Research Description and Directions

Mark A. Fitch

Recently I have been researching problems in discrete geometry. In particular I have been studying arrangements of points and lines. Erdös once posed the following: Show that \( n \) non-collinear points in the real plane determine at least \( n \) connecting lines. This is a problem in projective geometry. The equivalent problem in affine geometry is Show that \( n \) non-collinear points in the real affine plane determine at least \( n - 1 \) distinct slopes. This problem was first considered by P. R. Scott [8] and finally solved by Peter Ungar [9]. After Ungar proved this conjecture, research has centered on two areas: finding and characterizing configurations achieving the minimum number of slopes, which are called slope critical configurations, and answering the question in other settings such as finite planes.

A single theorem characterizing all point configurations attaining a minimum number of slopes, called slope critical configurations, seems unlikely. Four infinite families have been discovered, but over 100 sporadic cases are also currently known. All slope critical configurations do share some characteristics, however. For example, all slope critical configurations consist of an odd number of points. Properties about the distribution of the points are also known.

Slope critical configurations consist of points distributed in a wheel-like fashion. Every critical configuration possesses a point called a centrex [6], which acts like the hub of a wheel. The other points satisfy a modified central symmetry property. Every point is collinear with another point on the other side of the centrex. The rays from the centrex that include other points of the configuration are called spokes. In ordinary central symmetry, the distance from a point to the center is the same as the distance from the other point to the center. In critical configurations, the ratio of distances is not one, but it is fixed for that configuration.

In [6] Jamison analyzed critical configurations by considering groups of consecutive spokes. A group of three consecutive spokes is called a 3-wedge or just a wedge. Under the assumption that a configuration is critical, wedges are best described by a set of parallelism and collinearity conditions that can also be described by a set of equations. Using the assumption that each spoke consists of only one point, Jamison shows that a wedge must be one of three distinct types [6].

Configurations can be built from collections of wedge types. Each configuration meets all the parallel and collinearity conditions of the wedge types it contains. Equivalently, the coordinates of its points are determined by the solutions to the simultaneous system of equations that encode the parallel and collinearity conditions. Because the wedges meet the condition of being slope critical, configurations constructed from wedges are called locally slope critical configurations.

The concept of a wedge can be generalized to \( k \)-wedges, or \( k \) consecutive spokes. As with 3-wedges, these \( k \)-wedges are divided into types defined by parallel and collinearity conditions. The additional spokes add constraints, reducing the number of feasible \( k \)-wedge types. Using \( k \)-wedges, I have shown that all non-centrally symmetric, slope critical configurations having
only one point per spoke have been found.

While using wedge analysis, I discovered a special family of locally slope critical point configurations: a subset of the double polygons. In general a double $(k, s, t)$-gon in the affine plane is defined by the following construction.

Begin by constructing a $k$-gon in the plane. Without loss of generality, this $k$-gon can be assumed to be a regular $k$-gon with its center at the origin and a vertex on the positive $x$-axis. Any other $k$-gon is affinely equivalent to this one. Label the points $p_0, p_1, p_2, \ldots, k - 1$ counterclockwise, starting with the vertex on the positive $x$-axis.

Next, construct the $k$ diagonals from each vertex, $i = 0, 1, 2, \ldots, k - 1$, to the vertex $i + s$. The additional $k$ points, $q_i$, are defined by the following intersections: the intersection of the line $p_ip_{i+s}$ with the line $p_{i+t}p_{i+t-s}$ is labeled $q_i$.

The diagram below demonstrates this construction for the $(5,2,1)$-gon, which is best known as the pentagram.

[Diagram of a (5,2,1)-gon with labeled points]

The other notable $(k, s, t)$-gon is the $(8,3,2)$-gon or Grünbaum double octagon. The pentagram and double octagon are the only two planar $(k, s, t)$-gons that are slope critical. The family of double polygons found through wedge analysis is the family of $(k, \frac{(k-1)}{2}, 1)$-gons with $k$ odd.
In addition to the algebraic and geometric analysis of slope critical configurations, I have also used allowable sequences. Ungar used this combinatorial structure, developed by Goodman and Pollack [3], to prove the slope conjecture. Jamison also used it to prove the existence of the centrex [7].

