Diophantine Equations of Matching Games I

Chun Yin Hui, Wai Yan Pong

January 17, 2011

Abstract

We solve a family of quadratic Diophantine equations associated to a simple kind of games. We show that the ternary case, in many ways, is the most interesting and the least arbitrary member of the family.

1 The Matching Games

An (n, d)-*matching game* $(n, d \geq 2)$ is a game in which the player draws d balls from a bag of balls of n different colors. The player wins if and only if the balls drawn are all of the same color. A game is *non-trivial* if there are at least d balls in the bag. It is *faithful* if there are balls in each of the n colors. A game is *fair* if the player has an equal chance of winning or losing the game. In this article, we only study the $(n, 2)$-matching games or simply the n-*color games*, leaving the study of the higher d case to [10].

An n-tuple $(a_1, \dots, a_n)$ where a_i is the number of the i-th color balls in the bag represents an n-color game. For $m \leq n$, an m-color game $(a_1, \dots, a_m)$ can be regarded as the n-color game $(a_1, \dots, a_m, 0, \dots, 0)$. The only trivial n-color fair games, are the *zero game* $(0, \dots, 0)$ and, up to permutation, the game $(0, 0, \dots, 1)$.

By considering the number of ways for the player to win the game, one sees that the n-color fair games are exactly the non-negative integral solutions of

$$\binom{\sum_{i=1}^n x_i}{2} = 2\left(\sum_{i=1}^n \binom{x_i}{2}\right)$$

or equivalently,

$$F_n(x_1, \dots, x_n) := \left(\sum_{i=1}^n x_i\right)^2 - \sum_{i=1}^n x_i - 4\sum_{i \neq j} x_i x_j = 0 \tag{1}$$

The paper will be organized in the following way: We give a brief treatment of the 2-color games in Section 2. The results there illustrate the kind of questions that we try to answer in the general case. In Section 3, we give a "parametric" solution to Equation (1). It is unclear, however, from this method which choice of the parameters will yield the fair games. We tackle this problem in Section 4 by giving a graph structure to the solutions. We show that the components of this graph are trees and give an algorithm for finding their roots. This yields all solutions recursively. Furthermore, we characterize the components containing the fair games. For $n = 3$, we show that the graph consists of two trees with the nontrivial 3-color fair games forming a full binary tree. We then study what are the possible coordinates of fair games in Section 5. In Section 6, we establish some partial results concerning the asymptotic behavior of 3-color fair games. We conclude the article with some odds and ends of our study in Section 7.

The following conventions will be used throughout this article:

- All variables and unknowns range over the integers unless otherwise stated.
- The cardinality of a set A is denoted by $|A|$.
- For $\boldsymbol{a} \in \mathbb{Z}^m$, the (Euclidean) norm of $\boldsymbol{a}$ is denoted by $\|\boldsymbol{a}\|$. For $A \subseteq \mathbb{Z}^m$ and $k \geq 0$, $A(k)$ denotes the set of elements of A with norm at most k. We define the *height* of $\boldsymbol{a}$ so be $|1 + \sum a_i|$.
- For any integer d and $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{Z}^m$, we say that $\boldsymbol{a}$ and $\boldsymbol{b}$ are congruent modulo d, written as $\boldsymbol{a} \equiv \boldsymbol{b} \mod d$, if $a_i \equiv b_i \mod d$ for all $1 \leq i \leq m$.
- Denote by $\mathcal{S}_n$ and $\mathcal{F}_n$ the set of integral and non-negative integral solutions, identified up to permutations, of Equation (1) respectively. Elements of $\mathcal{F}_n$ are the n-color fair games. We often use an increasing (or decreasing) tuple to represent an element of $\mathcal{S}_n$. Denote by $\mathcal{C}_n$ the set of coordinates of $\mathcal{F}_n$.
- For any n-tuple $\boldsymbol{x} = (x_1, \ldots, x_n)$ and I a subset of the indices, we write $\boldsymbol{x}_I$ for the tuple obtained from $\boldsymbol{x}$ by omitting the variables indexed by the elements of I. We write $\boldsymbol{x}_i$ for $\boldsymbol{x}_{\{i\}}$ and $\boldsymbol{x}_{ij}$ for $\boldsymbol{x}_{\{i,j\}}$, etc.
- Let $s(\boldsymbol{x})$ and $p(\boldsymbol{x})$ be the symmetric polynomials of degree 1 and 2, respectively, i.e.

$$s(\boldsymbol{x}) = \sum\nolimits_{i=1}^{n} x_i, \qquad p(\boldsymbol{x}) = \sum\nolimits_{1 \leq i < j \leq n} x_i x_j$$

  We often omit writing out the variables explicitly, so we write s_i for $s(\boldsymbol{x}_i)$, s_{ij} for $s(\boldsymbol{x}_{ij})$, etc. We understood $s \equiv 0$ on zero variables and $p \equiv 0$ on either 0 or 1 variable.

Thanks go to Jackie Barab from whom the second author first learned about the 2-color games [1]. We thanks Bjorn Poonen for referring [7] to us. We would also like to thank Thomas Rohwer and Zeev Rudnick for bringing [3] and [11], respectively, to our attention. Finally, we thank Michael Larsen for reading a draft of this article and giving us several valuable comments.

[1] They are used in teaching 3rd and 4th graders in California about probability

2 The 2-color games

As a warm-up, we analyze the 2-color games first. In this case, Equation (1) becomes $(x_1 - x_2)^2 - (x_1 + x_2) = 0$ and is easy to solve: let $m = x_2 - x_1 \geq 0$, and then $x_1 = m(m-1)/2$ and so $x_2 = (m+1)m/2$. This shows that

Theorem 2.1. *The 2-color fair games are pairs of consecutive triangular numbers. In particular, $\mathcal{C}_2$ is the set of triangular numbers.*

Using Theorem 2.1, a few simple computations tell us the number of fair 2-color games of norm bounded by a given number.

Corollary 2.2. *For $k \geq 0$,*

1. $|\mathcal{F}_2(k)| = [\sqrt{r(k)}] + 1$. *Hence $|\mathcal{F}_2(k)|$ is asymptotic to $2^{1/4}\sqrt{k}$.*
2. $|\mathcal{C}_2(k)| = [r(\sqrt{k})]$. *Hence $|\mathcal{C}_2(k)|$ is asymptotic to $\sqrt{2k}$.*

Here $r(k) = (-1 + \sqrt{1 + 8k^2})/2$ and $[x]$ is the largest integer $\leq x$.

