Unit Distance Colouring Problems

Michael Payne

January 30, 2007

Abstract

Various aspects of the chromatic number of the plane problem are studied with particular emphasis on some more recent work. The first main section is dedicated to surveying some lower bounds on the chromatic number under different restrictions to the allowable colour sets. The next section discusses the influence of set theory on the chromatic number of infinite graphs embedded in a Euclidean space. The first example of a unit distance graph with ambiguous chromatic number is given. Such graphs are highly relevant to the chromatic number of the plane problem. Lastly, the related problem of constructing dense distance excluding sets is studied, focusing particularly on periodic sets.

1 Introduction

The chromatic number of the plane is the original example of what we may generally refer to as unit distance colouring problems. The problem is as follows

Chromatic Number of the Plane. What is the minimum number of colours required to colour the Euclidean plane $\mathbb{R}^2$ so that no two points distance 1 apart receive the same colour?

We denote this number χ. To be precise, a colouring is a function from $\mathbb{R}^2$ to a finite set of colours (or labels) that assigns each point one colour. The most obvious way to colour the plane is to divide it into 'tiles' of some sort and give each one a colour. However, as we will see, there are other less obvious ways of constructing colourings that may give better results. The problem can also be stated in graph theoretic terms:

Graph Theoretic Version. What is the chromatic number of the graph whose vertices are all the points of $\mathbb{R}^2$ with edges between any two points that are distance 1 apart?

We call this graph the *unit distance plane* (or UDP) following Soifer [14]. This way of viewing the problem has led to many interesting insights, so it can rightly be considered as much a problem in graph theory as one in geometry.

The problem was first posed in 1950 by Edward Nelson, now a professor at Princeton, when he was an 18 year old student at the University of Chicago [15]. Various noteworthy mathematicians were involved in ensuring its ongoing popularity including Paul Erdös and Martin Gardner who included it in his column in Scientific American [9]. For a comprehensive history of the problem's creation see [15].

Despite the age of this problem, very little progress has been made since the initial bounds on χ were discovered shortly after the problem's creation. This fact is testament to the difficulty of the problem and, in the absence of progress on the main problem, a number of restricted versions and related questions have been studied. Some of these will be described in the next section. This thesis will discuss some of the latest ideas that have emerged in relation to this problem, hopefully revealing some of the causes of the great difficulty that it presents.

The best known bounds for the chromatic number of the plane are

$$4 \leq \chi \leq 7.$$

The lower bound can be proven by finding a graph with straight unit length edges that has chromatic number 4. Such a graph, when placed anywhere on a coloured plane, must have no monochromatic edges if the colouring is to be distance 1 excluding. The graph shown in Figure 1 has chromatic number 4 and is due to Leo and William Moser [15]. It is known as the Moser Spindle.

This is representative of the main way (if not the only way) in which a general lower bound on χ can be proven. Any graph that can be drawn on the plane with straight unit length edges (a *unit distance graph*) is a subgraph of the unit distance plane. Hence the chromatic number of the plane must be greater than or equal to that of such a graph.

At this point it is a good idea to explicitly introduce some terminology regarding colourings and colour sets:

Definition. We say a set *excludes* distance d if no two points in the set are distance d apart. Conversely, if there are two points in the set separated by d we say the set *realises* distance d.

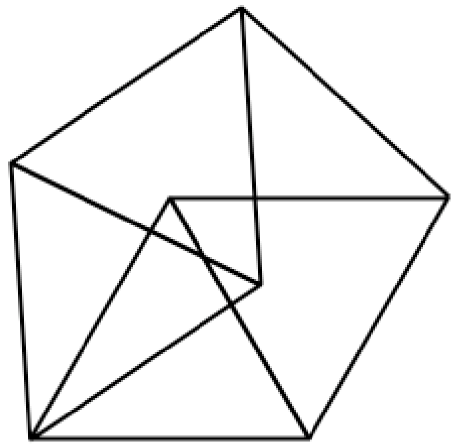

Figure 1: The Moser Spindle.

A colouring, which is a finite collection of sets that cover the plane, excludes distance d if all its colour sets exclude d.

The upper bound of 7 is proven by giving a satisfactory colouring of the plane using 7 colours. This colouring is constructed by tessellating the plane with regular hexagons. Then colour one hexagon colour 1 and the six surrounding ones six more colours (see Figure 2). This group of 7 hexagons then tessellates the plane, and it can easily be seen that distance 1 is not realised providing the hexagons are of diameter slightly less than 1. In fact this colouring has an excluded interval, i.e. there is an interval on $\mathbb{R}$ for which all those distances are excluded. This contrasts with some colourings which have exactly one excluded distance.

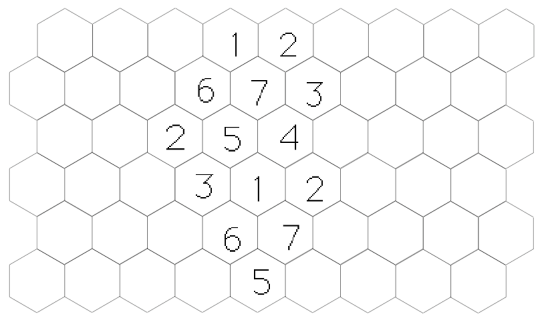

Figure 2: The Hexagon Colouring.

Of course, the chromatic number problem can be extended to other spaces. In fact, we could ask for the chromatic number of any space equipped with a notion of distance, that is any metric space. The chromatic number of the real line is known to be 2. This can be seen by colouring each half open interval $[n, n+1)$ for $n \in \mathbb{Z}$ with alternating colours. Significant work has been done on higher dimensional euclidean spaces. For instance it is known that

$$6 \leq \chi(\mathbb{R}^3) \leq 15$$

where the lower bound is due to Nechushtan [11] and the upper bound to D. Coulson [3].

It seems, however, that considering higher dimensional spaces brings even greater complexity to the problem without offering any truly new insights – the two dimensional problem is still the heart of the matter, and without solving it we are unlikely to be successful in higher dimensions.

2 Restrictions to the Problem

A number of restrictions to the type of allowable colourings of the plane have been studied, often resulting in better lower bounds on χ for those colourings.

2.1 Tiles bounded by Jordan curves

In 1973 Woodall presented a result for colourings made up of regions bounded by Jordan curves [18]. This class of colourings includes all those that we might invent if we set out to divide the plane into irregular 'tiles' and give each tile a colour. A Jordan curve is a closed curve that is simple, i.e. it has no self intersection. Woodall imagined us creating a colouring of the plane by first drawing a 'map' on it, then colouring the regions of the map. A map is a proper embedding of a graph G in the plane, or in other words, a set of points (the vertices of the graph) which are joined by simple curves (the edges) each of which starts and ends at a vertex, and does not cross any other edge or vertex. We suppose that any bounded region in $\mathbb{R}^2$ contains only finitely many vertices and parts of finitely many edges. The connected components of $\mathbb{R}^2 \backslash G$ are therefore bounded by a Jordan curve made up of a number of edges. Each such component is given a colour. The points that make up the edges and vertices of G are given the colour of some neighbouring region, so it's possible that the set of a particular colour be neither open nor closed. Woodall's proof shows that for such colourings 5 colours are insufficient, so χ in this case is either 6 or 7.

2.2 Polygonal Tiles

A similar result proven by D. Coulson considered colourings made up of arbitrary polygons (with some minimum area) and found the same bounds on χ [4]. Since such colourings could approximate a colouring of the type considered by Woodall to arbitrarily high accuracy, it is not surprising that they both found the same bounds. Coulson's proof is considerably shorter and more readable than Woodall's. As in Woodall's result, the tiles may include none, part or all of their boundaries.

Another result by Coulson shows that the number of colours required for a regular lattice based polygon tiling is 7. Lattice based polygonal tilings consist of polygons which are all congruent and are the 'nearest neighbour regions' of the lattice. A more detailed discussion of lattices can be found in Section 5. The hexagon colouring is an example of this class of colourings. In fact, Coulson showed that in general for $\mathbb{R}^n$ such colourings require at least $2^{n+1} - 1$ colours and found a 15-colouring for $\mathbb{R}^3$ [3]. In higher dimensions lattice based colourings that are optimal (in the sense of this lower bound) are yet to be found.

2.3 Thomassen's 7 Colour Result

Thomassen considered colourings using regions bounded by Jordan curves with the further restrictions that no two separate regions within distance 1 of each other received the same colour, and that the colour sets were closed [17]. With these requirements he showed that 7 colours were necessary. The proof uses mainly graph theoretic arguments about the dual graph of the map formed by the Jordan curves. This contrasts with the essentially geometric arguments used by Woodall and Coulson in their six colour results.

The first of Thomassen's restrictions, that regions within distance 1 of each other always receive different colours, seems quite natural. It is hard to imagine why placing separate tiles of the same colour close to each other would be necessary. On the other hand, the requirement that the sets be closed is perhaps more restrictive. It is equivalent to demanding that there be not just a single excluded distance, but an excluded interval. In any case, it seems that the hexagon tiling using 7 colours is probably optimal among a fairly large class of colourings. Perhaps it is the best tile based colouring of any sort, though the possibility of some irregular 6-colouring remains open.

What is clear is that to improve significantly on the hexagon colouring will require a quite complicated construction. The lower bound on χ of 4 seems a long way off. In the next subsection we will see that if we restrict ourselves to Lebesgue measurable sets then 5 colours are needed. This leaves non-measurable sets as the only possibility for a 4-colouring. The construction of non-measurable sets requires the Axiom of Choice. We will address the role of set theory in Section 3.

