So he is forced to have a lean distribution logistics that works efficiently. Due to this reason he, on the one hand, never replenishes and, on the other hand, tries to reduce the distribution complexity. To achieve this goal the supply of the branches is based on the delivery of lots, e. g., pre-packed assortments of single products in various sizes. Every branch can only be supplied with an integral multiple of one lot-type from a rather small number of available lot-types. So he has to face an approximation problem: which (integral) multiples of which (integral) lot-types should be supplied to a branch in order to meet a (fractional) mean demand as closely as possible?

We call this specific demand approximation problem the *lot-type design problem (LDP)*.

1.1 Related Work

The model we suggest for the LDP is closely related to the extensively studied p-median- and the facility location problem. These problems appear in various applications as some kind of clustering problems. Loads of heuristics have been applied onto them. Nevertheless the first constant-factor approximation algorithm, based on LP rounding, was given not until 1999 by Charikar, Guha, Tardos, and Shmoys [5]. We will give some more detailed treatment or literature of approximation algorithms and heuristics for the p-median- and the facility location problem in Subsection 4.1.

1.2 Our contribution

In cooperation with our business partner, we identified the lot-type design problem as a pressing real-world task. We

[1]Constantin Gaul, University of Bayreuth, Department of Mathematics, 95440 Bayreuth, Germany.
E-mail adress: costa.gaul@gmx.de
[2]Sascha Kurz, University of Bayreuth, Department of Mathematics, 95440 Bayreuth, Germany.
E-mail adress: sascha.kurz@uni-bayreuth.de
[3]Jörg Rambau, University of Bayreuth, Department of Mathematics, 95440 Bayreuth, Germany.
E-mail adress: joerg.rambau@uni-bayreuth.de

present an integer linear program (ILP) formulation of the LDP that looks abstractly like a p-median problem with an additional cardinality constraint. We call this problem the *cardinality constrained* p*-median problem (Card-*p*-MP)*. To the best of our knowledge, the Card-p-MP has not been studied in the literature so far.

Although the ILP model can be solved by standard software on a state-of-the-art PC in reasonable time, the computation times are prohibitive for the use in the field, where interactive decision support on a laptop is a must for negotiations with the supplier. Therefore, we present a very fast primal any-time heuristics, that yields good solutions almost instantly and searches for improvements as long as it is kept running. We demonstrate on real data that the optimality gaps of our heuristics are mostly way below 1 %. At the moment these heuristics are in test mode.

1.3 Outline of the paper

In Section 2 we will briefly describe the real world problem, which we will formalize and model in Section 3. In Section 4 we will present its abstract version, the *cardinality constrained* p*-median problem (Card-*p*-MP)*. Besides a formalized description we relate it to several other well known optimization problems like the matching problem, the facility location problem, or the p-median problem. In Section 5 we present a primal heuristic for the Card-p-MP, which we apply onto our real world problem. We give some numerical data on the optimality gap of our heuristic before we draw a conclusion in Section 6.

2 The real world problem

Our industry partner is a fashion discounter with over 1 000 branches. Products can not be replenished, and the number of sold items per product and branch is rather small. There are no historic sales data for a specific product available, since every product is sold only for one sales period. The challenge for our industry partner is to determine a suitable total amount of items of a specific product which should be bought from the supplier. For this part the knowledge and experience of the buyers employed by a fashion discounter is used. We seriously doubt that a software package based on historic sales data can do better.

But there is another task being more accessible for computer aided forecasting methods. Once the total amount of sellable items of a specific product is determined, one has to decide how to distribute this total amount to a set of branches B in certain apparel sizes with in general different demands. There are some standard techniques how to estimate branch- and size-dependent demand from historic sales data of related products, being, e. g., in the same commodity group. We will address the problem of demand forecasting very briefly in Subsection 3.1. But let us assume for simplicity that we know the exact (fractional) branch and size dependent mean demands for a given new product or have at least good estimates.

Due to cost reasons, our industry partner organizes his distribution process for the branches using a central warehouse. To reduce the number of necessary handholds in the distributing process he utilizes the concept of lots, by which we understand a collection of some items of one product. One could have in mind different sizes or different colors at this point. To reduce the complexity of the distribution process also the number of used lot-types, e. g., different collections of items, is limited to a rather small number.

One could imagine that the branch- and size-dependent demand for a specific product may vary broadly over the large set of branches. This is at least the case for the branches of our industry partner. The only flexibility to satisfy the demand in each single branch is to choose a suitable lot-type from the small sets of available lot-types and to choose a suitable multiplier, e. g., how many lots of a chosen lot-type a specific branch should get. One should keep in mind that we are talking about small multipliers here, e. g., small branches will receive only one lot, medium sized branches will receive two lots, and very big branches will receive three lots of a lot-type with, say, six items.

The cost reductions by using this lot-based distribution system are paid with a lack of possibility to approximate the branch and size-dependent demand. So one question is, how many different lot-types one should allow in order to be able to approximate the branch- and size-dependent demand of the branches up to an acceptable deviation on the one hand and to avoid a complex and cost intensive distribution process in the central warehouse on the other hand. But also for a fixed number of allowed lot-types the question of the best possible approximation of the demand by using a lot-based supply of the branches arises. In other words we are searching for an optimal assignment of branches to lot-types together with corresponding multipliers so that the deviation between the theoretical estimated demand and the planned supply with lots is minimal. This is the main question we will focus on in this paper.

3 Mathematical modeling of the problem

In this section we will prescind the real world problem from the previous section and will develop an formulation as a well defined optimization problem. Crucial and very basic objects for our considerations are the set of branches $\mathcal{B}$, the set of sizes $\mathcal{S}$ (in a more general context one could also think of a set of variants of a product, like, e. g., different colors), and the set of products $\mathcal{P}$.

In practice, we may want to sell a given product $\pi \in \mathcal{P}$ only in some branches $\mathcal{B}_\pi \subseteq \mathcal{B}$ and only in some sizes $\mathcal{S}_\pi \subseteq \mathcal{S}$ (clearly there are different types of sizes for, e. g., skirts or socks). To model the demand of a given branch $\mathrm{b} \in \mathcal{B}_\pi$ for a given product $\pi \in \mathcal{P}$ we use the symbol $\eta_{\mathrm{b},\pi}$, by which we understand a mapping $\varphi_{\mathrm{b},\pi}$ from the set of sizes $\mathcal{S}_\pi$ into a suitable mathematical object. This object may be a random variable or simply a real number representing the mean demand. In this paper we choose the latter possibility.

For the sake of a brief notation we regard $\eta_{b,\pi}$ as a vector $\begin{pmatrix}\varphi_{b,\pi}(s_{i_1}) & \varphi_{b,\pi}(s_{i_2}) & \dots & \varphi_{b,\pi}(s_{i_r})\end{pmatrix} \in \mathbb{R}^r$, where we assume that $\mathcal{S} = \{s_1, \dots, s_t\}$ and $\mathcal{S}_\pi = \{s_{i_1}, \dots, s_{i_r}\}$ with $i_j < i_{j+1}$ for all $j \in \{1, \dots, r-1\}$.

3.1 Estimation of the branch- and size-dependent demand

For the purpose of this paper, we may assume that the demands $\eta_{b,\pi}$ are given, but, since this is a very critical part in practice, we would like to mention some methods how to obtain these numbers. Marketing research might be a possible source. Another possibility to estimate the demand for a product is to utilize historic sales information. We may assume that for each product π which was formerly sold by our retailer, each branch $b \in \mathcal{B}$, each size $s \in \mathcal{S}$ and each day of sales d we know the number $\tau_{b,\pi}(d,s)$ of items which where sold in branch b of product π in size s during the first d days of sales. Additionally we assume, that we have a set $\mathcal{U} \subseteq \mathcal{P}$ of formerly sold products which are in some sense similar (one might think of the set of jeans if our new product is also a jeans) to the new product $\widehat{\pi}$. By $\mathcal{U}_{b,s}$ we denote the subset of products in $\mathcal{U}$, which were traded by a positive amount in size s in branch b and by $\chi_{b,s}(\pi)$ we denote a characteristic function which equals 1 if product π is distributed in size s to branch b, and equals 0 otherwise. For a given day of sales d the value $\widehat{\eta}_{b,\widehat{\pi},d}(s)$

$$:= \frac{c}{|\mathcal{U}_{b,s}|} \sum_{u \in \mathcal{U}_{b,s}} \frac{\tau_{b,u}(d,s) \cdot \sum\limits_{b' \in \mathcal{B}_{\widehat{\pi}}} \sum\limits_{s' \in \mathcal{S}_{\widehat{\pi}}} \chi_{b',s'}(u)}{\sum\limits_{b' \in \mathcal{B}_{\widehat{\pi}}} \sum\limits_{s' \in \mathcal{S}_{\widehat{\pi}}} \tau_{b',u}(d,s')} \qquad (1)$$

might be a useable estimate for the demand $\eta_{b,\widehat{\pi}}(s)$, after choosing a suitable scaling factor $c \in \mathbb{R}$ so that the total estimate demand

$$\sum_{b \in \mathcal{B}_{\widehat{\pi}}} \sum_{s \in \mathcal{S}_{\widehat{\pi}}} \widehat{\eta}_{b,\widehat{\pi},d}(s)$$

over all branches and sizes equals the total requirements. We would like to remark that for small days of sale d the quality of the estimate $\widehat{\eta}_{b,\widehat{\pi},d}(s)$ suffers from the fact that the stochastic noise of the consumer behavior is to dominating and for large d the quality of the estimate suffers from the fact of stockout-substitution.

There are parametric approaches to this problem in the literature (like Poisson-type sales processes). In the data that was available to us, we could not verify the main assumptions of such models, though (not even close).

In our real world data set we have observed the fact that the sales period of a product (say, the time by which 80 % of the supply is sold) varies a lot depending on the product. This effect is due to the attractiveness of a given product (one might think of two T-shirts which only differ in there color, where one color hits the vogue and the other color does not). To compensate this effect we have chosen the day of sales d in dependence of the product $u \in \mathcal{U}_{b,s}$. More precisely, we have chosen d_u so that in the first d_u days of sales a certain percentage of all items of product u where sold out over all branches and sizes.

Another possibility to estimate the demand is to perform the estimation for the branch-dependent demand aggregated over all sizes and the size-dependent demand for a given branch separately.

More sophisticated methods of demand estimation from historic sales based on small data sets are, e. g., described in [19, 20]. Also research results from forecasting NOS (never-out-of-stock) items, see, e. g., [1, 17, 24] for some surveys, may be utilized. Also quite a lot of software-packages for demand forecasting are available, see [31] for an overview.

3.2 Supply of the branches by lots

To reduce handling costs in logistic and stockkeeping our business partner orders his products from its external suppliers in so called lots. These are assortments of several items of one product in different sizes which form an entity. One could have a set of T-shirts in different sizes in mind which are wrapped round by a plastic foil. The usage of lots has the great advantage of reducing the number of picks during the distribution process in a high-wage country like Germany, where our partner operates.

Let us assume that the set of sizes for a given product π is given by $\mathcal{S}_\pi = \{s_{i_1}, \dots, s_{i_r}\}$ with $i_j < i_{j+1}$ for all $j \in \{1, \dots, r-1\}$. By a lot-type l we understand a mapping $\varphi : \mathcal{S}_\pi \to \mathbb{N}$, which can also be denoted by a vector $\begin{pmatrix}\varphi(s_{i_1}) & \varphi(s_{i_2}) & \dots & \varphi(s_{i_r})\end{pmatrix}$ of non-negative integers.

By $\mathcal{L}$ we denote the set of applicatory lot-types. One could imagine that a lot of a certain lot-type should not contain too many items in order to be manageable. In the other direction it should also not contain too few items in order to make use of the cost reduction potential of the lot idea. Since the set of applicatory lot-types may depend on a the characteristics of a certain product π we specialize this definition to a set $\mathcal{L}_\pi$ of manageable lot-types. (One might imagine that a warehouseman can handle more T-shirts than, e. g., winter coats; another effect that can be modeled by a suitable set of lot-types is to enforce that each size in $\mathcal{S}_\pi$ is supplied to each branch in $\mathcal{B}_\pi$ by a positive amount due to juridical requirements for advertised products.)

To reduce the complexity and the error-proneness of the distribution process in a central warehouse, each branch $b \in \mathcal{B}_\pi$ is supplied only with lots of one lot-type $l_{b,\pi} \in \mathcal{L}_\pi$. We model the assignment of lot-types $l \in \mathcal{L}_\pi$ to branches $b \in \mathcal{B}_\pi$ as a function $\omega_\pi : \mathcal{B}_\pi \to \mathcal{L}_\pi$, $b \mapsto l_{b,\pi}$. Clearly, this assignment ω_π is a decision variable which can be used to optimize some target function. The only flexibility that we have to approximate the branch-, size- and product dependent demand $\eta_{b,\pi}$ by our delivery in lots is to supply an integral multiple of $m_{b,\pi}$ items of lot-type $\omega_\pi(b)$ to branch b. Again, we can denote this connection by a function $m_\pi : \mathcal{B}_\pi \to \mathbb{N}$, $b \mapsto m_{b,\pi}$. Due to practical reasons, also the total number $|\omega_\pi(\mathcal{B}_\pi)|$ of used lot-types for a given product is limited by a certain number κ.

3.3 Deviation between supply and demand

With the notation from the previous subsection, we can represent the replant supply for branch b with product π as a vector $m_\pi(b) \cdot \omega_\pi(b) \in \mathbb{N}^r$. To measure the deviation between the supply $m_\pi(b) \cdot \omega_\pi(b)$ and the demand $\eta_{b,\pi}$ we may utilize an arbitrary vector norm $\|\cdot\|$. Mentionable vector norms in our context are the sum of absolute values

$$\left\|\begin{pmatrix} v_1 & v_2 & \dots & v_r \end{pmatrix}\right\|_1 := \sum_{i=1}^{r} |v_i|\,,$$

the maximum norm

$$\left\|\begin{pmatrix} v_1 & v_2 & \dots & v_r \end{pmatrix}\right\|_\infty := \max\{|v_i| : 1 \leqslant i \leqslant r\},$$

and the general k-norm

$$\left\|\begin{pmatrix} v_1 & v_2 & \dots & v_r \end{pmatrix}\right\|_k := \sqrt[k]{\sum_{i=1}^{r} |v_i|^k}$$

for real numbers $k > 0$, which is also called the Euclidean norm for $k = 2$. With this we can define the deviation

$$\sigma_{b,l,m} := \|\eta_{b,\pi} - m \cdot l\|_\star$$

between demand $\eta_{b,\pi}$ and supply $m \in \{1, \dots, M\} =: \mathcal{M} \subset \mathbb{N}$ times lot-type $l \in \mathcal{L}_\pi$ for each branch $b \in \mathcal{B}_\pi$ and an arbitrary norm $\|\cdot\|_\star$ for a given product $\pi \in \mathcal{P}$. It depends on practical considerations which norm to choose. The $\|\cdot\|_1$-norm is very insensitive in respect to outliers in contrast to the $\|\cdot\|_\infty$-norm which is absolutely sensitive with respect to outliers. A possible compromise may be the Euclidean norm $\|\cdot\|_2$, but for most considerations we choose the $\|\cdot\|_1$-norm because of its robustness. (We do not trust every single exact value in our demand forecasts that much.)

For given functions m_π and ω_π we can consider the deviation vector

$$\Sigma_\pi := \begin{pmatrix} \sigma_{b_1,\omega_\pi(b_1),m_\pi(b_1)} \\ \sigma_{b_2,\omega_\pi(b_2),m_\pi(b_2)} \\ \vdots \\ \sigma_{b_q,\omega_\pi(b_q),m_\pi(b_q)} \end{pmatrix}^T$$

if the set of branches is written as $\mathcal{B}_\pi := \{b_1, \dots, b_q\}$. To measure the total deviation of supply and demand we can apply an arbitrary norm $\|\cdot\|_\star$, which may be different from the norm to measure the deviation of a branch, onto Σ_π. In this paper we restrict ourselves on the $\|\cdot\|_1$-norm, so that we have

$$\|\Sigma_\pi\|_1 = \sum_{b \in \mathcal{B}_\pi} \sigma_{b,\omega_\pi(b),m_\pi(b)}.$$

3.4 The cardinality condition

For a given assignment ω_π of lot-types to branches and corresponding multiplicities m_π then quantity

$$I := \sum_{b \in \mathcal{B}_\pi} m_\pi(b) \cdot \|\omega_\pi(b)\|_1 \quad \in \mathbb{N}$$

gives the total number of replant distributed items of product π over all sizes and branches. From a practical point of view we introduce the condition

$$\underline{I} \leqslant I \leqslant \bar{I}, \tag{2}$$

where $\underline{I}$, $\bar{I}$ are suitable integers. One might imagine that our retailer may buy a part of already produced products so that there is a natural upper bound $\bar{I}$ or that there are some minimum quantities. Another interpretation may be that the buying department of our retailer has a certain idea on the value of I but is only able to give an interval $\left[\underline{I}, \bar{I}\right]$.

During our cooperation with our busines partner we have learned that in practice you do not get what you order. If you order exactly I items of a given product you will obtain I plus minus some certain percentage items in the end. (And there actually exists a certain percentage up to which a retailer accepts a deviation between the original order and the final delivery by its external suppliers as a fulfilled contract.)

Besides these and other practical reason to consider an interval $\left[\underline{I}, \bar{I}\right]$ for the total number of items of a given product, there are very strong reasons not to replace Inequalities (2) by an equation, as we will explain in the following. Let us consider the case where our warehouse (or our external suppliers in a low-cost-country) is only able to deal with a single lot-type per product. This is the case $\kappa = 1$. Let us further assume that there exists a rather small integer k (e. g. $k = 20$) fulfilling $\|l\|_1 \leqslant k$ for all $l \in \mathcal{L}_\pi$. If I contains a prime divisor being larger than k, then there exist no assignments multiplicities $m_\pi \in \mathbb{N}$ (ω_π is a constant function due to $\kappa = 1$) which lead to a feasible solution of our problem. These number-theoretic influences are somewhat ugly. In some cases the lead to the infeasibility of our problem or to bad solutions with respect to the quality of the demand-supply approximation in comparison to a relaxed version of the problem, where the restrictions on I are weaker. One could have in mind the possibility of throwing one item into the garbage if this will have a large impact on the quality of the demand-supply approximation.