3 Solving Equation (1)

It would be nice to know in advance that Equation (1) is solvable. The following simple observation tells us just that.

Theorem 3.1. *There are infinitely many faithful n-color fair games.*

Proof. Regarding the polynomial F_n in (1) as a quadratic in x_k, we have

$$F_n(\boldsymbol{x}) = x_k^2 - (2s_k + 1)x_k + F_{n-1}(\boldsymbol{x}_k). \tag{2}$$

Thus if $(a_1, \dots, a_k, \dots, a_n)$ is a solution then so is $(a_1, \dots, b_k, \dots, a_n)$ where $b_k = 2s_k(\boldsymbol{a}) + 1 - a_k$. In particular, if $(a_1, \dots, a_{n-1})$ is a fair game, then so are $(a_1, \dots, a_{n-1}, 0)$ and $(a_1, \dots, a_{n-1}, 1 + 2\sum_{i<n} a_i)$. The latter game is faithful if $(a_1, \dots, a_{n-1})$ is. Since there are infinitely many faithful 2-color fair games (Theorem 2.1), the theorem follows by induction on n. □

Since the two roots of (2) sum to $2s_k + 1$, they can be expressed as $s_k + m + 1$ and $s_k - m$ for some $m \geq 0$. Thus solving

$$\begin{aligned} (s_k + m + 1)(s_k - m) = F_{n-1}(\boldsymbol{x}_k) &= s_k^2 - s_k - 4p_k \\ 2s_k + 4p_k &= m^2 + m \end{aligned} \tag{3}$$

will solve (1) and vice versa. After adding $1 + 4s_{ijk}^2 + 2s_{ijk} - 4p_{ijk}$ (i, j, k pairwise distinct) to both sides of (3), the left-hand side factorizes:

$$\begin{aligned} (2x_i + 2s_{ijk} + 1)(2x_j + 2s_{ijk} + 1) &= m^2 + m + 1 + 4s_{ijk}^2 + 2s_{ijk} - 4p_{ijk} \\ &= m^2 + m + 1 + 2(s_{ijk}^2 + s_{ijk} + \|\boldsymbol{x}_{ijk}\|^2) \end{aligned} \tag{4}$$

Equation (4) gives us a way to solve Equation (1): Denote by $J(\boldsymbol{x}_{ijk}, m)$ the right-hand side of (4). Choose $\boldsymbol{a} \in \mathbb{Z}^{n-3}$ and $m \geq 0$ arbitrarily[2]. According to (4), we can solve for x_i and x_j by factorizing the odd number $J(\boldsymbol{a}, m)$ into a product two odd numbers. By (2), we can then solve for x_k and hence a solution of (1). Moreover, it is clear from the discussion above that any solution of (1) arises from such a factorization. For the record, we have

Theorem 3.2. *Fix $n \geq 3$. For any $\boldsymbol{a} \in \mathbb{Z}^{n-3}$, $m \geq 0$ and $0 \leq b \leq c$ such that $J(\boldsymbol{a}, m) = (2b+1)(2c+1)$, the following are solutions to Equation* (1)*:*

$$\begin{array}{l}
(b - s(\boldsymbol{a}),\ c - s(\boldsymbol{a}),\ b + c + 1 - s(\boldsymbol{a}) + m,\ \boldsymbol{a}) \\
(b - s(\boldsymbol{a}),\ c - s(\boldsymbol{a}),\ b + c - s(\boldsymbol{a}) - m,\ \boldsymbol{a}) \\
(-(c+1) - s(\boldsymbol{a}),\ -(b+1) - s(\boldsymbol{a}),\ -(b + c + s(\boldsymbol{a}) + 1) + m,\ \boldsymbol{a}) \\
(-(c+1) - s(\boldsymbol{a}),\ -(b+1) - s(\boldsymbol{a}),\ -(b + c + s(\boldsymbol{a}) + 2) - m,\ \boldsymbol{a})
\end{array}$$

Moreover, up to a permutation every solution of Equation (1) *is in one of these forms.*

There is a less tricky way to derive Equation (4). We give the idea here but leave the details to the reader. Equation (3) can be viewed as a curve on the $x_i x_j$-plane (i, j, k pairwise distinct). One can express the curve having an integral point by expressing that the corresponding quadratic in x_i is solvable in terms of $\boldsymbol{x}_{ijk}$ and $d := x_j - x_i$. The expression $1 + 4s_{ijk}^2 + 2s_{ijk} - 4p_{ijk}$ then flows out naturally.

Even though the method above solves Equation (1), it is unclear which choice of the parameters will produce fair games. For example, $J(2, 3) = 33$ does not produce any 4-color fair game. We will take up this issue in the next section.

4 Solutions as a graph

Starting from a solution $(a_1, \ldots, a_n)$ of Equation (1), we obtain another one by replacing a_k with $b_k := 2\sum_{i \neq k} a_i + 1 - a_k$ (see Theorem 3.1). This suggests that we can view $\mathcal{S}_n$ as a graph by putting an edge between two elements of $\mathcal{S}_n$ if they differ at only one coordinate[3]. An immediate question would be: can we generate every fair game from some fixed game, say the zero game? In other words, is $\mathcal{F}_n$ connected as a graph? We have seen that $\mathcal{F}_2$ is connected (Theorem 2.1) and we will show that the same is true for $\mathcal{F}_3$. However, $\mathcal{F}_n$ fails to be connected for $n \geq 4$.

Let us begin with a crucial observation. For any $\boldsymbol{a} \in \mathcal{S}_n$ and any three pairwise distinct indices i, j, k, according to (4), for some $m \geq 0$,

$$(2s_{jk}(\boldsymbol{a}) + 1)(2s_{ik}(\boldsymbol{a}) + 1) = 2(s_{ijk}(\boldsymbol{a})^2 + s_{ijk}(\boldsymbol{a}) + \|\boldsymbol{a}_{ijk}\|^2) + m^2 + m + 1.$$

Since the right-hand-side is always positive, we conclude that

[2]When $n = 3$, we only need to choose m.