2.4 Measurable Sets

Another less restricted result will be of particular interest to us in the following section. In 1981 K.J. Falconer found a lower bound on the chromatic number for colourings using Lebesgue measurable sets [8]. This class of colourings is much more general than those considered above. Falconer showed that in $\mathbb{R}^n$ the number of colours needed was $n+2$, so in particular at least 5 colours are necessary in $\mathbb{R}^2$.

To explain this proof in its entirety would take considerable space, so we will begin by stating without proof a number of lemmas. The idea of the proof is to show that we can place a certain configuration of points in ways such that a 4 colouring using measurable sets can be shown to be impossible. A configuration C is a finite or countable set of points in $\mathbb{R}^n$ to which we apply rigid motions. It is somewhat similar to an embedding of a graph, and the configuration we will use in the proof is related to the Moser spindle. The configuration used for $\mathbb{R}^2$ will be a diamond shape consisting of two equilateral triangles of side 1 sharing a common edge. We will suppose there is a 4-colouring and show it must contain two points of the same colour separated by unit distance. The basic idea of the proof is to place one of the end points p of the diamond in the boundary between two of the colour sets and rotate it around that point. See Figure 3. We show that the circle traced by the opposite point of the diamond must be almost all in two colour sets. From this we show that there are two points of the same colour at unit distance.

There are a number of definitions we will need. The Lebesgue density of a set S at a point x is defined as $\lim_{\epsilon \to 0} \frac{\mu(A \cap B_\epsilon(x))}{\mu(B_\epsilon(x))}$ where $B_r(x)$ is an open ball of radius r centred at x. The metrical boundary ∂S of S is the set of points where the Lebesgue density of S either fails to exist or has a value other than 0 or 1. The essential part of S is the set of points at which the Lebesgue density is equal to 1 and is denoted by $\tilde{S}$. Note that this is not the same as the interior of S. A point may be in $\tilde{S}$ but not in S. In all cases μ denotes Lebesgue measure although in some instances we will use circular measure, which is 1-dimensional Lebesgue measure on the angle which parametrises a circle.

The first lemma simply states that a measurable set of positive measure has a boundary (which is a set of measure 0) and that its essential part is Borel.

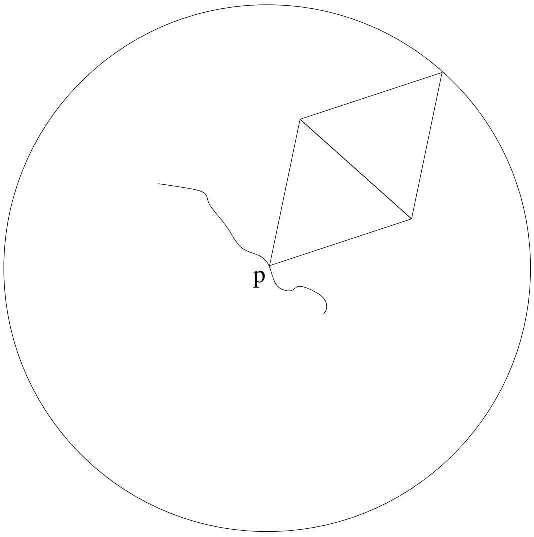

Figure 3: The diamond configuration is rotated about a point p in the boundary between two colours.

Lemma 2.4.1. *Let S be a Lebesgue measurable subset of $\mathbb{R}^n$ with $\mu(S) > 0$ and $\mu(\mathbb{R}^n \backslash S) > 0$, then ∂S is non-empty, $\mu(\partial S) = 0$ and $\tilde{S}$ is a Borel set.*

The second lemma essentially says that we can position a configuration with exactly one point in a given set of measure 0. We will use it to position a configuration with one point in the collective boundary of our colouring and the other points in the essential part of some colour.

Lemma 2.4.2. *Let B be a non-empty subset of $\mathbb{R}^n$ with $\mu(B) = 0$ and C be a configuration of points in $\mathbb{R}^n$. Then given a point $x \in C$ there exists a rigid motion m such that $m(C) \cap B = \{m(x)\}$. Furthermore, almost all rotations of $m(C)$ about $m(x)$ have this property.*

We now suppose that we have a colouring of the plane using disjoint measurable sets $S_1, ..., S_k$. The next lemma says that if one point of a configuration is in the boundary of some of the sets and the other points are in the essential part of some set, then we can slightly translate the configuration so as to put the first point in any of the adjoining regions and keep the other points in their respective sets.

Lemma 2.4.3. *Let $x_1, ..., x_m$ be a configuration of points in $\mathbb{R}^n$ and suppose that $x_1 \in \partial S_1 \cap ... \cap \partial S_r$ and $x_j \in \tilde{S}_{f(j)}$ for $j = 2, ..., m$. Then for each $i = 1, ..., r$ we can find a small translation t such that $t(x_1) \in S_i$ and $t(x_j) \in S_{f(j)}$.*

The next lemma will be used in the final step to give the existence of two points of the same colour at unit distance from each other. We will give a proof of this lemma since it forms a large part of the proof of the theorem.

Lemma 2.4.4. *Let T be a circle in $\mathbb{R}^2$ of radius $r > 1/2$ such that $\theta = 2\arcsin\left(\frac{1}{2r}\right)$ is an irrational multiple of π. Suppose almost all the points on T (in the sense of circular measure) lie in $\tilde{S}_1$ or $\tilde{S}_2$. Then at least one of S_1 or S_2 realises distance 1.*

Proof. We may take T as being centred at the origin. We let

$$t(\phi) = (r\cos\phi, r\sin\phi)$$

be points on T and suppose that the sets S_1 and S_2 don't realise distance 1. Note that θ is the angle subtended by two points on the circle T at unit distance from each other, so points of the form $t(\phi)$ and $t(\phi+\theta)$ for any ϕ must not both lie in the same colour set. Nor can they lie in the essential part of the same set, otherwise we could use Lemma 2.4.3 to get two points at unit distance in that set.

Now consider subsets of T of the form $P_\phi \equiv \{t(\phi + n\theta) : n \in \mathbb{Z}\}$. Since the set of points on T that aren't in $\tilde{S}_1 \cup \tilde{S}_2$ has circular measure zero, the union of the P_ϕ corresponding to these points must also have measure zero, since each P_ϕ has measure zero. For points $t(\phi)$ not in this union (almost all of T) we have

$$t(\phi) \in \tilde{S}_j \iff t(\phi + 2n\theta) \in \tilde{S}_j$$

for $n \in \mathbb{Z}$ and $j = 1, 2$. Since by assumption θ is an irrational multiple of π the sets P_ϕ are dense in T (by Kronecker's theorem). Hence we can find n such that $2n\theta$ is arbitrarily close to a multiple of 2π. So the sets $\tilde{S}_1 \cap T$ and $\tilde{S}_2 \cap T$ are periodic with arbitrarily small period. They are also Borel since $\tilde{S}_1$ and $\tilde{S}_2$ are Borel by Lemma 2.4.1, and are therefore measurable subsets of T.

Now a measurable set of arbitrarily small period has either zero or full measure (this is a well known result, see for example [1]) so almost all points of T belong to $\tilde{S}_1$ or to $\tilde{S}_2$, suppose without loss of generality $\tilde{S}_1$. Then we may select two points of $\tilde{S}_1 \cap T$ at unit distance and so by Lemma 2.4.3 S_1 realises distance 1. □

We are now ready to prove the main theorem. We will only give the proof for $\mathbb{R}^2$ since it simplifies the exposition and the generalisation to $\mathbb{R}^n$ is reasonably straightforward.

Theorem 2.4.5. *Suppose $\mathbb{R}^2$ is covered by 4 disjoint measurable sets $S_1, ..., S_4$. Then one of the sets contains a pair of points distance 1 apart.*

Proof. We begin by supposing that the sets $S_1, ..., S_4$ exclude distance 1 and derive a contradiction. Let C be a configuration of 4 distinct points $x_1, ..., x_4 \in \mathbb{R}^2$ that form two equilateral triangles of side 1 that share a common edge (x_2, x_3). Also let $B = \bigcup_{i=1}^{4} \partial S_i$. By Lemma 2.4.2 we can position C so that $C \cap B = \{x_1\}$ and for almost all rotations ρ of C around x_1 we have $\rho(C) \cap B = \{x_1\}$. Now x_1 must lie in the boundaries of at least two sets, say S_1 and S_2. For any rotation ρ such that $\rho(C) \cap B = \{x_1\}$ we must have that $\rho(x_2)$ and $\rho(x_3)$ lie in one each of $\tilde{S}_3$ and $\tilde{S}_4$. This is because if for instance $\rho(x_2) \in \tilde{S}_1$ then Lemma 2.4.3 would allow us to slightly shift $\rho(C)$ so as to put both x_1 and x_2 in S_1, contradicting the assumption. Furthermore, $\rho(x_4)$ must lie in $\tilde{S}_1$ or $\tilde{S}_2$ by a similar argument.

