Solitaire of Independence

Ville Salo[1][0000-0002-2059-194X] and Juliette Schabanel[2][0009-0002-3647-4892]

[1] University of Turku, Turku, Finland
[2] École Normale Supérieure PSL, Paris, France

Abstract. In this paper, we study a reversible process (more precisely, a groupoid/group action) resembling the classical 15-puzzle, where the legal moves are to "move the unique hole inside a translate of a shape S". Such a process can be defined for any finite subset S of a group, and we refer to such a process as simply "solitaire". We develop a general theory of solitaire, and then concentrate on the simplest possible example, solitaire for the plane $\mathbb{Z}^2$, and S the triange shape (equivalently, any three-element set in general position). In this case, we give a polynomial time algorithm that puts any finite subset of the plane in normal form using solitaire moves, and show that the solitaire orbit of a line of consecutive ones – the line orbit – is completely characterised by the notion of a so-called fill matrix. We show that the diameter of the line orbit, as a graph with edges the solitaire moves, is cubic. We show that analogous results hold for the square shape, but indicate some shapes (still on the group $\mathbb{Z}^2$) where this is less immediate.

Keywords: solitaire of independence · TEP subshift · subshift of finite type · bipermutive cellular automata

1 Introduction

In this paper, we introduce the *solitaire of independence*, which is a game played on subsets of a group, alternatively an abstract rewriting system, groupoid or group action. Specifically, if G is a discrete group, and $S \Subset G$ is finite, then on subsets of G we can play a variant of the classical "15-puzzle" by, for each $g \in G$, allowing a move from $x \subset G$ to $y \subset G$ if the difference between x and y is precisely that they lack a different element of gS (but both intersect gS in a set of size $|S| - 1$). Or in words, we are allowed to move a unique hole in any set gS.

For example, if $G = \mathbb{Z}^2$, and we consider the triangle shape $T = \{(0,1), (0,0), (1,0)\}$, then the action consists in arbitrarily permuting the three patterns depicted in Figure 1 at some coordinates where one of them appears in the pattern.

A simple illustration of a valid sequence of moves for this shape on subsets of the set $\{(a,b) \mid a, b \geqslant 0, a + b \leqslant 3\} \subset \mathbb{Z}^2$ is shown in Figure 2.

On the website [6], one can play the solitaire on $\mathbb{Z}^2$ with any small shape.

In the present paper, we develop a basic theory of solitaire processes, and illustrate it with a detailed look at the triangle shape, for which we can give a

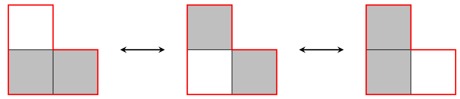

Fig. 1: The action of the triangle shape. Grey denotes 1, white denotes 0.

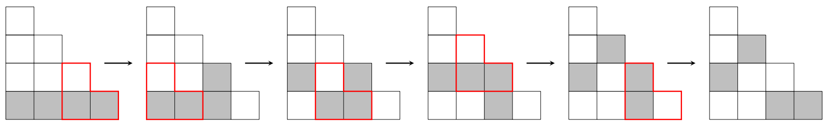

Fig. 2: A simple sequence of valid moves of triangle solitaire starting from the line $L_4 = \{(a, b) \mid a, b \geqslant 0, a + b \leqslant 3\}$.

full description of the family of sets reachable from a given set by playing the solitaire. In particular, for this example our main theorem is the following:

Theorem 1. *For the triangle solitaire process, for any finite set $P \Subset \mathbb{Z}^2$, there is a sequence of solitaire moves with cubic length in $|P|$ which puts P in normal form, and such a sequence can be found in polynomial time.*

By normal form, we mean that we can find a canonical representative of the orbit, such that we can compare orbits by comparing the representatives. The normal form we use is a disjoint (and "non-touching") union of lines with additional elements lined up on top.

We have implemented a variant of this algorithm (with a different idea but the same asymptotic complexity), and it is available in [9].

The general theory of solitaire is tightly connected with another process called the filling process described in Section 3.2, which is a simpler (confluent) process where we inductively construct the largest set that could possibly be involved in the solitaire game, by filling in the missing holes in sets gS (i.e. if $|gS \cap P| = |S| - 1$, then gS may be added to P).

Due to confluence, the filling process gives rise to a closure operator, and our primary interest is in understanding the connection between patterns that fill a particular set, and orbits in the solitaire process. Of specific interest is the case where in fact all sets that are minimal in cardinality, and fill a given set, are in the same solitaire orbit. We refer to this case as a *nice solitaire theory*. We do not know, at present, general conditions under which solitaire processes are nice. However, we obtain that the triangle shape, as well as the square shape $\{0, 1\}^2$, have nice solitaire theories on the group $\mathbb{Z}^2$.

We also explain the connection with *TEP subshifts* from [7], which are a generalization of spacetime subshifts of bipermutive cellular automata, and more generally *permutive subshifts* (which also generalize the polygonal subshifts of [2]). In particular, to any subshift we associate its "independent sets" and the sets "spanning" the contents of a given area (which are analogs of notions from linear algebra and matroid theory with the same names). We show that the solitaire process preserves these sets in the case of TEP and permutive subshifts. (The fact it acts on the independent sets gives the process its full name "solitaire of independence".)

The present paper is an extended version of the conference paper [10] where we presented the results about the triangle shape in isolation. The result here about the diameter of orbits is slightly sharper.

2 Conventions and basic definitions

For the most part, we give definitions as we go, but we list some general conventions here, and also collect some definitions that are used throughout.

We have $0 \in \mathbb{N}$. An *alphabet* is just a finite set A whose elements are called *letters* or *symbols*. Groups G are always countable (and discrete), with the main interest being in finitely-generated groups. By $\subset$ we mean $\subseteq$.

For a function $f : A \to B$ and $C \subset A$, write $f|_C$ for the restriction of f to C. For functions that are thought of as "patterns" (see below), we also write f_c for $f(c)$. If F is a family of functions (or patterns), then $F|_A = \{f|_A \mid f \in F\}$.

We use some conventions from formal language theory and combinatorics on words, in particular if A is a finite alphabet, A^* denotes the words on the alphabet (or elements of the free monoid), and concatenation of $u, v \in A^*$ is written simply as uv.

We use standard big-O notation: for $f, g : \mathbb{N} \to \mathbb{N}$, we have $g \in O(f)$ or "g is $O(f)$" (resp. $\Omega(f)$) if $g(n) \leqslant Mf(n)$ (resp. $\geqslant$) for large n and a constant M. We have $\Theta(f) = O(f) \cap \Omega(f)$.

If X is a set, *closure operator* is a map $\tau : \mathcal{P}(X) \to \mathcal{P}(X)$ which is *extensive* meaning $\forall A \subset X : A \subset \tau(A)$, *monotone* meaning $\forall A, B \subset X : A \subset B \implies \tau(A) \subset \tau(B)$, and *idempotent* meaning $\forall A \subset X : \tau(\tau(A)) = \tau(A)$.

The fixed points or *closed sets* of a closure operator form a *closure system*, i.e. X is closed and an arbitrary intersection of closed sets is closed.

A set of sets $\mathcal{S} \subset \mathcal{P}(X)$ is a *down set* if $A \in \mathcal{S}$ and $B \subset A$ imply $B \in \mathcal{S}$. *Up sets* are defined symmetrically.

The *support* of a permutation is the set of elements that it moves. Note that this set is closed under the permutation.

If A is an alphabet, G a group a *pattern* is $p : P \to A$ with $P \subset G$ (not necessarily finite). When playing the solitaire, we also talk about patterns as subsets of G (subsets are of course in bijection with patterns over the alphabet $A = \{0, 1\}$). If $a \in A$ and $g \in G$, write a^g for the unique pattern p of type $p : \{g\} \to \{a\}$,

3 The general theory of solitaire

3.1 Solitaire moves

We begin by defining the solitaire process on a group. We give a slightly more general definition than was discussed in the introduction, where the unique hole can only be moved if it is one of the elements in a particular subset $C \Subset S$.

While our main interest is indeed in the case $C = S$, cases $C \subsetneq S$ also arise naturally in applications (see [7] and Section 8.5) and do not add any difficulties in proofs.

Definition 1. *Let G be a discrete countable group. If $C \Subset S \Subset G$ are finite sets, then a* (C, S)-solitaire move *at $g \in G$ is a pair $(x, y) \in \mathcal{P}(G)^2$ such that*

$$|x \cap gS| = |y \cap gS| = |S| - 1 \wedge x \triangle y \in \mathcal{P}_2(gC).$$

In general, a (C, S)-solitaire move *is a (C, S)-solitaire move at any $g \in G$, and (S, S)-solitaire moves are called simply S-solitaire moves. A (C, S)-solitaire move (x, y) is usually written as $x \to_{C,S} y$ or even $x \to y$, and $x \to^* y$ denotes the transitive closure of this relation.*

Note that despite the arrow being directed, the moves are invertible. One can think of the (C, S)-solitaire moves in many ways. Of course, one can think of them as a relation, or as forming the edges of an undirected graph with vertices $\mathcal{P}(G)$, or as the objects of an abstract rewriting system (with the edges giving the legal rewrite rules).

Slightly less obviously, but usefully, one can think of them as a group action: for each $g \in G$, we have an action of the group of permutations on C, which permutes the hole in gC, in configurations $x \in \mathcal{P}(G)$ with $|x \cap gS| = |S| - 1 \wedge |x \cap gC| = |C| - 1$ (and fixes all other configurations). The free product of the groups P_c then acts on $\mathcal{P}(G)$, and this action is by homeomorphisms when $\mathcal{P}(G)$ is given the product topology.

Another, useful way is to think of them as a groupoid. Recall that a groupoid is a small category (meaning the objects form a set) where every morphism is an isomorphism. Every equivalence relation can be thought of as a groupoid, where the objects are the elements of the set, and morphisms are equivalent pairs. Such a groupoid is principal, meaning there is at most one (iso)morphism between any pair of objects.

The advantage of the groupoid point of view over the group action point of view is that the action of a certain sequence of elements now always applies the same permutation – it may simply not be applicable. This point of view is crucial in the proof of Lemma 14 (which was in fact one of the reasons that prompted us to develop this theory properly).

We note that a groupoid structure is a standard way to think about the classical 15-puzzle, which corresponds roughly to the case $G = \mathbb{Z}^2, C = S = \{0, 1\}^2$ of the solitaire (and which we study in Section 6.1).

Due to the group action and groupoid interpretation, it is natural to call equivalence classes of the relation $\to^*$ *orbits*. We denote the orbit of $x \in \mathcal{P}(G)$ as $\mathcal{O}(x)$. We say a set $X \subset Y$ is *closed under solitaire* if $\forall x \in X : \forall y \in Y : x \to^* y \implies y \in X$.

Remark 1. We note that the theory of solitaire might as well be built on finitely-generated groups. Namely, let $H = \langle S^{-1} S \rangle$, i.e. the subgroup of G generated by all $a^{-1}b$ such that $a, b \in S$. Then it is easy to see that each solitaire move only touches elements in a single coset. Thus, the cosets behave completely independently.

3.2 Filling process

Definition 2. *Let G be a discrete countable group. If $C \Subset S \Subset G$ are finite sets, then a (C,S)-*filling move *at $g \in G$ is a pair $(x,y) \in \mathcal{P}(G)^2$ such that*

$$x \neq y \wedge |x \cap gS| = |S| - 1 \wedge y = x \cup gC.$$

*In general, a (C,S)-*filling move *is a (C,S)-filling move at any $g \in G$, and (S,S)-filling moves are called simply S-filling moves. A (C,S)-filling move (x,y) is usually written as $x \Rightarrow_{C,S} y$ or even $x \Rightarrow y$, and $x \Rightarrow^* y$ denotes the transitive closure of this relation.*

Note that the filling process is not reversible, however:

Lemma 1. *The filling process has the* diamond property*, meaning whenever $x \Rightarrow y, x \Rightarrow z$ with $y \neq z$, there exists w such that $y \Rightarrow w, z \Rightarrow w$. In particular, the system is* confluent*, meaning whenever $x \Rightarrow^* y, x \Rightarrow^* z$, there exists w such that $y \Rightarrow^* w, z \Rightarrow^* w$.*

Proof. Suppose $x \Rightarrow y, x \Rightarrow z$ with $y \neq z$. Then $x \neq y$, $x \neq z$ and there exist $g \in G$ such that

$$|x \cap gS| = |S| - 1 \wedge y = x \cup gC$$

and $g' \in G$ such that

$$|x \cap g'S| = |S| - 1 \wedge z = x \cup g'C.$$

Take $w = y \cup z$. Observe that if $y \cap g'S = g'S$, then $y \supset z$ so in fact $y = w = z$ (which we assumed does not hold). Thus we must have $|y \cap g'S| = |x \cap g'S| = |S| - 1$ and thus $y \Rightarrow y \cup g'C = x \cup gC \cup g'C = w$ as desired. Symmetrically, $z \Rightarrow w$.