In Equation (1) for the demand estimation we have used a certain number $\tilde{I}$ for the total number of items to scale the demands $\eta_{b,\pi}$ by a factor c. From a more general point of view it may also happen that the total demand

$$\sum_{b \in \mathcal{B}_\pi} \sum_{s \in \mathcal{S}_\pi} \eta_{b,\pi}(s)$$

is not contained in the interval $\left[\underline{I}, \bar{I}\right]$. In this case the $\|\cdot\|_1$-norm may not be very appropriate. In our estimation process, however, the demand forecasts in fact yield demand percentages rather than absolute numbers. The total volume is then used to calculate the absolute (fractional) mean demand values, so that in our work-flow the total demand is always in the target interval.

3.5 The optimization problem

Summarizing the ideas and using the notations from the previous subsections we can formulate our optimization prob-

lem in the following form. We want to determine an assignment function $\omega_\pi : \mathcal{B}_\pi \to \mathcal{L}_\pi$ and multiplicities $m_\pi : \mathcal{B}_\pi \to \mathcal{M} = \{1, \dots, M\} \subset \mathbb{N}$ such that the total deviation between supply and demand

$$\sum_{b \in \mathcal{B}_\pi} \sigma_{b,\omega_\pi(b),m_\pi(b)} \tag{3}$$

is minimized with respect to the conditions

$$|\omega_\pi(\mathcal{B}_\pi)| \leqslant \kappa \tag{4}$$

and

$$\underline{I} \leqslant \sum_{b \in \mathcal{B}_\pi} m_\pi(b) \cdot \|\omega_\pi(b)\|_1 \leqslant \bar{I}. \tag{5}$$

We use binary variables $x_{b,l,m}$, which are equal to 1 if and only if lot-type $l \in \mathcal{L}_\pi$ is delivered with multiplicity $m \in \mathcal{M}$ to branch b, and binary variables y_l, which are 1 if and only if at least one branch in $\mathcal{B}_\pi$ is supplied with lot-type $l \in \mathcal{L}_\pi$. With this, we can easily model out problem as an integer linear program:

$$\begin{aligned}
\min \quad & \sum_{b \in \mathcal{B}_\pi} \sum_{l \in \mathcal{L}_\pi} \sum_{m \in \mathcal{M}} \sigma_{b,l,m} \cdot x_{b,l,m} && (6)\\
\text{s.t.} \quad & \sum_{l \in \mathcal{L}_\pi} \sum_{m \in \mathcal{M}} x_{b,l,m} = 1 \qquad \forall b \in \mathcal{B}_\pi && (7)\\
& \sum_{b \in \mathcal{B}_\pi} \sum_{l \in \mathcal{L}_\pi} \sum_{m \in \mathcal{M}} m \cdot \|l\|_1 \cdot x_{b,l,m} \leqslant \bar{I} && (8)\\
& \sum_{b \in \lfloor_\pi} \sum_{l \in \mathcal{L}_\pi} \sum_{m \in \mathcal{M}} m \cdot \|l\|_1 \cdot x_{b,l,m} \geqslant \underline{I} && (9)\\
& \sum_{m \in \mathcal{M}} x_{b,l,m} \leqslant y_l \qquad \forall b \in \mathcal{B}_\pi, l \in \mathcal{L}_\pi && (10)\\
& \sum_{l \in \mathcal{L}_\pi} y_l \leqslant \kappa && (11)\\
& x_{b,l,m} \in \{0,1\} \quad \forall b \in \mathcal{B}_\pi, l \in \mathcal{L}_\pi, m \in \mathcal{M} && (12)\\
& y_l \in \{0,1\} \qquad \forall l \in \mathcal{L}_\pi && (13)
\end{aligned}$$

The objective function (6) represents the sum (3), since irrelevant tuples (b, l, m) may be downtroddened by $x_{b,l,m} = 0$. Condition (7) states that we assign for each branch b exactly one lot-type with a unique multiplicity. The cardinality condition (5) is modeled by conditions (8) and (9) and the restriction (4) on the number of used lot-types is modeled by condition (11). The connection between the $x_{b,l,m}$ and the y_l is fixed in the usual Big-M condition (10). We would like to remark that the LP-relaxation of this ILP formulation is very strong above all in comparison to the more direct ILP formulation, where we assume the branch deviation between supply and demand is measured by the $\|\cdot\|_1$-norm:

$$\begin{aligned}
\min \quad & \sum_{b \in \mathcal{B}_\pi} \sum_{s \in \mathcal{S}_\pi} z_{b,s}\\
\text{s.t.} \quad & \eta_{b,\pi}(s) - \alpha_{b,s} \leqslant z_{b,s} \qquad \forall b \in \mathcal{B}_\pi, s \in \mathcal{S}_\pi\\
& \alpha_{b,s} - \eta_{b,\pi}(s) \leqslant z_{b,s} \qquad \forall b \in \mathcal{B}_\pi, s \in \mathcal{S}_\pi\\
& \sum_{l \in \mathcal{L}_\pi} \sum_{m \in \mathcal{M}} x_{b,l,m} = 1 \qquad \forall b \in \mathcal{B}_\pi\\
& \sum_{b \in \mathcal{B}_\pi} \sum_{l \in \mathcal{L}_\pi} \sum_{m \in \mathcal{M}} m \cdot \|l\|_1 \cdot x_{b,l,m} \leqslant \bar{I}\\
& \sum_{b \in \mathcal{B}_\pi} \sum_{l \in \mathcal{L}_\pi} \sum_{m \in \mathcal{M}} m \cdot \|l\|_1 \cdot x_{b,l,m} \geqslant \underline{I}\\
& \sum_{m \in \mathcal{M}} x_{b,l,m} \leqslant y_l \qquad \forall b \in \mathcal{B}_\pi, l \in \mathcal{L}_\pi\\
& \sum_{l \in \mathcal{L}_\pi} y_l \leqslant \kappa\\
& \sum_{l \in \mathcal{L}_\pi} \sum_{m \in \mathcal{M}} m \cdot l[s] \cdot x_{b,l,m} = \alpha_{b,s} \quad \forall b \in \mathcal{B}_\pi, s \in \mathcal{S}_\pi\\
& x_{b,l,m} \in \{0,1\} \qquad \forall b \in \mathcal{B}_\pi, l \in \mathcal{L}_\pi, m \in \mathcal{M}\\
& y_l \in \{0,1\} \qquad \forall l \in \mathcal{L}_\pi\\
& \alpha_{b,s} \in \mathbb{R}_0^+ \qquad \forall b \in \mathcal{B}_\pi, s \in \mathcal{S}_\pi,
\end{aligned}$$

where $l[s]$ is the entry in vector l corresponding to size s.

We would like to remark that our strong ILP formulation of the problem of Subsection 3.5 can be used to solve all real world instances of our business partner in at most 30 minutes by using a standard ILP solver like `CPLEX 11`. Unfortunately, this is not fast enough for our real world application. The buyers of our retailer need a software tool which can produce a near optimal order recommendation in real time on a standard laptop. The buying staff travels to one of the external suppliers to negotiate several orderings. When they get to the details, the buyer inserts some key data like $\underline{I}$, $\bar{I}$, $\mathcal{B}_\pi$, $\mathcal{S}_\pi$, and $\mathcal{L}_\pi$ into his laptop and immediately wants a recommendation for an order in terms of multiples of lot-types. For this reason, we consider in Section 5 a fast heuristic, which has only a small gap compared to the optimal solution on a test set of real world data of our business partner.

4 The cardinality constrained p-median problem

In the previous section we have modeled our real world problem from Section 2. Now we want to abstract from this practical problem and formulate a more general optimization problem which we will relate to several well known optimization problems.

For the general cardinality constrained p-median problem let p be an integer, $\mathcal{S}$ a set of chooseable items, $\mathcal{D}$ a set of demanders, a demand function $\delta : \mathcal{D} \to \mathbb{R}^+$, and $\left[\underline{I}, \bar{I}\right] \subseteq \mathbb{N}$ an interval. We are looking for an assignment $\omega : \mathcal{D} \to \mathcal{S}$ with corresponding multipliers $m : \mathcal{D} \to \mathbb{N}$, such that the sum of distances

$$\sum_{d \in \mathcal{D}} \|\delta(d) - m(d) \cdot \omega(d)\|$$

is minimized under the conditions

$$|\omega(\mathcal{D})| \leqslant p$$

and

$$\underline{I} \quad \leqslant \quad \sum_{d \in \mathcal{D}} m(d) \cdot |\omega(d)| \quad \leqslant \quad \bar{I}.$$

Let us now bring this new optimization problem in line with known combinatorial optimizations problems. Since we have to choose an optimal subset of $\mathcal{S}$ to minimize a cost function subject to some constraints the cardinality constrained p-median problem belongs to the large class of generic selection problems.

Clearly, it is closely related to the p-median problem. The only characteristics of our problem that are not covered by the p-median problem are the multipliers m and the cardinality condition. If we relax the cardinality condition we can easily transform our problem into a classical p-median problem. For every element $d \in \mathcal{D}$ and every element $s \in \mathcal{S}$ there exists an optimal multiplier $m_{d,s}$ such that $\|\delta(d) - m_{d,s} \cdot s\|$ is minimal.

If we do not bound $|\omega(\mathcal{D})|$ from above but assign costs for using elements of $\mathcal{S}$ instead, which means using another lot-type in our practical application, we end up with the facility location problem. Clearly we also have some kind of an assignment-problem, since the have to determine an assignment ω between the sets $\mathcal{D}$ and a subset of $\mathcal{S}$.

One can also look at our problem from a completely different angle. Actually we are given a set of $|\mathcal{B}|$ real-valued demand-vectors, which we want to approximate by a finite number of integer-valued vectors using integral multiples. There is a well established theory in number theory on so called Diophantine approximation [4, 21] or simultaneous approximation, which is somewhat related to our approximation problem. Here one is interested in simultaneously minimizing

$$\left\| \alpha_i - \frac{p_i}{q} \right\|$$

for linearly independent real numbers α_i by integers p_i and q [22, 27]. One might use some results from this theory to derive some bounds for our problem. One might also have a look at [9].

For a more exhaustive and detailed analysis of the taxonomy of the broad field of facility-location problems and their modeling we refer to [26].

4.1 Approximation algorithms and heuristics for related problems

Facility location problems and the p-median problem are well known and much research has been done. Since, moreover, these problems are closely related to our optimization problem, we would like to mention some literature and methods on approximation algorithms and heuristics for these problems.

Lin and Vitter [23] have developed a filtering and rounding technique which rounds fractional solutions of the standard LP for these problems to obtain good integer solution. For the metric case some some bounds for approximation quality are given. Based on this work some improvements were done in [28], were the authors give a polynomial-time 3.16-approximation algorithm for the metric facility location problem, and in [5, 6], where the authors give a polynomial-time $\frac{20}{3}$-approximation algorithm for the metric p-median problem and a 9.8-approximation algorithm for the p-facility location problem.

Besides rounding techniques of LP-solutions also greedy techniques have been applied to the facility location problem and the p-median problems. Some results are given in [12, 15, 16]. Since these problems are so prominent in applications the whole broadness of heuristics are applied onto it. Examples are scatter search [8, 10], local search [2, 18], and neighborhood search [11, 14].

Good overviews for the broad topic of approximation algorithms and heuristics for the facility location and the p-median problem are given in [7, 25, 28, 29].

Besides results for the metric case there are also results for the non-metric case, see, e. g., [30].

Unfortunately, none of the theoretical guarantees seem to survive the introduction of the cardinality constraint in general.

5 A practical heuristic for the cardinality constrained p-median problem

As already mentioned in Section 3 solving our ILP formulation of our problem is too slow in practical applications. So there is a real need for a fast heuristic which yields good solutions, which is the topic of this section.

In Section 4 we have analyzed our problem from different theoretical point of views. What happens if we relax some conditions or fix some decisions. A very important decision is: which lot-types should be used in the first place? Here one should have in mind that the cardinality $|\mathcal{L}_\pi|$ of the set of feasible lot-types is very large compared to the number κ of lot-types which can be used for the delivery process of a specific product π.

5.1 Heuristic selection of lot-types

For this selection problem of lot-types we utilize a scoring method. For every branch $b \in \mathcal{B}_\pi$ with demand $\eta_{b,\pi}$ there exists a lot-type $l \in \mathcal{L}_\pi$ and a multiplicity $m \in \mathbb{N}$ such that $\|\eta_{b,\pi} - m \cdot l\|$ is minimal in the set $\left\{ \|\eta_{b,\pi} - m' \cdot l'\| \; : \; l' \in \mathcal{L}_\pi, m' \in \mathbb{N} \right\}$. So for every branch $b \in \mathcal{B}_\pi$ there exists a lot-type that fits best. More general, for a given $k \leqslant |\mathcal{L}_\pi|$ there exist lot-types $l_1, \ldots, l_k$ such that l_i fits i-best if one uses the corresponding optimal multiplicity. Let us examine this situation from the point of view of the different lot-types. A given lot-type $l \in \mathcal{L}_\pi$ is the i-best fitting lot-type for a number $\rho_{l,i}$ of branches in $\mathcal{B}_\pi$. Writting these numbers $\rho_{l,i}$ as a vector $\rho_l \in \mathbb{N}^k$ we obtain score vectors for all lot-types $l \in \mathcal{L}_\pi$.

Now we want to use these score vectors ρ_l to sort the lot-types of $\mathcal{L}_\pi$ in decreasing *approximation quality*. Using the lexicographic ordering $\preceq$ on vectors we can determine a bijective rank function $\lambda : \mathcal{L}_\pi \to \{1, \dots, |\mathcal{L}_\pi|\}$. (We simply sort the score vectors according to $\preceq$ and for the case of equality we choose an arbitrary succession.) We extend λ to subsets $\mathcal{L}' \subseteq \mathcal{L}_\pi$ by $\lambda(\mathcal{L}') = \sum_{l \in \mathcal{L}'} \lambda(l) \in \mathbb{N}$.

To fix the lot-types we simply loop over subsets $\mathcal{L}' \subseteq \mathcal{L}_\pi$ of cardinality κ in decreasing order with respect to $\lambda(\mathcal{L}')$. In principle we consider all possible selections $\mathcal{L}'$ of κ lot-types, but in practise we stop our computations after an adequate time period with the great advantage that we have checked the in some heuristic sense most promising selections $\mathcal{L}'$ first.

Now we have to go into detail how to efficiently determine the p best fitting lot-types with corresponding optimal multiplicities for each branch $b \in \mathcal{B}_\pi$. We simply loop over all branches $b \in \mathcal{B}_\pi$ and determine the set of the p best fitting lot-types separately. Here we also simply loop over all lot-types $l \in \mathcal{L}_\pi$ and determine the corresponding optimal multiplier m by binary search (it is actually very easy to effectively determine lower and upper bounds for m from $\eta_{b,\pi}$ and l) due to the convexity of norm functions. Using a heap data structure the sorting of the p best fitting lot-types can be done in $O(|\mathcal{L}_\pi|)$ time if $p \log p \in O(|\mathcal{L}_\pi|)$, which is not a real restriction for practical problems. We further want to remark that we do not have to sort the score vectors completely since in practice we will not loop over all $\binom{|\mathcal{L}_\pi|}{\kappa}$ possible selections of lot-types. If one does not want to use a priori bounds (meaning that one excludes the lot-types with high rank λ) one could use a *lazy* or delayed computation of the sorting of λ by utilizing again a heap data structure.

5.2 Adjusting a delivery plan to the cardinality condition

If we determine assignments ω_π with corresponding multipliers m_π with the heuristic being described in Subsection 5.1 in many cases we will not satisfy the cardinality condition (2) since it is totally unaccounted by our heuristic. Our strategy to satisfy the cardinality condition (2) is to adjust m_π afterwards by decreasing or increasing the calculated multipliers unless condition (2) is fulfilled by pure chance.

Here we want to use a greedy algorithm and have to distinguish two cases. If $I(\omega_\pi, m_\pi)$ is smaller then $\underline{I}$, then we increase some of the values of m_π, other wise we have $I(\omega_\pi, m_\pi) > \bar{I}$ and we decrease some of the values of m_π. Our procedure works iteratively and we assume that the current multipliers are given by $\widetilde{m}_\pi$. Our stopping criteria is given by $\underline{I} \leqslant I(\omega_\pi, \widetilde{m}_\pi) \leqslant \bar{I}$ or that there are no feasible operations left. We restrict our explanation of a step of the iteration to the case where we want to decrease the values of $\widetilde{m}_\pi$. For every branch $b \in \mathcal{B}_\pi$ the reduction of $\widetilde{m}_\pi(b)$ by one produces costs

$$\Delta_b^- = \sigma_{b,\omega_\pi(b),\widetilde{m}_\pi(b)-1} - \sigma_{b,\omega_\pi(b),\widetilde{m}_\pi(b)}$$

if the reduction of $\widetilde{m}_\pi(b)$ by one is allowed (a suitable condition is $\widetilde{m}_\pi \geqslant 1$ or $\widetilde{m}_\pi \geqslant 2$) and $\Delta_b^- = \infty$ if we do not have the possibility to reduce the multiplier $\widetilde{m}_\pi(b)$ by one. A suitable data structure for the Δ_b^- values is a heap, for which the update after an iteration can be done in $O(1)$ time. If we reach $I(\omega_\pi, \widetilde{m}_\pi) < \underline{I}$ at some point, we simply discard this particular selection ω_π and consider the next selection candidate.

We would like to remark that for $\kappa = 1$ and an arbitrary norm $\|\cdot\|_\star$ the described heuristic produces the optimal solution if we check all lot-types $l \in \mathcal{L}_\pi$, which can be seen as follows: Since we loop over all $l \in \mathcal{L}_\pi$ in one step the optimal lot-type is chosen. Without the cardinality condition assigning to each branch its locally optimal multiplicity would be globally optimal. Now let us assume that the number of items I in our assignment is larger than the upper bound $\bar{I}$ of the cardinality condition. Due to a simple swapping argument and the convexity of a norm function we conclude that in an optimal solution no multiplier of a branch is larger then the locally optimal multiplier. Thus every optimal solution arises from our initial distribution plan, insulting the cardinality condition, by deleting exactly $\left\lceil \frac{I-\bar{I}}{l} \right\rceil$ lot packages. Due to the convexity of the norm function the greedy way of deleting lot packages of our heuristic ends up with an optimal solution.