The distance from x_1 to x_4 is $\sqrt{3}$ so we consider the circle of this radius centred at x_1, i.e. the circle traced by $\rho(x_4)$ as we rotate C. Almost all points on this circle must lie in either $\tilde{S}_1$ or $\tilde{S}_2$. Chords of length 1 on this circle subtend an angle of $2 \arcsin\left(\frac{1}{2\sqrt{3}}\right)$ at the centre. Hence to complete the proof we just need to show that this is an irrational multiple of π and apply Lemma 2.4.4. To do this first note that $\arcsin\left(\frac{1}{2\sqrt{3}}\right) = 2\arccos\left(\frac{5}{6}\right)$ so it is enough to show that $\theta' \equiv \arccos\left(\frac{5}{6}\right) = \arcsin\left(\frac{\sqrt{11}}{6}\right)$ is an irrational multiple of π. If this were not true then there would be an integer m such that $m\theta'$ was a multiple of 2π and

$$\begin{aligned} \left(\frac{5}{6} + i\frac{\sqrt{11}}{6}\right)^m &= 1 \\ \Rightarrow \left(5 + i\sqrt{11}\right)^m &= 6^m \end{aligned}$$

But $\mathbb{Z}[\sqrt{-11}]$ is a unique prime factorisation domain with ± 1 as the only units so this cannot be true for any m. □

To generalise this proof to $\mathbb{R}^n$ we would use a unit simplex (e.g. tetrahedron in $\mathbb{R}^3$) in place of triangles in constructing the configuration C. We also use a simple corollary of Lemma 2.4.4 that replaces the circle T with an $(n-1)$-sphere almost all in two sets. Then we can take a plane section through this $(n-1)$-sphere and proceed as for the 2-dimensional case.

3 Set theory and chromatic number

3.1 The Soifer-Shelah Graph

A recent series of papers by Alexander Soifer and Saharon Shelah have demonstrated the interesting way in which the axioms chosen for set theory may affect the chromatic number of infinite graphs embedded in space. This led the authors to the somewhat surprising conclusion that the chromatic number of the plane may depend on the version of set theory we adopt. To understand this properly let us first review their example of a graph that has what we might call 'ambiguous' chromatic number. Later we will consider the implications of this idea for the chromatic number of the plane.

The first graph presented by Soifer and Shelah was a graph on the real line rather than the plane. As its vertices it has all the points of the line, and points are connected by an edge if the distance between them is $q + \sqrt{2}$ for some rational number q. We will henceforth denote this graph S. Formally we say that S has $\mathbb{R}$ as its vertices and the edges are the pairs $\{(s,t) : s - t - \sqrt{2} \in \mathbb{Q}\}$. Note that the order of each vertex is $\aleph_0$. In subsequent papers Soifer and Shelah extended their original idea to make similar graphs on the plane and in $\mathbb{R}^n$. We will consider only the 1-dimensional case as it has all the essential properties of the more general graphs.

Soifer and Shelah found a two colouring for their graph S on the real line that relied on the Axiom of Choice. This axiom states that there exists a choice function on any collection of non-empty sets. A choice function is a function that assigns to each set in the collection an element of that set. In other words, it chooses one element from each set. In the colouring of the graph S this axiom will be used to choose representatives of the components of the graph. The important point is, as we will see below, that there are uncountably many components in S, so we are making use of the Axiom of Choice in a strong way. It has long been known that it is not necessary to accept the full Axiom of Choice to make a consistent set of axioms for set theory [10]. The Axiom of Choice is independent of the other axioms of set theory, much like the parallels postulate in geometry. The 'other' axioms are the standard Zermelo-Frankel axioms which are universally accepted. The system of axioms containing the Zermelo-Frankel axioms along with the Axiom of Choice is denoted ZFC.

Let us now describe how to construct a 2-colouring of S using the Axiom of Choice.

Theorem 3.1.1. *Under the system of axioms ZFC the chromatic number of S is 2.*

Proof. We first divide the graph into connected components by defining an equivalence relation $\sim$ on $\mathbb{R}$ by

$$a \sim b \Longleftrightarrow a - b = q + n\sqrt{2}$$

for some $q \in \mathbb{Q}, n \in \mathbb{Z}$. It is simple to see that $\sim$ does define an equivalence relation. If two points are connected by a path in S they will be in the same equivalence class since the length of each edge in the path is of the form $q + \sqrt{2}$ for some $q \in \mathbb{Q}$. The converse is also clearly true, that if $a \sim b$ then there is a path from a to b in S. Hence each equivalence class contains exactly one connected component of S.

We now use the Axiom of Choice by claiming that we can choose a representative of each of the connected components. We colour each of the representatives colour 0 say. Now suppose the representative of a particular component is x. Then each point in that component can be uniquely expressed as $x+q+n\sqrt{2}$ for some $q \in \mathbb{Q}, n \in \mathbb{Z}$. To see that the representation is unique suppose we have

$$\begin{aligned} x + q_1 + n_1\sqrt{2} &= x + q_2 + n_2\sqrt{2} \\ \Rightarrow q_1 - q_2 &= (n_2 - n_1)\sqrt{2} \end{aligned}$$

We must conclude that both sides are equal to zero since otherwise we would have an integer multiple of $\sqrt{2}$ equal to a rational. Using this representation we can colour each point in the component according to the parity of n. Since the distance between adjacent vertices is of the form $q + \sqrt{2}$ they will always receive different colours. We colour each connected component in this way. □

In this proof we used the Axiom of Choice to choose a representative of the connected components of the graph S. The number of components in S determines the 'strength' of choice axiom that we require in this step. The connected components of S are translations of $\mathbb{Q} + \mathbb{N}\sqrt{2}$, so each has countably many vertices. Since the reals are uncountable there must be uncountably many components. This is because if the number of components were countable then their union would also be countable, but their union is $\mathbb{R}$. Therefore to choose the representatives of the components we need the full uncountable version of the Axiom of Choice.

As already mentioned, it is not necessary to accept the 'full power' Axiom of Choice. For example we can restrict it to countable collections of sets, and only claim that there exists a choice function on these collections. We denote this restricted Axiom of Choice by $AC_{\aleph_0}$ where $\aleph_0$ is the cardinality of the natural numbers, countable infinity.

An axiom that we can add to our system if we reject the full Axiom of Choice is the axiom of Lebesgue measurability (LM), that every set of real numbers is Lebesgue measurable. This is because the full AC allows us to construct sets that are not Lebesgue measurable. Note that this means that AC and LM are incompatible, but not that the rejection of AC implies LM. Solovay proved that $ZF + AC_{\aleph_0} + LM$ was a consistent set of axioms [16].

Under this alternative axiomatisation of set theory it can be shown that the graph S can have no countable chromatic number. Before proving this we will first prove the following lemma:

Lemma 3.1.2. *For a Lebesgue measurable set $A \subset \mathbb{R}$ if $\mu(A) > 0$ then A contains a pair of adjacent vertices of S.*

Proof. Suppose that A is Lebesgue measurable with $\mu(A) > 0$. Here μ represents 1-dimensional Lebesgue measure. Then we can find an interval I such that

$$\frac{\mu(A \cap I)}{\mu(I)} > \frac{9}{10}.$$

This follows from the Lebesgue Density Theorem which states that almost every point in a Lebesgue measurable set has Lebesgue density equal to 1. Since A has positive measure it must contain points with density equal to 1. Around any such point we can find an interval with the desired property.

We now choose a rational q such that $\sqrt{2} < q < \sqrt{2} + \frac{\mu(I)}{10}$ and translate the set A by $q - \sqrt{2}$, that is we form the set $A' = \{x - (q - \sqrt{2}) : x \in A\}$. Then we have

$$\frac{\mu(A' \cap I)}{\mu(I)} > \frac{8}{10}$$

since the part of A translated out of I has measure at most $\mu(I)/10$.

These two lower bounds on the density of A and A' on I imply that there must be some $x \in A \cap A' \cap I$. To show this we assume that the intersection $A \cap A' \cap I$ is empty. Therefore $A \cap I$ and $A' \cap I$ are disjoint. So by additivity of μ for measurable sets we have

$$\mu(I) = \mu(A \cap I) + \mu(A' \cap I) + \mu(I \backslash (A \cup A'))$$

so using the density bounds we have

$$1 > \frac{9}{10} + \frac{8}{10} + \frac{\mu(I \backslash (A \cup A'))}{\mu(I)}$$

which is a contradiction. Hence there is an $x \in A \cap A' \cap I$.

Now since $x \in A'$ it has a preimage (under the translation) $y = x + (q - \sqrt{2})$ which is in A. So $x, y \in A$ and x and y are adjacent as required. $\square$

We can now prove the main claim.

Theorem 3.1.3. *Under the system of axioms $ZF+AC_{\aleph_0}+LM$ the chromatic number of S is neither finite, nor equal to $\aleph_0$.*

Proof. Suppose we have a countable colouring $A_1, A_2, A_3, \ldots$ of $\mathbb{R}$ where each A_n is Lebesgue measurable (all sets are measurable by assumption). Since $\bigcup_n A_n = \mathbb{R}$ there must be some A_i with positive measure. Hence by Lemma 3.1.2 A_i contains two adjacent vertices. But A_i was supposed to be a colour set, so cannot contain two adjacent vertices, a contradiction. □

3.2 A new example

Using as a starting point a colouring of the 'rational unit distance plane' described by Woodall in 1973 [18] we may construct another example of a graph with ambiguous chromatic number. An important feature of this new graph will be that it is a subgraph of the unit distance plane. Before describing it let us first describe Woodall's colouring. The rational unit distance plane is the unit distance graph induced by rational 2-space $\mathbb{Q}^2$. It has all the pairs of rationals as its vertices and two pairs joined if the Euclidean distance between them is 1.

Theorem 3.2.1. *The rational plane $\mathbb{Q}^2$ can be coloured with 2 colours so that no two points distance 1 apart receive the same colour.*

Proof. We assume that all rationals are expressed in their lowest terms and begin by noting that if a rational vector $(p/q, r/s)$ is unit length, that is

$$\begin{aligned} (p/q)^2 + (r/s)^2 &= 1 \\ \Rightarrow (ps)^2 + (rq)^2 &= (qs)^2 \end{aligned}$$

then q and s are odd and one of p and r is odd and the other even. This follows from the fact that in any Pythagorean triple a, b, c such that $a^2 + b^2 = c^2$, c is always odd (assuming a, b and c don't share a common factor).