The fact confluence follows from the diamond property is well-known, and follows from completing the paths $x \Rightarrow^m y, x \Rightarrow^n z$ to an m-by-n rectangle, by successive applications of the diamond property. □

Definition 3. *Let G be a discrete countable group. If $C \Subset S \Subset G$ are finite sets, then the* (C, S)-filling closure *of $x \in \mathcal{P}(G)$ is defined as*

$$\varphi_{C,S}(x) = \bigcup \{y \mid x \Rightarrow^*_{C,S} y\}.$$

Remark 2. Recall that a *directed set* is a set together with a partial order such that any two elements have a common upper bound. The previous lemma shows that for $x \subset G$, the set $F_x = \{y \subset G \mid x \Rightarrow^*_{C,S} y\}$ is a directed set under $\Rightarrow^*$ and thus under inclusion, since the filling process can only increase sets. The set F_x defined above is precisely the limit of this set (seen as a net indexed by itself) in the product topology of $\mathcal{P}(G)$. (For the definition of a limit of a net see any standard reference [4].)

Lemma 2. *The (C,S)-filling closure operator φ is a closure operator.*

Proof. By definition, we need to show $x \subset \varphi(x)$ (extensivity), $x \subset y \implies \varphi(x) \subset \varphi(y)$ (monotonicity) and $\varphi(\varphi(x)) = \varphi(x)$ (idempotency) for all sets x, y. But in fact, all of these properties are obvious. □

It follows from the previous lemma that the set of sets $x \in \mathcal{P}(G)$ which are their own filling closure form a closure system, and we call the fixed points *filling closed sets*. In particular, the following is a well-known consequence:

Lemma 3. *The filling closure $\varphi(x)$ is the intersection of all filling closed sets that contain x.*

The connection between the solitaire and the filling process is now the following:

Lemma 4. *Let G be a discrete countable group, and let $C \Subset S \Subset G$ be finite sets. Let $x \in \mathcal{P}(G)$ be arbitrary. Then for all $y \in \mathcal{O}(x)$, we have $\varphi_{C,S}(y) = \varphi_{C,S}(x)$.*

Proof. It suffices to show that if $x \in \mathcal{P}(G)$ and $x \to_{C,S} y$, then $y \subset \varphi(x)$. Namely then by idempotency $\varphi(y) \subset \varphi(x)$ as well.

Suppose thus $x \to_{C,S} y$, meaning

$$|x \cap gS| = |y \cap gS| = |S| - 1 \wedge x \triangle y \in \mathcal{P}_2(gC).$$

It suffices to show that $x \Rightarrow x \cup y$ by a move at g, i.e.

$$|x \cap gS| = |S| - 1 \wedge x \cup y = x \cup gC,$$

because then $y \subset x \cup y \subset \varphi(t)$. Of course $|x \cap gS| = |S| - 1$ is immediate. We have $x \cup gC = x \cup y$ because the only difference between x and y is that they lack a different element of gC. □

We note a useful property of the filling operator (which is a general property of closure operators).

Lemma 5. *Suppose $\varphi(x) = \varphi(y)$. Then $\varphi(x \cup z) = \varphi(y \cup z)$ for any z.*

Proof.

$$\varphi(x \cup z) = \varphi(\varphi(x \cup z)) \supset \varphi(y \cup z)$$

since $z \subset x \cup z \subset \varphi(x \cup z)$ and $y \subset \varphi(x) \subset \varphi(x \cup z)$. □

3.3 Linear sets, finite fillings and finite solitaires

Our interest is specifically on the solitaire orbits of finite sets. For the following definitions, we concentrate on the case $C = S$.

Definition 4. *Let G be a group. A subset $S \Subset G$ is* linear *if there exists $a, b, c \in G$ with b of infinite order, such that $S \subset a\langle b\rangle c$ where $\langle b\rangle = \{b^n \mid n \in \mathbb{Z}\}$.*

Note that $\langle b \rangle$ is the (cyclic) group generated by the element b, and b being of infinite order means precisely that this group is infinite. Recall that a *left coset* of a subgroup $H \leqslant G$ is a set of the form aH.

Lemma 6. *If S is a linear shape, then there exists a shape S' which is contained in a cyclic group, such that the set of S-solitaire moves coincides with the set of S'-solitaire moves.*

Proof. We have

$$S \subset a\langle b \rangle c = ac\langle c^{-1}bc \rangle.$$

Then $S' = (ac)^{-1}S \subset \langle c^{-1}bc \rangle$. Clearly S'-solitaire moves are equivalent to S-solitaire moves, as the set of moves only depends on the set of left translates of the shape. □

Definition 5. *Let G be a group and $S \Subset G$. We say G has the* finite S-filling property *if the S-filling closure of every finite set is finite. We say G has the* finite filling property *if this holds for all non-linear $S \Subset G$. We say G has the* finite S-solitaire property *if the S-solitaire orbit of every finite set is finite, and* finite solitaire property *if this holds for all non-linear S.*

The assumption of non-linearity is necessary:

Lemma 7. *If $S \Subset G$ is linear, and $|S| \geqslant 2$, then there is a finite set with infinite S-filling closure and infinite S-solitaire orbit.*

Proof. By Lemma 6 above we may suppose $S \subset \langle b \rangle$ with b of infinite order. Let m be maximal and $n \geqslant m$ minimal such that $S \subset T = \{b^m, \dots, b^n\}$. Then $T \setminus \{b^m\}$ is easily seen to have infinite S-filling orbit and infinite S-solitaire orbit. □

The following is immediate from Lemma 4.

Lemma 8. *If G has the finite filling property, then it has the finite solitaire property.*

The following is a simple geometric observation.

Proposition 1. *The groups $\mathbb{Z}^d$ have the finite filling property. In particular, it has the finite solitaire property.*

Proof. The case $d = 1$ is trivial (and rather meaningless).

Suppose $S \Subset \mathbb{Z}^d$ is non-linear. We start by showing that we may assume $\vec{0}$ is in the interior of the convex hull C of S. First, we may temporarily assume by translating that $\vec{0} \in S$. We may then assume that S generates $\mathbb{Z}^d$ (not necessarily positively), by possibly replacing $\mathbb{Z}^d$ by the group generated by S. Note that still $d \geqslant 2$ by the assumption that S is non-linear.

Because S generates $\mathbb{Z}^d$, C has nonempty interior, as it is well-known that a convex set without interior lies in a proper affine set (and thus we could've taken a smaller d). Now replace S by kS for large $k \in \mathbb{N}$ (i.e. dilate the shape; note that this preserves the finite filling property because kS-solitaire consists of independent S-solitaires on cosets of $k\mathbb{Z}^d$, see also Remark 1), and then translate it, after which we may actually assume that $\vec{0} \in C^\circ$ in the first place place (after this, we may no longer have $\vec{0} \in S$).

By basic convex geometry, such a set is a bounded intersection of a finite number of closed half-spaces, i.e. $C = \bigcap_i \ell_i((-\infty, r_i])$ where $\ell_i : \mathbb{R}^d \to \mathbb{R}$ are linear and nontrivial. Note that $r_i > 0$ since $\vec{0}$ is in the interior of C. Furthermore, the half-spaces $\ell_i((-\infty, r_i])$ can be taken to satisfy that each $\ell^{-1}(r_i) \cap C$ is a face (i.e. a set spanned by vertices) of C of dimension $d-1$.

In particular, the faces are spanned by d vertices of C, which must be elements of S, thus, $|\ell_i^{-1}(r_i) \cap S| \geqslant d \geqslant 2$ for all i. In particular, it follows that $H = \ell_i^{-1}((-\infty, r])$ is closed under S-filling for all $r \in R$. Namely, we have $\varphi(S) \subset (-\infty, r_i]$ and $\varphi(s) = r_i$ for at least two distinct $s \in S$. Thus the intersection $\vec{v} + S \cap H$ is always of cardinality $|S|$ (when $r < r_i$) or of cardinality at most $|S| - 2$ (when $r > r_i$), and thus Definition 2 does not apply for any $\vec{v} \in \mathbb{Z}^d$.

Now let $T \Subset \mathbb{Z}^d$ be arbitrary. Since T is finite, it is bounded and thus $\ell(T)$ is bounded for any $\ell \in L$. Thus, $T \subset D = \bigcap_i \ell_i((-\infty, t_i])$. As we showed, each $\ell_i((-\infty, t_i])$ is closed under S-filling, thus so is their intersection.

It suffices to show that D is bounded. If it is not, then there exist arbitrarily large (in norm) $\vec{v}$ such that $\ell_i(\vec{v}) \leqslant t_i$ for all i. By scaling, we obtain that there exist arbitrarily large $\vec{v}$ such that $\ell_i(\vec{v}) \leqslant r_i$ for all i. Such vectors $\vec{v}$ are then in C, so in fact S itself is unbounded, a contradiction. $\square$

4 Excess

Much of our interest is in understanding the connection between the filling process and the solitaire process. We measure the difference between a pattern and a minimal pattern with the same filling closure by the notions of excess and excess sets.

In the case of the triangle, we will see that for each filling closed set, all patterns with minimal cardinality that generate it lie in the same solitaire orbit.

4.1 Excess in a global sense

Fix $C \Subset S \Subset G$ for this section, and suppose G has finite S-fillings.

If $P \Subset G$, then the *rank* $\mathrm{rank}(P)$ of P is the minimal cardinality of $R \Subset G$ such that $\varphi(R) = \varphi(P)$. Note that the rank of P is at most $|P|$. Note also that the only candidate subsets $R \Subset G$ we need to consider are subsets of $\varphi(P)$.

Definition 6. *The* excess *of $P \Subset G$ is defined as $e(P) = |P| - \mathrm{rank}(P)$.*

This is excess in a global sense, since we are counting how much smaller we could make P by replacing it by another (possible quite different-looking) set, without changing its filling closure.

Note that every filling closed set F has by definition at least one subset with numerical excess 0.

Lemma 9. *Excess is monotone: if $P \subset Q \Subset G$, then $e(P) \leqslant e(Q)$.*

Proof. Let R be a set of size $\operatorname{rank}(P)$ with filling $\varphi(P)$. Then $\varphi(Q) = \varphi(P \cup (Q \setminus P)) = \varphi(R \cup (Q \setminus P))$ by Lemma 5. Thus $\operatorname{rank}(Q) \leqslant \operatorname{rank}(P) + |Q \setminus P|$ so

$$e(Q) = |Q| - \operatorname{rank}(Q) \geqslant |P| + |Q \setminus P| - \operatorname{rank}(P) - |Q \setminus P| = e(P).$$

□

4.2 Excess in a local sense

Definition 7. *Let $P \Subset G$. We say $Q \subset P$ is an* excess set *if $\varphi(P \setminus Q) = \varphi(P)$. The set of excess sets of P is denoted $E(P)$. The* visible excess *of P is the maximal cardinality of an excess set in P, denoted $\hat{E}(P)$. The* phantom excess *of P is $e(P) - \hat{E}(P)$.*

Note that excess sets are defined for a given set P (and of course they depend on the choice of C, S). Clearly if P contains an excess set of cardinality k, then the $e(P) \geqslant k$. Thus, excess is always at least as large as visible excess, and phantom excess is always nonnegative.

It is tempting to try to prove that phantom excess is always zero, or even that all maximal excess sets have the same cardinality, which equals excess. As we will see in Section 4, this fails badly even on the group $\mathbb{Z}^2$, and for the simplest possible non-linear shape.