[3]Incidentally, this is the same graph structure that was put on the solutions on the Markoff Equation [5, 9].

Proposition 4.1. *For any $\boldsymbol{a} \in \mathcal{S}_n$, the numbers $2s_{ij}(\boldsymbol{a})+1$ $(1 \le i < j \le n)$ all have the same sign. In particular, the coordinates of an element of $\mathcal{S}_3$ are either all non-negative or all negative.*

We define the *sign* of $\boldsymbol{a}$ as the common sign of the $2s_{ij}(\boldsymbol{a})+1$ $(1 \le i < j \le n)$. Note that it is the same of the sign of $s(\boldsymbol{a})+1$ since $\sum_{i<j} s_{ij} = \binom{n-1}{2}s$. Let $\mathcal{S}_n^+$ and $\mathcal{S}_n^-$ be the sets of positive and negative elements of $\mathcal{S}_n$, respectively. Since any two neighbors in $\mathcal{S}_n$ share $n-1$ coordinates, they must have the same sign, therefore

Proposition 4.2. *$\mathcal{S}_n^+$ and $\mathcal{S}_n^-$ are disjoint union of components of $\mathcal{S}_n$.*

Our next result shows how height varies among neighbors.

Proposition 4.3. *At most one neighbor of any vertex of $\mathcal{S}_n$ can have a smaller height. Moreover, any two neighbors must have different height.*

Proof. Fix any $\boldsymbol{a} \in \mathcal{S}_n$. Let $\boldsymbol{b}_k = (a_1, \ldots, b_k, \ldots, a_n)$ where $b_k = 2s_k(\boldsymbol{a})+1-a_k$ $(1 \le k \le n)$ be its neighbors. Rearranging the coordinates if necessary, we assume $a_1 \le a_2 \le \cdots \le a_n$.
Case 1: $\boldsymbol{a} \in \mathcal{S}_n^+$. Then for $k \ne n$,

$$b_k = 2s_{kn}(\boldsymbol{a}) + (a_n - a_k) + a_n + 1 > a_n \ge a_k$$

so $s(\boldsymbol{b}_k) > s(\boldsymbol{a}) \ge 0$.

Case 2: $\boldsymbol{a} \in \mathcal{S}_n^-$. Then for $k \ne 1$,

$$b_k = 2s_{1k}(\boldsymbol{a}) + (a_1 - a_k) + a_1 + 1 < a_1 \le a_k$$

so $s(\boldsymbol{b}_k) < s(\boldsymbol{a}) < 0$.
This completes the proof of the first statement since in both cases we have $\mathrm{ht}(\boldsymbol{b}_k) > \mathrm{ht}(\boldsymbol{a})$ for all but perhaps one k. The second statement follows readily from the fact that each $a_k + b_k$ is an odd number. □

We say that a vertex of $\mathcal{S}_n$ is a *root* if all its neighbors have a greater height. We would like to point out that replacing height by norm in the definition of root will yield the same concept since Equation (1) can be rewritten as

$$\binom{s+1}{2} = \|\boldsymbol{x}\|^2. \tag{5}$$

Theorem 4.4. *The components of $\mathcal{S}_n$ are rooted trees.*

Proof. Proposition 4.3 implies that for any subgraph H of $\mathcal{S}_n$, a vertex of maximal height in H cannot have two neighbors in H. This shows that $\mathcal{S}_n$ must be acyclic. Moreover, every component of $\mathcal{S}_n$ has a unique vertex of minimal height. If not, take a path with two vertices of minimal height as endpoints. Since neighbors in $\mathcal{S}_n$ have different heights, the path has length at least 2 but then a vertex of maximal height in the path will have two neighbors, a contradiction. □

Fair games are positive solutions of (1) and yet a positive solution, for example (-1,1,2,2), need not even represents a game. However, for any $\boldsymbol{a} \in \mathcal{S}_n^+$, a neighbor of $\boldsymbol{a}$ greater in height will have the different coordinate non-negative (see the proof of Proposition 4.3). Thus by going up in height along any branch, we see that

Proposition 4.5. *Every component of $\mathcal{S}_n^+$ contains fair games.*

Similarly, by going down in height, we see that every component of $\mathcal{S}_n^-$ contains solutions with all negative coordinates.

Our next goal is to locate the roots of $\mathcal{S}_n$. Once this is achieved, we will have an effective way of generating all fair games since each of them is connected to some positive root.

Proposition 4.6. *Suppose $\boldsymbol{r} \in \mathcal{S}_n^+$ ($\mathcal{S}_n^-$, resp.) is a root and i, j, k pairwise distinct where k is the index of a maximal (minimal, resp.) coordinate of $\boldsymbol{r}$. Then $\boldsymbol{r}$ is obtained from a factorization of $J(\boldsymbol{r}_{ijk}, m)$ for some $0 \leq m \leq B(\boldsymbol{r}_{ijk})$ where $B(\boldsymbol{r}_{ijk})$ is an explicit bound give in terms of $\boldsymbol{r}_{ijk}$.*

Proof. We argue for S_n^+ only. The proof for S_n^- is similar. According to (4), for some $m \geq 0$,

$$(2s_{ik} + 1)(2s_{jk} + 1) = 2(s_{ijk}^2 + s_{ijk} + \|\boldsymbol{r}_{ijk}\|^2) + m^2 + m + 1.$$

Since $\boldsymbol{r}$ is a root, r_k is the smaller root of Equation (2), i.e. $r_k = s_k - m$. And since $r_k \geq r_\ell$ ($\ell \neq k$), so $s_{\ell k} \geq s_{\ell k} + r_\ell - r_k = m$. Thus

$$\begin{aligned} (2m+1)^2 &\leq 2(s_{ijk}^2 + s_{ijk} + \|\boldsymbol{r}_{ijk}\|^2) + m^2 + m + 1 \\ 3m^2 + 3m &\leq 2(s_{ijk}^2 + s_{ijk} + \|\boldsymbol{r}_{ijk}\|^2). \end{aligned}$$

That means $0 \leq m \leq B(\boldsymbol{r}_{ijk})$ where $B(\boldsymbol{r}_{ijk})$ expresses the larger root of the quadratic $3x^2 + 3x - 2(s_{ijk}^2 + s_{ijk} + \|\boldsymbol{r}_{ijk}\|^2)$. □

Let us summarize how to find the roots of $\mathcal{S}_n$: for each $\boldsymbol{a} \in \mathbb{Z}^{n-3}$, we compute the finite set consisting of those solutions given by the factorizations of $J(\boldsymbol{a}, m)$ where $0 \leq m \leq B(\boldsymbol{a})$. We then check which element in this finite set is a root. While Proposition 4.6 guarantees that every root of $\mathcal{S}_n$ can be found this way, our next result shows that we do have to check for every $\boldsymbol{a} \in \mathbb{Z}^3$.