We define an equivalence relation $\sim$ on $\mathbb{Q}^2$ by saying that for two points (a, b) and (c, d), $(a, b) \sim (c, d)$ if and only if $a - c$ and $b - d$ both have odd denominators when expressed in their lowest terms. Therefore if two points are distance 1 apart then they must be in the same equivalence class. To see that $\sim$ really is an equivalence note that it is clearly reflexive (we define 0 as having odd denominator) and symmetric. Transitivity comes from the fact that if $a - c$ and $c - e$ have odd denominators then so does their sum $a - e$.

Clearly the equivalence classes of $\sim$ are disconnected. They are also translations of each other because $\mathbb{Q}^2$ is homogeneous. Therefore it suffices to give a colouring for just one of the equivalence classes. We do this for the class containing the origin. The points in this class all have odd denominators in both their coordinates, so we can divide them into two categories. Using o to represent odd and e for even the two categories are: (1) points of the form $(e/o, e/o)$ or $(o/o, o/o)$ and (2) points of the form $(e/o, o/o)$ or $(o/o, e/o)$. We give one colour to each of these categories. It is straightforward to check that two points from the same category can not be distance 1 apart since their difference won't be of the required form, i.e. one odd and one even numerator. □

Before proceeding we may wish to ask how many equivalence classes there are. Also, it is not clear whether each equivalence class is a single connected component of the graph. It may be the case that each one consists of a number of disconnected components. I am unable to answer either of these questions at this time, but we can at least say that the number of classes and components must be countable since $\mathbb{Q}^2$ is itself countable. Therefore we only need the weaker version of the Axiom of Choice when choosing representatives of the equivalence classes.

Now what would happen if we translated the rational unit distance plane to all points in the real plane? We construct a new graph G that has all the points of $\mathbb{R}^2$ as it vertices, but this time two vertices are joined only if they are unit distance apart *and* their coordinates differ by rational amounts. This graph is a subgraph of the unit distance plane. Note that as in the case of the graph S, the order of each vertex is $\aleph_0$. This contrasts with the full unit distance plane which has uncountable vertex order. We might call G the 'rationally connected unit distance plane'. We will also see that G has many disconnected components while UDP has only one component.

Theorem 3.2.2. *In the system of axioms ZFC the chromatic number of the rationally connected unit distance plane G is 2.*

Proof. We define an equivalence relation $\sim$ on $\mathbb{R}^2$ as follows. For two points $p_1, p_2 \in \mathbb{R}^2$

$$p_1 \sim p_2 \iff p_1 - p_2 \in \mathbb{Q}^2.$$

The equivalence classes of $\sim$ are copies of $\mathbb{Q}^2$ translated in the plane. Each class is clearly disconnected from others in the graph G because if there is a path from p_1 to p_2 in G then $p_1 \sim p_2$. We can therefore colour each class independently of the others. We use the Axiom of Choice to select representatives of each equivalence class and mark them as the 'origin' in their class. We then translate a 2-colouring of $\mathbb{Q}^2$ (as described in the proof of Theorem 3.2.1) to each representative, thus colouring the whole of G. □

We could denote the space of translations of $\mathbb{Q}^2$ as $\mathbb{R}^2/\mathbb{Q}^2$, a quotient group under vector addition. Since each element of $\mathbb{R}^2/\mathbb{Q}^2$ has measure 0 there must be uncountably many elements because their union is all of $\mathbb{R}^2$. So as was the case for the Soifer-Shelah graph S, this colouring requires the use of the full power Axiom of Choice to pick representatives of the copies of $\mathbb{Q}^2$.

It is interesting to note that the quotient group $\mathbb{R}/\mathbb{Q}$ is used in the construction of the Vitali set, a prototypical example of a non-measurable set. The Vitali set is a set of representatives of the elements of $\mathbb{R}/\mathbb{Q}$ taken from the interval $[0, 1]$. In fact a full set of representatives can be found on any interval of positive length. So in the construction of our colouring we have initiated the colouring on a non-measurable set. It is not surprising then that the colouring is itself non-measurable as we will see below.

Now we consider the graph G under the set of axioms $ZF + AC_{\aleph_0} + LM$. We wish to show that no 2-colouring can exist. Clearly the chromatic number of this graph under these axioms is at most 7 since the hexagon tiling is a measurable colouring. Our approach is to show that any measurable set with positive measure contains the endpoints of a path of length 3 in G. In the special case of 2-colourings such a pair of points must have different colours. We can then proceed in a similar fashion to Soifer and Shelah's proof of Theorem 3.1.3 given above. The first lemma we need says that there is a path of length 3 from any point to another point arbitrarily close to the first point.

Lemma 3.2.3. *For any point $p \in \mathbb{R}^2$ and any $\epsilon > 0$ we can find $q \in \mathbb{Q}$ with $|q| < \epsilon$ such that there is a path of length 3 in G from p to $p + (q, 0)$.*

Proof. We use the fact that the rational points are dense on the unit circle to choose an angle α such that $(\cos\alpha, \sin\alpha) \in \mathbb{Q}^2$ and

$$\left|\cos\alpha - \frac{1}{2}\right| < \frac{\epsilon}{3}$$

The path starting at p and passing the following 3 points has the desired property.

$$\begin{aligned} p_1 &= p + (\cos\alpha, \sin\alpha) \\ p_2 &= p + (\cos\alpha - 1, \sin\alpha) \\ p_3 &= p + (2\cos\alpha - 1, 0) \end{aligned}$$

From the previous inequality

$$|2\cos\alpha - 1| < \frac{2\epsilon}{3} < \epsilon$$

so $q = 2\cos\alpha - 1$. □

The path given in this proof is a 'triangle' that is not quite joined up at p. See Figure 4.

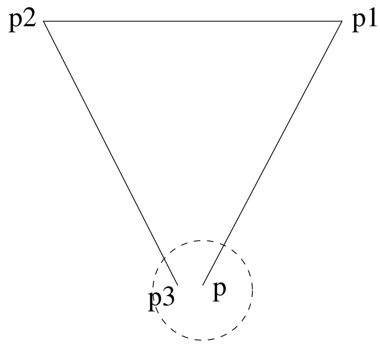

Figure 4: A path of length 3 in G.

The following lemma serves the same purpose as Lemma 3.1.2 in the proof of Theorem 3.1.3.

Lemma 3.2.4. *For a Lebesgue measurable set $A \subset \mathbb{R}^2$ if $\mu(A) > 0$ then A contains a pair of vertices of G that are joined by a path of length 3.*

Proof. Suppose that A is Lebesgue measurable with $\mu(A) > 0$ (μ is now 2-dimensional Lebesgue measure). Then there is some rectangle R in $\mathbb{R}^2$ (with sides parallel to the axes) such that

$$\frac{\mu(A \cap R)}{\mu(R)} > \frac{9}{10}.$$

As in the proof of Lemma 3.1.2, this follows from the Lebesgue Density Theorem.

Let l be the length of R in the x direction. We now choose a rational q such that $|q| < \frac{l}{10}$ and two points of the form (x, y) and $(x + q, y)$ are joined by a path of length 3. Lemma 3.2.3 allows us to do this. We now form the translated set $A' = \{(x + q, y) : (x, y) \in A\}$. We have

$$\frac{\mu(A' \cap R)}{\mu(R)} > \frac{8}{10}$$

since the part of A translated out of R has measure at most $\mu(R)/10$.

Using the same argument as used in the proof of Lemma 3.1.2 we know that $A \cap A' \cap R$ is non-empty (recall that this relies on the measurability of

A). We take a point u in this intersection. Since $u \in A'$ it has a preimage $v \in A$ such that $u = v + (q, 0)$. So $u, v \in A$ and u and v are connected by a path of length 3 in G. □

We can now prove the main theorem.

Theorem 3.2.5. *Under the set of axioms $ZF + AC_{\aleph_0} + LM$ the chromatic number of G is at least 3 and not more than 7.*

Proof. Suppose we have a two colouring of G, that is two sets A_1 and A_2 that cover the plane each containing the points of one colour. Clearly at least one, say A_1, has positive measure. A_1 must also be measurable since all sets are measurable by assumption. So by Lemma 3.2.4 A_1 contains a pair of points connected by a path of length 3 in G. But in a 2-colouring of any graph points connected by paths of length 3 must have the opposite colour, so A_1 cannot be a colour set.

The upper bound of 7 is given by the hexagon colouring which can clearly be used to colour G since G is a subgraph of the unit distance plane. □

Hence we have shown that, like the Soifer-Shelah graph S, the rationally connected unit distance plane G has a different chromatic number depending on the axioms chosen for set theory. There are at least two interesting properties of G that set it apart from S. Firstly, its chromatic number takes a finite value under both sets of axioms, unlike the graph S. Secondly, it is an infinite subgraph of the unit distance plane. There are however several important differences between G and the unit distance plane. For a start, G has infinitely many disconnected components while the UDP has only one. Also, G does not contain the Moser spindle, nor even any triangles. In what follows we will consider some more examples that address some of these differences, but first let us discuss the deeper implications of these cases of 'ambiguous' chromatic number.