Lemma 10. *The excess sets of P form a down set.*

Proof. If Q is an excess set of P and $Q' \subset Q$, then $\varphi(P) = \varphi(P \backslash Q) \subset \varphi(P \backslash Q') \subset \varphi(P)$. □

Lemma 11. *Visible excess is monotone: if $P \subset Q \Subset G$, then $\hat{E}(P) \leqslant \hat{E}(Q)$.*

Proof. If $R \subset P$ is an excess set then $\varphi(P) = \varphi(P \setminus R)$, so

$$\varphi(Q) = \varphi(P \cup (Q \setminus P)) = \varphi((P \setminus R) \cup (Q \setminus P)) = \varphi(Q \setminus R)$$

so R is an excess set for Q, implying $\hat{E}(Q) \geqslant \hat{E}(P)$. □

Lemma 12. *If a pattern contains a subpattern with no excess and with the same filling closure, then it has no phantom excess.*

Proof. If $P \subset Q$ where P has no excess and $\varphi(P) = \varphi(Q)$, then P can be used to calculate excess for Q, and we have $e(Q) = |Q \setminus P|$. On the other hand, $Q \setminus P$ is clearly an excess set of this cardinality, thus visible excess matches excess. □

Interestingly, visible excess is *not* preserved under the solitaire process. We will see an example in Section 5.4.

4.3 Monotonicity of solitaire, and "elements are not there"

Fix $C \Subset S \Subset G$ for this section.

We show that the solitaire process is monotone in the sense that after adding elements to a set, we can still apply all moves we could previously. Intuitively, we can always pretend not to see some of the elements in a pattern, and play solitaire on the subpattern. The elements we pretended not to see are then permuted in some (possibly nontrivial) way.

Recall that we may think of a solitaire move as applying a permutation of C in gC (through the natural identification $c \leftrightarrow gc$), and we are allowed to apply this permutation to a pattern P if $g(S \setminus C) \subset P$ and $|gC \cap P| = |C| - 1$. We note that it does not hurt if we also allow applying the permutation when $gS \subset P$, as in this case the permutation fixes the pattern anyway.

Definition 8. *We say that* solitaire applies *at g to pattern P if $g(S \setminus C) \subset P$ and $|gC \cap P| \geqslant |C| - 1$. Then for any permutation $\pi : C \to C$ we have the* associated solitaire move *$g\pi$ defined by*

$$P \to P' = (P \setminus gC) \cup g\pi(g^{-1}P \cap C).$$

We also use function notation $g\pi(P) = P'$.

Note that $g^{-1}P \cap C$ is in bijection with $P \cap gC$, and this bijection is the natural one, $c \leftrightarrow gc$. Thus we are indeed applying π to the subset of P intersecting gC, and finally moving it back to a subset of gC.

Lemma 13. *Let $P \subset G$. If solitaire applies at g to pattern P, then it applies to any $P \cup R$. Furthermore, $g\pi(P \cup R) \supset g\pi(P)$.*

Proof. Permutations at g stay applicable when the set increases: the only possible change in the intersection with gS is that $(P \cup R) \cap gS$ becomes gS (because already $P \cap gS$ misses at most one element). Both $g\pi(P \cup R)$ and $g\pi(P)$ are produced by applying the same permutation in the same set of cells. □

Lemma 14. *Suppose $P \to^* P'$. Then there exists a bijection $\pi : G \setminus P \to G \setminus P'$ with support contained in $\varphi(P)$, such that $P \cup Q \to^* P' \cup \pi(Q)$ for all $Q \subset G \setminus P$.*

Proof. Since $P \to^* P'$, there is a sequence of patterns $P = P_0 \to P_1 \to \cdots \to P_m = P'$, thus there is a sequence of permutations $g_1, g_2, \ldots, g_m$ and a sequence of permutations $\pi_1, \ldots, \pi_m$ of C such that solitaire applies at g_i in P_{i-1}, and $P_i = g_i\pi_i(P_{i-1})$.

Let $R \cap P = \varnothing$, and define $Q = Q_0 = P \cup R$. Induction and the previous lemma show that the sequence $Q_i = g_i\pi_i(Q_{i-1})$ is well-defined (because solitaire applies at g_i to pattern Q_{i-1}), and Q_i contains P_i.

The entire evolution of the pattern comes from a sequence of permutations applied to G, proving the claim about there existing of a single permutation of π.

For the claim about the support, observe that each P_i is contained in the filling closure $\varphi(P)$. The only way solitaire could apply at g_i in P_{i-1}, without g_iS being contained in $\varphi(P)$, is that actually $g_i\pi_i$ moves an element outside of $\varphi(P)$, which is impossible. □

In particular:

Lemma 15. *Suppose $P \to^* P'$. Then for any Q, we have $P \cup Q \to^* R$ for some pattern R containing P'.*

Proof. Take $R = P' \cup \pi(Q)$ in the previous lemma. □

4.4 Transporting excess

Fix again $C \Subset S \Subset G$ for solitaire purposes.

Sometimes a filling closed set F contains a filling subset P (i.e. $F = \varphi(P)$), which allows us to freely perform transformations on the excess sets. Intuitively, P can be moved around F by the solitaire freely enough that it can "drag other elements along" anywhere in F, and ultimately permute them arbitrarily.

Definition 9. *Suppose $P, R \Subset G$. The R-*restricted orbit *of P is the set of patterns $Q \Subset G$ obtained from P by applying only solitaire moves only at g inG such that $gS \subset R$.*

Definition 10. *Suppose $P \Subset G$ and $F = \varphi(P)$. We say P* transports k-excess inside R *if all of the patterns $P \sqcup Q \subset R$ where $|Q| \leqslant k$ are in the same R-restricted S-orbit. We say P* transports k-excess *if it transports k-excess inside F. We say P transports excess (inside R) if it transports k-excess (inside R) for all k.*

A *super-P* pattern is $Q \supset P$ such that $\varphi(Q) = \varphi(P)$. In terms of this notion, P transports k-excess if and only if every for every $\ell \leq |P| + k$, every super-P pattern with cardinality ℓ is in the same orbit, and transports excess if and only if every super-P pattern with the same cardinality is in the same orbit. Note that in this definition P itself is allowed to have excess, although in our applications it will not. To explain the terminology, it is the excess *on top of* the excess of P itself that its transported.

By definition, if P transports k-excess inside F, then it transports $(k-1)$-excess. We show that in at least very special situations, the converse is true.

Lemma 16. *Suppose $P \Subset G$ and $F = \varphi(P)$. Suppose $F \setminus P = \{p_1, \ldots, p_n\}$ and P transports 1-excess in $R_k = P \cup \{p_1, \ldots, p_k\}$ for all k. Then P transports excess.*

Proof. By induction on k and n, we show that P transports n-excess in R_k for all k, n. For $k = 0$, this is trivial, and for a fixed k, it is trivially true for $n = 0$, since in these cases P is the only pattern to consider.

Consider then a pattern Q of size $n+1$ inside R_k, such that the claim holds for patterns up to size n, and holds for all smaller values of k. If $n + 1 = k$, then Q is the only pattern of its size, and the claim holds trivially, so suppose $n + 1 < k$. If Q does not contain p_k, we may think of it as being contained in R_{k-1}, inside which we can transport $(n+1)$-excess freely. Thus, it suffices to show that p_k can be moved inside R_{k-1} without involving elements outside R_k.

Actually, it is easier to show that we can move an element into p_k. Namely, since $n+1 \geqslant 1$, we know that Q contains some p_i. Since P transports 1-excess, we have $P \cup \{p_i\} \to^* P \cup \{p_n\}$.

By Lemma 15, we then have $P \cup Q \to^* P \cup R$ for some $R \ni p_k$. In fact, by Lemma 14, we can choose $R \setminus \{p_k\}$ freely by choosing Q freely, and the latter is possible since P transports $(n+1)$-excess inside R_{k-1}. □

Lemma 17. *The solitaire preserves the set of patterns that transport excess.*

Proof. Suppose P transports excess, and suppose $Q \to^* P$. Let $Q \subset R \sqcup Q \subset \varphi(Q)$. By Lemma 15, we have $Q \cup R \to^* P'$ for some pattern P_R containing P. Since P transports excess, all such P_R with equal cardinality are in the same orbit, therefore all patterns $Q \cup R$ with the same cardinality are in the same orbit. □

4.5 Nice solitaire theories

Definition 11. *Let G be a group, $C \Subset S \Subset G$. We say (C, S, G) has a* nice solitaire theory *if the following holds: if P, Q have the same filling closure, and neither of them has excess, then $P \to^*_{C,S} Q$.*

In our case studies below, we will show that $G = \mathbb{Z}^2$ has nice solitaire theory for $C = S$ either the triangle shape $\{(0,0),(1,0),(0,1)\}$ or the square shape $\{0,1\}^2$. In both cases, excess can also be fully understood.

Note that, at least if G has finite fillings, then it is never true that all patterns with the same filling closure and the same excess are in the same orbit. Namely, let Q be any pattern with excess, and P any smaller pattern with $\varphi(P) = Q$ (for example, $P = S \setminus \{c\}$, $Q = S$ for some $c \in C$). Then it is easy to see that for almost all $g \in G$, the filling closures of $gQ \cup P$ and $gP \cup Q$ are equal, but they are not in the same solitaire orbit.

5 Triangle solitaire on the plane

The *triangle* is the set $T = \{(0,0),(1,0),(0,1)\}$. A T-solitaire move is called a *triangle move*, and the T-solitaire process is referred to as *triangle solitaire*.

We have $T = T_2$ where $T_n = \{(a,b) \in \{0,\dots,n-1\}^2 \mid a+b \leqslant n-1\}$ is the more general *n-triangle*. Sometimes, we refer to a generic n-triangle as simply a triangle, and "the triangle" or "triangle shape" to refer to T.

In this section, we completely characterize the orbits of the triangle solitaire, and prove various things about our favorite orbit, namely *line orbit*, which is the orbit of the *line* $L_n = \{0,\dots,n-1\} \times \{0\}$. Sometimes "a line" can also refer to a translate of the line. By the *edges* of the n-triangle we refer to the intersections of the edges of its convex hull with the lattice $\mathbb{Z}^2$; the line is one of the three edges of the n-triangle.

The set L_n has the n-triangle as its filling closure. One of our main results is that all sets of cardinality at most n whose filling closure is T_n are in the solitaire orbit of L_n.

A useful graph structure to consider on $\mathbb{Z}^2$ is the triangular lattice. Specifically, the *neighbourhood* of a point $\vec{v} \in \mathbb{Z}^2$ is the one depicted in Figure 3, it corresponds to the points which can be involved in a triangle move with x. The neighbourhood of a pattern A, denoted $N(A)$ is the union of the neighbourhoods of its points. Two patterns A and B *touch* if $A \cap N(B) \neq \varnothing$ or equivalently $B \cap N(A) \neq \varnothing$.

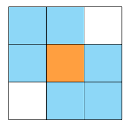

Fig. 3: The neighbours of the orange cell are the blue ones.

Note that by squishing vertically by a factor of $1/\sqrt{2}$, and then shearing horizontal lines by 0.5, the triangular lattice becomes invariant under rotation by 120 degrees, and triangles become equilateral. Although we find it more convenient to work with coordinates in $\mathbb{Z}^2$, we apply this observation in symmetry considerations.

Note that when expressed in coordinates $\mathbb{Z}^2$, the 120-degree rotation counterclockwise corresponds to the automorphism of $\mathbb{Z}^2$ given by the matrix $M = \left(\begin{smallmatrix} -1 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ (when multiplying column vectors from the left). Formally, we may take the phrase "up to rotation" to refer to an applications of M or M^{-1}.

We begin with the observation that the choice of orientation of the line does not matter:

Proposition 2. *For every n, the three edges of T_n are in the same orbit.*

Proof. The first line of Figure 4 explains by example how to transform the horizontal edge into the diagonal, and the second one how to transform the diagonal into the vertical edge (of course this is just a rotated inverse of the first transformation). □

This proof shows that the number of solitaire moves needed to go from one edge to another is at most $O(n^2)$. We will see later that $\Omega(n^2)$ steps are necessary, and that the diameter of the orbit of the line is $\Omega(n^3)$.

5.1 The filling process for the triangle

We now specialize the filling process for the triangle. This specific process has been studied previously, in [5].

Since $\mathbb{Z}^2$ has the finite filling property and T is non-linear, the filling process is terminating, and thus the filling closure $\varphi(P)$ is reached in finitely many steps.

Let us say that P *fills* if $\varphi(P) = T_n$, where $n = |P|$. In [5], filling sets P were called *fill matrices*. We will show that they in fact correspond to the elements of the line orbit.