Since this adjustment step can be performed very fast one might also take some kind of general swapping techniques into account. Since for these techniques there exists an overboarding amount of papers in the literature we will not go into detail here, but we would like to remark that in those cases (see Subsection 5.3) where the optimality gap of our heuristic lies above 1 % swapping can improve the solutions of our heuristic by a large part.

5.3 Optimality gap

To substantiate the usefullness of our heuristic we have compared the quality of the solutions given by this heuristic after one second of computation time (on a standard laptop) with respect to the solution given by `CPLEX 11`.

Our business partner has provided us historic sales information for nine different commodity groups each ranging over a sales period of at least one and a half year. For each commodity group we have performed a test calculation for $\kappa \in \{1, 2, 3, 4, 5\}$ distributing some amount of items to almost all branches. By I we denote the cardinality interval, by $|\mathcal{S}_\pi|$ the number of sizes, and by $|\mathcal{B}_\pi|$ the number of branches.

Commodity group 1:
$I = [10630, 11749]$, $|\mathcal{S}_\pi| = 5$, $|\mathcal{B}_\pi| = 1119$:

	$\kappa = 1$	$\kappa = 2$	$\kappa = 3$	$\kappa = 4$	$\kappa = 5$
CPLEX	4033.34	3304.10	3039.28	2951.62	2891.96
heuristic	4033.34	3373.95	3076.55	3011.49	2949.31
gap	0.000%	2.114%	1.226%	2.028%	1.983%

Table 1: Optimality gap in the $\|\cdot\|_1$-norm for our heuristic on commodity group 1.

Commodity group 2:
$I = [10000, 12000]$, $|\mathcal{S}_\pi| = 5$, $|\mathcal{B}_\pi| = 1091$:

	$\kappa = 1$	$\kappa = 2$	$\kappa = 3$	$\kappa = 4$	$\kappa = 5$
CPLEX	2985.48	2670.04	2482.23	2362.75	2259.57
heuristic	2985.48	2671.72	2483.52	2362.90	2276.32
gap	0.000%	0.063%	0.052%	0.006%	0.741%

Table 2: Optimality gap in the $\|\cdot\|_1$-norm for our heuristic on commodity group 2.

Commodity group 3:
$I = [9785, 10815]$, $|\mathcal{S}_\pi| = 5$, $|\mathcal{B}_\pi| = 1030$:

	$\kappa = 1$	$\kappa = 2$	$\kappa = 3$	$\kappa = 4$	$\kappa = 5$
CPLEX	3570.33	3022.27	2622.82	2488.10	2413.55
heuristic	3570.33	3023.91	2625.29	2492.07	2417.65
gap	0.000%	0.054%	0.094%	0.160%	0.170%

Table 3: Optimality gap in the $\|\cdot\|_1$-norm for our heuristic on commodity group 3.

Commodity group 4:
$I = [10573, 11686]$, $|\mathcal{S}_\pi| = 5$, $|\mathcal{B}_\pi| = 1119$:

	$\kappa = 1$	$\kappa = 2$	$\kappa = 3$	$\kappa = 4$	$\kappa = 5$
CPLEX	4776.36	4364.63	4169.94	4023.60	3890.87
heuristic	4776.36	4365.47	4170.23	4024.55	3892.35
gap	0.000%	0.019%	0.007%	0.024%	0.038%

Table 4: Optimality gap in the $\|\cdot\|_1$-norm for our heuristic on commodity group 4.

Commodity group 5:
$I = [16744, 18506]$, $|\mathcal{S}_\pi| = 5$, $|\mathcal{B}_\pi| = 1175$:

	$\kappa = 1$	$\kappa = 2$	$\kappa = 3$	$\kappa = 4$	$\kappa = 5$
CPLEX	4178.71	3418.37	3067.74	2874.70	2786.69
heuristic	4178.71	3418.87	3068.25	2875.21	2787.21
gap	0.000%	0.015%	0.017%	0.018%	0.019%

Table 5: Optimality gap in the $\|\cdot\|_1$-norm for our heuristic on commodity group 5.

Commodity group 6:
$I = [11000, 13000]$, $|\mathcal{S}_\pi| = 4$, $|\mathcal{B}_\pi| = 1030$:

	$\kappa = 1$	$\kappa = 2$	$\kappa = 3$	$\kappa = 4$	$\kappa = 5$
CPLEX	2812,22	2311,45	2100,78	1987,46	1909,21
heuristic	2812,22	2311,87	2101,25	1987,93	1909,63
gap	0.000%	0.018%	0.022%	0.024%	0.022%

Table 6: Optimality gap in the $\|\cdot\|_1$-norm for our heuristic on commodity group 6.

Commodity group 7:
$I = [15646, 17293]$, $|\mathcal{S}_\pi| = 5$, $|\mathcal{B}_\pi| = 1098$:

	$\kappa = 1$	$\kappa = 2$	$\kappa = 3$	$\kappa = 4$	$\kappa = 5$
CPLEX	4501.84	3917.96	3755.20	3660.32	3575.55
heuristic	4501.84	3918.46	3755.70	3660.84	3576.04
gap	0.000%	0.013%	0.013%	0.014%	0.014%

Table 7: Optimality gap in the $\|\cdot\|_1$-norm for our heuristic on commodity group 7.

Commodity group 8:
$I = [11274, 12461]$, $|\mathcal{S}_\pi| = 5$, $|\mathcal{B}_\pi| = 989$:

	$\kappa = 1$	$\kappa = 2$	$\kappa = 3$	$\kappa = 4$	$\kappa = 5$
CPLEX	3191.66	2771.89	2575.37	2424.31	2331.67
heuristic	3191.66	2772.33	2575.81	2424.75	2332.11
gap	0.000%	0.016%	0.017%	0.018%	0.019%

Table 8: Optimality gap in the $\|\cdot\|_1$-norm for our heuristic on commodity group 8.

Commodity group 9:
$I = [9211, 10181]$, $|\mathcal{S}_\pi| = 5$, $|\mathcal{B}_\pi| = 808$:

	$\kappa = 1$	$\kappa = 2$	$\kappa = 3$	$\kappa = 4$	$\kappa = 5$
CPLEX	3616.71	3215.17	2981.02	2837.66	2732.29
heuristic	3616.71	3215.53	3009.01	2860.85	2758.39
gap	0.000%	0.011%	0.939%	0.817%	0.955%

Table 9: Optimality gap in the $\|\cdot\|_1$-norm for our heuristic on commodity group 9.

Besides these nine test calculations we have done several calculations on our data sets with different parameters, we have, e. g., considered case with fewer sizes, fewer branches, smaller or larger cardinality intervals, larger κ, or other magnitudes for the cardinality interval. The results are from a qualitative point of view more or less the same, as for the presented test calculations.

6 Conclusion and outlook

Starting from a real world optimization problem we have formalized a new general optimization problem, which we call cardinality constrained p-median problem. It turns out that this problem is related to several other well known standard optimization problems. In Subsection 3.5 we have given an integer linear programming formulation which has a very strong LP-relaxation. Nevertheless this approach is quit fast (computing times below one hour), there was a practical need for fast heuristics to solve the problem. We have presented one such heuristic which performs very well on real world data sets with respect to the optimality gap.

Some more theoretic work on the cardinality constrained p-median problem and its relationships to other classical op-

timization methods may lead to even stronger integer linear programming formulations or faster branch-and-bound frameworks enhanced with some graph theoretic algorithms.

We leave also the question of a good approximation algorithm for the cardinality constrained p-median problem. Having the known approximation algorithms for the other strongly related classical optimization problems in mind, we are almost sure that it should be not too difficult to develop good approximation algorithms for our problem.

For the practical problem the uncertainties and difficulties concerning the demand estimation have to be faced. There are several ways to make solutions of optimization problems more robust. One possibility is to utilize robust optimization methods. Another possibility is to consider the branch- and size dependent demands as stochastic variables and to utilize integer linear stochastic programming techniques. See, e. g., [3] or more specifically [29]. These enhanced models, however, will challenge the solution methods a lot, since the resulting problems are of a much larger scale than the one presented in this paper. Nevertheless, this is exactly what we are looking at next.

Bibliography

[1] J. S. Armstrong (ed.), *Principles of forecasting: A handbook for researchers and practitioners*, Kluwer, 2001.

[2] V. Arya, N. Garg, R. Khandekar, K. Munagala, A. Meyerson, and V. Pandit, *Local search heuristics for k-median and facility location problems*, 33rd Annual ACM Symposium on Theory of Computing (STOC), 2001, 21–29.

[3] J. R. Birge and F. Louveaux, *Introduction to stochastic programming*, Springer Series in Operations Research and Financial Engineering, Springer, 1997.

[4] J. W. S. Cassels, *An introduction to diophantine approximation*, Cambridge, 1965.

[5] M. Charikar, S. Guha, E. Tardos, and D. Shmoys, *A constant-factor approximation algorithm for the k-median problem*, 31st Annual ACM Symposium on the Theory of Computing, 1999, 1–10.

[6] M. Charikar, S. Guha, E. Tardos, and D. Shmoys, *A constant-factor approximation algorithm for the k-median problem*, JCSS **65** (2002), no. 1, 129–149.

[7] G. Cornuéjols, G. L. Nemhauser, and L. A. Wolsey, *The uncapacitated facility location problem*, Discrete Location Theory (P. Mirchandani and R. Francis, eds.), John Wiley and Sons, Inc., New York, 1990, 119–171.

[8] J. A. Díaza and E. Fernández, *Hybrid scatter search and path relinking for the capacitated p-median problem*, European J. Oper. Res. **169** (2006), no. 2, 570–585.

[9] A. Frank and E. Tardos, *An application of simultaneous diophantine approximation in combinatorial optimization*, Combinatorica **7** (1987), no. 1, 49–65.

[10] F. García, B. Melian, J. A. Moreno, and J. M. Moreno-Vega, *Scatter search for multiple objective p-facility location problems*, Workshop on Multiobjective Metaheuristics, 2002.

[11] D. Ghosh, *Neighborhood search heuristics for the uncapacitated facility location problem*, European J. Oper. Res. **150** (2003), no. 1, 150–162.

[12] S. Guha and S. Khuller, *Greedy strikes back: improved facility location algorithms*, ninth annual ACM-SIAM symposium on Discrete algorithms, 1998, 649–657.

[13] M. T. Hajiaghayi, M. Mahdian, and V. S. Mirrokni, *The facility location problem with general cost functions*, Networks **42** (2003), no. 1, 6 p.

[14] P. Hansen and N. Mladenović, *Variable neighborhood search for the p-median*, Location Science **5** (1997), no. 4, 207–226.

[15] K. Jain, M. Mahdian, E. Markakis, A. Saberi, and V. V. Vazirani, *Greedy facility location algorithms analyzed using dual fitting with factor-revealing LP*, J. ACM **50** (2003), no. 6, 795–824.

[16] K. Jain and V. V. Vazirani, *Primal-dual approximation algorithms for metric facility location and k-median problems*, 40th Annual Symposium on Foundations of Computer Science, 1999, 2 p.

[17] A. Kok and M. Fisher, *Demand estimation and assortment optimization under substitution: Methodology and application*, Oper. Res. (to appear).

[18] M. R. Korupolu, C. G. Plaxton, and R. Rajaraman, *Analysis of a local search heuristic for facility location problems*, J. Algorithms **37** (2000), no. 1, 146–188.

[19] S. Kurz and J. Rambau, *Demand forecasting for companies with many branches, low sales numbers per product, and non-recurring orderings*, Proceedings of the Seventh International Conference on Intelligent Systems Design and Applications, 2007, 196–201.

[20] S. Kurz, J. Rambau, J. Schlüchtermann, and R. Wolf, *The top-dog index: A new measurement for the demand consistency of the size distribution in prepack orders for a fashion discounter with many small branches*, (submitted).

[21] S. Lang, *Introduction to diophantine approximations (new expanded edition)*, Springer-Verlag, 1995.

[22] D. Leviatan and V. Temlyakov, *Simultaneous approximation by greedy algorithms*, Adv. Comput. Math. **25** (2006), no. 1-3, 73–90.

[23] J.-H. Lin and J. S. Vitter, *e-approximations with minimum packing constraint violation (extended abstract)*, twenty-fourth annual ACM symposium on Theory of computing, 1992, 771–782.

[24] S. Makridakis, S. Wheelwright, and R. Hyndman, *Forecasting:methods and applications*, Wiley, 2004.

[25] G. L. Nemhauser and L. A. Wolsey, *Integer and combinatorial optimization*, John Wiley and Sons, Inc., New York, 1988.

[26] C. S. ReVelle, H. A. Eiselt, and M. S. Daskin, *A bibliography for some fundamental problem categories in discrete location science*, European J. Oper. Res. **184** (2008), 817–848.

[27] W. M. Schmidt, *Simultaneous approximation to algebraic numbers by rationals*, Acta Math. **125** (1970), no. 1, 189–201.

[28] D. B. Shmoys, E. Tardos, and K. Aardal, *Approximation algorithms for facility location problems (extended abstract)*, twenty-ninth annual ACM symposium on Theory of computing, 1997, 265–274.

[29] C. Swamy, *Approximation algorithms for clustering problems*, Ph.D. thesis, Cornell University, 2004.

[30] N. E. Young, *K-medians, facility location, and the chernoff-wald bound*, eleventh annual ACM-SIAM symposium on Discrete algorithms, 2000, 86–95.

[31] J. Yurkiewicz, *Software survey: Forecasting 2000*, OR/MS Today **27** (2000), no. 1.

Chapter 15

On the Hegselmann-Krause conjecture in opinion dynamics

SASCHA KURZ[1] AND JÖRG RAMBAU[2]

ABSTRACT. We give an elementary proof of a conjecture by R. Hegselmann and U. Krause in opinion dynamics, concerning a symmetric bounded confidence interval model: If there is a truth and all individuals take each other seriously by a minimal positive amount, then all truth seekers will converge to the truth. Here truth seekers are the individuals which are attracted by the truth by a positive amount. In the absence of truth seekers it was already shown by Hegselmann and Krause that the opinions of the individuals converge.

2000 MSC: 40A99; 62P25, 91D10, 37N99.
Key words and phrases: opinion dynamics, consensus/dissent bounded confidence, non-linear dynamical systems.

1 Introduction

In this article we consider a symmetric bounded confidence interval model to model opinion dynamic and prove a conjecture of R. Hegselmann and U. Krause. Suppose there is a set $[n] := \{1,\dots,n\}$ of individuals with opinion $x_i(t) \in [0,1]$ at time t for all $i \in [n]$, $t \in \mathbb{N}$. The abstract truth is modeled as a constant over time denoted by $h \in [0,1]$. The opinion of an individual $i \in [n]$ is influenced in a time step t only by those individuals which have a similar opinion, more precisely which have an opinion in the *confidence set* of $x_i(t)$.

Definition 1.1 *For $x \in [0,1]$ and a parameter $\varepsilon \geqslant 0$ we define the **confidence set of value** x **at time** t as*

$$I_x^\varepsilon(t) := \{j \in [n] \mid |x - x_j(t)| \leqslant \varepsilon\}.$$

As a shorthand we define $I_i^\varepsilon(t) := I_{x_i(t)}^\varepsilon(t)$ for any $i \in [n]$.

The update of the opinions is modeled as a weighted arithmetic mean of opinions in the confidence set and a possible attraction towards the truth.

Definition 1.2 *Let $n \in \mathbb{N}$, $h \in [0,1]$, $\varepsilon \in [0,1]$, $\alpha \in (0,1]$, $\beta \in \left(0,\frac{1}{2}\right]$, $\alpha_i(t) \in [\alpha,1]$ or $\alpha_i(t) = 0$ for all $t \in \mathbb{N}$, $\beta_{ij}(t) \in [\beta, 1-\beta]$ with $\sum_{j=1}^{n} \beta_{ij}(t) = 1$ for all $i \in [n]$ and for all $t \in \mathbb{N}$, and starting positions $x_i[0] \in [0,1]$. These parameters define a weighted arithmetic mean symmetric bounded confidence model (WASBOCOD) for opinion dynamics. Now we define the update of the opinions by*

$$x_i(t+1) := \alpha_i(t)\cdot h + \left(1-\alpha_i(t)\right)\frac{\sum_{j\in I_i^\varepsilon(t)} \beta_{ij}(t)x_j(t)}{\sum_{j\in I_i^\varepsilon(t)} \beta_{ij}(t)}. \quad (1)$$

*The value h is called the **truth**. **Individuals** are members of the index set $[n]$. **Truth seekers** are members of the set $K := \{k \in [n] \mid \alpha_k(t) \geqslant \alpha\, \forall t \in \mathbb{N}\}$. All other individuals, i. e. those with $\alpha_i(t) = 0$ for all t, are called **ignorants**; their set is denoted by $\overline{K}$.*

The main result we wish to prove is the following:

Theorem 1.3 *(Generalized Hegselmann-Krause Conjecture) All truth seekers in an (WASBOCOD) Ω converge to the truth h. Formally, for each $\gamma > 0$ and each setting Ω there exists a $T(\gamma,\Omega)$ so that we have $|x_k(t) - h| < \gamma$ for all $k \in K$ and all $t \geqslant T(\gamma,\Omega)$.*

This is a somewhat weak convergence since it depends on the complete setting Ω and not only on the structural parameters ε, α, β, and n. Actually, such a stronger convergence can not be expected in general:

Example 1.4 *Consider a (WASBOCOD) with truth $h = \varepsilon$, $\varepsilon > 0$, $\alpha_1(t) = \alpha$, $\alpha_2(t) = 0$, $\beta_{ij}(t) = \frac{1}{2}$ (for all $t \in \mathbb{N}_0$), $\beta = \frac{1}{2}$, $x_1(0) = 2\varepsilon$, $x_2(0) = \tilde{\varepsilon}$, where $\varepsilon > \tilde{\varepsilon} > 0$. Let $T \in \mathbb{N}_0$ be the smallest integer such that $(1-\alpha)^T\varepsilon \leqslant \tilde{\varepsilon}$. Then by induction we have $x_1(t) = \varepsilon + (1-\alpha)^t\varepsilon$ and $x_2(t) = x_2(0) = \tilde{\varepsilon}$ for all $t \leqslant T$. So truth seeker 1 converges to the truth, but at time $T+1$ we have $x_1(T+1) = \alpha\varepsilon + \frac{1-\alpha}{2}\left(\varepsilon + (1-\alpha)^T\varepsilon + \tilde{\varepsilon}\right) \leqslant \frac{1+\alpha}{2}\cdot\varepsilon + \tilde{\varepsilon}$.*

[1]Sascha Kurz, University of Bayreuth, Department of Mathematics, 95440 Bayreuth, Germany.
E-mail adress: sascha.kurz@uni-bayreuth.de
[2]Jörg Rambau, University of Bayreuth, Department of Mathematics, 95440 Bayreuth, Germany.
E-mail adress: joerg.rambau@uni-bayreuth.de

This example shows that in general we can not expect convergence in a strong sense, meaning that for every $\gamma > 0$ there is a $T(\gamma, \varepsilon, \alpha, \beta, n)$ that for all $t \geqslant T(\gamma, \varepsilon, \alpha, \beta, n)$ we have $|x_k(t) - h| < \gamma$ for each truth seeker $k \in K$, since we may choose $\tilde{\varepsilon}$ arbitrarily small. But we may have an **interrupted convergence**:

Definition 1.5 *In our situation we say that the convergence of the truth seekers $k \in K$ is (1-fold) **interrupted convergent**, if there exist two functions $T_1^s(\gamma, \varepsilon, \alpha, \beta, n)$ and $T_2^s(\gamma, \varepsilon, \alpha, \beta, n, T_1^e)$, so that for each (WASBOCOD) Ω, with structural parameters ε, α, β and n, there exists an $T_1^e \in \mathbb{N}_0$ fullfilling*

$$\forall k \in K,\ \forall t \in [T_1^s(\gamma, \varepsilon, \alpha, \beta, n), T_1^e] : |x_k(t) - h| < \gamma,$$
$$\forall k \in K,\ \forall t \geqslant T_2^s(\gamma, \varepsilon, \alpha, \beta, n, T_1^e) : |x_k(t) - h| < \gamma.$$

With this definition we can sharpen Theorem 1.3 so that it becomes a corollary of the following:

Theorem 1.6 *All truth seekers in an (WASBOCOD) Ω are (1-fold) interrupted convergent.*

Originally Hegselmann and Krause considered the (WASBOCOD) model for $\alpha_i(t) \in \{0, \alpha\}$ and $\beta_{ij}(t) = \frac{1}{n}$ for all $t \in \mathbb{N}_0$. In the case of complete absence of truth seekers they have already proved, that the opinion of each individual converges, as can be expected, not necessarily to the truth. In fact in general the individuals form several clusters, where two individuals of different clusters converge to different opinions.