To understand allowable sequences, consider the derivation of the allowable sequence for a planar configuration. Choose a set of $n$ points in the affine plane and label the points arbitrarily $0, 1, 2, \ldots , n - 1$. Choose a directed line in general position relative to the points. The projection of the points onto the line determines an ordering of the labels. Begin rotating the set of points clockwise. The rotation will reorder the projections onto the line, generating another linear ordering of the labels. Different orderings of the labels will be generated until the configuration has rotated $2\pi$. This list of orderings of the labels will be called the circular sequence of the configuration.

In general, an allowable sequence on $n$ points is a sequence of permutations of the numbers $0, 1, 2, \ldots , n - 1$ satisfying the following conditions.

1. The beginning permutation is $\pi_0 = 0, 1, 2, \ldots , n - 1$. The ending permutation is $\pi_k = n - 1, n - 2, n - 3, \ldots , 0$.

2. Each pair of points reverses positions exactly once in the entire sequence.

3. Moves between orderings consist of the reversal of one or more non-overlapping, increasing substrings.

As a result of difficulties in describing properties of $(k, s, t)$-gons and the symmetries using the original definition of allowable sequences, I developed terminology and a new equivalent combinatorial structure. The important issue is to distinguish among the list of orderings of the marks that represent points, the permutations (functions) that permute the marks from one ordering to the next, and the permutations that are defined by permuting indices rather than marks.

The geometric derivation of allowable sequences defines a list of orderings of $n$ marks, usually the integers $0, 1, \ldots , n - 1$. Thus, without loss of generality, an allowable sequence begins with the ordering $0, 1, 2, \ldots , n - 1$ and ends the first half period with the ordering $n - 1, n - 2, n - 3, \ldots , 0$. These orderings directly encode only the left to right ordering of the points at a given step.

The change from one ordering of marks to the next defines a permutation of those marks. These permutations are functions on the marks, normally the first $n$ non-negative integers. If the first two orderings are $\Omega_0 = [0, 1, 2, 3]$ and $\Omega_1 = [1, 0, 3, 2]$, the first permutation on the marks is defined by $\Pi_0 = (0 1) (2 3)$.

The change from one ordering to the next can also be defined by a permutation on the indices. For example, the marks $0, 1, 2, 3$ are in the positions, read left to right, $1, 0, 3, 2$ in the second ordering. This ordering demonstrates that the marks in positions 0 and 1 switched and the marks in positions 2 and 3 switched. Under the assumption that the marks are in standard order in the first move, the first permutation on the indices and the
first permutation on the marks will be the same. Additional permutations on the marks and
permutations on the indices will not be the same in general.

The orderings of the marks and both types of permutations are interchangeable. Both types
of permutations generate a list of orderings of the marks given an initial ordering. For any
given list of orderings of the points, a list of permutations of either sort can be constructed
to describe it. The indices permutations are, however, much easier to use when describing
symmetries and working with \((k, s, t)\)-gons.

Essentially rotational symmetry in allowable sequences is defined by a repetition of the
permutations. In a planar configuration with rotational symmetry of order \(k\), the rotation
(a transformation) can be applied \(k\) times, producing the same point configuration. In
allowable sequences, described as permutations of the indices, a configuration with rotational
symmetry of order \(k\) has the property that the list of permutations of length \(m\) consists of
a set of \(m/k\) permutations that are repeated \(k\) times. Thus the transformation of moving
\(m/k\) elements through the list produces the same list of permutations.

Reflectional symmetry is defined by the permutations matching, read forward or backward
from a given permutation. Reflection over a line changes the left to right ordering of points.
In particular, if the line of reflection is vertical, then the points are reversed. Also note, that
rotating a point configuration counterclockwise around a point becomes clockwise rotation
after reflecting the configuration over a line through that point. Allowable sequences encode
left to right ordering and direction of rotation. If a planar point configuration has reflectional
symmetry, then the reflection, a transformation, can be applied and the same configuration
is produced. Thus after a reversal of the left to right ordering, reading the allowable sequence
backward is the same as reading it forward, in original left to right order.