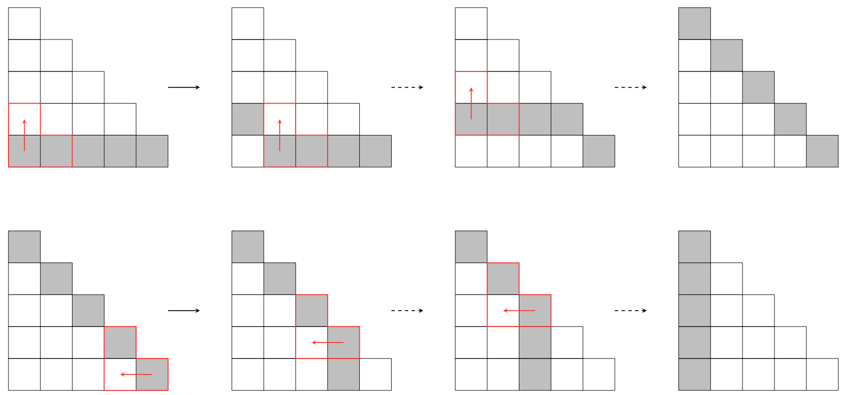

Fig. 4: How to transform one edge into another for $n = 5$.

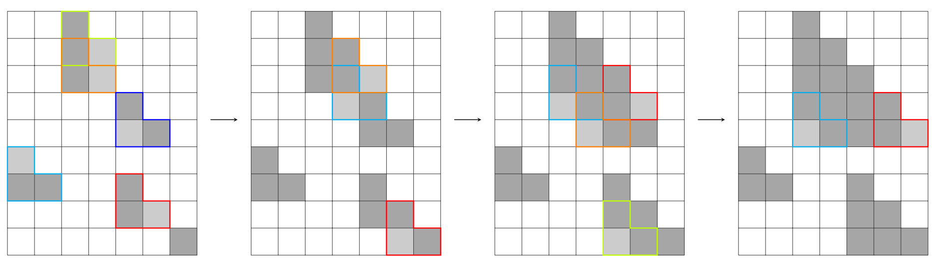

Fig. 5: An example of the filling process.

Lemma 18. *For any pattern P, there are unique integers $k_1, \ldots, k_r$ and vectors $\vec{v_1}, \ldots, \vec{v_r}$ such that $\varphi(P) = \bigcup_{i=1}^{r} \vec{v_i} + T_{k_i}$, $\sum_{i=1}^{r} k_i \leqslant |P|$ and $N(\vec{v_i} + T_{k_i}) \cap (\vec{v_j} + T_{k_j}) = \varnothing$ for each $i \neq j$.*

We refer to $\varphi(P) = \bigcup_{i=1}^{r} \vec{v_i} + T_{k_i}$ as the *fill decomposition* of P. (Note that this refers to the formal union rather than the set, but it is easy to see that the set determines the values v_i and k_i up to a permutation.)

Proof. We prove this by induction on $|P|$. The case $|P| = 1$ is trivial.

Now assume the result is true for patterns of size at most n and let P be a pattern of size $n+1$. Then if $x \in P$, $P \setminus \{x\}$ satisfies the induction hypothesis so we can write $\varphi(P \setminus \{x\}) = \bigcup_{i=1}^{r} \vec{v_i} + T_{k_i}$. We now have three cases to consider.

First, if $x \in \varphi(P - x)$ then $\varphi(P) = \varphi(P \setminus \{x\})$, and the inequality on the size of P clearly continues to hold, as the right-hand size increase, but the left-hand side does not.

Second, if x is not in the neighbourhood of $\varphi(P \setminus \{x\})$ then no additional filling can be done with it, therefore $\varphi(P) = \varphi(P \setminus \{x\}) \cup \{x\}$ and, as $\{x\}$ is a triangle (translate of T_1), we have the appropriate decomposition, and both sides of the inequality are increased by 1.

Finally, assume $x \in N(\vec{v_1} + T_{k_1})$. Then we can extend $\vec{v_1} + T_{k_1}$ as in Figure 6. By doing so we may lose the property that the triangles do not touch, but if some do so we can merge them by repeating the extension process. Notice that if two triangles are merged, then the new triangle cannot be larger than the sum of the sizes of the initial triangle, so the inequality on the triangles' sizes is still satisfied. (Merges may be triggered recursively, but nevertheless no merge can increase the sum of triangle sizes.) $\square$

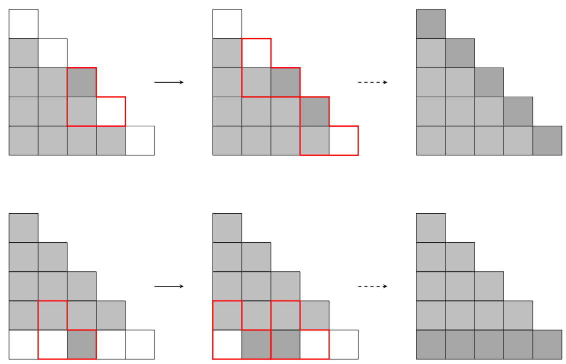

Fig. 6: How to extend a triangle with a top neighbour or a subdiagonal neighbour. The right neighbour case is symmetric to the top neighbour case.

Define the *triangle excess* of P as the difference $e_T(P) = |P| - \sum_{i=1}^{r} k_i$.

Lemma 19. *Triangle excess is equal to excess.*

Proof. Excess is defined as

$$e(P) = |P| - \operatorname{rank}(\varphi(P)),$$

and triangle excess is defined as

$$e_T(P) = |P| - \sum_{i=1}^{r} k_i.$$

where k_i come from the fill decomposition of P. To show these equivalent is the same as showing $\operatorname{rank}(\varphi(P)) = \sum_{i=1}^{r} k_i$.

Since $\varphi(\bigcup_{i=1}^{r} \vec{v}_i + L_{k_i}) = \varphi(P)$, we have $\operatorname{rank}(\varphi(P)) \leqslant \sum_{i=1}^{r} k_i$. On the other hand, if we have fill decomposition $\varphi(R) = \varphi(P) = \bigcup_i \vec{v}_i + T_{k_i}$ then $|R| \geqslant \sum_i k_i$ by Lemma 18, so indeed $\operatorname{rank}(\varphi(P)) \geqslant \sum_i k_i$ as desired. $\square$

5.2 Characterisation of the solitaire orbit through the filling

We can mimic the filling process using the solitaire, by using lines in place of triangles, and pretending the excess elements are not there, as in Section 4.3.

In line with the terminology in Section 4.4, we say a pattern P is a *superline* if $\varphi(P) = \vec{v} + T_n$ for some n, and $P \supset \vec{v} + L_n$.

Lemma 20. *Let P be a superline, and suppose $\vec{v}$ touches P. Then $P \cup \{\vec{v}\}$ contains a superline in its solitaire orbit.*

Proof. Suppose P contains $\vec{u} + L_m$ and $\varphi(P) = \vec{u} + T_m$. By translational symmetry we may assume $\vec{u} = (0, 0)$. First suppose the exact equality $P = L_m$ holds. If $\vec{v} \in T_m$, the claim is clear as P is already a superline.

Otherwise, up to rotational symmetry (and because of Proposition 2), we may assume $\vec{v}$ is below P, i.e. $\vec{v} = (k, -1)$ for $0 \leqslant k \leqslant m$. Figure 7 shows how we can turn $P \cup \{\vec{v}\}$ into the line $(0, -1) + L_{m+1} = \{0, \ldots, m\} \times \{-1\}$ by applying solitaire moves, again proving the claim.

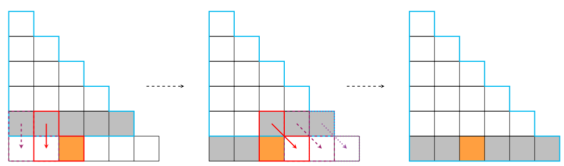

Fig. 7: How to extend a line with a bottom neighbour.

Now consider the general case $P \supset L_m$ and $\varphi(P) = T_m$, where we may again assume $\vec{v} \notin T_m$. We have $\varphi(P) = \varphi(L_m)$ so $\varphi(P \cup \{\vec{v}\}) = \varphi(L_m \cup \{\vec{v}\})$ by Lemma 5. Denote $T' = \varphi(L_m \cup \{\vec{v}\})$.

We have $L_m \cup \{\vec{v}\} \to^* Q$ for some superline Q with of course $\varphi(Q) = T'$, so by Lemma 14 we have $P \cup \{\vec{v}\} \to^* Q'$ for some pattern $Q' \supset Q$, and $\varphi(Q') = T'$. Thus Q' is also a superline. □

Lemma 21. *Suppose that P, Q are (not necessarily disjoint) superlines, and their filling closures touch. Then $P \cup Q$ contains a superline in its solitaire orbit.*

Proof. Up to rotational symmetry, translation, and exchanging the roles of P and Q, we may assume $\varphi(P) = T_m$ and $\varphi(Q) = \vec{v} + T_n$ where $\vec{v} = (a, b)$ with $a \geqslant 0$, $b \leqslant -1$, and $n + b \leqslant m - a$. To see this, we recommend the reader draw a picture and argue geometrically: if $n + b > m - a$, then the top corner of Q sticks above the diagonal of P, in which case the right corner of P will be inside Q or just next to it, and we may exchange the roles of P and rotate and translate to get to the situation $n + b \leqslant m - a$.

Since the filling closures touch, we have $n \geqslant -b$ and $a \leqslant m$. Since Q is a superline and all edges of the triangle are in the same orbit, $Q' = \vec{v} + \{0\} \times \{0, \ldots, n - 1\}$ is contained in some pattern in the orbit of Q. Note that the assumption on n is precisely equivalent to the fact that vectors in Q' with positive second coordinate are contained in T_n. Now we can simply apply the previous lemma successively to the vectors in Q' with negative second coordinate. □

Lemma 22. *Let P be a pattern with fill decomposition $\varphi(P) = \bigcup_i \vec{v}_i + T_{k_i}$. Then the orbit of P contains a union of superlines whose filling closures do not touch, and where the corresponding lines are precisely the lines $\vec{v}_i + L_{k_i}$.*

Proof. We start by observing that every pattern P can be thought of as a union of superlines $\bigcup_i \vec{v}_i + T_{k_i}$, for example by taking $k_i = 1$ for all i, and having the $\vec{v}_i$ enumerate P.

Let $P' = \bigcup_{i=1}^{r} Q_i$ be in the orbit of P with Q_i superlines, such that r is minimal. Then the filling closures of distinct Q_i do not touch: suppose $\varphi(Q_j)$ and $\varphi(Q_k)$ do touch for $j \neq k$. Then by Lemma 21, $Q_j \cup Q_k$ contains a superline Q in its solitaire orbit. By Lemma 15, $\bigcup_{i=1}^{r} Q_i$ then contains a set $R \supset Q' = Q \cup \bigcup_{i \in [1,r] \setminus \{j,k\}} Q_i$ in it solitaire orbit. Each element of R which is not in Q' can be added in one of the superlines Q_i or Q, since the union of their filling closures is $\varphi(P)$. Thus, we have found a pattern in the orbit of P which can be written as a union of strictly less than r superlines.

For the final claim, we observe that the filling closure of P must be precisely the union of the filling closures of the superlines, which determines the lines to be those corresponding to the fill decomposition. ☐

Next, we show that the non-touching superlines can be put in normal form.

Lemma 23. *The line transports excess.*

Proof. By Lemma 16 it suffices to show that there is a linear order $p_1, p_2, \ldots, p_m$ on $\varphi(L_n) \setminus L_n = T_n \setminus L_n$ such that L_n transports 1-excess in $R_k = P \cup \{p_1, \ldots, p_k\}$ for all k. Note that $m = |T_n \setminus L_n| = n(n-1)/2$. We use the order $(a, b) \leqslant (c, d)$ if $b < d$ or $b = d \wedge a < c$. Note that extending this formula to all of T_n, $L_n = \{0, \ldots, n-1\} \times \{0\}$ would form the minimal elements.

Figure 8 now explains the procedure for retrieving an element from anywhere in the triangle to a position of the form $(k, 1)$, without involving any positions larger than the retrived element. Figure 9 in turn explains how to move an element from position $(k, 1)$ to position $(0, 1)$. This concludes the proof. ☐

Since the line transports excess, we obtain that all superlines with the same filling closure and same cardinality are in the same orbit. The representatives informed by the previous proof are as follows: for $0 \leqslant k \leqslant n(n-1)/2$, we denote by $P_{n,k}$ the shape composed of a line of length n to which k points were added by filling the triangle under the line from top to bottom, and at each height from left to right. Examples are shown in Figure 10.