We give an example without truth seekers where the individuals will converge to five different clusters.

Example 1.7 *Consider a (WASBOCOD) with $\alpha_i(t) = 0$ (no truth seekers), $\beta = \beta_{ij}(t) = \frac{1}{n}$, $n = 12$, and $\alpha = h = \frac{1}{2}$ for formal reasons. The starting positions are given by*

$$x_1(0) = x_2(0) = 0,\ x_3(0) = \varepsilon,\ x_4(0) = 2\varepsilon,$$
$$x_5(0) = 3\varepsilon,\ x_6(0) = x_7(0) = 4\varepsilon,\ x_8(0) = 5\varepsilon,$$
$$x_9(0) = 6\varepsilon,\ x_{10}(0) = 7\varepsilon, \text{ and } x_{11}(0) = x_{12}(0) = 8\varepsilon,$$

see Figure 1.

In Table 1 we give the complete dynamics of the opinions of all 12 individuals over time until the opinion of every individual has converged. For brevity we write x_i instead of $x_i(t)$ and drop the constant ε by writing 2 instead of 2ε. After three time steps we have reached a stable state, see Figure 2.

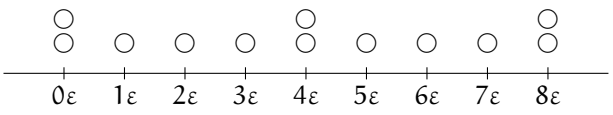


Figure 1: Starting positions of the individuals in Example 1.7.

	x_1				x_6				x_{11}
t	x_2	x_3	x_4	x_5	x_7	x_8	x_9	x_{10}	x_{12}
0	0	1	2	3	4	5	6	7	8
1	$\frac{1}{3}$	$\frac{3}{4}$	2	$\frac{13}{4}$	4	$\frac{19}{4}$	6	$\frac{29}{4}$	$\frac{23}{3}$
2	$\frac{17}{36}$	$\frac{17}{36}$	2	$\frac{15}{4}$	4	$\frac{17}{4}$	6	$\frac{271}{36}$	$\frac{271}{36}$
3	$\frac{17}{36}$	$\frac{17}{36}$	2	4	4	4	6	$\frac{271}{36}$	$\frac{271}{36}$

Table 1: Dynamics of Example 1.7.

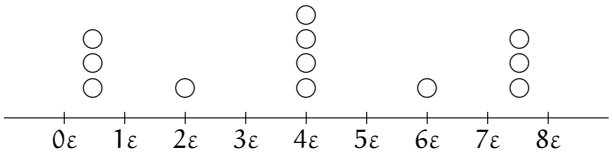


Figure 2: Final positions of the individuals in Example 1.7.

We remark that for symmetric weights $\beta_{ij}(t) = \beta_{ji}(t)$ one can easily show that in the absence of truth seekers the dynamics becomes stable after a finite number of time steps. In the case of asymmetric weights $\beta_{ij}(t) \neq \beta_{ji}(t)$ we only have convergence, but need not reach a stable state after an arbitrary, problem dependent, but finite number of time steps, as illustrated in the following example.

Example 1.8 *Consider a (WASBOCOD) with $\alpha_i(t) = 0$ (no truth seekers), $n = 2$, $x_1(0) = 0$, $x_2(0) = \varepsilon$, $\beta_{11}(t) = \frac{2}{3}$, $\beta_{12}(t) = \frac{1}{3}$, $\beta_{21}(t) = \frac{1}{2}$, $\beta_{22}(t) = \frac{1}{2}$ for all $t \geqslant 0$, and $\beta = \alpha = h = \frac{1}{3}$ for formal reasons.*

One can easily verify, e. g. by induction, that we have

$$x_1(t) = \left(\frac{2}{5} - \frac{2}{5 \cdot 6^t}\right) \cdot \varepsilon \text{ and } x_2(t) = \left(\frac{2}{5} + \frac{3}{5 \cdot 6^t}\right) \cdot \varepsilon$$

for all $t \in \mathbb{N}_0$. So we have $|x_1(t) - x_2(t)| = \frac{1}{6^t} \cdot \varepsilon > 0$ but clearly the opinions of the two individuals converge to $\frac{2}{5}$.

All stated insights with the absence of truth seekers were known so far. It becomes a bit more interesting if we allow truth seekers, i. e. if we consider a general (WASBOCOD).

Example 1.9 *Consider a (WASBOCOD) with $\alpha_1(t) = \alpha$, $\alpha_i(t) = 0$ for $i \neq 1$, $\beta_{ij}(t) = \frac{1}{n}$, $h = \frac{1}{2}\varepsilon$, $x_1(0) = 0$, and $x_i(0) = \varepsilon$ for $i \neq 1$. The opinion u_t of the truth seeker 1 at time t and the opinion v_t of the other ignorants at time $t > 0$ are given by*

$$u_t = \left[\frac{1}{2} - \alpha\left(\frac{1}{2} - \frac{1}{n}\right)\left(1 - \frac{\alpha}{n}\right)^{t-1}\right]\varepsilon,$$
$$v_t = \left[\frac{1}{2} + \left(\frac{1}{2} - \frac{1}{n}\right)\left(1 - \frac{\alpha}{n}\right)^{t-1}\right]\varepsilon$$

respectively. This can be verified e. g. by induction. We see that the opinion of the truth seekers, and here also those of the ignorants, converge to the truth $h = \frac{1}{2}\varepsilon$ (not all opinions of ignorants must converge to the truth as one can see by adding some further ignorants with $\tilde{x}_i(0) = 3\varepsilon$).

As our analysis of the previous example was rather technical we also depict the situation for special values $n = 6$ and $\alpha = \frac{2}{3}$ in Figure 3 where we depict the truth seeker by a filled circle and the ignorants by an empty circle.

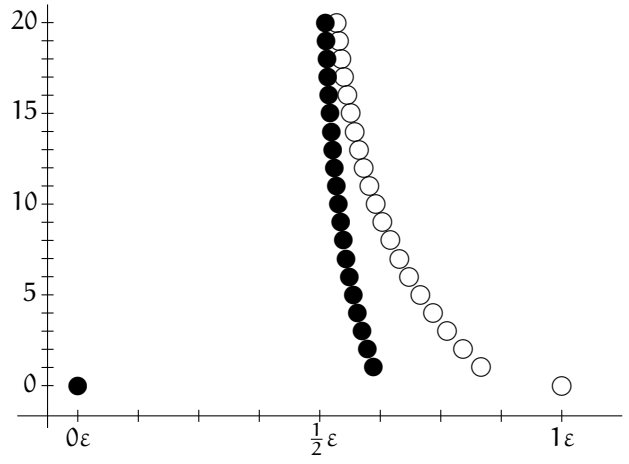


Figure 3: Dynamics of example 1.9.

One can easily imagine more complicated configurations as in Example 1.9 where one has little chance and willingness to describe the situation analytically. Our result Theorem 1.3 states that whatever the parameters of a (WASBOCOD) are the opinions of the truth seekers converge to the truth. This settles an open conjecture of Hegselmann and Krause.

2 The crucial objects

To get a first impression of what we may expect in terms of convergence we consider a lonely truth seeker, i. e. $n = 1$.

Lemma 2.1 *For a lonely truth seeker* $i = 1$ *we have*

$$|x_i(t+r) - h| \leqslant |x_i(t) - h| \cdot (1-\alpha)^r.$$

PROOF.

$$\begin{aligned} |x_i(t+1) - h| &= |x_i(t) - h| \cdot (1 - \alpha_i(t)) \\ &\leqslant |x_i(t) - h| \cdot (1-\alpha). \end{aligned}$$

□

Clearly this bound is sharp. Similar to this very special situation of a lonely truth seeker is the case $\varepsilon = 0$, so that we now assume $\varepsilon > 0$ for the remaining part of this article.

To describe the state of the discrete time dynamical system we look at the truth seekers with the most extreme opinion.

Definition 2.2 *We define* $\tilde{u}(t) \in K$ *as the lexicographically smallest truth seeker which fullfills* $x_{\tilde{u}(t)}(t) \geqslant h$ *and* $x_{\tilde{u}(t)}(t) \geqslant x_k(t)$ *for all* $k \in K$. *If there is no truth seeker with opinion greater or equal to the truth* h *we set* $\tilde{u}(t) = 0$, *where* $x_0(t') := h$ *for all* $t' \in \mathbb{N}_0$. *Similar we define* $\tilde{l}(t)$ *as the lexicographically smallest truth seeker which fullfills* $x_{\tilde{l}(t)}(t) \leqslant h$ *and* $x_{\tilde{l}(t)}(t) \leqslant x_k(t)$ *for all* $k \in K$. *Again we set* $\tilde{l}(t) = 0$ *if there is no such truth seeker.*

Due to the *symmetrical* or *fair* definition of the confidence set, the intermediate interactions between the individuals can be described as a simple graph with loops.

Definition 2.3 *The* ***confidence graph*** $G(t)$ *with vertex set* $V(t)$ *and edge set* $E(t)$, *of a configuration* $x(t) = (x_1(t), \ldots, x_n(t)) \in \mathbb{R}^n$ *is defined as follows:*

$$\begin{aligned} V(t) &:= [n] \cup \{0\}, \\ E(t) &:= \{\{i,j\} \mid |x_i(t) - x_j(t)| \leqslant \varepsilon\}. \end{aligned}$$

For $i \in V(t)$ *let* $C_i(t)$ *be the set of vertices in the connectivity component of vertex* i *in* $G(t)$.

Because we want to keep track of the individuals which can influence the truth seekers in the future, we give a further definition for individuals, which is similar to Definition 2.2 for truth seekers.

Definition 2.4 *We define* $\hat{u}(t) \in C_{\tilde{u}(t)}(t)$ *as the lexicographically smallest individual with* $x_{\hat{u}(t)}(t) \geqslant x_c(t)$ $\forall c \in C_{\tilde{u}(t)}(t)$ *and* $\hat{l}(t) \in C_{\tilde{l}(t)}(t)$ *as the lexicographically smallest individual with* $x_{\hat{l}(t)}(t) \leqslant x_c(t)$ $\forall c \in C_{\tilde{l}(t)}(t)$ *for all* $t \in \mathbb{N}_0$.

The opinions of $\hat{u}(t)$ and $\hat{l}(t)$ form an interval $\left[x_{\hat{l}(t)}(t), x_{\hat{u}(t)}(t)\right]$ called the **hope interval** which is crucial for our further investigations. To prove the main theorem we will show that the length of this hope interval converges to zero.

In Figure 4 we have depicted a configuration to illustrate Definition 2.2 and Definition 2.4. Here we have $\tilde{l} = 4$, $\tilde{u} = 9$, $\hat{l} = 2$, and $\hat{u} = 12$. Individual 1 is *lost* and not contained in the hope interval, because there is no path in G from 1 to $\tilde{l} = 4$. So we already know that the opinion of individual 1 will not converge to the truth.

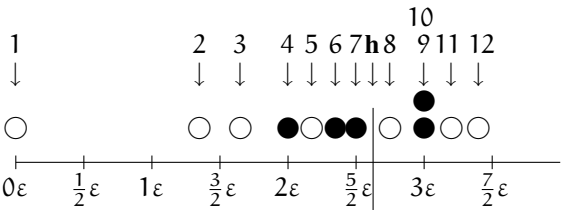


Figure 4: Illustration of Definition 2.2 and Definition 2.4.

In the configuration depicted in Figure 5 we have $\tilde{l} = 2$, $\tilde{u} = 0$, $\hat{l} = 2$, and $\hat{u} = 5$.

We remark that due to the possibility of asymmetric weights β_{ij} the sequence of the opinions of the individuals may reorder during the time steps. Therefore let us consider e. g. three ignorants with starting positions $x_1(0) = 1\varepsilon$, $x_2(0) = \frac{3}{2}\varepsilon$, and $x_3(0) = 2\varepsilon$. The weights may be given as

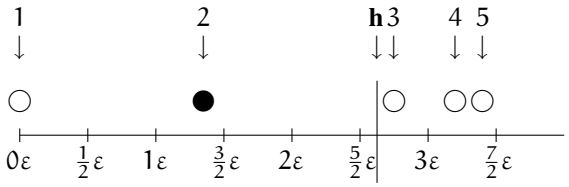


Figure 5: Illustration of a special case in Definition 2.2.

$\beta_{11}(0) = 0.01$, $\beta_{12}(0) = 0.01$, $\beta_{13}(0) = 0.98$, $\beta_{21}(0) = 0.98$, $\beta_{22}(0) = 0.01$, $\beta_{23}(0) = 0.01$, $\beta_{31}(0) = 0.4$, $\beta_{32}(0) = 0.4$, and $\beta_{33}(0) = 0.2$. After one time step the new opinions are given by $x_1(1) = 1.985\varepsilon$, $x_2(1) = 1.015\varepsilon$, and $x_3(1) = 1.4\varepsilon$. We remark that it is possible to achieve every ordering of the three opinions in one time step by choosing suitable weights β_{ij} in this example. Nevertheless we have the following trivial lemma:

Lemma 2.5 *Let i be an ignorant, $l \in I_i^\varepsilon(t)$ be an individual with smallest opinion and $u \in I_i^\varepsilon(t)$ be an individual with largest opinion then we have $x_i(t+1) \in [x_l(t), x_u(t)]$.*

PROOF. Follows from update equation (1). □

For truth seekers we have a similar lemma:

Lemma 2.6 *Let i be a truth seeker, $l \in I_i^\varepsilon(t)$ be an individual with smallest opinion and $u \in I_i^\varepsilon(t)$ be an individual with largest opinion. For $x_i(t) \leqslant h$ we have $x_i(t+1) \in [x_l(t), \max(h, x_u(t))]$ and for $x_i(t) \geqslant h$ we have $x_i(t+1) \in [\min(h, x_l(t)), x_u(t)]$.*

Our aim is to prove that the length of the hope interval converges to zero. So at first we show that the length does not increase after an iteration of Equation (1).

Lemma 2.7 *For all time steps $t \in \mathbb{N}_0$ we have $x_{\hat{u}(t+1)}(t+1) \leqslant x_{\hat{u}(t)}(t)$ and $x_{\hat{l}(t+1)}(t+1) \geqslant x_{\hat{l}(t)}(t)$.*

PROOF. We only prove the last inequality since the proof is similar for the first inequality. Due to Definition 2.4 we have $x_{\hat{l}(t+1)}(t+1) \leqslant h$ and $x_{\hat{l}(t)}(t) \leqslant h$. By $\mathcal{L}(t)$ we denote the set of individuals with opinion strictly smaller than $x_{\hat{l}(t)}(t)$, this is $\mathcal{L}(t') := \left\{ i \in [n] \mid x_i(t') < x_{\hat{l}(t')}(t') \right\}$ for all $t' \geqslant t$. We remark that by definition $\mathcal{L}(t')$ does not contain a truth seeker. We set $\mathcal{U}(t') := [n] \backslash \mathcal{L}(t')$, which contains the remaining individuals.

Let u be an individual in $\mathcal{L}(t)$ with the largest opinion. By applying Lemma 2.5 we get $x_i(t+1) \leqslant x_u(t)$ for all $i \in \mathcal{L}(t)$. Now let l (e. g. $l = \hat{l}(t)$) be an individual in $\mathcal{U}(t)$ with smallest opinion then by applying Lemma 2.5 and Lemma 2.6 we receive $x_i(t+1) \geqslant x_l(t)$ for all $i \in \mathcal{U}(t)$. Thus we have $\hat{l}(t+1) \in \mathcal{U}(t)$ and so $x_{\hat{l}(t+1)}(t+1) \geqslant x_{\hat{l}(t)}(t)$ follows. □

In the remaining part of this article we prove that the length of the hope interval $\left| x_{\hat{u}(t)}(t) - x_{\hat{l}(t)}(t) \right|$ converges (in some special sense) to zero, as t tends to infinity.