3.3 Implications for χ

After demonstrating the existence of graphs whose chromatic number depends on the axiomatisation of set theory, Soifer and Shelah went on to make an interesting observation regarding the chromatic number of the plane. We have already considered Falconer's proof that if we demand that our colouring consist of Lebesgue measurable sets then χ is at least 5. Therefore in the system of axioms $ZF + AC_{\aleph_0} + LM$ we must have $\chi \geq 5$. However if we allow arbitrary colourings, adopting the usual axioms ZFC, it is possible that $\chi = 4$. There is an important theorem due to Erdös and de Bruijn [7] which is highly relevant here:

Theorem 3.3.1. *Let k be an integer and G a graph. If every finite subgraph of G is k-colourable then G is also k-colourable.*

This means that if there is no finite unit distance graph with chromatic number greater than 4, then $\chi = 4$. The proof of this theorem relies heavily on the uncountable Axiom of Choice. Considerable efforts have been made to find such finite unit distance graphs without success. Therefore there seems to be a real possibility that χ depends on whether we accept the uncountable Axiom of Choice. It could even be the case that $\chi = 4$ if we do and $\chi = 7$ if we don't!

This statement requires some qualification. Rejecting AC does not require us to accept LM, it merely allows us to adopt it. Let us illustrate this point with respect to the graph G. We have seen that if we have AC then the chromatic number of G is 2. However, if we merely reject AC we can only say that the 2-colouring of G that we gave is no longer valid. This does not exclude the possibility of another 2-colouring that didn't rely on AC. The proof that the chromatic number of G was greater than 2 relied on us first accepting LM. So we cannot say that the chromatic number of G depends only on whether we accept AC.

How surprising then is the possible ambiguity of the chromatic number of the plane? It may not be as troubling as it first appears. When we choose the system of axioms with LM we are really placing a restriction on the allowed colourings. This restriction is somewhat similar to the restrictions we considered in Section 2, it is just imposed in a different way. By taking LM as an axiom we are really assuming that all colourings must be Lebesgue measurable. However, we know that if we make a different assumption, namely AC, we can prove the existence of sets that are not Lebesgue measurable.

The unit distance plane graph UDP is induced by the geometry of the plane, but the geometry of the plane is not fully encoded in the graph UDP. In the graph UDP the only basic relationship between vertices is that of being connected by an edge, which is only the case when the points corresponding to those vertices are distance 1 apart in the plane. On the other hand, each point in the plane has a simple relationship with every other point, namely a vector difference between points. The axiom LM adds constraints to the colouring by linking (to some degree) the colour of points that are 'close' to each other.

This local interdependence can be made explicit. Recall that the Lebesgue density theorem states that almost all points in a measurable set have Lebesgue density 1. In a measurable colouring almost all points of the space are in the essential part of some colour (i.e. have density 1 in some colour). This follows easily from Lemma 2.4.1. Thus almost all points must be surrounded by a small neighbourhood which is 'mostly' one colour. By 'mostly' we mean that the colour covers more than $\frac{9}{10}$ (say) of the neighbourhood.

The other restrictions we considered in Section 2 added similar local interdependence conditions. The restriction to any sort of tiled colouring has the effect that almost all points must be interior to the tiles. This means that around almost all points there must be a small neighbourhood that is entirely (not just mostly) the colour of that point. This is exactly the type of dependence that is not encoded in the UDP graph. It only specifies pairs of points that can't have the same colour.

It seems that we must conclude that there really is a difference between the geometric and graph theoretic formulations of the chromatic number of the plane problem. If we just use the geometry of the plane to induce the UDP graph and then ask only 'what is the chromatic number of this graph?' then we would probably choose the ZFC system as it allows truly arbitrary graph colouring (i.e. *any* function from the vertices to the set of colours). With the UDP graph extracted and the geometry of the plane forgotten the LM axiom loses its relevance since it refers to sets in the plane. On the other hand if we are interested in the geometric version of the problem and allow the plane to maintain the full richness of its geometry, it would be reasonable (although not necessary) to accept the LM axiom and conclude that $\chi \geq 5$. We may do this because we feel that measurable colourings were more what we had in mind when we posed the problem in the first place.

4 Some More Subgraphs of UDP

In this section we will describe some more interesting infinite subgraphs of the unit distance plane. The ultimate reason for doing this is to see if we can get any closer to a colouring of UDP. Unfortunately due to time constraints the analysis and discussion of these graphs must be left largely incomplete. As we go through each graph there are a number of questions that we should keep in mind.

- Can we give a colouring with less than 7 colours?
- Does it have ambiguous chromatic number?
- How many connected components?
- Does it contain 3-cycles, the triangle lattice or the Moser spindle?

A note on some terminology that will be used: We will often call a point in $\mathbb{Q}^2$ simply a 'rational', and points in $\mathbb{R}^2 \backslash \mathbb{Q}^2$ 'irrationals'. We will say a point is 'rationally connected' if it is connected to points that differ from it by a rational unit vector and 'fully connected' if it is connected to all points on the unit circle around it.

4.1 Rationals Fully Connected

The rationally connected unit distance plane G can be thought of as containing arbitrary translations and 'rational rotations' of the unit edge (joining the origin to $(1,0)$ say). By rational rotations we mean rotations so that the differences in coordinates are rationals, not the angle of the rotation. We can call the edges 'rational unit vectors' since they are unit vectors in $\mathbb{Q}^2$. Note that viewed in this way the full UDP consists of arbitrary rotations and arbitrary translations of the unit edge.

We could ask what happens if we switch 'arbitrary' and 'rational' in the above description of G? That is, we form the graph that contains rational translations and arbitrary rotations of the unit edge. This means every point of $\mathbb{Q}^2$ is fully connected to the unit circle around it, but other points are only connected to the points of $\mathbb{Q}^2$ that lie at distance 1 from them. Call this new graph 'rationals fully connected' or RFC. To be explicit, the vertex set V of RFC is $\mathbb{Q}^2$ union with points that are distance 1 from a rational. We will discuss below whether this is all of $\mathbb{R}^2$. RFC has edges between two points if they are distance 1 apart and at least one of them is in $\mathbb{Q}^2$.

We can give a 3-colouring of RFC as follows. We use colours 1 and 2 to colour the points of $\mathbb{Q}^2$ as per Woodall's colouring. We then give colour 3 to all the irrational points in V. These points are only connected to points of $\mathbb{Q}^2$ which are coloured either 1 or 2. Note we don't need the uncountable Axiom of Choice for this colouring. RFC does contain 3-cycles, so cannot be 2-coloured and therefore its chromatic number is 3. Hence RFC does not have ambiguous chromatic number.

The triangles in RFC all contain two rational points and one irrational point. This lets us see that RFC does not contain the infinite triangle lattice. Nor does it contain the Moser spindle, as we would hope since we have given a 3-colouring.

A very interesting question is whether V is equal to all of $\mathbb{R}^2$. This is equivalent to asking whether every point in $\mathbb{R}^2$ is distance 1 from some rational. In other words, does every translation of the unit circle pass through a rational point? This question lies in the area of algebraic geometry. It is an example of the general question of the existence of rational points on curves which is highly non-trivial.

Lastly we may ask how many connected components RFC has. The answer is 1.

Theorem 4.1.1. *RFC has 1 connected component.*

Proof. First recall that every irrational point in V is connected to a point in $\mathbb{Q}^2$. Therefore if we can show that there is a path joining any two points in $\mathbb{Q}^2$ we are done. Note that the rational unit distance plane RUDP is a subgraph of RFC. Take two points $p, q \in \mathbb{Q}^2$ and find a path in RUDP from p to another point $p' \in \mathbb{Q}^2$ that is less than distance 2 from q. It is clear that this is possible. Now there must be a point x that is distance 1 from both p' and q. Since $p', q \in \mathbb{Q}^2$ they are connected to all points distance 1 from them so they are both connected to x. Hence we have a path from p to q. □

4.2 Rationals Fully Connected ∪ G

To add even more edges to the previous example we can take the rationally connected unit distance plane G and make each rational fully connected to the unit circle around it. So this graph has all $\mathbb{R}^2$ as its vertices with two points connected if they are distance 1 apart and either at least one of them is rational or they differ by a rational unit vector. Let's call this graph FRED. We can say that FRED = RFC ∪ G.

We can give a 4-colouring of FRED without much trouble. Give each point two binary labels, the first being the colour it receives in the 2-colouring of G, and the second indicating if it is rational or not. This colouring works because the only edges added to G to make FRED are between rationals and irrationals. They will differ in the second label. Vertices joined by the original edges of G will differ in the first label.

It is not clear that there could be no 3-colouring of FRED, so his chromatic number (using AC) may be 3 or 4. If we could show that FRED contained the Moser spindle then the issue would be settled, but this is not easy. Note that any triangle in FRED must have either one or two rational vertices. In the case of there being one, the opposite edge must be a rational unit vector. Given the constraints on triangles it seems very unlikely that FRED contains the Moser spindle, but a proof of this would be quite long. One would need to build up a list of all the constraints imposed by the relationships between vertices and edges of the spindle and show that they could not all be satisfied.

Note that any edge in FRED must either be a rational unit vector or have one end as a rational. There are some nice properties of the set of rational unit vectors that would help us in understanding the constraints on embeddings of the Moser spindle and other graphs in FRED. Most importantly the set A of angles α such that $(\cos\alpha, \sin\alpha) \in \mathbb{Q}^2$ (the set of 'rational rotations' in our terminology) is closed under addition and therefore forms a group. This follows from the double angle formulae. Therefore if we suppose that some edge in the Moser spindle is a rational unit vector this automatically determines if the other edges are. For two edges to both be rational unit vectors the difference in the angles that they make with the x-axis must be in A.

It is possible that FRED has ambiguous chromatic number, though this would be much harder to prove than for G. The 4-colouring we have given is based on the 2-colouring of G so relies on AC and is non-measurable. Supposing it is the best colouring, to show FRED has ambiguous chromatic number would require us to show that no measurable 4-colouring can exist. We might try to use a proof similar to Falconer's proof for UDP but this would not work because it relies on points distance 1 apart on a circle of radius $\sqrt{3}$ being connected.