Using permutations on indices, configurations with reflectional symmetry can be described
in relatively few permutations. The reflectional property indicates that a basic pattern is
repeated a fixed number of times to generate the full allowable sequence. For example the
two permutations \(\Psi_0 = (01)(23)\) and \(\Psi_1 = (1 2)\) describe a square. These two are repeated
four times to generate the full allowable sequence. The difficult question is determining
if a given pattern, consisting of multiple permutations on indices and a fixed number of
repetitions, actually produces an allowable sequence. I have shown that if a known allowable
sequence is embedded in the pattern, proving that the pattern generates another allowable
sequence is easier. This theorem is used to prove that slope critical \((k, s, t)\)-gons exist for
many values of \(k, s,\) and \(t\) by listing only four permutations on indices. The theorem used
is somewhat restrictive. I would like to investigate the design of critical allowable sequences
from other allowable sequences in a more general setting.

In addition to the affine slope problem, I am working on sum covers. A subset \(S\) of an
Abelian group \((A, +)\) is a sum cover provided every element \(x \in A\) can be expressed in at
least one way as a sum \(x = s + t\) where \(s\) and \(t\) are in \(S\). If \(s \neq t\) is also required, \(S\) is a
strict sum cover. For example, the set \(\{0, 1, 2\}\) is a sum cover of \(\mathbb{Z}_5\), because \(0=0+0,\ 1=0+1,\ 2=0+2,\ 3=1+2,\ and\ 4=2+2.\) However, this set is not a strict sum cover because there is no
way to produce \(0\) except as the sum \(0+0.\) A strict sum cover of \(\mathbb{Z}_5\) would be \(\{0, 1, 2, 3\},\ as\ 0=2+3\) is the sum of two unique elements.
These covers had arisen from a study of sets of points in a finite, affine plane that generate all possible slopes [1]. Essentially, a finite plane is an $n \times n$ grid of points. The coordinates are integer coordinates $0, 1, 2, \ldots, n - 1$. Arithmetic is done modulo $n$.

The goal of the study was to find sets of minimum cardinality that generate all possible slopes. This is the slope problem over finite planes. The points on a parabola generate all possible slopes over any plane. Consider two points on a parabola: $(a, a^2)$, $(b, b^2)$. The slope of the line through these points is $\frac{b^2 - a^2}{b - a} = \frac{(b-a)(b+a)}{b-a} = b + a$. Thus, finding a set of points that generates all points is related to finding a set of numbers that generates all possible sums. Strict sum covers are necessary because sums of the form $a + a$ represent tangent lines to the parabola, not secants.

The questions that I have investigated involve minimum sum covers over $\mathbb{Z}_n$. For an arbitrary modulus $n$, $\mathbb{Z}_n$ has the structure of an algebraic ring. As a result, general theorems for all $\mathbb{Z}_n$ are not expected. My research efforts have centered on finding bounds for the minimum cardinality and developing constructions that produce near minimum covers.

Initial work involved finding minimum sum covers for small moduli through search algorithms. The number of possible sum covers of length $k$ over the ring $\mathbb{Z}_n$ is, of course, $n \choose k$. As a result an exhaustive search for minimum sum covers quickly becomes infeasible. To overcome this difficulty, I have applied various theoretical results to trim the search tree.

First define an equivalence relation using holomorphies. A holomorphy of an Abelian group $(A, +)$ is a mapping of the form $h(x) := \phi(x) + t$, where $\phi$ is an automorphism of $A$, and $t$ is any fixed element of $A$. Two sets are considered equivalent if there exists a holomorphy that maps one set onto the other set. For example, over the ring $\mathbb{Z}_8$ the sets $\{0, 1, 2\}$ and $\{0, 3, 6\}$ are equivalent, because the holomorphy $t \mapsto 3t + 0$ maps the first set onto the second.