3 Proof of the Hegselmann-Krause Conjecture

At first we show that after a finite number T of time steps, depending only on n, ε, α, and β, the hope interval $\left[x_{\hat{l}(T)}(T), x_{\hat{u}(T)}(T) \right]$ is contained in the interval $\left[h - \varepsilon - \frac{\varepsilon\alpha\beta}{12}, h + \varepsilon + \frac{\varepsilon\alpha\beta}{12} \right]$. Therefore we give:

Definition 3.1 ***A good iteration*** *is an iteration where for a bounded r one of the following conditions is fullfilled:*

(1) the number of individuals in the hope interval decreases,

(2) the opinion of $\hat{l}(t+r)$ reaches or passes $h - \varepsilon - \frac{\varepsilon\alpha\beta}{12}$,

(3) the opinion of $\hat{u}(t+r)$ reaches or passes $h + \varepsilon + \frac{\varepsilon\alpha\beta}{12}$,

(4) $\left| x_{\hat{u}(t+r)}(t+r) - x_{\hat{u}(t)}(t) \right| \geqslant \frac{\varepsilon\alpha\beta^2}{12}$,

(5) $\left| x_{\hat{l}(t+r)}(t+r) - x_{\hat{l}(t)}(t) \right| \geqslant \frac{\varepsilon\alpha\beta^2}{12}$.

Clearly, there is only a finite number of good iterations. We may choose $T = r \cdot \left(n + 2 \cdot 1 + 2 \cdot \frac{12}{\varepsilon\alpha\beta^2} \right)$. We formulate the next lemmas only for the lower bound $x_{\hat{l}(t)}(t)$ because analog arguments hold for $x_{\hat{u}(t)}(t)$. As a shorthand we define $d(i,j,t) := |x_i(t) - x_j(t)|$. For each point in time t we define the sets

$$\mathcal{N}(t) := \left\{ i \in [n] \mid d\left(\hat{l}(t), i, t\right) \in \left[0, \frac{\varepsilon\alpha\beta}{12}\right) \right\},$$

$$\mathcal{M}(t) := \left\{ i \in [n] \mid d\left(\hat{l}(t), i, t\right) \in \left[\frac{\varepsilon\alpha\beta}{12}, \varepsilon\right] \right\}, \text{ and}$$

$$\mathcal{F}(t) := \left\{ i \in [n] \mid x_i(t) - x_{\hat{l}(t)}(t) > \varepsilon \right\}.$$

Lemma 3.2 *If $\mathcal{M}(t) \neq \emptyset$ then there is a good iteration after 1 step.*

PROOF. We assume that there is an individual $j \in \mathcal{M}(t)$ with $d\left(\hat{l}(t), j, t\right) \in \left[\frac{\varepsilon\alpha\beta}{12}, \varepsilon\right]$. For the evaluation of Equation (1) for elements of $\mathcal{N}$, $\mathcal{M}$, or $\mathcal{F}$ we do not need to consider the opinion of individuals in $[n] \backslash (\mathcal{N} \cup \mathcal{M} \cup \mathcal{F})$. Let i be an element of $\mathcal{N}$ with opinion $x_i(t) = x_{\hat{l}(t)} + \delta$, where $0 \leqslant \delta < \frac{\varepsilon\alpha\beta}{12}$. Let us first assume that i is an ignorant. Due to individual j we have $x_i(t+1)$

$$\geqslant x_i(t) - \underbrace{\delta(1-2\beta)}_{\text{individuals in } \mathcal{N}\backslash\{i\}} + \underbrace{0 \cdot \beta}_{i} + \underbrace{\left(\frac{\varepsilon\alpha\beta}{12} - \delta\right) \cdot \beta}_{j}$$

$$\geqslant x_{\hat{l}(t)} + \frac{\varepsilon\alpha\beta^2}{12}.$$

For a truth seeker we similarly receive $x_i(t+1)$

$$\geqslant x_i(t) + \alpha\varepsilon + (1-\alpha)\left(-\delta(1-2\beta) + \left(\frac{\varepsilon\alpha\beta}{12} - \delta\right)\cdot\beta\right)$$

$$\geqslant x_{\hat{l}(t)} + \frac{\varepsilon\alpha\beta^2}{12}.$$

Now let i be an element of $\mathcal{M} \cup \mathcal{F}$ with $x_i(t) = x_{\hat{l}(t)} + \delta$ where $\delta \geqslant \frac{\varepsilon\alpha\beta}{12}$. In any case ($i$ being a truth seeker or an ignorant) we have

$$x_i(t+1) \geqslant x_{\hat{l}(t)} + \delta - \underbrace{\delta(1-\beta)}_{\text{individuals with smaller opinion than } i} + \beta \cdot 0 \geqslant x_{\hat{l}(t)} + \frac{\varepsilon\alpha\beta^2}{12}.$$

□

Lemma 3.3 *If $x_{\hat{l}(t)} < h - \varepsilon - \frac{\varepsilon\alpha\beta}{12}$ then after at least 3 time steps we have a good iteration.*

PROOF. Due to Lemma 3.2 we can assume $\mathcal{M}(t) = \mathcal{M}(t+1) = \mathcal{M}(t+2) = \emptyset$. We can also assume

$$\begin{aligned} \left| x_{\hat{l}(t)}(t) - x_{\hat{l}(t+1)}(t+1) \right| &< \frac{\varepsilon\alpha\beta^2}{12}, \\ \left| x_{\hat{l}(t+1)}(t+1) - x_{\hat{l}(t+2)}(t+2) \right| &< \frac{\varepsilon\alpha\beta^2}{12}, \text{ and} \\ d\left(\hat{l}(t+1), 0, t+1\right) &> \varepsilon + \frac{\varepsilon\alpha\beta}{12} \end{aligned}$$

since otherwise we have a good iteration in at most 2 time steps. At first we claim $\mathcal{N}(t+1) \cap K = \emptyset$. If at time t there is a truth seeker $i \in \mathcal{N}(t) \cap K$ then we have

$$\begin{aligned} x_i(t+1) &\geqslant x_{\hat{l}(t)}(t) + \alpha\varepsilon - \frac{(1-\alpha)(1-\beta)\varepsilon\alpha\beta}{12} \\ &\geqslant x_{\hat{l}(t)}(t) + \frac{\varepsilon\alpha\beta^2}{12} + \frac{\varepsilon\alpha\beta}{12} \\ &\geqslant x_{\hat{l}(t+1)}(t+1) + \frac{\varepsilon\alpha\beta}{12}. \end{aligned}$$

So the only possible truth seeker who has a chance to move into the set $\mathcal{N}(t+1)$ could be those of the set $\mathcal{F}(t)$. So let truth seeker i be in the set $\mathcal{F}(t) \cap K$, with $x_i(t) = x_{\hat{l}(t)}(t) + \delta$, where $\varepsilon < \delta < \varepsilon + \frac{\varepsilon\alpha\beta}{12}$. (Truth seekers where $\delta \geqslant \varepsilon + \frac{\varepsilon\alpha\beta}{12}$ are ruled out by Lemma 2.6.) We have

$$\begin{aligned} x_i(t+1) &\geqslant \underbrace{x_{\hat{l}(t)}(t) + \varepsilon}_{\leqslant x_i(t)} - (1-\alpha)(1-2\beta) \\ &\geqslant x_{\hat{l}(t)}(t) + \varepsilon\alpha \\ &\geqslant x_{\hat{l}(t)}(t) + \frac{\varepsilon\alpha\beta^2}{12} + \frac{\varepsilon\alpha\beta}{12} \\ &\geqslant x_{\hat{l}(t+1)}(t+1) + \frac{\varepsilon\alpha\beta}{12}. \end{aligned}$$

Similarly we can deduce $\mathcal{N}(t+2) \cap K = \emptyset$. Now we can assume that the individuals of $\mathcal{N}(t+1)$, who are all ignorants, are in the hope interval at time $t+1$, since otherwise we would have a good iteration after 1 time step. So there exist individuals $i \in \mathcal{N}(t+1)$ and $j \in \mathcal{F}(t+1)$ with $|x_i(t+1) - x_j(t+1)| \leqslant \varepsilon$. We set $x_i(t+1) = x_{\hat{l}(t+1)}(t+1) + \delta$, where $0 \leqslant \delta \leqslant \frac{\varepsilon\alpha\beta}{12}$ and calculate

$$\begin{aligned} & x_i(t+2) \\ &\geqslant x_i(t+1) - (1-2\beta)\delta + \underbrace{\beta \cdot 0}_{i} + \underbrace{\beta\left(\varepsilon - \frac{\varepsilon\alpha\beta}{12}\right)}_{j} \\ &\geqslant x_{\hat{l}(t+1)}(t+1) + \frac{\varepsilon\alpha\beta^2}{12} + \frac{\varepsilon\alpha\beta}{12} \\ &\geqslant x_{\hat{l}(t+2)}(t+2) + \frac{\varepsilon\alpha\beta}{12}. \end{aligned}$$

For the other direction we have

$$\begin{aligned} x_i(t+2) &\leqslant x_i(t+1) - \underbrace{\beta\delta}_{\hat{l}(t+1)} + (1-2\beta)\varepsilon \\ &\leqslant x_{\hat{l}(t+1)}(t+1) + \frac{\varepsilon\alpha\beta}{12} + \varepsilon - 2\beta\varepsilon \\ &\leqslant x_{\hat{l}(t+1)}(t+1) + \frac{\varepsilon\alpha\beta^2}{12} + \varepsilon \\ &\leqslant x_{\hat{l}(t+2)}(t+2) + \varepsilon. \end{aligned}$$

Thus $i \in \mathcal{M}(t+2)$, which results in a good iteration in three time steps. □

Thus we can conclude:

Lemma 3.4 *After a finite number of steps T we have $x_{\hat{l}(T)}(T) \geqslant h - \varepsilon - \frac{\varepsilon\alpha\beta}{12}$ and $x_{\hat{u}(T)}(T) \leqslant h + \varepsilon + \frac{\varepsilon\alpha\beta}{12}$.*

Due to Lemma 2.1 there can not exist a general bound on the convergence that does not depend on α. We consider the two side lengths $l_2(t) := \left| x_{\hat{u}(t)}(t) - h \right|$ and $l_1(t) := \left| x_{\hat{l}(t)}(t) - h \right|$ of the hope interval. Clearly $l_1(t)$ and $l_2(t)$ are not increasing due to Lemma 2.7. For $t \geqslant T$ we have $l_1(t), l_2(t) \leqslant \varepsilon + \frac{\varepsilon\alpha\beta}{12}$

Lemma 3.5 *If $l_1(t) + l_2(t) \leqslant \varepsilon$ then we have*

$$(l_1(t+1) + l_2(t+1)) \leqslant (l_1(t) + l_2(t)) \cdot \left(1 - \frac{\beta}{2}\right).$$

PROOF. Let us assume w.l.o.g. that $l_1(t) \geqslant l_2(t) > 0$. If i is an ignorant with $x_i(t) = h - l_1(t) + \delta$ then we have $x_i(t+1)$

$$\begin{aligned} &\geqslant h - l_1(t) + \delta - (1-2\beta)\delta + \beta\left(l_1(t) + l_2(t) - \delta\right) \\ &\geqslant h - (1-\beta)l_1(t). \end{aligned}$$

For a truth seeker i with $x_i(t) = h - l_1(t) + \delta$ we have $x_i(t+1)$

$$\begin{aligned} &\geqslant h - l_1(t) + \delta - \alpha(\delta - l_1(t)) - (1-\alpha)(1-2\beta)\delta + \\ &\quad (1-\alpha)\beta\left(l_1(t) + l_2(t) - \delta\right) \\ &\geqslant h - l_1(t) + \beta\delta(1-\alpha) + \alpha l_1(t)(1-\beta) + \\ &\quad \beta l_2(t)(1-\alpha) + \beta l_1(t) \\ &\geqslant h - (1-\beta)l_1(t). \end{aligned}$$

Thus we have $l_1(t+1) \leqslant (1-\beta)l_1(t)$. Using $l_2(t+1) \leqslant l_2(t)$ and $l_2(t) \leqslant l_1(t)$ we conclude

$$\begin{aligned} l_1(t+1) + l_2(t+1) &\leqslant (1-\beta)l_1(t) + l_2(t) \\ &\leqslant \left(1 - \frac{\beta}{2}\right)\left(l_1(t) + l_2(t)\right). \end{aligned}$$

□

This states that once the length of the hope interval becomes at most ε its lengths converges to zero.

Lemma 3.6 *If there exists an individual* i *with* $\frac{\alpha\beta l_1(t)}{12} \leqslant d\left(\hat{l}(t), i, t\right) \leqslant \varepsilon$, *then we have* $l_1(t+1) \leqslant l_1(t) \cdot \left(1 - \frac{\alpha\beta^2}{12}\right)$. *If there exists an individual* i *with* $\frac{\alpha\beta l_2(t)}{12} \leqslant d\left(\hat{u}(t), i, t\right) \leqslant \varepsilon$, *then we have* $l_2(t+1) \leqslant l_2(t) \cdot \left(1 - \frac{\alpha\beta^2}{12}\right)$.

PROOF. Due to symmetry it suffices to prove the first statement. Let j be an ignorant with $x_j(t) = h - l_1(t) + \delta$, where $\delta \geqslant 0$. We have $x_j(t+1)$

$$\begin{aligned} &\geqslant h - l_1(t) + \delta - (1-2\beta)\delta + \underbrace{\beta\left(\frac{\alpha\beta l_1(t)}{12} - \delta\right)}_{i} \\ &\geqslant h - \left(1 - \frac{\alpha\beta^2}{12}\right) l_1(t). \end{aligned}$$

For a truth seeker j with $x_j(t) = h - l_1(t) + \delta$, $\delta \geqslant 0$ we have $x_j(t+1)$

$$\begin{aligned} &\geqslant h - l_1(t) + \delta + \alpha(l_1(t) - \delta) - (1-\alpha)(1-2\beta)\delta + \\ &\quad \underbrace{(1-\alpha)\beta\left(\frac{\alpha\beta l_1(t)}{12} - \delta\right)}_{i} \\ &\geqslant h - l_1(t) + \beta\delta(1-\alpha) + \alpha l_1(t)\left(1 - \frac{\alpha\beta^2}{12}\right) + \\ &\quad \frac{\alpha\beta^2 l_1(t)}{12} \\ &\geqslant h - \left(1 - \frac{\alpha\beta^2}{12}\right) l_1(t). \end{aligned}$$

□

For transparency we introduce the following six sets:

$$\begin{aligned} \mathcal{N}_1(t) &= \left\{ i \in [n] \mid d\left(\hat{l}(t), i, t\right) < \frac{\alpha\beta l_1(t)}{12} \right\}, \\ \mathcal{N}_2(t) &= \left\{ i \in [n] \mid d\left(\hat{u}(t), i, t\right) < \frac{\alpha\beta l_2(t)}{12} \right\}, \\ \mathcal{M}_1(t) &= \left\{ i \in [n] \mid \frac{\alpha\beta l_1(t)}{12} \leqslant d\left(\hat{l}(t), i, t\right) \leqslant \varepsilon \right\}, \\ \mathcal{M}_2(t) &= \left\{ i \in [n] \mid \frac{\alpha\beta l_2(t)}{12} \leqslant d\left(\hat{u}(t), i, t\right) \leqslant \varepsilon \right\}, \\ \mathcal{F}_1(t) &= \left\{ i \in [n] \mid d\left(\hat{l}(t), i, t\right) > \varepsilon,\ x_i(t) \leqslant h + l_2(t) \right\}, \\ \mathcal{F}_2(t) &= \left\{ i \in [n] \mid d\left(\hat{u}(t), i, t\right) > \varepsilon,\ x_i(t) \geqslant h - l_1(t) \right\}. \end{aligned}$$

With this the individuals of the hope interval are partitioned into

$$\mathcal{N}_1(t) \cup \mathcal{M}_1(t) \cup \mathcal{F}_1(t) = \mathcal{N}_2(t) \cup \mathcal{M}_2(t) \cup \mathcal{F}_2(t).$$

Lemma 3.7 *If for* $k \in \{1,2\}$ *there exists an ignorant* $i \in \mathcal{N}_k(t)$ *and an individual* $j \in \mathcal{F}_k(t)$ *with* $|x_i(t) - x_j(t)| \leqslant \varepsilon$ *then* $l_k(t+2) \leqslant l_k(t) \cdot \left(1 - \frac{\alpha\beta^2}{12}\right)$.

PROOF. If $l_k(t+1) > l_k(t) \cdot \left(1 - \frac{\alpha\beta^2}{12}\right)$, then it is easy to check that the influence of individual j suffices to put ignorant i in set $\mathcal{M}_k(t+1)$. In this case we can apply Lemma 3.6 □

Lemma 3.8 *If* $\mathcal{N}_k(t+1) \cap K \neq \emptyset$ *then* $l_k(t+1) \leqslant l_k(t) \cdot \left(1 - \frac{\alpha}{2}\right)$.

PROOF. Due to symmetry it suffices to consider $k = 1$. So let i be a truth seeker with $i \in \mathcal{N}_1(t+1)$. We set $x_i(t) = h - l_1(t) + \delta$ and calculate $x_i(t+1)$

$$\begin{aligned} &\geqslant h - l_1(t) + \delta + \alpha(l_1(t) - \delta) - (1-\alpha)(1-\beta)\delta \\ &\geqslant h - (1-\alpha) l_1(t). \end{aligned}$$

□

Lemma 3.9 *We have* $l_k(t+3) \leqslant l_k(t) \cdot \left(1 - \frac{\alpha\beta^2}{12}\right)$ *for at least one* $k \in \{1,2\}$.