Theorem 2 (Characterisation of the orbits). *If P is a finite pattern then there are integers $n_1, \ldots n_r$ and $k_1, \ldots k_r$ and vectors $\vec{v_1}, \ldots \vec{v_r}$ (uniquely determined by P) such that $P \rightarrow^* \bigcup_{i=1}^{r} (\vec{v_i} + P_{n_i,k_i})$, and the filling closures $\varphi(P_{n_i,k_i} + \vec{v_i})$ do not touch each other. Furthermore, $\sum_{i=1}^{r} n_i = |P| - e(P)$ and $\sum_{i=1}^{r} k_i = e(P)$.*

Proof. The claim $P \rightarrow^* \bigcup_{i=1}^{r} (\vec{v_i} + P_{n_i,k_i})$ follows by first observing that the orbit contains a union of superlines whose filling closures do not touch, and then using

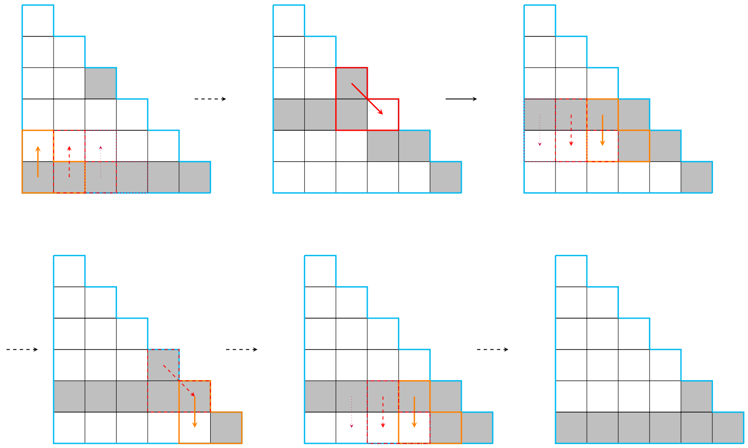

Fig. 8: Fetching an excess point.

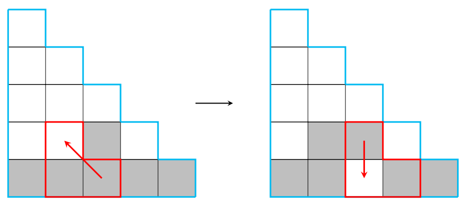

Fig. 9: Pushing excess to the left.

the fact the line transports excess to put each superline in normal form P_{n_i,k_i}. For uniqueness, observe that (up to permutation) the vectors $\vec{v_i}$ and values n_i are determined by the fill decomposition, and k_i by the cardinalities of the sets $P \cap \vec{v_i} + T_{k_i}$.

For the last claim, we recall that $e(P)$ is equal to the triangle excess defined as $e_T(P) = |P| - \sum_{i=1}^r n_i$, thus $\sum_{i=1}^r n_i = |P| - e(P)$. Since solitaire preserves cardinality,

$$|P| = |\bigcup_{i=1}^{r} (\vec{v_i} + P_{n_i,k_i})| = \sum_i (n_i + k_i)$$

so $e(P) = \sum_i k_i$. □

Corollary 1. *Let P be a pattern with fill decomposition $\varphi(P) = \bigcup_{i=1}^r \vec{v_i} + T_{k_i}$. Then P has no excess if and only if P is in the solitaire orbit of $\bigcup_i \vec{v_i} + L_i$.*

Corollary 2. *If P is a pattern, then $P \in \mathcal{O}(P_{n,k})$ if and only if $\varphi(P) = T_n$ and $e(P) = k$.*

Note that we now know the orbits for solitaire processes on the plane $\mathbb{Z}^2$ with all shapes of size 3. Indeed, such a shape is either linear (and thus easy to

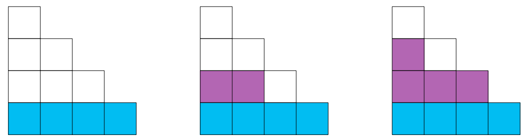

Fig. 10: From left to right: $P_{4,0}$, $P_{4,2}$ and $P_{4,4}$. The purple cells are the excess.

analyse), or it is a triangle shape on a finite index subgroup of $\mathbb{Z}^2$, and the orbits in different cosets of this subgroup are completely independent (Remark 1), and are individually described by the triangle solitaire.

5.3 Size of the line orbit

One can build an element of the orbit of the line by choosing a corner, then choosing a point on each line parallel to the edge opposed to the corner. This gives $3n! - 3$ elements of the orbit as the only patterns that can be created this way from two different corners are the lines.

If P is in the line orbit, then the number of points in P in the first k columns is at most k, and [3] gives the count $c\left(\frac{e}{2}\right)^n (n-1)^{n-\frac{5}{2}}$ with $c = \frac{4+2W(-2e^{-2})}{e^3\sqrt{2\pi}}$ for patterns with this property where W is the inverse function of $f : z \mapsto ze^z$ (Lambert's W function).

Combining these, we get the following bounds on the size of the line orbit.

Theorem 3. *There are constants* c_1 *and* c_2 *such that* $c_1 e^{-n} n^{n+\frac{1}{2}} \leqslant |\gamma(L_n)| \leqslant c_2\left(\frac{e}{2}\right)^n (n-1)^{n-\frac{5}{2}}$.

5.4 Excess sets for the triangle

We show that excess sets do not behave as intuitively as one might hope, even for our simple triangle example. Namely, it is tempting to think that if a set has excess, then we can remove some of its points to remove the excess, i.e. that the size of a maximal excess set in P always matches the excess $e(P)$. This is not true.

Theorem 4. *For the triangle solitaire,*

1. *there exist patterns with arbitrarily large excess, which contain no excess sets (in particular, phantom excess is unbounded),*
2. *visible excess (equivalently phantom excess) can vary by an unbounded amount within one solitaire orbit,*
3. *there exist patterns where there are maximal excess set of different cardinality.*

Proof. For the first claim, consider the pattern P in Figure 11. It is clear that its filling closure is the triangle, so $e(P) = 1$. However, this pattern has no nonempty excess sets. To see this, observe that the process of joining triangles always merges two existing triangles to a larger one, namely two T_1s to T_2, then two T_2s to T_4, and then two T_4s to T_7. If any element is removed, then one of these pairs cannot be joined.

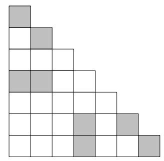

Fig. 11: Here, $e(p) = 1$ but $E(P) = \{\varnothing\}$ so $\hat{E}(P) = 0$.

We see that in fact when any element is removed, the entire bottom left corner of the triangle is left empty. This allows us to continue the construction inductively: We can add another triangle below whose *top* does not become filled if any element is removed. Then together these triangles will fill a large triangle to the bottom right from P, but if any element is removed, then this triangle will stay empty in the filling closure. We obtain more and more excess at each step, but still have no nonempty excess sets.

For the second claim, observe that the pattern we constructed (with arbitrarily large phantom excess) has in its orbit a pattern containing the corresponding line, and by Lemma 12 any such pattern has excess equal to its visible excess, since the line itself has no excess.

For the third claim, consider Figure 12. □

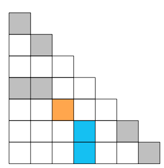

Fig. 12: The blue and orange sets are both maximal excess sets but do not have the same cardinality.

6 A brief look at other shapes on the plane

6.1 The square shape

In this section we study the solitaire action induced by the 2×2 square shape $\{0, 1\}^2$. We show that the theory can be developed entirely analogously as for the triangle.

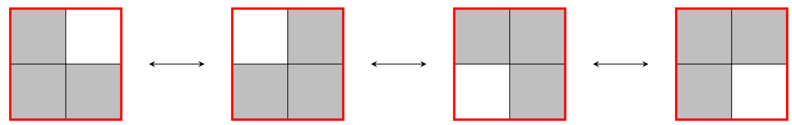

Fig. 13: The action of the square shape.

The action of the square rotates a square with a single empty cell as depicted in Figure 13.

The filling process for the square shape of course completes squares which are missing a single point. By analogy with the triangle shape, we expect the orbit of a finite pattern to be contained in non touching rectangles which can be computed by filling the initial pattern, and those orbits to be recognisable with only the shape of the filling and by counting the amount of excess elements in each rectangle.

With respect to the square, a good notion of *neighbours* of a point are the eight points surrounding it, and the neighbourhood of a pattern is the set of points which have two adjacent points of the pattern as neighbours. (Note that this is precisely the condition for being able to apply a square solitaire move.)

Examples are given in Figure 14. The notion of touching is then again defined

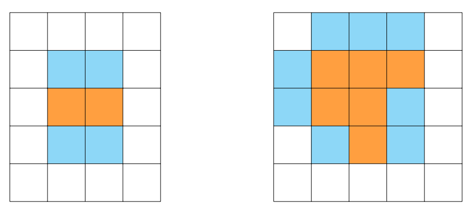

Fig. 14: The neighbours of the orange pattern are the blue cells.

as A touches B if $A \cap (B \cup N(B)) \neq \varnothing$.

Rectangles and crosses. As already indicated, the natural analog of the triangle T_n for the square solitaire is a *rectangle* $R_{m,n} = \{0, \ldots, m-1\} \times \{0, \ldots, n-1\}$. A natural analog of a line L_n for square solitaire is in turn a horizontal line intersecting a vertical line. We call such a pattern a *cross*. Notice that the square-shape filling of a cross is a rectangle of width the length of the horizontal line and height the length of the vertical line.

Lemma 24. *All crosses that fill the same rectangle are in the same orbit.*

Proof. Using the moves described in Figure 15, one can shift the vertical line to the left. By reversing the moves, the line shifts to the right. By rotating them, the horizontal line can be shifted up or down. □

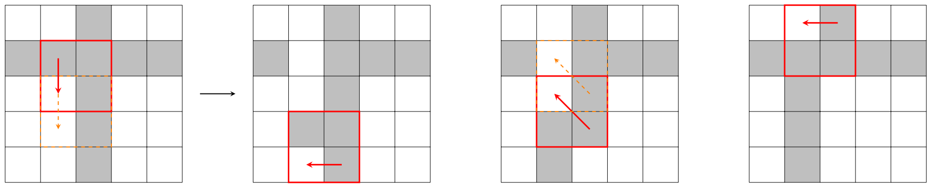

Fig. 15: How to shift the vertical line of a cross to the left. All the other movements are symmetrical.

Shape of the square filling

Lemma 25. *With respect to the square shape, the filling of any pattern P is a set of non touching rectangles $R_1, \ldots, R_k$ whose sizes satisfy $\sum_{i=1}^{k}(w(R_i) + h(R_i) - 1) \leqslant |P|$, where $w(R_i)$ and $h(R_i)$ are respectively the width and height of R_i.*

Again, we refer to $\bigcup R_i$ as the fill decomposition of P, and its uniqueness is clear.

Proof. Let P be a pattern. First, notice that a maximal connected subpattern Q of $\varphi(P)$ has no neighbour because otherwise a cell could be filled.

Now consider a maximal connected subpattern Q of $\varphi(P)$ that is not a rectangle. Then Q has an inner corner, thus Q has a neighbour. Hence the result on the shape of the filling.

We now prove the inequality on the size of the rectangles by induction on $|P|$.

If P is a single point, then $\varphi(P) = P$ is a rectangle of length and height 1 so the inequality holds.

Now assume the inequality holds for all patterns of size n, and let P be a pattern of size $n + 1$. Let $x \in P$, $|P \setminus \{x\}| = n$, so the rectangles that compose $\varphi(P \setminus \{x\})$ satisfy the inequality. If x is in one of those rectangles, then $\varphi(P) = \varphi(P \setminus \{x\})$, and we have the inequality. If x is not in the neighbourhood of a rectangle then $\varphi(P) = \varphi(P \setminus \{x\}) \cup \{x\}$, and again the inequality holds.

Lastly, if x is in the neighbourhood of rectangle R_i, for example x is adjacent to the right size of R_i, then the filling adds a column to rectangle R_i, augmenting its length by 1.