Proposition 4.7. *Every $(n-3)$-tuple of integers can be extended to a root in $\mathcal{S}_n$. More precisely, for any $\boldsymbol{a} \in \mathbb{Z}^{n-3}$, n-tuples*

$$\begin{aligned} \boldsymbol{r}_+ &:= (s(\boldsymbol{a})^2 + \|\boldsymbol{a}\|^2,\ s(\boldsymbol{a})^2 + \|\boldsymbol{a}\|^2,\ -s(\boldsymbol{a}),\ \boldsymbol{a}) \quad \textit{and} \\ \boldsymbol{r}_- &:= (-(s(\boldsymbol{a})+1)^2 - \|\boldsymbol{a}\|^2,\ -(s(\boldsymbol{a})+1)^2 - \|\boldsymbol{a}\|^2,\ -(s(\boldsymbol{a})+1),\ \boldsymbol{a}) \end{aligned}$$

are a positive and a negative root of $\mathcal{S}_n$, respectively.

Proof. First by Theorem 3.2, they are solutions corresponding to the trivial factorization of $J(\boldsymbol{a}, 0)$ (in the notation there, $b = s(\boldsymbol{a})^2 + s(\boldsymbol{a}) + \|\boldsymbol{a}\|^2$ and $c = 0$.). Clearly, $\boldsymbol{r}_+$ is positive while $\boldsymbol{r}_-$ is negative. The neighbor of $\boldsymbol{r}_+$ obtained by varying its largest coordinate $s(\boldsymbol{a})^2 + \|\boldsymbol{a}\|^2$ has an even larger coordinate, namely $s(\boldsymbol{a})^2+\|\boldsymbol{a}\|^2+1$. Thus $\boldsymbol{r}_+$ is indeed a root (see the proof of Theorem 4.3). A similar argument show that $\boldsymbol{r}_-$ is a root as well. □

Since each component of $\mathcal{S}_n$ has exactly one root, an immediate consequence of Proposition 4.7 is that

Theorem 4.8. *For $n \geq 4$, $\mathcal{S}_n^+, \mathcal{S}_n^-$ each has infinitely many components.*

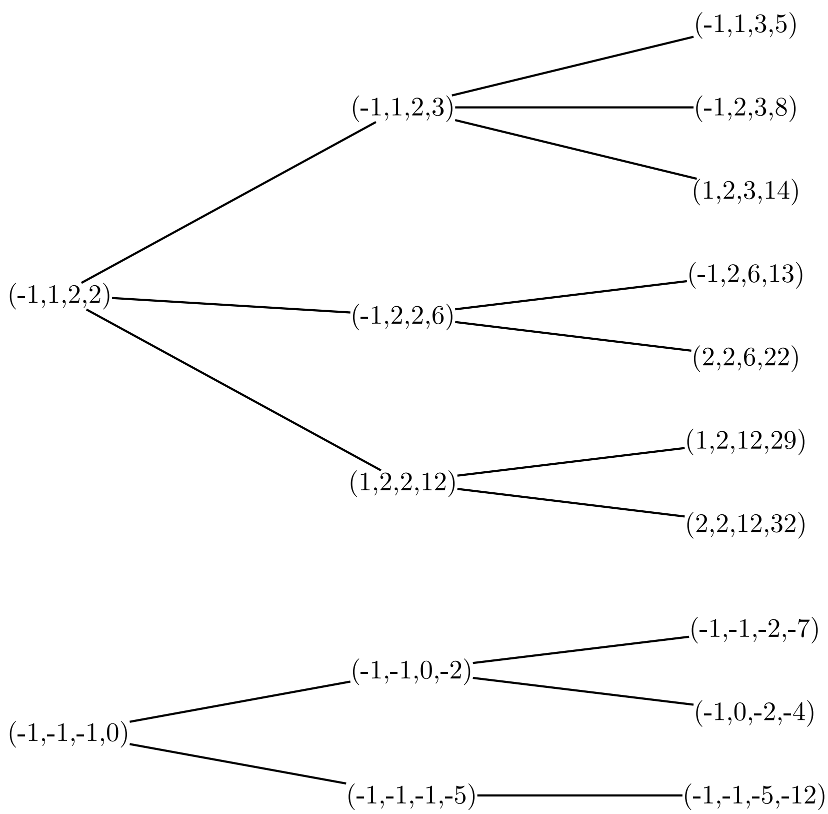

Figure 1: A positive tree and a negative tree in $\mathcal{S}_4$

On the contrary, by examining the proof of Proposition 4.6, one readily checks that $(0, 0, 0)$ is the only root of $\mathcal{S}_3^+$. By Theorem 3.2, the map $\boldsymbol{a} \mapsto -(\boldsymbol{a}+\mathbf{1})$ where $\mathbf{1} = (1, 1, 1)$ is a graph isomorphism between $\mathcal{S}_3^+$ and $\mathcal{S}_3^-$. Moreover, every element of $\mathcal{S}_3^+$ is actually a fair game according to Proposition 4.1. Thus,

Theorem 4.9. *$\mathcal{S}_3^+$ and $\mathcal{S}_3^-$ are the two components of $\mathcal{S}_3$. Moreover, $\mathcal{S}_3^+ = \mathcal{F}_3$.*

A straight-forward computation shows that every vertex of S_3 with distinct coordinates has two distinct children (i.e. neighbors with a bigger norm). Moreover, each of its children also has distinct coordinates. Hence,

Theorem 4.10. *The non-trivial 3-color fair games form an infinite full binary tree with* $(0, 1, 3)$ *as root.*