PROOF. Due to Lemma 3.8 we can assume $\mathcal{N}_k(t+1) \cap K = \emptyset$. At time $t+1$ there must be a truth seeker i. W.l.o.g. we assume $x_i(t) \leqslant h$ and $i = \tilde{l}(t+1)$. Due to Lemma 3.6 we can assume $i \in \mathcal{F}_1(t+1)$. Now let j_1 be the ignorant with smallest opinion fulfilling $d(i, j_1, t+1) \leqslant \varepsilon$. If $j_1 \in \mathcal{N}_1(t+1)$ then we can apply Lemma 3.7 with j_1 and i. Otherwise we let j_2 be the ignorant with smallest opinion fulfilling $d(j_1, j_2, t+1) \leqslant \varepsilon$. So we have $d(j_2, i, t+1) > \varepsilon$ and $j_2 \in \mathcal{N}_1(t+1)$. Thus we can apply Lemma 3.7 with j_2 and j_1. □

Lemma 3.10 *If* $l_k(t) > \varepsilon$ *then we have* $\varepsilon - l_k(t+3) \leqslant (\varepsilon - l_k(t)) \cdot \left(1 - \frac{\alpha\beta^2}{12}\right)$ *or* $l_k(t+3) \leqslant \varepsilon$.

PROOF. Due to Lemma 3.8 we can assume $\mathcal{N}_k(t+1) \cap K = \emptyset$ and due to Lemma 3.6 we can assume $\mathcal{M}_k(t+1) = \emptyset$. Due to symmetry we only consider the case $k = 1$. Let $i \in \mathcal{N}_1(t+1)$ the ignorant with largest opinion $x_i(t+1)$, meaning that $d\left(\hat{l}(t+1), i, t+1\right)$ is maximal. If there exists an individual $j \in \mathcal{F}_1(t+1)$ with $d(i, j, t+1) \leqslant \varepsilon$, then we can apply Lemma 3.7. If no such individual j exists then we must have $d(i, 0, t+1) \leqslant \varepsilon$ or $l_1(t+1) = 0$. So only the first case remains. We set $\delta = d\left(\hat{l}(t+1), i, t+1\right) \geqslant \varepsilon - l_1(t+1)$. Let $h \in \mathcal{N}_1(t+1)$ be an ignorant with $x_h(t+1) = x_{\hat{l}(t+1)}(t+1) + \mu$, where $0 \leqslant \mu \leqslant \delta$. For time $t+2$ we get $x_h(t+2)$

$$\begin{aligned} &\geqslant x_{\hat{l}(t+1)}(t+1) + \mu - (1-2\beta)\mu + \beta(\delta - \mu) \\ &\geqslant x_{\hat{l}(t+1)}(t+1) + \beta\delta \\ &\geqslant x_{\hat{l}(t+1)}(t+1) + \beta(\varepsilon - l_1(t+1)). \end{aligned}$$

□

From Lemma 3.9 and Lemma 3.10 we conclude:

Lemma 3.11 *There exists a finite number* $T(\varepsilon, n, \alpha, \beta)$ *so that we have*

$$l_1(t) + l_2(t) \leqslant \varepsilon + \frac{\varepsilon\alpha^2\beta^3}{60}$$

and

$$\min(l_1(t), l_2(t)) \leqslant \frac{\varepsilon\alpha^2\beta^3}{60}$$

for all $t \geqslant T(\varepsilon, n, \alpha, \beta)$.

Let $k \in \{1, 2\}$ be the value where the maximum of $l_1(T)$ and $l_2(T)$ is attained. Then we have

(1) $\varepsilon - \frac{\varepsilon\alpha^2\beta^3}{60} \leqslant l_k(T) \leqslant \varepsilon + \frac{\varepsilon\alpha^2\beta^3}{60}$,

(2) $l_{3-k}(T) \leqslant \frac{\varepsilon\alpha^2\beta^3}{60}$.

Due to Lemma 3.6 and Lemma 3.7, at time $T + 3$ we either have $l_1(T + 3) + l_2(T + 3) \leqslant \varepsilon$ and can apply Lemma 3.5, or at time T the opinions of the individuals are contained the two intervals

$$\begin{aligned} I_1 &= \left[h - \varepsilon - \frac{\varepsilon\alpha^2\beta^3}{60}, h - \varepsilon + \frac{\varepsilon\alpha^2\beta^3}{60}\right] \quad \text{and} \\ I_2(t) &= \left[h - \frac{\varepsilon\alpha^2\beta^3}{60} \cdot \left(1 - \frac{\beta}{2}\right)^{t-T}, \right. \\ &\qquad \left. h + m \cdot \frac{\varepsilon\alpha^2\beta^3}{60} \cdot \left(1 - \frac{\beta}{2}\right)^{t-T}\right], \end{aligned}$$

and $x_k(T) \in I_2(T)$ for all truth seekers $k \in K$.

Lemma 3.12 *For each* $t \geqslant T$ *we have*

$$l_1(t + 3) + l_2(t + 3) \leqslant \varepsilon$$

or the individuals of the hope interval all have opinions in $I_1 \cup I_2(t)$ *and* $x_k(t) \in I_2(t)$ *for all truth seekers* $k \in K$.

PROOF. We prove by induction on t. The induction base is given for $t = T$. If there exist individuals i and j with $x_i(t) \in I_1$, $x_j(t) \in I_2(t)$, and $d(i, j, t) \leqslant \varepsilon$ then due to Lemma 3.6 and Lemma 3.7 we would have $l_1(t + 3) + l_2(t + 3) \leqslant \varepsilon$. Otherwise the individuals with opinions in $I_2(t)$ influence each other pairwise and there is no influence from another individual with an opinion not in $I_2(t)$. With a similar calculation as in the proof of Lemma 3.5, we get $x_i(t + 1) \in I_2(t + 1)$ for all i with $x_i(t) \in I_2(t)$. □

From the previous lemmas we can conclude Theorem 1.3 and Theorem 1.6. After a finite time $T(\varepsilon, n, \alpha, \beta)$ we are in a *nice* situation as described in Lemma 3.11. If we have $l_1(T + 3) + l_2(T + 3) \leqslant \varepsilon$ then we have an ordinary convergence of the truth seekers being described in Lemma 3.5. Otherwise we have $d(k, 0, T) \leqslant \frac{\varepsilon\alpha^2\beta^3}{60}$ for all truth seekers $k \in K$. Due to Lemma 3.12 and Lemma 3.5 either we have

$$d(k, 0, t) \leqslant \frac{\varepsilon\alpha^2\beta^3}{60} \cdot \left(1 - \frac{\beta}{2}\right)^{t-T}$$

for all truth seekers $k \in K$ and all $t \geqslant T$, or there exists an $S \in \mathbb{N}_0$, such that

(1) $d(k, 0, t) \leqslant \frac{\varepsilon\alpha^2\beta^3}{60} \cdot \left(1 - \frac{\beta}{2}\right)^{t-T}$ for all $T \leqslant t \leqslant S$,

(2) $d(k, 0, t) \leqslant \varepsilon \left(1 - \frac{\beta}{2}\right)^{t-S-3}$ for all $t \geqslant S + 3$.

The latter case is 1-fold interrupted convergence. Thus the Hegselmann-Krause Conjecture is proven.

4 Remarks

In this section we would like to generalize the Hegselmann-Krause Conjecture and show up which requirements can not be weakened.

Lemma 4.1 *A finite number* n *of individuals and symmetric confidence intervals are necessary for a convergence of the truth seekers.*

PROOF. Infinitely many ignorants can clearly hinder a truth seeker in converging to the truth. If the confidence intervals are not symmetric then it is easy to design a situation where some ignorants are influencing a truth seeker which does not influence the ignorants, so that the truth seeker has no chance to converge to the truth. □

Lemma 4.2 *The condition* $\beta_{ij}(t) \geqslant \beta > 0$ *is necessary for a convergence of the truth seekers.*

PROOF. If we would only require $\beta_{ij}(t) > 0$, then we could have the following example: $n = 2$, $x_1(0) = 1 - \frac{1}{5}\varepsilon$, $x_2(0) = 1 - \varepsilon$, $\alpha_1(t) = \frac{1}{5}$, $\alpha_2(t) = 0$, $\beta_{11}(t) = \left(\frac{1}{2}\right)^{t+1}$, $\beta_{12}(t) = 1 - \left(\frac{1}{2}\right)^{t+1}$, $\beta_{21}(t) = \left(\frac{1}{2}\right)^{t+1}$, $\beta_{22}(t) = 1 - \left(\frac{1}{2}\right)^{t+1}$, and $h = 1$. By a straight forward calculation we could see that $|x_1(t) - h| \geqslant \frac{1}{2}\varepsilon$ for $t \geqslant 1$. □

We remark that conditions like $\beta_{ij}(t) + \beta_{ij}(t + 1) \geqslant 2\beta$ would also not force a convergence of the truth seekers in general. One might consider an example consisting of two ignorants with starting positions $h \pm \frac{7}{10}\varepsilon$ and a truth seeker k with starting position $h - \frac{1}{5}\varepsilon$. We may choose suitable $\beta_{ij}(t)$ and $\alpha_i(t)$ so that we have $|h - x_k(t)| \geqslant \frac{1}{5}\varepsilon$ for all t, $h - x_k(t) \geqslant \frac{1}{5}\varepsilon$ for even t and $x_k(t) - h \geqslant \frac{1}{5}\varepsilon$ for odd t.

For the next lemma we need a generalization of Definition 1.5.

Definition 4.3 *In our situation we say that the truth seekers are* r*-fold interrupted convergent, if there exists* $r + 1$ *functions* $T_i^s\left(\gamma, \varepsilon, \alpha, \beta, n, T_{i-1}^e\right)$, $i = 1, \ldots, r + 1$ *so that for each (WASBOCOD)* Ω *with structural parameters* ε, α, β *and* n *there exist* $T_i^e \in \mathbb{N}_0$, $i = 1, \ldots, r$ *fullfilling*

$$\begin{aligned} &\forall k \in K, \forall t \in \left[T_i^s\left(\gamma, \varepsilon, \alpha, \beta, n, T_{i-1}^s\right), T_i^e\right] : \\ &|x_k(t) - h| < \gamma \end{aligned}$$

for $i = 1, \ldots, r$, *where* $T_0^e = 0$, *and*

$$\begin{aligned} &\forall k \in K, \forall t \geqslant T_{r+1}^s\left(\gamma, \varepsilon, \alpha, \beta, n, T_r^e\right) : \\ &|x_k(t) - h| < \gamma. \end{aligned}$$

Lemma 4.4 *The condition* $\alpha_i(t) = 0$ *for all* $i \in \overline{K}$ *is necessary for Theorem 1.6. If it is dropped then the truth seekers are not* $(|\overline{K}| - 1)$*-fold convergent in general.*

PROOF. At first we remark that clearly it suffices to have $\alpha_i(t) = 0$ for all $i \in \overline{K}$ only for all $t \geqslant T$, where T is a fix integer. W.l.o.g. we assume $T = 0$ and consider the following example: $h = 1$, $x_i(0) = 1 - 2i\varepsilon$, $1 \in K$, $1 \neq i \in$

$\overline{K}$, $\beta_{ij}(t) = \beta$, $\alpha_i(t) = \alpha$ for the truth seekers, and $\alpha_i(t) = 0$ for the ignorants until we say otherwise. Let there be a given $\gamma > 0$ being sufficiently small. There exists a time T_1 until $x_1(T_1) < 1-\gamma$. Up to this time no other individual has changed its opinion. After time $T_1 + 1$ we suitably choose $\alpha_2(t)$ so that we have $\frac{1}{2}\varepsilon \leqslant x_1(\tilde{T}_1) - x_2(\tilde{T}_1) \leqslant \varepsilon$. So at time $\tilde{T}_1 + 1$ the convergence of truth seeker 1 is interrupted the first time. After that we may arrange it that x_1 and x_2 get an equal opinion and will never differ in there opinion in the future. Now there exists a time T_2 until $x_2(T_2) = x_1(T_2) < 1 - \gamma$ and we may apply our construction described above again. Thus every ignorant $i \in \overline{K}$ may cause an interruption of the convergence of the truth seekers. □

Conjecture 4.5 *If we drop the condition $\alpha_i(t) = 0$ for all $i \in \overline{K}$ in Theorem 1.6 then we have $(|\overline{K}|)$-fold convergence of the truth seekers.*

The Hegselmann-Krause Conjecture might be generalized to opinions in $\mathbb{R}^m$ instead of $\mathbb{R}$ when we use a norm instead of $|\cdot|$ in the definition of the update formula. Using our approach to prove this m-dimensional conjecture would become very technical, so new ideas and tools are needed. We give an even stronger conjecture:

Conjecture 4.6 *The m-dimensional generalized Hegselmann-Krause Conjecture holds and there exists a function $\phi(\Omega, \gamma)$ so that the truth seekers in an arbitrary generalized (WASBOCOD) Ω are $\phi(\Omega, \gamma)$-fold interrupted convergent in ε, α, β, and n.*

Bibliography

[1] S. Fortunato, *The Krause-Hegselmann Consensus Model with Discrete Opinions*, Int. J. Mod. Phys. C **15** (2004), 1021–1029.

[2] S. Fortunato, *Damage spreading and opinion dynamics on scale-free networks*, Phys. A **348** (2005), 683–690.

[3] S. Fortunato, *On the Consensus Threshold for the Opinion Dynamics of Krause-Hegselmann*, Int. J. Mod. Phys. C **16** (2005), 259–270.

[4] S. Fortunato, V. Latora, A. Pluchino, and A. Rapisarda, *Vector opinion dynamics in a bounded confidence consensus model*, Int. J. Mod. Phys. C **16** (2005), no. 10, 1535–1551.

[5] S. Fortunato, V. Latora, A. Pluchino, and A. Rapisarda, *Vector Opinion Dynamics in a Bounded Confidence Consensus Model*, Int. J. Mod. Phys. C **16** (2005), 1535–1551.

[6] S. Fortunato and D. Stauffer, *Computer Simulations of Opinions*, ArXiv Condensed Matter e-prints (2005).

[7] R. Hegselmann and U. Krause, *Opinion dynamics and bounded confidence: models, analysis and simulation*, Journal of Artificial Societies and Social Simulation (JASSS) **5** (2002), no. 3.

[8] R. Hegselmann and U. Krause, *Truth and cognitive division of labour: First steps towards a computer aided social epistemology*, Journal of Artificial Societies and Social Simulation (JASSS) **9** (2006), no. 3.

[9] U. Krause, *Time-variant consensus formation in higher dimensions*, Elaydi, Saber (ed.) et al., Proceedings of the 8th international conference on difference equations and applications (ICDEA 2003), Masaryk University, Brno, Czech Republic, July 28–August 1, 2003. Boca Raton, FL: Chapman & Hall/CRC. 185-191 (2005), 2005.

[10] U. Krause, *Arithmetic-geometric discrete systems*, J. Difference Equ. Appl. **12** (2006), no. 2, 229–231.

[11] W. Liebrand, A. Nowak, and R. Hegselmann (eds.), *Computer modeling of social processes*, Sage Publications, 1998.

[12] J. Lorenz, *A stabilization theorem for dynamics of continuous opinions*, Phys. A **355** (2005), 217–223.

[13] J. Lorenz, *Consensus strikes back in the Hegselmann-Krause model of continuous opinion dynamics under bounded confidence*, Journal of Artificial Societies and Social Simulation (JASSS) **9** (2006), no. 1.

[14] K. Malarz, *Truth seekers in opinion dynamics models*, ArXiv Physics e-prints (2006).

[15] D. Stauffer, *The Sznajd Model of Consensus Building with Limited Persuasion*, Int. J. Mod. Phys. C **13** (2002), 315–317.

[16] D. Stauffer, *How to Convince Others? Monte Carlo Simulations of the Sznajd Model*, AIP Conf. Proc. 690: The Monte Carlo Method in the Physical Sciences (J. E. Gubernatis, ed.), November 2003, 147–155.

[17] D. Stauffer, *Sociophysics simulations II: opinion dynamics*, AIP Conf. Proc. 779: Modeling Cooperative Behavior in the Social Sciences (P. Garrido, J. Maroo, and M. A. Muñoz, eds.), July 2005, 56–68.

[18] D. Stauffer and M. Sahimi, *Discrete simulation of the dynamics of spread of extreme opinions in a society*, Phys. A **364** (2006), 537–543.

Zusammenfassung

Diese kumulative Habilitationsschrift handelt von diskreten Strukturen, den zugehörigen Algorithmen und Anwendungsproblemen in denen diskrete Strukturen vorkommen bzw. zur Lösung nützlich sind. Als Leitfrage im Hintergrund stand:

„Wie kann man auf diskreten Strukturen optimieren?“

Da dies eine sehr umfassende Frage ist haben wir uns im Rahmen dieser Arbeit, auf einige Anwendungsbeispiele und ausgewählte diskrete Strukturen beschränkt.

1 Polyominoes

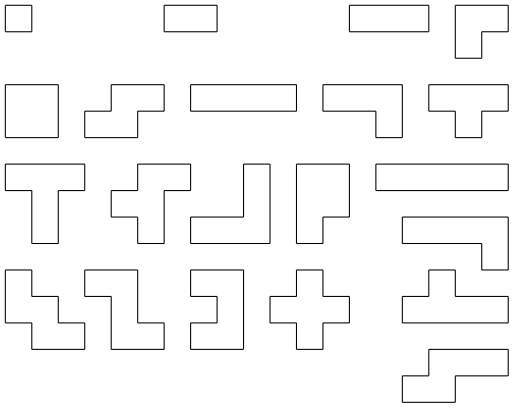


Abbildung 1: Polyominoes aus höchstens fünf Einheitsquadraten.

Ein Polyomino ist eine Seite-an-Seite-Anlagerung von Einheitsquadraten, siehe Abbildung 1. Hierbei werden Polyominoes, die sich durch Translationen, Rotationen bzw. Spiegelungen ineinander überführen lassen, als äquivalent betrachtet.

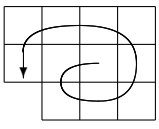


Abbildung 2: Spiralkonstruktion.

Seit längerem ist bekannt, dass der minimale Umfang $p(n)$ eines Polyominoes aus n Einheitsquadraten durch $2\left\lceil 2\sqrt{n}\right\rceil$ gegeben ist. Diese untere Schranke wird z. B. von der sogenannten Spiralkonstruktion angenommen, siehe Abbildung 2. Ausgehend von einem Einheitsquadrat werden hier spiralförmig Einheitsquadrate angebaut. Bestimmt man für kleine Anzahlen n an Einheitsquadraten alle Polyominoes mit minimalem Umfang $p(n)$, so stellt man fest, dass es einige Beispiele gibt, die nicht durch die Spiralkonstruktion erzeugt werden, siehe Abbildung 3.