Last of all we ask if FRED has just a single connected component like RFC. This depends on whether the vertex set V of RFC is the whole of $\mathbb{R}^2$. If it is then the answer is trivially yes. If it is not, so there are some irrationals that aren't distance 1 from any rational, then the answer is no. In FRED a point p not in V is only directly connected to other irrationals. Might there not be a path through the translation t of $\mathbb{Q}^2$ that contains p to a point p' distance 1 from a rational $q \in \mathbb{Q}^2$? This is impossible because $(p' - p) \in \mathbb{Q}^2$ so we could find the point

$$q' = q - (p' - p) \in \mathbb{Q}^2$$

which is distance 1 from p because $p - q' = p' - q$. Therefore if one point in $t(\mathbb{Q}^2)$ is not distance 1 from any rational then no point is.

4.3 Rational Rotations of the Triangle Lattice

Another way to add triangles to G is to connect every point to points at an angle of $\pm\pi/3$ from a 'rational rotation'. This means points are not only connected to points at an angle $\alpha \in A$, but also to points at angles $\alpha \pm \pi/3$. The effect is to give us lots of copies of the triangle lattice. More precisely, we get all the rotations of the triangle lattice where one of the basis vectors is a rational unit vector, translated to every point in the plane. Call this graph RRTL.

We can also think of RRTL as 3 copies of G in different rotations, call them $G, G_{+\pi}$ and $G_{-\pi}$. We could easily 8 colour it using 3 binary labels, but this isn't very interesting. It is not immediately obvious how we might find a better than 7-colouring. Hence it is hard to say if RRTL might have ambiguous chromatic number, but it is possible since it has a similar structure to G.

The Moser spindle contains two 'diamond' configurations with a common point. It can be found that the angle between the two copies of the diamond is $\theta = 2\arcsin(\sqrt{3}/6)$. Clearly θ is not in A since $\sin\theta = 1/\sqrt{33}$ is irrational. Therefore for RRTL to contain a Moser spindle we would need $\theta \pm \pi/3$ to be in A. But

$$\sin\left(\theta \pm \frac{\pi}{3}\right) = \sin\theta\cos\frac{\pi}{3} \pm \cos\theta\sin\frac{\pi}{3}$$

so these are not in A either. This means that if one of the diamonds has a rational unit vector in it the other cannot. So RRTL does not contain the Moser spindle and may have chromatic number as low as 3.

It is likely that RRTL has many connected components. Here is an idea for a proof. Each point is contained in one component of each of the three rotations of G. If RRTL had only one component then we would be able to find paths from any component of G to any other. Starting at a point p if we leave by an edge in G we stay in the same G-component. If we leave by an edge not in G we change G-component, but there are only countably many possibilities as the vertex order is countable. We need to show that paths starting at p can only end in countably many different G-components. Since there are uncountably many components in G we would conclude that some are unreachable from p. Suppose there is a canonical ordering of the edges coming out of a vertex. Then a finite path beginning at p can be defined by a finite string of integers. Hence the number of finite paths starting at p is countable so the number of different G-components reached is also countable.

If our definition of connectedness is that there is a path of finite length between any two points in a connected component then this proof should work. However if we allow paths of infinite length the argument breaks down. In showing that RFC and UDP have just one component we needed only to allow finite paths. The most obvious problem with infinite paths is deciding where they 'end'. If we said there was an infinite path from p to p' then we would probably mean that there was a subsequence of points in the path converging to p' (under the usual Euclidean metric). But since every edge in our graph has length 1 the sequence of points in an infinite path cannot itself be convergent. In fact, an infinite path could contain more than one convergent subsequence and thus 'end' at two or more distinct points. It would seem that our usual understanding of connectedness is better served by allowing only finite paths.

4.4 The Quadratic Field $\mathbb{Q}(\sqrt{3})$

D. Coulson suggested that I consider the unit distance graph induced by the field $\mathbb{Q}(\sqrt{3})$ crossed with itself. That is points in the plane with coordinates expressly as $q_1 + q_2\sqrt{3}$ for $q_1, q_2 \in \mathbb{Q}$. This is a neat way of adding triangles to the RUDP because if $\alpha \in A$ then the unit vector with angle $\alpha \pm \pi/3$ is in $(\mathbb{Q}(\sqrt{3}))^2$.

$(\mathbb{Q}(\sqrt{3}))^2$ is still a countable subset of $\mathbb{R}^2$. To add more points we could consider the graph formed by translating the unit distance graph on $(\mathbb{Q}(\sqrt{3}))^2$ everywhere. This graph contains RRTL as a subgraph. I have not studied this graph in any detail so will restrict myself to one more comment; it seems that it still does not contain the Moser spindle. This is because, as mentioned in the discussion of RRTL, the construction of the Moser spindle involves the angle $\theta = 2\arcsin(\sqrt{3}/6)$. Since $\sin\theta = 1/\sqrt{33}$ we see that the coordinates of the unit vector with angle θ will involve $\sqrt{11}$ and therefore not be in $\mathbb{Q}(\sqrt{3})$.

4.5 Summary

We have studied a few interesting subgraphs of the UDP in an attempt to find graphs that have more of the properties of UDP than our original example G. Our findings are summarised in the following table.

Graph:	G	RFC	FRED	RRTL
χ with AC:	2	3	$3 \leq \chi \leq 4$	$3 \leq \chi \leq 7$
χ with LM:	$3 \leq \chi \leq 7$	3	$3 \leq \chi \leq 7$	$3 \leq \chi \leq 7$
Components:	$\aleph_1$	1	Many (??)	Many (?)
3-cycles:	No	Yes	Yes	Yes
Triangle lattice:	No	No	Yes	Yes
Moser spindle:	No	No	No	No

Notably, we did not find any graphs that contained the Moser spindle. There is obviously a lot more that could be done in this area. The main thing to keep in mind when considering new subgraphs is whether they can be easily coloured with less than 7 colours. If we can't do this then they are unlikely to offer any insights as to how we might better colour the unit distance plane.

5 Dense distance excluding sets

Linked to the problem of the chromatic number of a $\mathbb{R}^n$ is the question of the highest density that a single colour might have in a colouring. In other words, what is the densest possible distance excluding set? We denote the maximum possible density by ρ. To be precise we can define the overall density d of a set S in $\mathbb{R}^n$ as follows

$$d(S) \equiv \lim_{r\to\infty} \frac{\mu(S \cap B_r(O))}{\mu(B_r(O))}$$

where $B_r(O)$ is an open ball of radius r centred at the origin and μ is Lebesgue measure. The maximum density ρ is then

$$\rho(\mathbb{R}^n) \equiv \sup_{S\subset\mathbb{R}^n} \{d(S) : S \text{ excludes distance } 1\}.$$

This number is related to the chromatic number restricted to measurable sets χ_m in the following way

$$\rho \geq \chi_m^{-1}.$$

This is because the overall density of the colour sets in an optimal colouring must be χ_m^{-1} on average. Upper bounds on ρ may therefore provide lower bounds on χ_m. We will however concentrate on lower bounds on ρ.

5.1 Lattices and spheres

Lower bounds on ρ are found by constructing dense distance excluding sets. The simplest way to do this is with periodic sets since we may calculate their overall density by calculating their density on a suitable finite region. It is certainly possible that some non-periodic sets might achieve greater density, but periodic sets serve as a reasonable starting point. We can construct periodic sets by translating some bounded set to each of the points in an infinite lattice.

We will need a few concepts from the theory of lattices. A lattice is the set of points that are linear combinations of a set of basis vectors with integer coefficients. The basis must be linearly independent and span the space, but is otherwise arbitrary. Once we have a lattice we can divide the space into 'nearest neighbour' or 'Voronoi' regions. These are the regions around each lattice point comprising points in the space for which that lattice point is the closest. The Voronoi regions of a lattice are all congruent. They are convex polytopes with faces that are perpendicular to the lattice vectors, lying halfway between lattice points. The vertices of these polytopes are the points which are locally as far as possible from any lattice points, and are called the 'holes' of the lattice.

If we construct a periodic set by translating some bounded set to each point of a lattice then the Voronoi regions each represent one period of the set. The set will therefore be similar on each Voronoi region and by calculating the density on one region we can find the overall density of the set.

Due to a result known as the isodiametric inequality an obvious place to start when looking for dense distance excluding sets is with n-spheres. The isodiametric inequality says that among all shapes of a given diameter the n-sphere (henceforth just 'sphere') has greatest volume. For this reason we can immediately construct a reasonably dense set by arranging open spheres of diameter 1 in the space so that the distance between neighbouring spheres is 1. Deciding the best way to arrange the spheres in space is equivalent to the sphere packing problem which asks the densest way to pack spheres so that they touch. This problem has been extensively studied and optimal packing lattices are known in dimensions up to 8. Regular lattices give the best known packings for infinite spaces, although some irregular packings are better on finite regions [2]. In $\mathbb{R}^n$ these sphere based sets will have a density of $1/2^n$ times the packing density.

5.2 Croft's set

In $\mathbb{R}^2$ the best packing lattice is the equilateral triangle lattice. In 1967 H.T. Croft presented a construction of a periodic set in $\mathbb{R}^2$ that improved on the density of open discs placed on the triangle lattice and is still the densest know set in the plane [5]. This paper is difficult to obtain and so the following description has been inferred from other incomplete descriptions such as that found in [6]. Croft's construction used the triangle lattice but with a slight modification to the circles. It was found that by increasing the radius of the circle towards the holes of the lattice and drawing the boundary in from the directions of the adjacent lattice points the area of the shape could be increased while still maintaining an excluded distance. The resulting shape was the intersection of an open disc and an open hexagon of slightly larger diameter, as shown in Figure 5. Note that the hexagon is the Voronoi region of the triangle lattice. As suggested by D. Coulson, we could call shapes of this type ‘polyhedral dice’, so crofts shape is the ‘hexagonal die’.