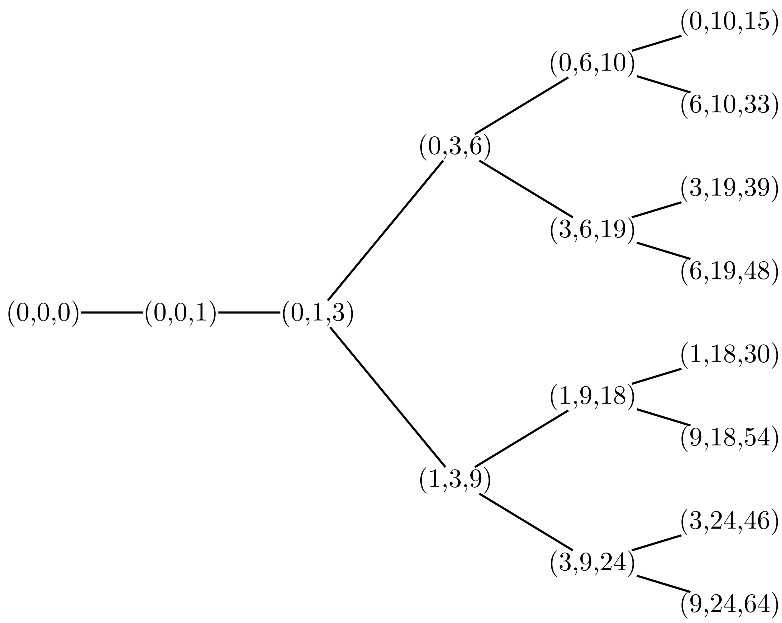

Figure 2: Part of $\mathcal{F}_3$

5 The set $\mathcal{C}_n$

Proposition 5.1. *For* $n \geq 4$, $\mathcal{C}_n$ *is the set of non-negative integers.*

Proof. For any $a \geq 0$, let $\boldsymbol{a}$ be the $(n-3)$-tuple with all coordinates equal a. Then the child $(4(s(\boldsymbol{a})^2 + \|a\|^2) + 3s(\boldsymbol{a}) + 1, s(\boldsymbol{a})^2 + \|\boldsymbol{a}\|^2, s(\boldsymbol{a})^2 + \|\boldsymbol{a}\|^2, \boldsymbol{a})$ of the positive root $\boldsymbol{r}_+$ in Proposition 4.7 is a fair game with a as a coordinate. Incidentally, this also shows that for $n \geq 4$, every natural number is a coordinate of some faithful n-color fair game. $\square$

This leaves us only $\mathcal{C}_3$ to study. It turns out that our analysis of $\mathcal{C}_3$ will yield another way of finding the 3-color fair games (Theorems 5.3 and 5.4). First, note that $\mathcal{C}_3$ is the set of $c \geq 0$ such that the curve defined by

$$(x_1 - x_2)^2 - (2c+1)(x_1 + x_2) + c(c-1) = 0 \tag{6}$$

has a non-negative point. Arguing mod 2, one sees that any integral point on the parabola

$$u^2 - (2c+1)v + c(c-1) = 0 \tag{7}$$

must have coordinates with the same parity. Thus, the transformation $u = x_1 - x_2,\ v = x_1 + x_2$ is a 1-to-1 correspondence between the integral points of these curves. Moreover, those (x_1, x_2)'s with $x_1, x_2 \geq 0$ correspond to the (u, v)'s with $u \leq v$. However, the inequality is automatic:

Proposition 5.2. *Solutions of Equation* (7) *are of the form*

$$(u, v) = \left(u, \quad \frac{u^2 + c(c-1)}{2c+1}\right)$$

where $u^2 \equiv -c(c-1) \mod (2c+1)$. *In particular,* (7) *is solvable if and only if* $-c(c-1)$ *is a quadratic residue mod* $(2c+1)$. *Moreover,* $|u| \leq v$ *for every integral solution* (u, v).

Proof. The first statement is clear by considering Equation (7) mod $(2c+1)$. Note that for any u,

$$-\frac{8c+1}{4} \leq u^2 \pm (2c+1)u + c(c-1).$$

Since $c \geq 0$,

$$\pm u - \frac{8c+1}{8c+4} \leq \frac{u^2 + c(c-1)}{2c+1}.$$

So in particular

$$|u| \leq \frac{u^2 + c(c-1)}{2c+1}$$

if both sides are integers. Therefore, $|u| \leq v$ for any integral solution of (7). □

Solving for x_1, x_2 in terms of u, v yields a parametrization of $\mathcal{F}_3$.

Theorem 5.3. *The 3-color fair games are of the form*

$$\left(\frac{u^2 + (2c+1)u + c(c-1)}{2(2c+1)}, \quad \frac{u^2 - (2c+1)u + c(c-1)}{2(2c+1)}, \quad c\right) \tag{8}$$

where $c \in \mathcal{C}_3$ *and* $u^2 \equiv -c(c-1) \mod (2c+1)$.

For $i = -1, 0, 1$, let P_i be the set of primes that are congruent to i mod 3. Let $P_i^{\geq 0}$ be the set of natural numbers whose prime factors are all in P_i. With this notation, we have

Theorem 5.4. $\mathcal{C}_3 = \{c \colon c \geq 0,\ 2c+1 \in P_1^{\geq 0} \cup 3P_1^{\geq 0}\}$.

Proof. By Proposition 5.2 and the discussion preceding it, $\mathcal{C}_3$ is the set of all c such that $-c(c-1)$ is a quadratic residue mod $2c+1$. Since $2c+1$ is odd, $-c(c-1)$ and $-4c(c-1)$ are either both squares or both non-squares mod $2c+1$. Note that $-4c(c-1)$ is congruent to -3 mod $2c+1$. Therefore $c \in \mathcal{C}_3$ if and only if -3 is a square mod $2c+1$. This condition is equivalent to: for every prime p, -3 is a square mod p^{v_p} where v_p is the exponent of p in $2c+1$. That means $v_3 = 0$ or 1 and -3 is a quadratic residue mod p for $p \neq 3$. The law of quadratic reciprocity then yields

$$\left(\frac{-3}{p}\right) = \left(\frac{-1}{p}\right)\left(\frac{3}{p}\right) = (-1)^{\frac{1}{2}(p-1)}\left(\frac{3}{p}\right) = \left(\frac{p}{3}\right).$$

So the last condition is equivalent to $p \equiv 1 \mod 3$. This completes the proof. ∎