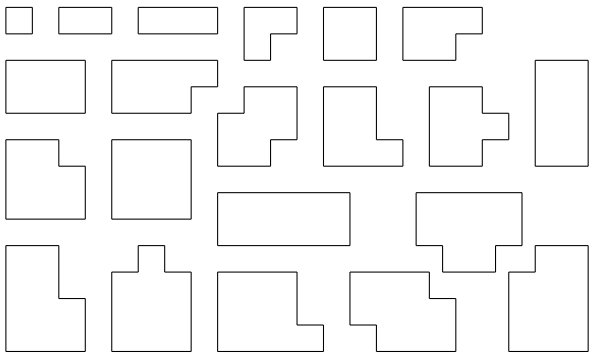


Abbildung 3: Polyominoes mit minimalem Umfang $p(n)$ für $n \leqslant 11$.

Bei dem Optimierungsproblem „minimaler Umfang eines Polyomino aus n Einheitsquadraten“ war also bisher der minimale Wert $p(n)$ bekannt. Man hatte aber keine komplette Beschreibung der Menge der minimalen Elemente.

In Kapitel 2 zeigen wir, dass jedes Polyomino mit minimalem Umfang entsteht, indem man bei einem geeignet gewählten $a \times b$-Rechteck, mit $2(a+b) = p(n)$ und $ab \geqslant n$, von den vier Ecken ausgehend $ab - n$ Einheitsquadrate entfernt. Für $n = 122$ braucht man z. B. nur die 11×12-, 10×13- und 9×14-Rechtecke zu betrachten. In Abbildung 4 haben wir die Mengen der an einer Ecke entfernten Einheitsquadrate dargestellt. Sie bilden sogenannte Ferrers-Diagramme von Partionen.

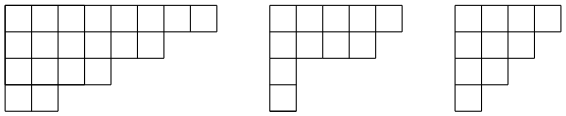


Abbildung 4: Gestalt der entfernten Einheitsquadrate an den Ecken.

Die Anzahl f_i der Ferrers-Diagramme aus i Quadraten bzw. die Anzahl der Partitionen von i lässt sich durch die

Erzeugendenfunktion

$$f(x) = \sum_{i=0}^{\infty} f_i x^i = \prod_{j=1}^{\infty} \frac{1}{1-x^j}$$

angeben.

Da wir die $ab - n$ Einheitsquadrate an allen vier Ecken des $a \times b$-Rechtecks entfernen können und symmetrische Konfigurationen identifizieren müssen, zählen wir durch $s(x) = 1 + \sum_{k=1}^{\infty} x^{k^2} \prod_{j=1}^{k} \frac{1}{1-x^{2j}}$ die symmetrischen Ferrers-Diagramme und wenden folgendes Lemma von Cauchy-Frobenius an.

Lemma 1.1 *Sei eine Gruppenaktion einer endlichen Gruppe G auf einer Menge S und eine Abbildung $w : S \longrightarrow R$ von S in einen kommutativen Ring R, der $\mathbb{Q}$ als Teilring enthält, gegeben. Falls w konstant auf den Bahnen von G auf S ist, dann gilt für jede Transversale $\mathcal{T}$ der G-Bahnen:*

$$\sum_{t \in \mathcal{T}} w(t) = \frac{1}{|G|} \sum_{g \in G} \sum_{s \in S_g} w(s),$$

wobei S_g die Elemente von S bezeichnet, welche von einem Gruppenelement g auf sich selber abgebildet werden:

$$S_g = \{s \in S | s = gs\}.$$

Fügt man alle technischen Details zusammen, so erhält man algorithmisch schnell auswertbare Formeln für die Anzahl der Polyominoes aus n Einheitsquadraten mit minimalem Umfang $p(n)$. Da dieses Abzählresultat in einem gewissen Sinne konstruktiv erzielt wurde, kann man mit diesem Vorgehen auch alle minimalen Beispiele algorithmisch effizient erzeugen. Man könnte aber noch mehr damit anstellen. Nach leichten Modifikationen kann man einen Algorithmus erhalten, der Polyominoes aus n Einheitsquadraten mit minimalem Umfang gleichverteilt zufällig erzeugt, ohne vorher alle derartigen Polyominoes erzeugen zu müssen.

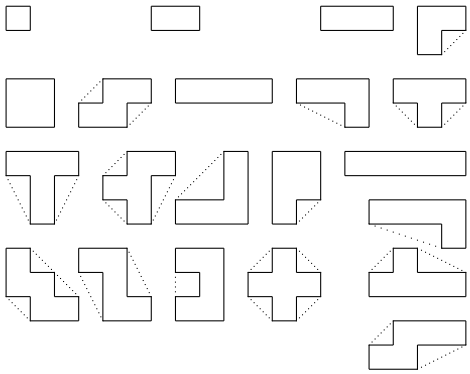


Abbildung 5: Die konvexe Hülle von Polyominoes aus bis zu 5 Einheitsquadraten.

In Kapitel 3 betrachten wir ein weiteres Optimierungsproblem auf Polyominoes. Diesmal soll der Flächeninhalt der konvexen Hülle eines Polyominoes aus n Einheitsquadraten maximiert werden, siehe Abbildung 5.

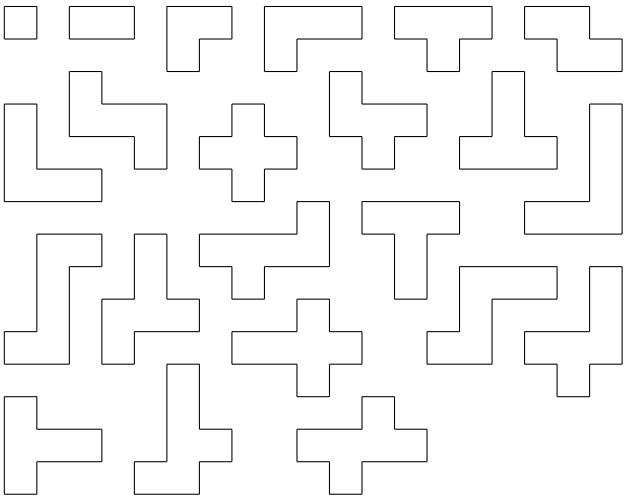


Abbildung 6: Polyominoes mit maximalem Flächeninhalt der konvexen Hülle aus bis zu 6 Einheitsquadraten.

In Abbildung 6 haben wir alle Polyominoes mit maximalem Flächeninhalt der konvexen Hülle aus bis zu 6 Einheitsquadraten dargestellt. Auch bei diesem Problem ergibt sich ein *Zoo* an extremalen Lösungen. Wiederum war der maximale Wert dieses Optimierungsproblems schon bestimmt, die Menge der extremalen Lösungen aber noch nicht charakterisiert. Der maximale Flächeninhalt von $n + \lfloor \frac{n-1}{2} \rfloor \lfloor \frac{n}{2} \rfloor$ wird z. B. von einer Winkelkonstruktion angenommen, siehe Abbildung 7.

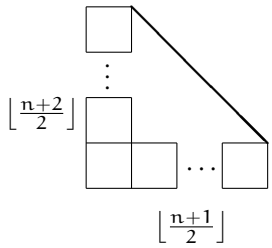


Abbildung 7: Winkelkonstruktion für ein Polyomino mit maximalem Flächeninhalt der konvexen Hülle.

In Kapitel 3 klassifizieren wir die Menge der Polyominoes aus n Einheitsquadraten mit maximalem Flächeninhalt der konvexen Hülle konstruktiv. Als Abzählresultat erhalten wir die zugehörige Erzeugendenfunktion

$$\frac{1 + x - x^2 - x^3 + 2x^5 + 8x^6 + 2x^7 + 4x^8 + 2x^9 - x^{10} + x^{12}}{(1-x^2)^2 (1-x^4)^2}.$$

Da die Eckpunkte eines Polyominoes auf einem quadratischen Gitter liegen, ist der Flächeninhalt der konvexen Hülle ein ganzzahliges Vielfaches von $\frac{1}{2}$. Der minimale Flächeninhalt der konvexen Hülle bei n Einheitsquadraten ist offensichtlich durch n gegeben und wird von

den rechteckigen Polyominoes angenommen. Es stellt sich die Frage, welche Flächeninhalte der konvexen Hülle bei Polyominoes aus n Einheitsquadraten angenommen werden können. Dies kann vollständig charakterisiert werden und wird ebenfalls in Kapitel 3 dargestellt.

Für Polyominoes gibt es viele Verallgemeinerungsmöglichkeiten. Eine sehr naheliegende betrachtet Seite-an-Seite-Anlagerungen von d-dimensionalen Einheitshyperwürfeln. Auch hier stellt sich die Frage nach dem maximalen Volumen der konvexen Hülle eines d-dimensionalen Polyominoes aus n Einheitshyperwürfeln. Verallgemeinert man die Winkelkonstruktion aus Abbildung 7 auf Dimension d, so erhält man ein Polyomino mit Volumen

$$\sum_{I\subseteq\{1,\dots,d\}} \frac{1}{|I|!} \prod_{i\in I} \left\lfloor \frac{n-2+i}{d} \right\rfloor$$

der konvexen Hülle.

Dass dieser Wert dem Maximum entspricht, war bisher eine offene Vermutung. In Kapitel 3 verwenden wir sogenannte Potentialfunktionen, eine Technik, die mehr aus der Online-Optimierung bekannt ist, um diese Vermutung zu beweisen. Hierzu führen wir Parameter $l_1, \dots, l_d, \nu_1, \dots, l_d$ ein, um d-dimensionale Polyominoes näher zu beschreiben. Dieses Vorgehen vereinfacht den bisherigen Beweis für Dimension $d = 2$ und macht es überhaupt möglich, die Vermutung für größere Dimensionen d zu beweisen.

2 Ganzzahlige Punktmengen

Seit Jahrhunderten beschäftigen sich Mathematiker mit geometrischen Objekten deren Seiten und Diagonalen ganzzahlig sind. Etwas allgemeiner versteht man unter einer ganzzahligen Punktmenge $\mathcal{P}$ eine Menge von n Punkten im m-dimensionalen Euklidischen Raum $\mathbb{E}^m$, bei der alle paarweisen Abstände ganzzahlig sind. Die Fortsetzungen der pythagoräischen Dreiecke zu Rechtecken sind Beispiele solcher ganzzahligen Punktmengen. Bezeichnet man den größten Abstand einer solchen Punktmenge als ihren Durchmesser, so stellt sich aus kombinatorischer Sicht sofort die Frage nach dem kleinstmöglichen Durchmesser einer ganzzahligen Punktmenge bei gegebener Dimension m und gegebener Anzahl an Punkten n.

In Abbildung 8 haben wir die minimalen Beispiele für Dimension $m = 2$ und $3 \leqslant n \leqslant 9$ Punkte dargestellt. Die besten bisher bekannten Schranken für den minimalen Durchmesser in Dimension $m = 2$ sind

$$c_1 n \leqslant d(2,n) \leqslant n^{c_2 \log\log n}$$

mit geeignet zu wählenden Konstanten c_1 und c_2. Durch vollständige Enumeration konnte gezeigt werden, dass die 2-dimensionalen Beispiele mit minimalem Durchmesser für $9 \leqslant n \leqslant 122$ Punkte eine sehr einfache geometrische Struktur haben: Sie bestehen aus einer Menge von $n-1$ kollinearen Punkten und einem Punkt außerhalb dieser Linie. Derartige Punktmengen entsprechen Faktorzerlegungen einer korrespondierenden Zahl, siehe z. B. Kapitel 10, so dass es möglich war, für ihren Durchmesser eine untere Schranke von $n^{c_3 \log\log n}$ herzuleiten.

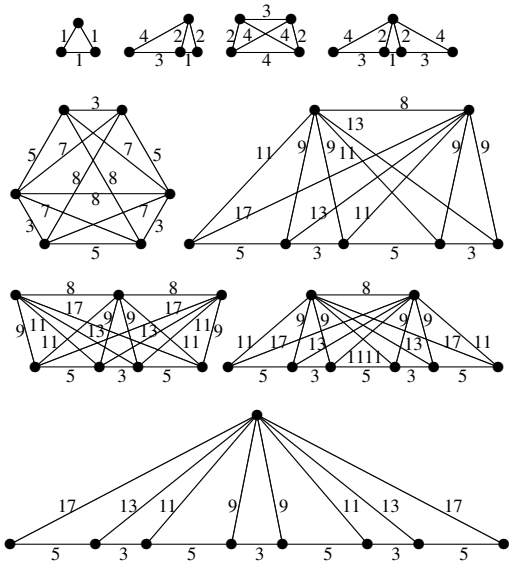


Abbildung 8: Ganzzahlige Punktmengen mit minimalem Durchmesser in der Ebene für $3 \leqslant n \leqslant 9$ Punkte.

Verbietet man, dass drei Punkte auf einer Linie liegen dürfen, so bestehen die ganzzahligen Punktmengen mit minimalem Durchmesser aus bis zu 36 Punkten aus Punktmengen auf jeweils einem Kreis. Auch hier lässt sich ein Zusammenhang zwischen Faktorzerlegungen über gewissen Ringen herstellen, siehe z. B. Kapitel 10.

Verbietet man nun zusätzlich, dass vier Punkte auf einem Kreis liegen dürfen, so waren bisher nur Beispiele aus bis zu 6 Punkten bekannt. In Kapitel 6 geben wir zwei derartige Beispiele aus sieben Punkten an und klären damit die seit längerem offene Existenzfrage konstruktiv.

Eine mögliche Anwendung von planaren ganzzahligen Punktmengen liegt in der Planung von Antennensystemen, wie dem VERY LARGE ARRAY (VLA) auf den Ebenen von San Agustin, 80 Kilometer westlich von Socorro, New Mexico, siehe Abbildung 9.

Abbildung 9: Very Large Array in New Mexico, USA (Image courtesy of NRAO/AUI).

Mit Hilfe solcher Antennensysteme werden Objekte im Weltraum näher analysiert. Ist der Abstand zwischen zwei Antennen hierbei kein ganzzahliges Vielfaches der verwendeten Wellenlänge, so tritt Interferenz auf. Da z. B. die am

VLA verwendeten Wellenlängen zwischen 7 Millimetern und 4 Metern liegen, ist es technisch machbar, auch solch riesige Antennen so zu positionieren, dass quasi keine Interferenz auftritt. Eine komplete Konfiguration von Antennen ohne Interferenzverlusten entspricht offensichtlich einer ganzzahligen Punktmenge. In der Praxis gibt es allerdings eine ganze Reihe weiterer Nebenbedingungen und anderer Ziele neben der Vermeidung von Interferenz.

Die bisher stärkste Strukturaussage über ganzzahlige Punktmengen betrifft die sogenannte Charakteristik einer ganzzahligen Punktmenge. Betrachten wir hierfür die Heronsche Formel $A_\Delta(a,b,c) =$

$$\frac{\sqrt{(a+b+c)(a+b-c)(a-b+c)(-a+b+c)}}{4}$$

für den Flächeninhalt A_Δ eines Dreiecks mit Seitenlängen a, b und c. Da diese ganzzahlig sind, existieren für $A_\Delta \neq 0$ eine eindeutig bestimmte rationale Zahl q und eine eindeutig bestimmte quadratfreie Zahl k mit $A_\Delta = q\sqrt{k}$. Dieses k heißt Charakteristik des Dreiecks Δ. Die Tatsache, dass jedes nicht-degenerierte Dreieck Δ in einer ganzzahligen Punktmenge dieselbe Charakteristik besitzt, erlaubt die effiziente Erzeugung ganzzahliger planarer Punktmengen durch Kombination von Dreiecken. Vor kurzem konnte dieser wichtige Satz auch auf den m-dimensionalen Raum übertragen werden, so dass m-dimensionale ganzzahlige Punktmengen effizient durch Kombination ganzzahliger Simplizes erzeugt werden können.

Möchten wir nun ganzzahlige Punktmengen im dreidimensionalen Euklidischen Raum konstruieren, so benötigen wir bei diesem Ansatz alle ganzzahligen Tetraeder bis zu einem gegeben Durchmesser d. In Kapitel 4 geben wir einen Algorithmus an, mit dem man eine derartige Liste effizient erstellen kann. Für alle Durchmesser $d \leqslant 1000$ bestimmen wir dort die Anzahl ganzzahliger Tetraeder mit Durchmesser d bis auf Isomorphie. Desweiteren betrachten wir an dieser Stelle eine interessante geometrische Wahrscheinlichkeit. Die Wahrscheinlichkeit $\mathcal{P}_2$, dass drei Zufallszahlen aus $[0,1]$ die Seiten eines Dreiecks sein können, also die Dreiecksungleichungen erfüllen, lässt sich leicht zu $\mathcal{P}_2 = \frac{1}{2}$ bestimmen. Als Verallgemeinerung kann man natürlich nach der Wahrscheinlichkeit $\mathcal{P}_3$, dass sechs Zufallszahlen aus $[0,1]$ die Seiten eines Tetraeders sein können, fragen. In Kapitel 4 bestimmen wir numerisch die Schranken

$$0,090 \leqslant \mathcal{P}_3 \leqslant 0,111.$$

Obwohl sich Mathematiker schon seit langer Zeit mit ganzzahligen Punktmengen beschäftigen, gibt es bisher nur sehr wenige theoretische Resultate. Die Lücke zwischen der besten bekannten unteren und der oberen Schranke für den minimalen Durchmesser $d(2,n)$ spricht Bände. In Kapitel 5 wird deswegen versucht, ein ähnliches Problem zu betrachten, in der Hoffnung die zugrunde liegende Struktur ganzzahliger Punktmengen in Euklidischen Räumen hinreichend genau zu approximieren. Anstatt von Punkten im $\mathbb{R}^m$ betrachten wir dort Punkte in Restklassenringen $\mathbb{Z}_n^m$. Analog zum Euklidischen Fall lässt sich dort ein quadrierter Euklidischer Abstand zwischen zwei Punkten $(u_1,\ldots,u_d)$ und $(v_1,\ldots,v_d)$ als

$$\sum_{i=1}^{m} (u_i - v_i)^2 = a \in \mathbb{Z}_n$$

definieren. Existiert nun ein Element $d \in \mathbb{Z}_n$ mit $d^2 = a$ so sagen wir, dass diese zwei Punkte einen ganzzahligen Abstand besitzen und können damit von ganzzahligen Punktmengen über $\mathbb{Z}_n^m$ sprechen. Noch etwas allgemeiner lässt sich dieses Konzept auch auf kommutative Ringe $\mathcal{R}$ anstatt $\mathbb{Z}_n$ übertragen.