The rectangles may intersect, triggering a sequence of merges. But if R_i intersects the neighbourhood of another rectangle R_j, then the filling merges them into a rectangle R that satisfies $w(R) \leqslant w(R_i) + w(R_j)$ and $h(R) \leqslant h(R_i) + h(R_j)$, and at least one of these inequalities is strict. From this, it follows that the inequality continues to hold when rectangles are merged. □

Characterisation of the orbits. If the filling of pattern P is $\bigsqcup_{i=1}^{r} R_i$ then its square excess is $e_S(P) = |P| - \sum_{i=1}^{r}(w(R_i) + h(R_i) - 1)$. We can see as in the case of the triangle, that square excess is equal to excess in the abstract sense.

For a, b, k integers, let $L_{a,b,k}$ be the L-shape of length a and height b with k points in the rectangle in generates, filled from left to right and bottom to top. Examples are given Figure 16.

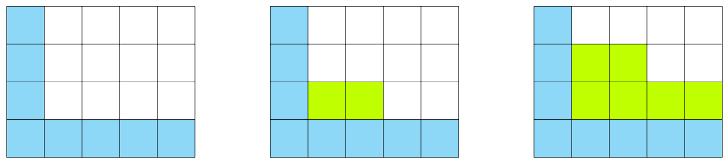

Fig. 16: From left to right, $L_{5,4,0}, L_{5,4,2}$ and $L_{5,4,6}$.

Lemma 26. *Every cross transports excess.*

Proof. By Lemma 17 and Lemma 24, it suffices to show this for one cross. We use the cross with the left and bottom edges of the rectangle, i.e. $L_{m,n,0}$.

Excess transportation can be now shown using Lemma 16 by using the lexicographic order, and is even easier than the corresponding process for the triangle. Namely, by the procedure in Lemma 24 we can move the vertical line of a cross so that it is next to an excess point, then move it down, then move the vertical line back to the left, and transport the excess point to get the final pattern $L_{m,n,1}$.

This process only uses elements below the excess point in lexicographic order, and we conclude from Lemma 16 that the cross transports excess. □

Theorem 5 (Characterisation of the orbits of the square). *If P is a pattern, then it is in the orbit of the pattern composed of $L_{a,b,k}$-shapes corresponding to the fill decomposition of $\varphi(P)$ and adding the excess present in the rectangle. I.e., if $\varphi(P) = \bigcup_{i=1}^{k}[(x_i, y_i), (x_i + a_i, y_i)] \times [(x_i, y_i), (x_i, y_i + b_i)]$ and $k_i = |P \cap R_i| - (a_i + b_i - 1)$ then $P \in \mathcal{O}\left(\bigcup_{i=1}^{k} (x_i, y_i) + L_{a_i,b_i,k_i}\right)$.*

Proof. This is proved analogously as for the triangle: we first show that we can apply the solitaire process to turn P into a disjoint union of supercrosses (patterns containing a cross, whose filling closure is the same as that of the cross) whose filling closures do not touch. The union of the filling closures of these super-crosses form the filling closure of P.

For this, we observe analogously to the case of the triangle, that two touching supercrosses can be merged into a supercross. Thus, a pattern in the orbit of P which is a union of a minimal number of supercrosses will in fact be a union of non-touching disjoint supercrosses.

Finally, we can use the previous lemma (the excess transportation property of crosses) to put the super-crosses in the stated normal form. □

6.2 General shapes S admit interaction without merging

Polygons and fillings. Consider a general non-linear shape $S \Subset \mathbb{Z}^2$, i.e. not contained in any affine line. We say $S \subset \mathbb{Z}^2$ is *discrete-convex* if it is the intersection of a geometrically convex subset of $\mathbb{R}^2$ and $\mathbb{Z}^2$.

Recall that $P \to Q$ is an S-solitaire move if there is a vector $x \in \mathbb{Z}^2$ such that $|P \cap (\vec{x}+S)| = |Q \cap (\vec{x}+S)| = |S|-1$ and $P \triangle Q \subset \mathcal{P}_2(\vec{x}+S)$. An S-filling step consists in adding to a pattern such that $|P \cap (\vec{x}+S)| = |S|-1$ the missing point of $\vec{x}+S$. This process is confluent and when it exists, we denote its limit $\varphi_S(P)$.

Previously, the filling of a pattern had the "same shape" as the shape used for the solitaire moves. The following definition aims at generalising this idea for arbitrary S.

Definition 12. *Given a shape S, an S-*polygon *is a pattern obtained by intersecting $\mathbb{Z}^2$ with a geometric polygon in $\mathbb{R}^2$ with the same edge directions as the convex hull of S, each edge being at least as long as the corresponding edge in S's convex hull (in terms of the number of lattice points). We define the S-*hull *of P as $H_S(P) = Q \cap \mathbb{Z}^2$ where $Q \subset \mathbb{R}^2$ is the smallest polygon with (at most) the same edge directions as S containing P.*

Notice that an S-hull is not necessarily an S-polygon since some edges might be too short.

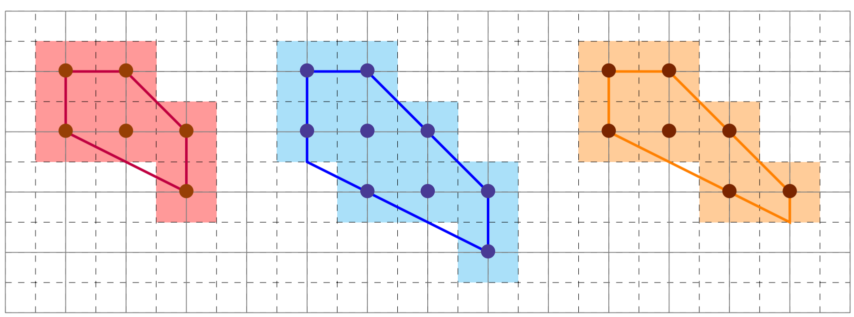

Fig. 17: With S the red shape, the blue pattern is an S-polygon but not the orange one because the right side is too short. Here the solid lines correspond to standard generators of $\mathbb{Z}^2$, and dotted lines show the dual lattice of $\mathbb{Z}^2$-centered unit squares.

The following two lemmas are essentially the two-dimensional version of Proposition 1. We give direct (slightly informal) proofs.

Lemma 27. *The S-hull of every set is well-defined, and is finite.*

Proof. For each edge e of S, there is a corresponding normal vector $\vec{v}$ (pointing outside), and points with positive dot product with $\vec{v}$ define a half-space H_e.

Given $P \subset \mathbb{R}^2$, we can see P as a subset of $\mathbb{R}^2$ and slide copies of all these half-planes H_e as close to P as possible, and finally take the convex polygon spanned by their edges. □

Unlike in the case of the triangle and square, there nevertheless does not always exist a smallest S-polygon containing a given finite set $P \subset \mathbb{Z}^2$, because the edges in the S-hull can be too short. This is in particular the case when P is smaller than S, an example is given in Figure 18.

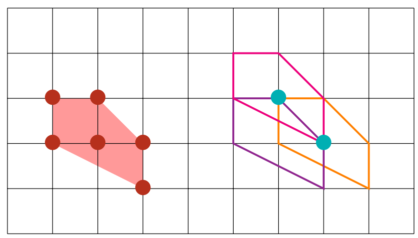

Fig. 18: With S the red shape, the blue pattern has 3 smallest S-polygons containing it.

Lemma 28. *The S-filling of pattern P is contained in its S-hull.*

Proof. Let L be a line parallel to an edge on S, then the filling process can add points on the outer side of the line only if there is already some point on the outer side. Therefore the filling process cannot add points outside the S-hull. □

In the case of the triangle shape and the square shape, we were able to identify a set of patterns (triangles and rectangles) and a notion of touching, such that non-touching patterns evolve entirely independently in the solitaire (if the patterns do not touch), or they merge into a larger pattern of the same kind.

In fact, these sets were precisely S-polygons: for a triangle, the sets T_n are precisely the triangle-polygons, and rectangles $\{0, \ldots, m-1\} \times \{0, \ldots, n-1\}$ are precisely the square-polygons.

The following example illustrates that, unlike in the case of the triangle and the square, for a general shape S we can have two S-polygons sets which do not "merge" (in that their union is closed under filling, and is not an S-polygon), yet the solitaire process can in a nontrivial way share a hole between the two sets.

Example 1. Consider the red shape S in Figure 19, and the pattern P formed by the orange, green and blue points. One can apply solitaire moves to the left (orange) of right (blue) part separately, moving the hole arbitrarily.

If we move the hole in the orange part to the left, then we can borrow a green point from the middle. This can freeze the process in the blue part, until the element is returned.

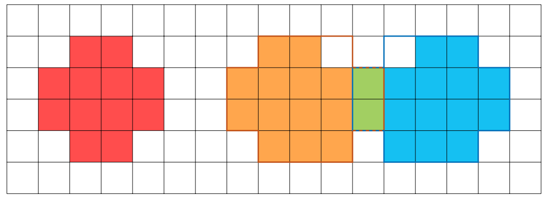

Fig. 19: An example of "non-merging shapes" with non-trivial interaction.

The filling closure of the colored area is the area with the orange and blue outlines, i.e. the filling process adds only two points. This limit does not fit our notion of an S-polygon.

Remark 3. For large S, one can form long chains out of such non-merging S-polygons, and even cycles and other planar graphs, where elements can be exchanged between the various parts, but which do not merge into a discrete-convex pattern.

We do not yet have clean descriptions of the processes described in the previous example and remark, but we can make are least a vague conjecture:

Conjecture 1. There is a notion of S-components such that the maximal S-components of a pattern closed under filling are either S-polygons or single points (those component being not necessarily disjoint), and the solitaire can be understood as in the triangle and square case, except that distinct S-components may share points.

Contours and Orbits. One thing that does work with general shapes is the notion of a contour from [7]. It provides a natural generalization of the line in triangle solitaire, and the cross (or an L-shape) in the case of square solitaire.

Here we find it helpful to work with more general groups G, to avoid getting caught up in geometric details, although our particular emphasis will be on the groups $\mathbb{Z}^d$.

Recall that a total order $<$ on (the elements of) a group G is *bi-invariant* if $a < b \iff ca < cb \iff ac < bc$ for all $a, b, c \in G$. Such orders exists for example for free abelian groups and free groups.

Definition 13. *Let us call $s \in S$ a* corner *if there is a bi-invariant order on G such that s is the maximal element of S (it is then minimal for the inverted order, which is also bi-invariant).*[3]

The following fact is well-known:

[3] It seems that here we need a bi-invariant order, while [7] only needs a left-invariant order.

Lemma 29. *The (bi-)invariant orders on $\mathbb{Z}^d$ are fully described as follows: First order vectors according to their dot product with a unit vector $\vec{v}$. If the kernel of the projection to the line spanned by $\vec{v}$ has non-trivial intersection with $\mathbb{Z}^d$, then this intersection is a finitely-generated abelian group of smaller rank. Order it inductively.*

In particular, our notion of corner agrees with the usual notion of a corner in convex geometry.

Definition 14. *Suppose G is a group, $S \Subset G$ and suppose $c \in S$ is a corner of S. If P is a pattern, its S-*contour *with respect to $c \in S$ is*

$$C_{S,c}(P) := \{x \in P \mid xc^{-1}S \not\subset P\}.$$

Note that the set of x such that $xc^{-1}S \not\subset P$ can be expressed as "those x such that the if g is the translation such that $gc = x$, we have $gS \not\subset P$", since $gc = x \iff g = xc^{-1}$.

Note also that if $c = e_G$ (which we)may assume without affecting the solitaire process by translating the shape, then this simplifies to $\{x \in P \mid xS \not\subset P\}$.

Examples of contours on the plane are shown in Figure 20.

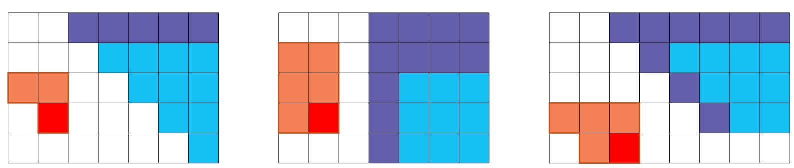

Fig. 20: The dark blue points are the contour of the light blue polygon for shape S in orange with respect to the red corner.