With the binary tree structure of $\mathcal{F}_3$ in mind, it is tempting to establish some relation between the equation for fair 3-color games with the Markoff equation (see [5] and [9] for more information on the Markoff equation). Unfortunately, we could not find any, however the naive analog of the Markoff conjecture (that the solution is uniquely determined by its largest coordinate) fails in our case, more precisely:

Theorem 5.5. *For $c \in \mathcal{C}_3$ and $c > 1$, the number of fair games with c as the largest coordinate is 2^{m-1} where m is the number of distinct prime factors of $2c+1$ other than 3.*

Proof. The coordinates of nontrivial 3-color fair games are distinct, so it follows from (8) that c is the largest coordinates if and only if

$$c \geq \frac{u^2 + (2c+1)|u| + c(c-1)}{2(2c+1)} + 1.$$

A simple calculation shows that the above inequality is equivalent to $|u| \leq c$. Hence, proving the theorem boils down to counting the number of solutions to the congruence $u^2 \equiv -c(c-1) \mod (2c+1)$, or equivalently $u^2 \equiv -3 \mod (2c+1)$ in a complete set of representatives $(-c \leq u \leq c)$. By Theorem 5.4, the prime factorization of $2c+1$ is of the form $3^{v_0} p_1^{v_1} \cdots p_m^{v_m}$ where $v_0 = 0$ or 1 and $p_j \equiv 1 \mod 3$ $(1 \leq j \leq m)$. Note that $m \geq 1$ since $c > 1$. The Chinese reminder theorem yields a 1-to-1 correspondence between solutions of $u^2 \equiv -3 \mod (2c+1)$ and the solutions of the system $u^2 \equiv -3 \mod p_j^{v_j}$ $(1 \leq j \leq m)$ together with the congruence $u^2 \equiv -3 \equiv 0 \mod 3$ depending on whether $v = 1$ or not. But last congruence has only one solution hence its presence will not affect the total number of solutions which is the product of the number of solutions of each congruence in the system. Since $p_j \neq 3$ $(1 \leq j \leq m)$, the number of solutions for $u^2 \equiv -3 \mod p_j^{v_j}$ is the same as that for $u^2 \equiv -3 \mod p_j$ which is precisely two since $p_j \equiv 1 \mod 3$. Therefore, we conclude that the system has exactly 2^m solutions. Finally, it is clear from (8) that $\pm u$ give rise to the same fair game except with the first two coordinates permuted. Thus, up to permutation of coordinates, there are $2^m/2 = 2^{m-1}$ fair games with c as the largest coordinate. $\square$

We conclude this section by computing the natural density of $\mathcal{C}_3$. The *natural density* of a set of natural numbers A is defined to be $\lim_{k\to\infty} |A(k)|/k$ whenever the limit exists. Statement (2) of Corollary 2.2 states that $\mathcal{C}_2$ has density zero. The description of $\mathcal{C}_3$ in Theorem 5.4 allows us to show that the same phenomenon occurs in the case $n = 3$ as well.

Theorem 5.6. *The natural density of $\mathcal{C}_3$ is zero.*

Proof. The map $a \mapsto 2a+1$ is a bijection between $\mathcal{C}_3$ and $P_1^{\geq 0} \cup 3P_1^{\geq 0}$. Therefore, the density of $\mathcal{C}_3$ is twice the density of $P_1^{\geq 0} \cup 3P_1^{\geq 0}$ if the latter exists. Since P_1 and $3P_1$ are disjoint and the density of $3P_1$ (if it exists) is one-third that of P_1, it suffices to show that P_1 has density 0. By Lemma 11.8 in [3], it is enough to show that the series $\sum_{p\in P_{-1}} 1/p$ diverges. But this assertion is an immediate consequence of Dirichlet Theorem of primes in arithmetic progressions [12, Chapter VI Theorem 2]) which, in particular, asserts that

$$\lim_{s\to 1^+} \frac{\sum_{p\in P_{-1}} 1/p}{\log 1/(s-1)} = \frac{1}{\phi(3)} = \frac{1}{2}.$$

$\square$

6 Asymptotic Behavior

Comparing to the binary case, determining the asymptotic behavior of $|\mathcal{F}_n(k)|$ ($n \geq 3$) seems to be much harder problem. In fact, we will only show that $|\mathcal{F}_3(k)|$ is $\Theta(k)$. Our strategy is to relate the equation for 3-color fair games to the Lorentzian form $L_3(w) = w_1^2 + w_2^2 - w_3^2$ then apply the results from [11] (for more information on distribution of integral points on affine homogenous varieties, see [6] and [2]). Let W be the set of integral solutions of $L_3(w) = -3$.

Lemma 6.1. *For any $(w_1, w_2, w_3) \in \mathsf{W}$, we have*

1. $|w_3| > |w_1|, |w_2|$.
2. *exactly one of the w_i is odd; moreover, it must be either w_1 or w_2.*
3. *If w_2 is the odd coordinate that $w_1 + w_3 \equiv w_1 - w_3 \equiv 2 \mod 4$.*
4. $w_1 \neq w_2$.

Proof. The first statement is immediate. The second statement is clear by arguing mod 8. Since w_2 is odd, $w_1^2 - w_3^2 \equiv 0 \mod 4$. Since w_3 is even, $w_1 + w_3 \equiv w_1 - w_3 \mod 4$. They must be both congruent to 2 mod 4; otherwise $w_1^2 - w_3^2 \equiv 0 \mod 8$ making -3 a square mod 8, contradiction. This establishes the third statement. The last statement is true since $2w_1^2 - w_3^2 = -3$ is not solvable mod 3. $\square$

Let $\approx$ be the equivalence relation on W identifying the elements (w_1, w_2, w_3) and (w_2, w_1, w_3). It follows from Lemma 6.1 (4) that the canonical map from W to $\mathsf{W}/\approx$ is 2-to-1. If we identify the equivalence classes with those elements of W with an odd second coordinate, then

Proposition 6.2. *The map given by*

$$w_1 = 2(x_2 - x_3), \quad w_2 = 2(x_1 - x_2 - x_3) - 1, \quad w_3 = 2(x_2 + x_3 + 1) \tag{9}$$

is a 1-to-1 correspondence between $\mathcal{S}_3$ and $\mathsf{W}/\approx$. Moreover, if the coordinates of the elements of $\mathcal{S}_3$ are listed in ascending order, then elements of $\mathcal{F}_3$ correspond to those elements of $\mathsf{W}/\approx$ with $w_1, w_2 \leq 0$ and $w_3 \geq 0$.