Anstatt ganzzahligen Punktmengen mit minimalem Durchmesser über $\mathbb{E}^m$ untersuchen wir in Kapitel 5 ganzzahlige Punktmengen über $\mathbb{Z}_n^m$ mit maximaler Kardinalität $\mathcal{I}(n,m)$. Interessanterweise zeigt sich für $n = 3,4$ ein Zusammenhang zur Kodierungstheorie über Hammingabstände.

Um starke theoretische Hilfsmittel zur Hand zur haben spezialisieren wir uns in Kapitel 7 auf den Fall wo der kommutative Ring $\mathcal{R}$ ein endlicher Körper $\mathbb{F}_q$ ist. Für $2 \nmid q$ können wir für die maximale Kardinalität einer ganzzahligen Punktmenge $\mathcal{I}(q,2) = q$ zeigen und in einigen Teilfällen die geometrische Struktur der extremalen Lösungen bestimmen. Definiert man Punkte auf einer Geraden bzw. Punkte auf einem Kreis durch eine polynomiale Gleichung über $\mathbb{F}_q$, lassen sich auch Punktmengen ohne drei Punkte auf einer Geraden bzw. ohne vier Punkte auf einem Kreis betrachten. Für $q \equiv 3 \pmod 4$ konnte in Kapitel 7 beispielsweise bewiesen werden, dass ganzzahlige Punktmengen mit maximaler Kardinalität ohne drei Punkte auf einer Gerade aus Punkten auf einem Kreis bestehen.

Ein wichtiger Punkt bei der Analyse einer diskreten Struktur ist die Bestimmung ihrer Automorphismengruppe. Und so bestimmen wir in Kapitel 7 die Menge aller Automorphismen von $\mathbb{F}_q^2$, welche ganzzahlige Abstände wieder auf ganzzahlige Abstände abbilden. Dieses Resultat wird in Kapitel 8 auf den m-dimensionalen Fall verallgemeinert. Dort wird für $q \equiv 3 \pmod 4$ die geometrisch reizvolle Vermutung $\mathcal{I}(q,3) \leqslant q+1$ aufgestellt.

Bildet man ganzzahlige Abstände über $\mathbb{F}_q^m$ auf Kanten in einem Graphen ab, so erhält man interessante Graphen $\mathfrak{G}_{m,q}$. Für Dimension $m = 2$ und $q \equiv 3 \pmod 4$ ist dieser Graph beispielsweise isomorph zum sogennanten Paleygraphen quadratischer Ordnung q^2. In Kapitel 9 konnte gezeigt werden, dass auch $\mathfrak{G}_{2,q}$ für $q \equiv 1 \pmod 4$ ein stark regulärer Graph ist. Als Verallgemeinerung beweisen wir in Kapitel 8, dass die Graphen $\mathfrak{G}_{m,q}$ aus einem 3-Klassen Assoziationsschema hervorgehen. Für gerade Dimensionen m vermuten wir, dass diese Graphen sogar stark regulär sind.

Da es einige Arbeiten über die inklusionsmaximalen Cliquen von Paleygraphen gibt, haben wir uns in Kapitel 9 mit inklusionsmaximalen ganzzahligen Punktmengen über $\mathbb{F}_q^2$ beschäftigt. Es gelang eine vollständige Klassifikation der inklusionsmaximalen Punktmengen für $q \leqslant 47$ und der Beweis der Maximalität einiger Klassen von Punktmengen.

Auch im Euklidischen Raum $\mathbb{E}^2$ kann man inklusionsmaximale ganzzahlige Punktmengen betrachten. Die zugehörigen Resultate beschreiben wir in Kapitel 10.

3 Minimale Orientierungen von Graphen

In der extremalen Graphentheorie betrachtet man Graphen welche in Bezug auf eine gegebene Eigenschaft extremal in einer Klasse von Graphen sind. Ein Beispiel eines Graphen G ist in Abbildung 10 dargestellt. Die Länge eines Weges zwischen zwei Knoten in G ist durch die Anzahl der verwendeten Kanten gegeben. Als Abstand zwischen zwei Knoten bezeichnet man nun die Länge eines kürzesten Weges zwischen diesen beiden Knoten. Damit lässt sich nun der Durchmesser eines Graphen G als der maximale Abstand eines Paares von Knoten aus G definieren.

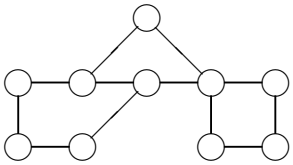


Abbildung 10: Ein Graph mit Durchmesser 5.

Unser Beispiel aus Abbildung 10 besitzt Durchmesser 5. Dies bedeutet, dass man von jeder Ecke in G zu jeder beliebigen anderen Ecke in G durch Benutzung von maximal 5 Kanten kommen kann, wenn man sich von Ecke zu Ecke fortbewegt.

Eine Verallgemeinerung von (ungerichtenen) Graphen, wie in Abbildung 10, sind gerichtete Graphen. Hier besitzt jede Kante eine Richtung, siehe Abbildung 11. Entsprechend kann man auch hier den Abstand zweier Knoten bzw. den Durchmesser eines Graphen definieren. In unserem Beispiel ergibt sich ein orientierter Durchmesser von 9.

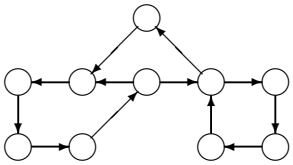


Abbildung 11: Ein gerichteter Graph mit Durchmesser 9.

Bis auf die Orientierungen der Kanten sind die Graphen aus Abbildung 10 und Abbildung 11 identisch. Wir sagen auch, dass der Graph aus Abbildung 11 aus dem Graphen aus Abbildung 10 durch Orientierung der Kanten hervorgeht. Besitzt ein Graph m Kanten so gibt es, ohne Berücksichtigung von Isomorphie, 2^m verschiedene Orientierungen dieses Graphen. In Abbildung 12 geben wir eine weitere Orientierung unseres Ausgangsgraphen an. Dieses Mal mit einem orientierten Durchmesser von 8.

Hier drängt sich natürlich ein Optimierungsproblem auf: Finde zu einem gegebenen ungerichteten Graphen G eine Orientierung mit minimalem orientierten Durchmesser $\overrightarrow{diam}_{min}(G)$.

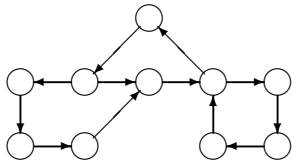


Abbildung 12: Ein gerichteter Graph mit Durchmesser 8.

Abgesehen von einer algorithmischen Lösung dieses Problems ist man auch an oberen Schranken für $\overrightarrow{diam}_{min}(G)$ in Abhängigkeit anderer Graphinvarianten interessiert.

In Bezug auf den Durchmesser $diam(G)$ des ungerichteten Graphen lässt sich eine unendliche Klasse von Beispielen konstruieren, bei denen der minimal orientierte Durchmesser $\overrightarrow{diam}_{min}(G)$ quadratisch in $diam(G)$ wächst.

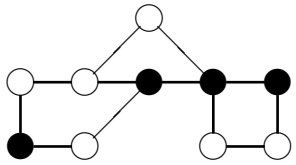


Abbildung 13: Dominierende Mengen eines Graphen.

Betrachten wir nun eine andere Grapheninvariante. Eine Teilmenge D der Eckenmenge $V(G)$ eines Graphens G heißt dominierende Menge, wenn jeder Knoten in $V(G)\backslash D$ einen Nachbarn in D besitzt. Die kleinste Kardinalität einer dominierenden Menge von G bezeichnen wir als $\gamma(G)$. In Abbildung 13 bilden sowohl die schwarzen als auch die weißen Knoten eine dominierende Menge des Graphen.

Kürzlich konnte für brückenfreie Graphen die obere Schranke

$$\overrightarrow{diam}_{min}(G) \leqslant 5\gamma(G) - 1$$

gezeigt werden.

In Kapitel 11 verbessern wir dieses Resultat zu

$$\overrightarrow{diam}_{min}(G) \leqslant 4\gamma(G)$$

und vermuten

$$\overrightarrow{diam}_{min}(G) \leqslant \left\lceil \frac{7\gamma(G)+1}{2} \right\rceil$$

als scharfe obere Schranke. Der geführte Beweis ist konstruktiv und lässt sich leicht in einen Algorithmus umwandeln, der aus einer gegebenen dominierenden Menge D in Polynomialzeit eine Orientierung von G mit Durchmesser höchstens $4|D|$ bestimmt.

4 Vektorapproximation bzw. Optimierung bei einem Textildiscounter

Die Motivation für den nächsten Block an Forschungsartikeln dieser Arbeit kommt aus einer sehr praktischen An-

wendung und wurde von einem aktuellen Kooperationspartner aus der Industrie initiiert. Betrachten wir einen Textildiscounter mit vielen Filialen, sehr geringen Verkaufszahlen pro Artikel ohne die Möglichkeit von Nachlieferungen zu einem späteren Zeitpunkt als dem Erstbelieferungszeitpunkt.

Da die Möglichkeit von Nachlieferungen fehlt, ist es besonders wichtig, den potentiellen Bedarf in den einzelnen Filialen zu schätzen, da man nur sehr teure Möglichkeiten (z. B. Preisreduzierungen) hat, Fehlschätzungen zu korrigieren.

Bedarfsschätzung von Sortimentsware ist ein sehr gut untersuchtes Gebiet, sowohl in der Theorie, als auch in der Praxis, mit einer Unmenge an Literatur. Für einmalig angebotene Werbeartikel oder sonstige Artikel mit einmaligen sehr kurzen Lebenszyklen konnten wir dagegen keine geeignete Literatur finden. Und so beschreiben wir in Kapitel 12 einen neuen Index, um die Abweichung zwischen Bedarf und Belieferung auf sehr kleinen Datenmengen stochastisch robust zu schätzen. Aufbauend auf dieser Schätzung schlagen wir eine adaptive Veränderung der Belieferung vor. In Kapitel 13 wird eine Feldstudie, welche die praktische Relevanz und Nützlichkeit dieser Methode belegt, beschrieben.

Filiale	**S**	**M**	**L**	**XL**
1	1,23	2,32	3,21	0,71
2	3,71	6,52	7,79	2,50
3	0,38	1,47	1,63	0,41
4	1,73	3,18	3,08	1,68
5	0,81	1,94	4,32	1,13
6	1,57	3,08	2,94	1,45
7	1,21	2,31	3,22	0,72
8	1,25	2,27	3,35	0,83
9	3,41	5,79	6,37	3,21

Tabelle 1: Liste von Bedarfsvektoren.

Hat man einmal gute filialgenaue Schätzungen für den zukünftigen Bedarf ermittelt, siehe Tabelle 1 für ein Beispiel, so hat man ein anderes Problem zu lösen. Aus Kostengründen ist die Belieferung unseres Partners lotbasiert. Dies bedeutet, dass es eine endliche Menge an Lottypen wie z. B.

$$\begin{pmatrix}1\\2\\2\\1\end{pmatrix}, \begin{pmatrix}1\\2\\3\\1\end{pmatrix}, \begin{pmatrix}1\\1\\2\\1\end{pmatrix}, \begin{pmatrix}1\\2\\4\\1\end{pmatrix}$$

gibt. Der erste Vektor steht hierbei für einen Lottypen bestehend aus einem Artikel in Größe S, zwei Artikeln in Größe M, zwei Artikeln in Größe L und einem Artikel in Größe XL.

Weiter wird jede Filiale mit einem ganzzahligen Vielfachen eines Lottypes beliefert. Um die Komplexität im Warenlager zu reduzieren, werden dabei nur eine kleine Anzahl κ an unterschiedlichen Lottypen verwendet. In Tabelle 2 haben wir eine Zuordnung von Filialen zu Lottypen mit zugehörigen Multiplikatoren für $\kappa = 2$ angegeben. Um den Abstand zwischen Bedarf und geplanter Belieferung zu messen, haben wir die Summe der absoluten Differenzen $\|\cdot\|_1$ verwendet.

Filiale	$\begin{pmatrix}1\\2\\2\\1\end{pmatrix}$	$\begin{pmatrix}1\\2\\3\\1\end{pmatrix}$	$\begin{pmatrix}1\\1\\2\\1\end{pmatrix}$	$\begin{pmatrix}1\\2\\4\\1\end{pmatrix}$	$\|\cdot\|_1$
1		1			1,05
2		3			2,94
3	1				2,11
4	2				2,33
5		1			1,70
6		1			2,16
7		1			1,02
8		1			1,04
9	3				1,20
$\sum$					15,55

Tabelle 2: Zuordnung und Multiplikator von Lottypen.

Wir erhalten ein diskretes Optimierungsproblem: Wähle eine kleine Anzahl κ an Lottypen aus einer gegebenen Liste aus Lottypen aus und weise jeder Filiale einen Lottyp und einen Multiplikator zu, so dass die Summe der Abweichungen zwischen Bedarf und geplanter Belieferung minimiert wird. Zusätzlich gibt es eine weitere Nebenbedingung aus der Praxis: Die Gesamtstückzahl der Belieferung muss in einem gegebenen Intervall $[\underline{I}, \overline{I}]$ liegen. Falls wir in unserem Beispiel das Intervall $[100, 120]$ wählen, so ist die Zuordnung aus Tabelle 2 nicht mehr zulässig. In Tabelle 3 geben wir eine zulässige Zuordnung an.

Filiale	$\begin{pmatrix}1\\2\\3\\1\end{pmatrix}$	$\begin{pmatrix}1\\2\\2\\1\end{pmatrix}$	$\|\cdot\|_1$
1		1	2,05
2	3		2,94
3		1	2,11
4		2	2,33
5	1		1,70
6	1		2,16
7	1		1,02
8	1		1,04
9		4	5,22
$\sum$			20,57

Tabelle 3: Zuordnung und Multiplikator von Lottypen welche eine Kardinalitätsrestriktion berücksichtigt.

In Kapitel 14 beschreiben wir eine Modellierung dieses Optimierungsproblems als ganzzahlig lineares Programm. Neben diesem exakten Algorithmus entwickeln wir eine

sehr schnelle Heuristik, welche das Problem auf realen Daten mit einer akzeptablen Optimalitätslücke löst.

5 Modellierung bzw. Optimierung von Meinungsbildungsdynamiken

Eine weitere Anwendung, die als Motivation für diese Arbeit gedient hat, kommt aus den Sozialwissenschaften. Hier ist man beispielsweise daran interessiert, wie sich die Meinung von Individuen in einer Gruppe über die Zeit entwickelt. Eine Möglichkeit so eine Meinungsbildungsdynamik zu modellieren ist das sogenannte Bounded-Confidence-Modell. Im eindimensionalen Fall beschreibt man die Meinung eines Individuums i zu einem (diskreten) Zeitpunkt t durch eine reelle Zahl $x_i^{(t)}$. Grundannahme des Modells ist, dass meine Meinung nur von Individuen beeinflusst wird, deren Meinung nahe genung an der eigenen Meinung ist. Eine Möglichkeit dies zu formalisieren ist die Einführung eines Konfidenzintervalls $I\left(x_i^{(t)}\right) = \left[x_i^{(t)} - \varepsilon, x_i^{(t)} + \varepsilon\right]$ mit einer reellen Zahl $\varepsilon > 0$. Hiermit können wir die Meinung $x_i^{(t+1)}$ von Individuum i zum Zeitpunkt $t+1$ als das arithmetische Mittel aller Meinungen $x_j^{(t)}$ zum Zeitpunkt t von Individuen j, deren Meinung im Konfidenzintervall $I\left(x_i^{(t)}\right)$ liegt, definieren.

Für dieses Modell konnte von Hegselmann und Krause festgestellt werden, dass die Meinungen der Individuen nach einer endlichen Anzahl an Zeitschritten in einen stabilen Zustand konvergiert sind.

Um die etwas philosophischere Frage nach der Suche nach der Wahrheit zu untersuchen, haben Hegselmann und Krause dieses Modell etwas erweitert. Zusätzlich gibt es nun eine *Wahrheit* $h \in \mathbb{R}$ und *Wahrheitssucher*, die in jedem Zeitschritt ihres Meinungsbildungsprozesses durch einen positiven Faktor α in der arithmetischen Mittelung von der Wahrheit angezogen werden. Die Autoren vermuteten, dass alle Wahrheitssucher mit ihrer Meinung Richtung Wahrheit konvergieren.

In Kapitel 15 analysieren wir, unter welchen Voraussetzungen welche Art von Konvergenz im Hegselmann-Krause-Modell vorliegt und beweisen ihre Vermutung. Erstaunlicherweise ist die Konvergenz im allgemeinen Fall nicht von so einfacher Gestalt, wie man zunächst vermuten könnte, so dass eine evtl. existierende Ljapunov Funktion ziemlich kompliziert sein muss. Wir beweisen die Vermutung deswegen mit elementaren Hilfsmittel, was den Beweis leider etwas technisch werden lässt.

Abgesehen vom sozialwissenschaftlichen Interesse an Meinungsbildungsdynamiken, gibt es auch Interesse aus dem Marketing. Das Hauptziel einer Marketingkampagne ist es doch, eine Menge an Käufern davon zu überzeugen, Produkte der werbenden Unternehmung zu kaufen. Die Berücksichtigung einer Meinungsbildungsdynamik in diesem Prozess ist ein sehr natürlicher Schritt. Das Bounded-Confidence-Modell lässt sich in dieser Richtung erweitern, indem man das Platzieren von Meinungen zu bestimmten Zeitpunkten erlaubt. Mathematisch ergibt sich ein diskretes Steuerungsproblem: Positioniere die Meinungen so, dass (nach einer gewissen Zeit) möglichst viele Individuen nahe einer vorgegebenen Wunschmeinung sind.