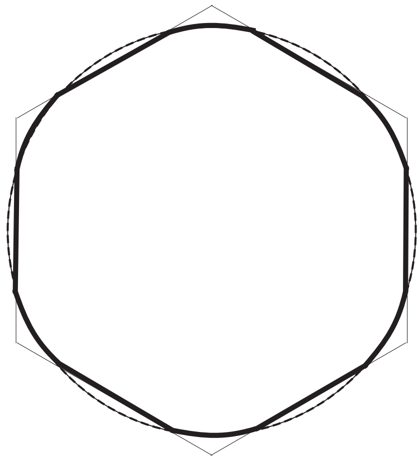

Figure 5: The hexagonal die.

To calculate the optimal size of disc and hexagon we first set the length of the lattice vectors to 2. If r_d is the radius of the disc and r_h is the inradius (i.e. minimum radius) of the hexagon then to have an excluded distance we need $2r_d + 2r_h = 2$. If θ is the angle subtended by the midpoint and end point of a chord then $\cos\theta = r_h/r_d$. Using these two equations we can find an expression for the area A of the figure in terms of θ alone, namely

$$A = \frac{24(\cos\theta\sin\theta + \pi/6 - \theta)}{(1+\cos\theta)^2}$$

and then confirm that the area is at a maximum when $\theta \approx 0.2633$ and this gives a density of 0.2294. We omit the derivation of the above equation since a similar but more complex calculation in $\mathbb{R}^3$ appears below.

5.3 The general method

The method used by Croft can be generalised to improve the density of other (perhaps all) sphere and lattice based distance excluding sets. The general method for creating a dense distance excluding set in $\mathbb{R}^n$ is to take the best packing lattice and at each point place a figure consisting of the intersection of a sphere and a scaled Voronoi region of the lattice. The idea is that by expanding towards the holes we can increase the volume of the figure. However this increases the diameter of the figure, so it must be ‘drawn in’ from the direction of the other figures on the lattice to compensate. This 'drawing in' involves cutting back a flat surface perpendicular to the lattice vectors, hence the use of the Voronoi region. Some calculus must be applied to find the optimal size of sphere and Voronoi region, but we can make a convincing argument as to why (and where) a maximum occurs.

Firstly note that when we expand the sphere by $\triangle r$ we must cut back along the lattice vector by the same amount. Therefore when we first start to expand and cut back there must be an increase in volume since the caps we cut off are much smaller in area than the expanded curved surface area. At the other extreme we have the situation of a scaled Voronoi region of maximal diameter being rounded back. Since the maximum diameter is achieved between vertices of the Voronoi region, rounding the vertices off by $\triangle r$ (which only removes a small volume) will allow us to expand the flat surfaces by the same amount. But this will add a larger volume since most of the surface is flat. Hence there will be an overall increase in volume.

This type of reasoning led D. Coulson to formulate a very neat characterisation of the optimal configuration for these polyhedral dice [12]. The volume is maximised when the curved surface area is equal to the flat surface area. This is because a small expansion of the sphere will add the same amount of volume as the corresponding contraction of the Voronoi region will remove, so the net effect will be negligible. This means that the change in volume with respect to change in radius is zero, i.e. $\frac{dV}{dr} = 0$. In general we could write the change in volume $\triangle V$ in terms of a small change in radius of the sphere $\triangle r$ as

$$\triangle V = (\triangle r) \times (\text{curved area}) - (\triangle r) \times (\text{flat area}) + O((\triangle r)^2)$$

where $O((\triangle r)^2)$ are error terms of higher order arising from the boundary between the curved and flat parts of the surface. Clearly if we set 'curved area' = 'flat area' we get $\triangle V \approx 0$.

5.4 Application to $\mathbb{R}^3$

D. Coulson and the author have recently demonstrated that this method works in $\mathbb{R}^3$ by constructing the densest known distance excluding set in 3-space [12]. The construction is based on the Face Centred Cubic Lattice (FCCL) which is the optimum packing lattice for $\mathbb{R}^3$. We use the FCCL with basis vectors of length 2, that is

$$\text{FCCL} = \left\{ z_1(\sqrt{2}, \sqrt{2}, 0) + z_2(0, \sqrt{2}, \sqrt{2}) + z_3(\sqrt{2}, 0, \sqrt{2}) : z_i \in \mathbb{Z} \right\}.$$

On each point of this lattice we place a figure which is the intersection of a sphere of radius r_s and a scaled Voronoi region of FCCL of inradius r_d such that $2r_s + 2r_d = 2$. This ensures that there is an excluded distance of $2r_s$ if the figure is open. The Voronoi region $\mathcal{V}$ of FCCL is a rhombic dodecahedron and is the convex hull of its vertices

$$\sqrt{2}\{\pm(1,0,0), \pm(0,1,0), \pm(0,0,1), (\pm 1/2, \pm 1/2, \pm 1/2)\}$$

with volume $4\sqrt{2}$. It can be seen in Figure 6 along with the shape that results from intersecting it with a sphere. We will call this final shape the Rhombic Dodecahedral Die or RDD.

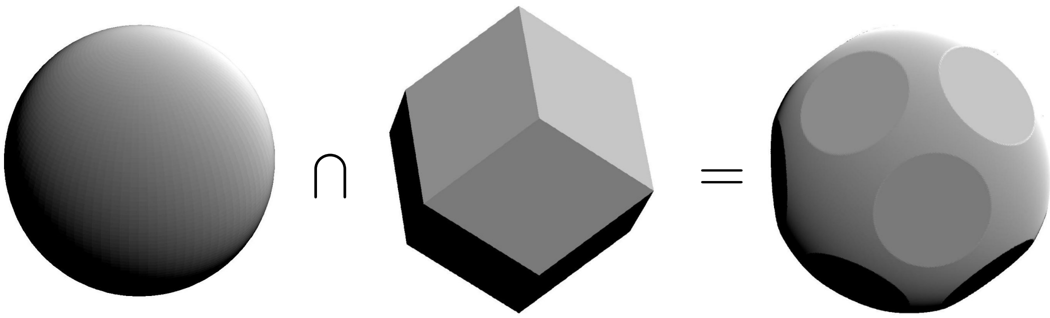

Figure 6: Building the Rhombic Dodecahedral Die

To find the optimum value for r_s it is necessary to express the volume of the RDD in terms of r_s. For $1/2 \leq r_s \leq 4 - 2\sqrt{3}$ the RDD is a sphere with 12 identical circular 'caps' cut off, similar to that shown in Figure 6. This means the volume can be calculated reasonably easily by first finding the volume of the caps. However when $r_s > 4 - 2\sqrt{3}$ the edges of the rhombic dodecahedron $\mathcal{V}$ emerge from the sphere, and the shape is more like $\mathcal{V}$ with its vertices 'rounded off'. In this situation the volume would be significantly more difficult to calculate. Fortunately, the maximum volume occurs when $r_s \leq 4 - 2\sqrt{3}$, as we will see below.

To calculate the volume of the caps we use a solid of rotation and the fact that $r_d = 1 - r_s$.

$$\begin{aligned} V\text{cap}(r_s) &= \pi \int_{r_d}^{r_s} r_s^2 - x^2 \, dx \\ &= \pi[\frac{2}{3}r_s^3 - r_s^2 r_d + \frac{1}{3}r_d^3] \\ &= \frac{\pi}{3}[4r_s^3 - 3r_s + 1] \end{aligned}$$

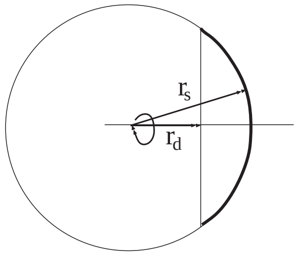

Figure 7: The 'caps' cut from the sphere of radius r_s.

Hence for $1/2 \leq r_s \leq 4-2\sqrt{3}$ the volume of the RDD can be written as

$$\begin{aligned} V_{\text{RDD}}(r_s) &= \frac{4}{3}\pi r_s^3 - 12 \times V_{\text{cap}}(r_s) \\ &= \frac{4\pi}{3}\left(r_s^3 - 3[4r_s^3 - 3r_s + 1]\right) \\ &= \frac{4\pi}{3}\left(9r_s - 11r_s^3 - 3\right) \end{aligned}$$

By differentiating this function we find that it achieves a maximum at $r_s = \sqrt{3/11}$ at which point

$$V_{\text{RDD}}(\sqrt{3/11}) = \frac{4\pi}{3}\left(6\sqrt{\frac{3}{11}} - 3\right) \approx 0.5588.$$

Since the volume of the Voronoi region of the lattice $\mathcal{V}$ is $4\sqrt{2}$ the overall density of this set in $\mathbb{R}^3$ is

$$V_{\text{RDD}}(\sqrt{3/11})/(V_{\mathcal{V}}) = \frac{\pi}{3\sqrt{2}}\left(6\sqrt{\frac{3}{11}} - 3\right) \approx 0.09878.$$

The density of the set consisting of open spheres of radius 1/2 placed on the FCC lattice will be $\frac{4\pi}{3\times 32\sqrt{2}} \approx 0.09256$. Hence the density of the RDD based set is about 6.7% greater than that of the sphere based set.