Proof. The rational inverse of the map in (9) is given by

$$x_1 = \frac{w_2 + w_3 - 1}{2}, \quad x_2 = \frac{w_1 + w_3 - 2}{4}, \quad x_3 = \frac{w_3 - w_1 - 2}{4}. \tag{10}$$

By Lemma 6.1 (2) and (3), it actually preserve integral points. This establishes the map in (9) is a 1-to-1 correspondence between $\mathcal{S}_3$ and $\mathsf{W}/\approx$. Moreover, the images of elements of $\mathcal{F}_3$ (as ascending triples) under (9) clearly satisfy $w_1, w_2 \leq 0$ and $w_3 \geq 0$. Conversely, by Lemma 6.1 (1), $|w_3| \geq |w_2| + 1$ and $|w_3| \geq |w_1| + 2$ since both w_1 and w_3 are odd (Lemma 6.1 (2)). Therefore, if $(w_1, w_2, w_3) \in \mathsf{W}$ with $w_1, w_2 \leq 0$ and $w_3 \geq 0$ then the corresponding (x_1, x_2, x_3) is in $\mathcal{F}_3$. $\square$

Theorem 6.3. *There exists positive constants c_1, c_2 such that $c_1 k \leq |\mathcal{S}_3(k)| \leq c_2 k$ for $k \gg 0$, i.e. $|\mathcal{S}_3(k)| = \Theta(k)$. Similarly, $|\mathcal{F}_3(k)| = \Theta(k)$.*

Proof. The idea is simple: the sphere of radius k centered at the origin maps to an ellipsoid (centered at $(0, -1, 2)$) under (9). For k sufficiently large, it is enveloped between spheres centered at the origin. Clearly, the radii of these spheres can be chosen as linear functions of the length of the axes of the ellipsoid which are in turn linear in k since the transformation given in (9) is affine. So by Proposition 6.2, the proof is complete once we show that $|\mathsf{W}(k)|$ is asymptotic to a linear function in k. And this last statement follows from Formula (3) in [11] which asserts that $|\mathsf{W}_3(k)| \sim (4\sqrt{6}/3)k$. $\square$

Remark 6.4.

(i) Since the maps and the equations are all explicit, one can provide the constants c_1 and c_2 in Theorem 6.3 explicitly. However, we will leave the computation for the interested readers.

(ii) In a sense, one gets a cleaner result without extra efforts if one is satisfied by counting the number of solutions inside the ellipsoids that are the images of spheres under the map given in (10). To be more precise, let $\mathcal{S}'_3(k)$ and $\mathcal{F}'_3(k)$ be the set elements of $\mathcal{S}_3$ and $\mathcal{F}_3$ inside the image of the sphere of radius k centered at the origin under the map given in (10). Then again using Formula (3), Table (1) in [11] and Proposition 6.2, one gets

$$|\mathcal{S}'_3(k)| = \frac{1}{2}|\mathsf{W}(k)| \sim \frac{2\sqrt{6}}{3}k, \quad |\mathcal{F}'_3(k)| \sim \frac{6}{8}\frac{1}{2}|\mathsf{W}(k)| = \frac{\sqrt{6}}{2}k.$$

(iii) Here is how we arrive to the map in (9): There is a general method of solving quadratic Diophantine equations given by Grunewald and Segal in [7] and [8][4]. The first step transforms the fair game equation into the equivalent system

$$Q_n(\boldsymbol{z}) = -n(n-2), \quad z_i \equiv 1 \mod 2(n-2). \qquad (1 \leq i \leq n) \tag{11}$$

where Q_n is the quadratic form in n-variables with diagonal entries 1 and off-diagonal entries -1. When $n = 3$, one checks readily that congruences in (11) are implies by $Q_3(\boldsymbol{z}) = -3$. And the Lorentzian form $L_3(\boldsymbol{w})$ is obtained by diagonalizing Q_3. Taking the composition of these transformations yields the map in (9).

(iv) In the ternary case, the description of the solution set given by Grunewald and Segal's method relates quite beautifully to ours. We encourage the reader to pursuit their original papers (see [4] for the necessary backgrounds). Just to give a little enticement, let us remark that $\mathcal{S}_3$ will correspond to a single orbit under the integral orthogonal group of a suitable quadratic form. While $\mathcal{S}_3^+$ (i.e. $\mathcal{F}_3$) and $\mathcal{S}_3^-$ will correspond to two orbits of a subgroup of the orthogonal group.

(v) Using the same idea to study $|\mathcal{F}_n(k)|$ for $n \geq 4$ becomes more problematic. First, it is not clear to us how to take into account of the congruences in (11). Moreover, even though Q_n and the Lorentzian form in n-variables have the same signature, namely $n-2$, the transformation taking one to the other in general does not preserve integral points.

7 Odds and Ends

The last section is dedicated to various results that do not quite fit in previous sections.

Proposition 7.1. *The sum of coordinates of any element of S_n is congruent to either 0 or 1 mod 4.*

Proof. Equation (1) is simply $s^2 - s = 4p$. □

Proposition 7.2. *Every vertex in the connected component of $\mathbf{0}$ in $\mathcal{S}_n$ is congruent mod 3 to either $\mathbf{0}$ or $\boldsymbol{e}_j := (0, \ldots, 1, \ldots, 0)$ for some $1 \leq j \leq n$.*

[4] However, their method is not uniform in the number of variables.

Proof. The proposition is clearly true for $\mathbf{0}$. Now suppose it is true for every vertex of distance m from $\mathbf{0}$. Let $\boldsymbol{a}' = (a_1, \ldots, a_i', \ldots, a_n)$ be a vertex of distance $m+1$ from $\mathbf{0}$ and is adjacent to $\boldsymbol{a} = (a_1, \ldots, a_i, \ldots, a_n)$ which is of distance m from $\mathbf{0}$. By the induction hypothesis, $\boldsymbol{a}$ is congruent to either $\mathbf{0}$ or $\boldsymbol{e}_j$ mod 3 for some $1 \leq j \leq n$. Since $a_i + a_i' = 1 + 2s_i(\boldsymbol{a}) \equiv 1 - s_i(\boldsymbol{a}) \mod 3$, we have the following 3 cases

1. $\boldsymbol{a}' \equiv \boldsymbol{e}_i \mod 3$ if $\boldsymbol{a} \equiv \mathbf{0} \mod 3$,
2. $\boldsymbol{a}' \equiv \boldsymbol{e}_j \mod 3$ if $\boldsymbol{a} \equiv \boldsymbol{e}_j \mod 3$ and $i \neq j$, or
3. $\boldsymbol{a}' \equiv \mathbf{0} \mod 3$ if $\boldsymbol{a} \equiv \boldsymbol{e}_j \mod 3$ and $i = j$.