Holomorphies are useful for search algorithms because a set and its holomorphic image generate the same number of slopes. For example, the set $\{0, 1, 2\}$ generates the slopes $\{0, 1, 2, 3, 4\}$, and the set $\{0, 3, 6\}$ generates the slopes $\{0, 1, 3, 4, 6\}$. Both sets generate five slopes but do not generate the same ones. To understand why holomorphies preserve the number of sums generated, consider the following. Let $a, b$ be elements of a ring. By definition of an automorphism, $\phi(a+b) = \phi(a) + \phi(b)$. Thus every unique sum in the original set remains a unique sum in the image set. Suppose $h(x) = x + t$. $a \mapsto a + t$, and $b \mapsto b + t$; hence, the sum $a + b$ is replaced by the sum $a + b + 2t$. Thus, the set of sums of the image set is just a shift of the set of sums of the original set.

Because the size of the sum set is invariant under holomorphies, the holomorphic image of a sum cover is also a sum cover. As a result a search algorithm needs to test only one set from each equivalence class. However, testing if one set is holomorphically equivalent to another is time consuming. For example, in the integer rings $\mathbb{Z}_n$ there are $\Phi(n) + n$ possible holomorphies in which $\Phi$ is the Euler phi function. Thus, if every possible set is compared to previously tested sets, the algorithm will take at least $\Phi(n) + n$ steps per set: far too much time. However, certain cases of holomorphic equivalence can be implemented with quick algorithms.
The search technique that I have used is a modified exhaustive search. The algorithm is given the modulus \(n\) and the current set size \(q\). It generates the sets in lexicographic order. The algorithm can skip over sets that are known to be holomorphically equivalent to another set that is lexicographically less.

Jamison has shown [1] that for moduli that are the product of four or fewer prime powers, every sum cover is equivalent to a sum cover containing 0 and 1. This includes all moduli less than 2310, which is far greater than any current search algorithm can attain to; therefore, sets without 0 and 1 are never tested.

Arithmetic progressions can also be used to reduce the search tree. An arithmetic progression is any finite sequence of the form \(a, a+b, a+2b, \ldots, a+kb\) for some length \(k\). The arithmetic progressions of interest here are those with \(b=1\). Testing the lengths of progressions requires little time in an algorithm and many sets with long progressions are holomorphically equivalent to a lexicographically lesser set.

My analysis of examples found by the search algorithm has led to a number of conjectures and open questions. For example, for most moduli there exists a sum cover with an arithmetic progression that is close to half the size of the cover. Although such an arithmetic progression does not exist in all minimum sum covers, small covers can always be constructed using this idea. This construction provides a means to establish an upper bound on the minimum size of a sum cover. The upper bound is \(\frac{k^2+2k+n+3}{k+2}\) where \(k = -2 + \sqrt{n+3}\). This is \(O(\sqrt{n})\). An equivalent proof giving the same upper bound is given in [2].

If a set has \(k\) elements, it determines at most \(k\) choose 2 sums. Thus a lower bound for the size of a sum cover is \(\left\lceil \frac{\sqrt{1+8n}}{2} \right\rceil\). Note that this is also \(O(\sqrt{n})\). Thus we know that sum covers are \(O(\sqrt{n})\) because they are bounded above and below by this bound [2].

Multiple avenues of continued research in sum covers exist. Although a formula for the size of a minimum sum cover probably does not exist, I would like to continue research to find tighter bounds on the size. Improving existing constructions or devising new constructions for small sum covers might provide me with the means to produce tighter bounds. Another area of possible research involves the size of a minimum sum cover. Not surprisingly minimum sum covers are sometimes smaller than minimum strict sum covers. The difference is known to take on at least the values 0, 1, and 2. I am interested in investigating if the difference can be larger than 2. Also, I am curious if reasons for the difference can be found. Generalizations of sum covers to other types of groups are also of interest to me.

References