At this point we might recall the relationship between the maximum possible density ρ and χ_m, namely $\chi_m \geq \rho^{-1}$. The density we have found for the RDD based set is between $\frac{1}{11}$ and $\frac{1}{10}$. While we can not draw any definite conclusions about χ_m from a lower bound on ρ, we can observe that if this set were the densest possible, or close to it, we would be able to conclude that $\chi_m(\mathbb{R}^3) \geq 11$. This is a large improvement over the lower bound of 6 given by Falconer's proof considered earlier.

D. Coulson also confirmed that the 'flat surface area' = 'curved surface area' characterisation holds for the RDD as indeed it does for Croft's hexagonal die [12].

5.5 Maximality of Polyhedral Dice

There is an interesting property of spheres which has an analogue in polyhedral dice. On the surface of a sphere each point has an antipodal pair with which the diameter is achieved. On a polyhedral die the points on the curved surface also have an antipodal pair on the other side of the figure. Points on the flat faces are paired with a point on the nearest flat surface of the next copy of the die. So all points except the points on the boundary between flat and curved surface are diametrically opposed to exactly one other point. This similarity between spheres and polyhedral dice may be evidence that polyhedral dice perform a similar role to spheres.

It would be interesting to know if polyhedral die based sets are maximal among distance excluding sets based on the same lattices. This would be similar to proving an isodiametric inequality for periodic spaces with the polyhedral dice as the maximal shapes. However, we would have to redefine the notion of diameter since the usual definition (the supremum of the distances between pairs of points in the set) would give an infinite value for periodic sets. We could call this p-diameter (for periodic set diameter) and the new definition would be something like

$$\text{p-diam}(S) \equiv \inf\{\mathbb{R}^+ \backslash \{|x-y| : x, y \in S\}\}.$$

That is, the infemum of the distances not realised by points in S. This would only exist if there was an excluded distance. We we will also use the term 'volume' to mean volume restricted to one period of the set.

Let us attempt to formulate a conjecture regarding the possible maximality of polyhedral dice. The strongest result we could hope for is this:

Conjecture 5.5.1. Among all sets that share the periodicity of a lattice Λ and have a finite p-diameter, the polyhedral die induced by Λ with equal curved and flat surface area has the greatest volume.

This conjecture is of a quite different form to the isodiametric inequality. It does not allow us to specify the p-diameter of the set, but just uses the fact that there is a maximum possible p-diameter. The maximum possible p-diameter is the length of the hole vectors h. An example of a shape with this p-diameter is the Voronoi region scaled to half size. But we know this shape is not maximal, so our Conjecture 5.5.1 claims that the maximum volume is not achieved with the maximum p-diameter.

To formulate a conjecture that is more like the isodiametric inequality we need to specify a p-diameter for our set. Let us assume that the lattice vectors of Λ have length l. We define a family of polyhedral dice as being the intersection of an open sphere of radius r_s and an open scaled Voronoi region of inradius r_p such that $2r_s + 2r_p = l$. This ensures that they all have an excluded distance and hence a p-diameter. The die of p-diameter d has $r_s = d/2$. A die is determined by its p-diameter and the lattice that induces it, so we will denote it by D_d^Λ. We denote the p-diameter of the die of maximum volume by d^*. Recall that d^* must be found on a case by case basis – this is the where the hard work is. But from our arguments above it is reasonable to assume that d^* exists and is somewhere between $l/2$ (where $D_{l/2}^\Lambda$ is a sphere) and h (where D_h^Λ is a Voronoi region). Our new conjecture would go along these lines:

Conjecture 5.5.2. Among all sets that share the periodicity of a lattice Λ and have p-diameter d or less, the set of greatest volume is D_d^Λ for $d \leq d^*$ and $D_{d^*}^\Lambda$ for $d > d^*$.

There is one problem with this conjecture that we must address. For small enough d it is possible that a set could have more than one path connected component in each period of Λ. This would obviously increase the volume. For example, for small d our die D_d^Λ is just a small sphere centred on the lattice point. But we could increase the volume of our set by adding more little spheres within the Voronoi region of Λ as long as the excluded distance was maintained. So in the above conjecture we should demand that the sets have just one path connected component per period. This is not much of a restriction, since we are mainly interested in larger p-diameter sets for which having just one component is clearly the best idea.

How might we prove this conjecture? The proof of the isodiametric inequality uses a process known as Steiner Symmetrisation [13]. The idea is to take a general set and progressively symmetrise it about each $(n-1)$-coordinate hyperplane. This means we take the 1-dimensional measure of the set on each line perpendicular to the plane and then replace it with an interval of the same measure on the line, centered on the point where the line meets the plane. We do this until the set is symmetrised in every coordinate direction. The important point is that this process can only decrease the diameter and that the resulting set has the same volume as the original. We conclude that the fully symmetrised set is contained in the ball of the same diameter, so must have volume less than or equal to that of the ball.

To adapt this proof to our case is not very easy. We would need a symmetrisation process that maintained volume and did not increase p-diameter. We would hope that such a process could be shown to produce sets that are contained in the polyhedral dice. But our definition of p-diameter is more complicated than the usual one. We would need to take care to show that by moving parts of the set around we did not destroy the excluded distance. The approach to take may be to perform the symmetrisation about the planes perpendicular to the lattice vectors. We may be able to show that if the p-diameter of the set is d then the measure on lines parallel to the lattice vectors must be at most $l - d$. Therefore after symmetrisation in all the directions of the lattice vectors we could say that the set was contained in the Voronoi region of inradius $l - d$. Since the symmetrisation process does not increase normal diameter, we should be able to say that the symmetrised set is also contained in the ball of diameter d. Hence the polyhedral die would emerge naturally enough.

If this argument went through we would have shown that the symmetrised set of p-diameter d was contained in D_d^Λ. The last part of the proof would be to note that for sets with $d > d^*$ there are larger sets with *smaller* p-diameter, namely $D_{d^*}^\Lambda$. We have not addressed the possibility that the lattice vectors may not all be the same length. The construction of the polyhedral dice still works in this case so it is likely that the proof would still go through, perhaps with minor modifications.

Acknowledgements

This thesis was written in partial fulfilment of the requirements of the Mathematics Honours course at Monash University. I would like to thank my supervisors Dr. Burkard Polster and Dr. Ian Wanless for their valuable suggestions and support. Thanks also to all the staff involved in running and teaching the honours program and to my friends and family for putting up with me.

My interest in the chromatic number of the plane problem began during a summer vacation scholarship program a the University of Melbourne under the guidance of Dr. Dave Coulson. I am indebted to him for introducing me to a very interesting problem, and also for giving me copies of Figures 1, 2, 5, 6 and 7. The last three of these are from our paper [12] which hopefully will appear shortly.

References

[1] R. Cignoli and J. Hounie. Functions with arbitrarily small periods. *Amer. Math. Monthly*, 85(7):582–584, 1978.

[2] J. H. Conway and N. J. A. Sloane. *Sphere packings, lattices and groups*, volume 290 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, New York, third edition, 1999. With additional contributions by E. Bannai,

R. E. Borcherds, J. Leech, S. P. Norton, A. M. Odlyzko, R. A. Parker, L. Queen and B. B. Venkov.

[3] D. Coulson. A 15-colouring of 3-space omitting distance one. *Discrete Math.*, 256(1-2):83–90, 2002.

[4] D. Coulson. On the chromatic number of plane tilings. *J. Aust. Math. Soc.*, 77(2):191–196, 2004.

[5] H. T. Croft. Incidence incidents. *Eureka (Cambridge)*, 30:22–26, 1967.

[6] Hallard T. Croft, Kenneth J. Falconer, and Richard K. Guy. *Unsolved problems in geometry.* Problem Books in Mathematics. Springer-Verlag, New York, 1994. Corrected reprint of the 1991 original [MR1107516 (92c:52001)], Unsolved Problems in Intuitive Mathematics, II.

[7] N. G. de Bruijn and P. Erdös. A colour problem for infinite graphs and a problem in the theory of relations. *Nederl. Akad. Wetensch. Proc. Ser. A.* **54** = *Indagationes Math.*, 13:369–373, 1951.

[8] K. J. Falconer. The realization of distances in measurable subsets covering $\mathbf{R}^n$. *J. Combin. Theory Ser. A*, 31(2):184–189, 1981.

[9] Martin Gardner. A new collection of brain teasers. *Sci. Amer.*, 203:172–180, Oct. 1960.

[10] Gregory H. Moore. *Zermelo's Axiom of Choice.* Springer-Verlag New York Inc., 1982.

[11] Oren Nechushtan. On the space chromatic number. *Discrete Math.*, 256(1-2):499–507, 2002.

[12] M. S. Payne and D. Coulson. A dense distance 1 excluding set in $\mathbb{R}^3$. *Forthcoming.*

[13] Marty Ross. Measure theory notes. *ICE-EM Summer School 2006.*

[14] Saharon Shelah and Alexander Soifer. Axiom of choice and chromatic number of the plane. *J. Combin. Theory Ser. A*, 103(2):387–391, 2003.

[15] Alexander Soifer. Chromatic number of the plane & its relatives. I. The problem & its history. *Geombinatorics*, 12(3):131–148, 2003.

[16] Robert M. Solovay. A model of set-theory in which every set of reals is Lebesgue measurable. *Ann. of Math. (2)*, 92:1–56, 1970.

[17] Carsten Thomassen. On the Nelson unit distance coloring problem. *Amer. Math. Monthly*, 106(9):850–853, 1999.

[18] D. R. Woodall. Distances realized by sets covering the plane. *J. Combin. Theory Ser. A*, 14:187–200, 1973.