This establishes the proposition by induction. □

Recall that for $n \geq 4$ any natural number, in particular 2, can be a coordinate of an n-color fair game (Proposition 5.1). Thus, Proposition 7.2 gives another proof of the fact that $\mathcal{F}_n$ is disconnected for $n \geq 4$.

The nontrivial 3-color fair games form a full binary tree with the nontrivial 2-color fair games embeds as a branch. It is easy to see that among the nodes of distance k from the root $(0, 1, 3)$, the one with the smallest norm is $\left(0, \binom{k+2}{2}, \binom{k+3}{2}\right)$ and the one with the largest norm is (m_k, m_{k+1}, m_{k+2}) where (m_i) is the sequence defined recursively by $(m_0, m_1, m_2) = (0, 1, 3)$ and $m_{i+3} = 2(m_{i+1} + m_{i+2}) + 1 - m_i$ for $i \geq 0$. It turns out that the m_i's have an intimate relation with the Fibonacci numbers.

Proposition 7.3. *Let f_i be the i-th Fibonacci number, then for any $k \geq 0$*

$$m_k = \begin{cases} f_k^2 & \text{if } k \text{ is odd} \\ f_k^2 - 1 & \text{if } k \text{ is even} \end{cases}$$

Proof. The Fibonacci numbers are defined inductively by $f_0 = f_1 = 1$, and $f_{i+2} = f_i + f_{i+1}$ $(i \geq 0)$. Note that

$$\begin{aligned} 2f_{i+1}f_{i+2} &= f_{i+1}(f_{i+2} + f_i + f_{i+1}) = f_{i+1}(f_{i+2} + f_i) + f_{i+1}^2 \\ &= (f_{i+2} - f_i)(f_{i+2} + f_i) + f_{i+1}^2 \\ &= f_{i+2}^2 + f_{i+1}^2 - f_i^2. \end{aligned}$$

Therefore,

$$\begin{aligned} f_{i+3}^2 &= (f_{i+1} + f_{i+2})^2 = f_{i+1}^2 + 2f_{i+1}f_{i+2} + f_{i+2}^2 \\ &= 2f_{i+1}^2 + 2f_{i+2}^2 - f_i^2. \end{aligned}$$

The proposition now follows from an easy induction. The base cases are immediate. Suppose $k \geq 3$ and the proposition is true for all $0 \leq i < k$. When k is odd, we have

$$\begin{aligned}
m_k &= 2(m_{k-2} + m_{k-1}) + 1 - m_{k-3} \\
&= 2(f_{k-2}^2 + f_{k-1}^2 - 1) + 1 - (f_{k-3}^2 - 1) \\
&= 2f_{k-2}^2 + 2f_{k-1}^2 - f_{k-3}^2 \\
&= f_k^2.
\end{aligned}$$

A similar computation shows that $m_k = f_k^2 - 1$ when k is even. □

We end the article with another curious "by-product" of our results.

Proposition 7.4. *For $m \geq 0$,*

(1) $m^2 + m + 1 \in P_1^{\geq 0} \cup 3P_1^{\geq 0}$.

(2) $2f_m^2 - (-1)^m \in P_1^{\geq 0} \cup 3P_1^{\geq 0}$.

Proof. Triangular numbers $m(m+1)/2$ $(m \geq 0)$ appear as coordinates of the 2-color (Theorem 2.1) and hence 3-color fair games. By Theorem 5.4, we have $m^2+m+1 = 2(m(m+1)/2)+1 \in P_1^{\geq 0} \cup 3P_1^{\geq 0}$. Similarly, Statement (2) follows from Proposition 7.3 and Theorem 5.4. □

References

[1] Baker, Alan. *A concise introduction to the theory of numbers.* Cambridge University Press, Cambridge, 1984.

[2] Borovoi, Mikhail. On representations of integers by indefinite ternary quadratic forms. J. Number Theory 90 (2001), no. 2, 281–293.

[3] Burris, Stanley N. *Number theoretic density and logical limit laws.* Mathematical Surveys and Monographs, 86. American Mathematical Society, Providence, RI, 2001.

[4] Cassels, J. W. S. *Rational quadratic forms.* London Mathematical Society Monographs, 13. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], London-New York, 1978.

[5] Cusick, Thomas W.; Flahive, Mary E. The Markoff and Lagrange spectra. Mathematical Surveys and Monographs, 30. American Mathematical Society.

[6] Duke, W.; Rudnick, Z.; Sarnak, P. Density of integer points on affine homogeneous varieties. Duke Math. J. 71 (1993), no. 1, 143–179.

[7] Grunewald, Fritz J.; Segal, Daniel. How to solve a quadratic equation in integers. Math. Proc. Cambridge Philos. Soc. 89 (1981), no. 1, 1–5.

[8] Grunewald, Fritz J.; Segal, Daniel. Some general algorithms. I. Arithmetic groups. Ann. of Math. (2) 112 (1980), no. 3, 531–583.

[9] Andrew D. Pollington, William Moran (ed.). Number theory with an emphasis on the Markoff spectrum. Lecture Notes in Pure and Applied Mathematics, 147. Marcel Dekker, Inc., New York, 1993.

[10] Pong W.Y.; Stroker. R. Diophantine Equations of Matching Games II, in preparation.

[11] Ratcliffe, John G.; Tschantz, Steven T. On the representation of integers by the Lorentzian quadratic form. J. Funct. Anal. 150 (1997), no. 2, 498–525.

[12] Serre, J.-P. A course in arithmetic. Translated from the French. Graduate Texts in Mathematics, No. 7. Springer-Verlag, New York-Heidelberg, 1973.