The following result is essentially from [7]. For the proof, we need the coencept of a TEP subshift discussed in Section 8 (here, we only cite a fact about TEP subshifts from [7], so it is not necessary to take a look at the definition).

Proposition 3. *Any two contours of a given pattern with the same shape have the same cardinality.*

Proof. By [7, Lemma 28], if we pick any TEP subshift with the given shape $S \Subset G$, then all contours of a domain $P \subset G$ can be filled arbitrarily and their contents uniquely determine a valid P-pattern. (In the terminology of Section 8, the contours of $P \Subset G$ are independent and P-spanning sets.)

The number of patterns with shape $C_{S,c}(P)$ is then precisely the logarithm (with base the size of the alphabet) of the number of legal patterns on the domain P. In particular, it does not depend on the choice of c. □

□

The above proof is rather indirect, as it involves the choice of an arbitrary TEP subshift. We show that under some conditions, one can apply the solitaire process to move between two contours.

Definition 15. *Suppose $c, c' \in S$ are two distinct corners of S. We say c and c' are* sweep swappable *if there exists a bi-invariant total order on G such that $c = \min S$ and $c' = \max S$ with respect to this order.*

Lemma 30. *Let $c, c' \in S$ be sweep swappable. Then for any $P \Subset G$, the contours $C_{S,c}$ and $C_{S,c'}$ are in the same solitaire orbit.*

Proof. Let $<$ denote a total bi-invariant order on G such that $c = \min S$ and $c' = \max S$. By translating S on the left (which preserves solitaire moves and contours) we may assume $c' = e_G$. Let $C, C' \Subset P$ be the contours corresponding to c, c' respectively. Order $S \setminus C$ in increasing order as $g_1, \ldots, g_n$. Now define $C_0 = C$ and inductively $C_i = (C_{i-1} \cup \{g_i\}) \setminus \{g_i c\}$. We claim that this is a valid solitaire move, namely the solitaire move at g_i that moves the hole from $g_i c$ to g_i.

For this, we need to show that $|g_i S \cap C_{i-1}| = |S| - 1$, and that $g_i \notin C_i$. The latter fact is clear since g_i is larger than any g_j added previously. For the former, observe first that the filling process following this solitaire process (i.e. the result of applying filling moves at each $g_j S$ for $j < i$, which can be proven well-defined by induction) would certainly have added the remaining $|S| - 1$ elements, since they are smaller, thus either on the list of the g_j, or in the initial contour C.

Thus it suffices to show that the application of a previous solitaire move at $g_j S$ taking us from $C_{j-1} \to C_j$ did not remove any of the elements in $g_i S$. But the removed element when applying a move at g_j is always $g_j c$. Since $c = \min S$, and the order is bi-invariant, $g_j c \leq g_j s \leq g_i s$ for any $s \in S$, with equality only if $j = i$ and $c = s$.

We claim that we now have $C_n \supset C'$, equivalently $P \setminus C_n \subset P \setminus C'$. To see this, suppose that $gc \in P$ but $gc \notin C_n$. From $gc \in P$ we have that $gc \in C_k$ for some k. From $gc \notin C_n$ we have that gc is of the form $g_i c$ for some i (since it was removed at some point). The fact it was indeed removed means that $gS \subset P$ so $gc \notin C'$. ◻

In particular, all contours are in the same solitaire orbit whenever all the corners are in the same equivalence class under the transitive closure of sweep swappability.

Lemma 31. *If $S \subset \mathbb{Z}^2$ is a shape whose convex hull does not have two pairs of parallel edges, then all its contours are in the same solitaire orbit.*

Proof. Consider the following representation of polygon S : let $\vec{u_1}, \ldots, \vec{u_k}$ be the vectors normal to the edges of S, oriented toward the exterior of S. Take a circle C with centre O and represent edge e_i by the intersection between C and the half line with origin O and direction $\vec{u_i}$, and each corner by the arc between the two edges. We claim that corners a and b are sweep swappable if a diameter of C joins there arcs.

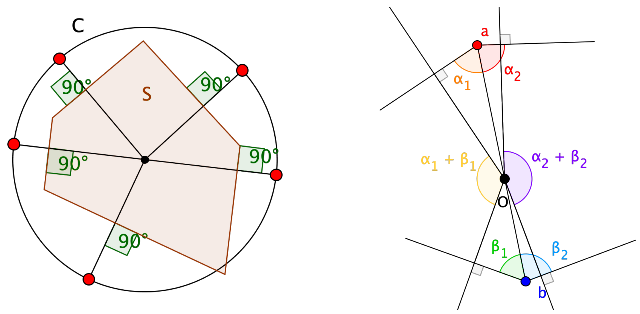

Fig. 21: On the left, an illustration on the circle representation of a polygon. On the right, the notations used for the proof.

Note that the open arcs corresponding to corner a consists of the possible directions of unit vectors $\vec{v}$ which make a a minimal (or maximal) element of the corresponding order Lemma 29. Thus if a diameter joins two arcs, then the corners are sweep swappable, thus the corresponding contours are in the same solitaire orbit for all patterns.

Now notice that if there are no opposite points (which corresponds to a pair of parallel edges) on the circle, then any border can be crossed by crossing the circle twice. One pair of opposite points is not a problem because we can simply move along the circle in the opposite direction. □

In the case of $\mathbb{Z}^2$, we can in fact describe a process that allows to transform one contour into another it cannot be swapped with.

Theorem 6. *In dimension 2, all the contours of a pattern are in the same orbit.*

Proof. If S has two parallel edges e and e', let c and c' be the extremities of respectively e and e' on the same side (such that the line from c to c' is not a diameter). Then the contours C and C' with respect to c and c' can be exchanged as follows. In C, as the other end of e', b', is diametrically opposed to c, an S-solitaire step can be done there. Follow edge e' from b' to c' and at each step move the point in corner b' to c. Then follow e' from c' to b' and move the point in corner c' to b'. Repeat at each line until you get C.

If looked at in the circle representation, this process allows to exchange corners whose section can be joined by a line parallel to a diameter, thus allow to cross it. Combined with Lemma 31, it follows that all contours are in the same orbit. □

7 Quantitative and algorithmic aspects of convex solitaires

7.1 Complexity of the identification of the orbit of a pattern

The characterisation of the orbits through filling and excess provides a polynomial time algorithm to identify to which orbit a given pattern belongs for the triangle and square solitaires.

The algorithm is the following for the triangle shape:

Algorithm 1 (Identify orbit) *Data: pattern P. Result: the canonical representative of the orbit of P.*

1. *Fill the pattern.*
2. *Divide the filling into triangles $\vec{v_1} + T_{k_1}, \dots, \vec{v_r} + T_{k_r}$.*
3. *Count the excess in each triangle, the canonical representative of the orbit of the pattern is $\bigcup_{i=1}^{r} \vec{v_i} + P_{k_i, e(P \cap (\vec{v_i} + T_{k_i}))}$.*

It can easily be adapted to the square shape, and some other well behaved smalls shapes.

This algorithm has a total time complexity of $O(n^2)$ where $n = |P|$: The first two steps are linear in the number of points in $\varphi(P)$ and $|\varphi(P)| \leqslant \frac{n(n+1)}{2}$ for the triangle shape so they run in time $O(n^2)$. Step 3 is then linear so the total time complexity of the algorithm is $O(n^2)$.

For shapes other than the triangle, it still holds for S-polygons that $|P| \leqslant |C_S(P)|^2$ so the bound remains the same.

7.2 Number of steps needed to put a pattern in normal form

One can use the argument in Lemma 20 and Lemma 21 to successively merge superlines following the process of merging triangles in the filling process.

A single step of adding a new element into an existing superline takes at most $O(n^2)$ steps: the process of turning one edge into another using the solitaire process clearly takes $O(n^2)$ steps, and adding an element once we have the correct edge takes $O(n)$ steps. Thus, after $O(n^3)$ steps we have turned any pattern of cardinality n into a union of superlines.

It suffices to show that a superline can be turned into the canonical representative $P_{n,k}$ in $O((n+k)^3)$ steps.

For this, the argument in Lemma 16 can be unraveled to give a practical procedure for fetching the excess one element at a time, at all times keeping the excess lined up, i.e. so that at all times the pattern is of the form $P_{n,k} \sqcup R$.

One can use the procedure in Figure 9 to move all excess to the left on the top line, and the one in Figure 22 to move excess from the top line to the line below, if it is not yet full. Figure 23 illustrates how one can fetch elements if excess is already lined up.

Algorithm 2 (Transform a pattern to its canonical form)

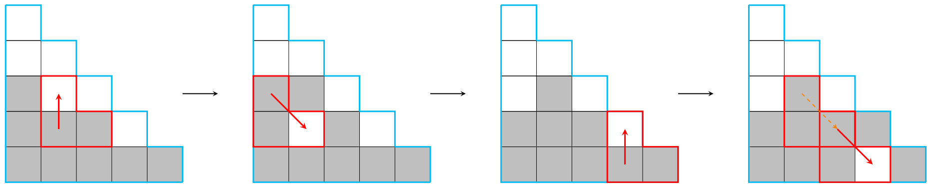

Fig. 22: Pushing excess to the right.

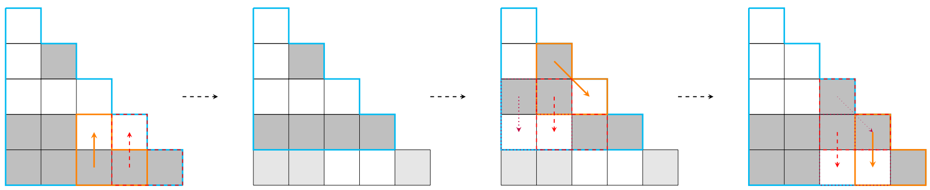

Fig. 23: Fetching an excess point with some excess already lined up.

1. *Merge the different components and form superlines using the process described in Lemma 21*
2. *Fetch the excess with the process described in Lemma 16.*

For the square shape, the same algorithm can be applied by replacing superlines by supercrosses and noticing than an excess point can very easily be moved along a line.

Indeed, naively implemented, this algorithm takes $O(n^2(n+k))$, where k denotes the excess: The first step takes $O(n^3)$, since each merging takes $O(n^2)$ time and we merge at most n times. The second step takes $O(n^2k)$ if we fetch the k many excess points one by one, as each fetch takes $O(n^2)$.

One can in fact achieve $O(n^2+nk)$. Such a process for the triangle (resp. square) is as follows:

1. Form a vertical line (resp. a bottom left cross).
2. Move the line from left to right, and after moving it by one column, use it to move down all the excess point to its left to the bottom of there column. You now have a horizontal line (resp. a bottom right cross) with piles of excess on it.
3. Move the excess points to the right or left to form horizontal lines on top of the bottom one.

Step one requires $O(n^2)$ moves. Step 2 requires again $O(n^2)$ moves for the line movement from left to right and at most n moves to move each excess point down since one point cannot go down by more than n so a total of $O(kn)$ moves. Step 3 takes $O(kn)$ moves for the same reason.

Theorem 7. *For the triangle and the square solitaire, the orbit of a pattern with n elements has diameter at most $O(n^3)$.*

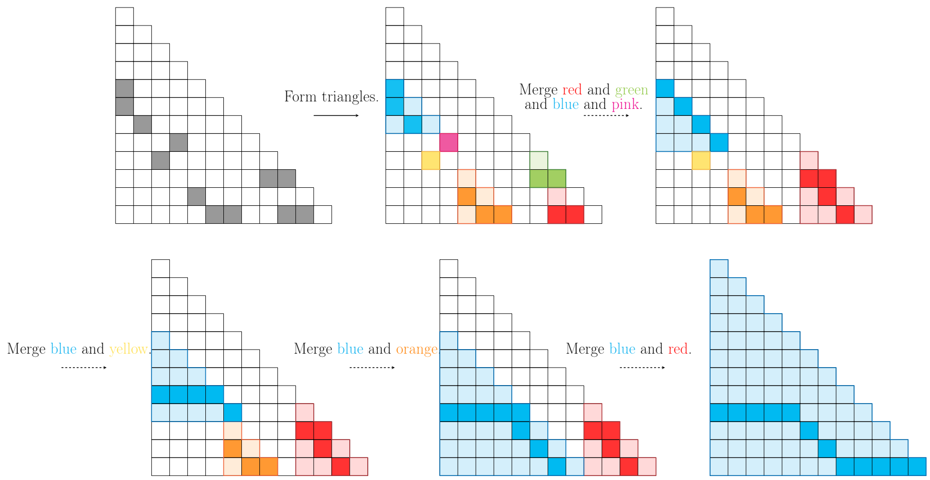

Fig. 24: An example of how to get back to the line from a random element of its orbit.

This is in fact optimal for the line orbit in the triangle solitaire.

Theorem 8. *The diameter of the orbit of the line of length n, seen as a graph, is $\Theta(n^3)$.*

Proof. We are going to build an infinite family of patterns that require $\Omega(n^3)$ steps to get back to the line.

Let P_0 be the empty pattern. P_{n+1} is inductively built by extending P_n as described in Figure 25 where the grey triangle is the triangle in which P_n's orbit is confined. In pattern P_{n+1}, $\Omega(n^2)$ steps are required to fetch the three coloured

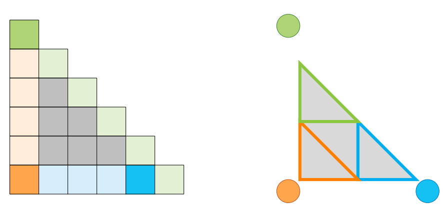

Fig. 25: Left: The extension of P_n into P_{n+1}. Right: A schematic representation of P_{n+1} used in the proof.

points.

Indeed, first notice that up to renaming points, the blue point has to move right to fetch the orange one, then the orange one will have to move down to fetch the green one and finally the green one will have to go up to prepare for the next extension. Now let us analyse the movement of the blue point, starting from the first moment it touches another triangle. To move a point right, one need a point in the column at its right. We'll prove that this means that at some point in the process, half of the points of the pattern have to be in the left half of the triangle.

Mark the blue point, and whenever an unmarked point is in the same column as a marked one, mark it. Consider column i from the left, and the first marked point x to reach it. Clearly the point x moves from column $i-1$ to column i, and this requires us to have an unmarked point in column i to allow this. This point was unmarked, and is now marked. Therefore, when the blue point has reached column $\frac{3n}{2}$ (the middle column), there are at least $\frac{3n}{2}$ marked points, all of which were in the blue triangle in Figure 25 at some point during the journey of the blue point.

The same reasoning on the orange and green point gives that at some point of the process, between the first moment the blue point moves, until the point where we are ready to move the blue point of the next level, $\frac{3n}{2}$ points were in the orange triangle and the same amount in the green one. As the pattern P_n only has $3n$ points, at least $\frac{3n}{2}$ need to be moved from one subtriangle to another, therefore there is a pair of triangles that will share at least $\frac{n}{2}$ points. Those at least $\frac{n}{2}$ points will need to be moved by a mean distance of at least $\frac{n}{4}$, which requires at least $\frac{n^2}{8}$ steps. Thus pattern P_n requires $\Omega(n^3)$ moves to be transformed into a line. □

Conjecture 2. The orbit of a contour of cardinality n of an S-polygon has diameter $\Theta(n^3)$

A good candidate for patterns far from a contour are the ones built by adding points around a pattern so that each point extends the filling closure of the pattern by a line, and adding the points on successive sides so that the contour has to be changed between the mergings. Such a pattern for the square is shown in Figure 26. The natural method to put it in normal form takes $\Theta(n^3)$ moves. We have not proved that this is optimal, but suspect this can be shown analogously to the triangle case.

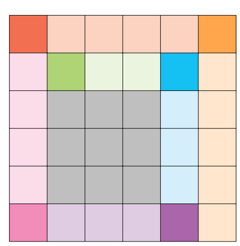

Fig. 26: A pattern we conjecture is at $\Theta(n^3)$ steps from the cross. P_n (gray) is extended into P_{n+1} by adding the green, blue, purple and pink points, red and orange are here to show how it keeps growing.

7.3 Distance between two contours

Proposition 4. *For any shape S, the transformation from one contour to another cannot be done in $o(n^2)$ steps, for some S-polygons of cardinality n.*

Proof. We define a metric on the set of finite subsets of $\mathbb{Z}^2$ with a given cardinality by $\Delta(A,B) = \min_{f:A\sim B}\sum_{a\in A} d(a,f(a))$ where d is the Euclidean distance on $\mathbb{Z}^2$. As any move with the solitaire moves exactly one point, by a distance at most the diameter of S, the number of moves needed to transform pattern A into pattern B is $\Omega(\Delta(A,B))$.

Let us first consider the triangle. Let $n \in \mathbb{N}$, denote $H_n = [|0; n-1|] \times 0$ and $V_n = 0 \times [|0, n-1]$ respectively the horizontal and the vertical line of length n with origin 0.

$$\Delta(H_n, V_n) = \min_{f:H_n\sim V_n} \sum_{x\in H_n} d(x, f(x)) = \min_{\sigma\in\mathfrak{S}} \sum_{k=0}^{n-1} \sqrt{k^2+\sigma(k)^2} \geqslant \sum_{k=0}^{n-1} k = \frac{n(n-1)}{2}$$

.

Let S be now a general convex shape, and suppose S has at least 4 sides (the case of 3 edges being similar). Notice that because of the triangular inequality, any bijection that reaches $\Delta(A,B)$ must fix every point of $A\cap B$. Consider a family of S-polygons P_n for $n \in I \subset \mathbb{N}$ infinite such that the contours of P_n have size n and there is a constant c such that every edge of P_n has length at least $\frac{n}{c}$ and two opposite edges have distance at least $\frac{n}{c}$. One can be built by choosing any S-polygon, then taking the family of polygons obtained by multiplying the lengths of its edges by an integer.

Let C and C' be two contours of P_n. P_n is convex, so the hypothesis on the length and distance of edges implies that any two non consecutive edges of P are at distance at least $\frac{n}{c}$ of each other.

- If C and C' differ by exactly one edge, then there has to be three consecutive edges of P, e, f, e' such that $e \in C\backslash C'$, $e' \in C'\backslash C$ and $f \notin C\cup C'$. Since e and e' are not consecutive so $\Delta(C,C') \geqslant \Delta(e,e') \geqslant d(e,e')\min(|e|,|e'|) \geqslant \left(\frac{n}{c}\right)^2$.
- Else, C and C' differ by two edges, $e, f \in C$ and $e', f' \in C'$. If they are not consecutive, the inequality is obtained just as above.
- If they are consecutive, for example f follows f', then by the triangular case $\Delta(f,f') \geqslant \alpha\left(\frac{n}{c}\right)^2$ with $\alpha > 0$, by the remark above $\Delta(e,f'), \Delta(e',f) \geqslant \left(\frac{n}{c}\right)^2$ and combining both (depending on whether e and e' are consecutive or not), $\Delta(e,e') \geqslant \alpha'\left(\frac{n}{c}\right)^2$. Thus $\Delta(C,C') \geqslant \Delta(e\cup f, e'\cup f') \geqslant \alpha''\left(\frac{n}{c}\right)^2$

□

This shows in particular that Proposition 2 is roughly optimal:

Corollary 3. *For the triangle solitaire, the distance between two distinct edges of T_n is $\Omega(n^2)$.*

In particular, while we like to to think of the line as the "center" of the line orbit, this makes sense only when we consider $O(n^2)$ to be a small distance (which makes some sense because the diameter $\Omega(n^3)$ is much larger). We do not know if there are elements in the line orbit which are essentially closer to their rotations (by the matrix M).

8 Connection to TEP and permutive subshifts

8.1 TEP and permutivity

Definition 16. *Let $C \Subset S \Subset G$ and let A be a finite alphabet. Suppose $P \subset A^S$ is such that $|P| = |A|^{|S|-1}$, and for each $c \in C$, the projection $\pi : P \to A^{S\setminus\{c\}}$ given by $\pi(p) = p|_{S\setminus\{c\}}$ is surjective. Then the set of $x \in A^G$ such that for all g, $g^{-1}x|_S \in P$ is called a (C, S)-*TEP subshift.

The TEP subshifts (with a specific choice of C for a given shape S, discussed below) were introduced in [7] by the first author. Here, we develop the connection to solitaire in slightly more generality. (In [7], the definition is in fact generalized in another direction, to "k-TEP subshifts", but they do not seem to admit a notion of solitaire.)

Much of the theory makes sense under the weaker assumption of permutivity, which we defined below.

Definition 17. *Let G be a group, and let $C \Subset S \Subset G$. Let A be a finite alphabet and $X \subset A^G$ be a closed set. We say X is (C, S)-permutive if for each $c \in C$ and $g \in G$ there exists a function $f_g : A^{S\setminus\{c\}} \to A$ such that for every $g \in G$ we have*

$$g^{-1}x|_S = p \implies f_g(p|_{S\setminus\{c\}}) = p_c.$$

An (S, S)-permutive subshift is called simply S-permutive.

Such a set X may not be a subshift. When all the functions are the same, i.e. $\forall g \in G : f_g = f$, the condition simplifies to

$$\forall p \in X|_S : f(p|_{S\setminus\{c\}}) = p_c,$$

which is a property of TEP subshifts (but the assumption about surjectivity of projection to $S \setminus \{c\}$ is missing).

A special case of permutivity was studied by [2] under the name polygonal subshifts. This special case is that the $S \Subset \mathbb{Z}^d$, C is the set of corners of S in the geometric sense, and X is a subshift.

The term *permutive* comes from the following lemma (whose proof is straightforward and omitted).

Lemma 32. *Let G be a group, and let $C \Subset S \Subset G$. Let A be a finite alphabet and $X \subset A^G$ be a closed set. If X is (C, S)-permutive, then for any $g \in G$ and distinct $c, c' \in C$, writing $T = G \setminus \{gc, gc'\}$, we have for all $x \in A^T$ a bijection $\pi : A \to A$ such that*

$$x \sqcup a^{gc} \in X|_{T\cup\{gc\}} \iff x \sqcup \pi(a)^{gc'} \in X|_{T\cup\{gc'\}}.$$

Furthermore, the permutation π only depends on $x|_{gS\setminus\{gc,gc'\}}$

8.2 Independent sets

Definition 18. *Let G be a group, and let $X \subset A^G$ be a closed set. The* topological independence sets $\mathrm{TI}(X)$ *are the sets $T \in \mathcal{P}(G)$ satisfying $X|_T = A^T$. If μ is a probability measure on X, then a finite set $T \Subset G$ is* equidistributed *if the $\mu([p]) = \mu([q])$ for any $p, q \in A^T$. The* μ-independence sets $\mathrm{I}_\mu(X)$ *are the closure of the equidistributed finite sets $T \Subset G$ in the product topology of $\mathcal{P}(G)$.*

Note that as stated, both types of independence sets depend on the alphabet A. One may always use the effective alphabet (symbols that actually appear in configurations) unless stated otherwise. Note that the assumption on μ in particular implies that if T is equidistributed then all patterns A^T are in the support of the marginal distribution of μ on T.

Theorem 9. *Let G be a group, let $C \Subset S \Subset G$ be as above, and let $X \subset A^G$ be a (C, S)-permutive closed set. Then the sets of topological independence* $\mathrm{TI}(X)$ *are closed under the (C, S)-solitaire. If μ is a Borel probability measure on X, then the μ-independence sets are closed under (C, S)-solitaire. Both sets are down sets under inclusion, and are topologically closed subsets of $\mathcal{P}(G)$. If X is shift-invariant, then* $\mathrm{TI}(X), I_\mu(X)$ *are shift-invariant sets.*

Proof. We start by proving the last three claims. First, the set $\mathrm{TI}(X)$ is obviously down, and $\mathrm{I}_\mu(X)$ is down because it is the closure of finite μ-independent sets, which in turn are down because for $T' \subset T \Subset G$, the measure on $A^{T'}$ is determined by the one on A^T by the natural projection from A^T to $A^{T'}$, where all fibers have equal cardinality.

Topological closure of $\mathrm{I}_\mu(X)$ is true by definition. For $\mathrm{TI}(X)$ it suffices to show that $X|_P = A^P$ if and only if $X|_Q = A^Q$ for any $Q \Subset P$. But this is immediate from the fact that $\pi : X \to X|P$ is a continuous map between two closed subsets of Cantor space (which are compact Hausdorff).

If X is shift-invariant, then shift-invariance of $\mathrm{TI}(X)$ and $\mathrm{I}_\mu(X)$ are trivial.

Now we show closure under solitaire. Let $T \subset G$ be a set of topological independence, meaning $X|_T = A^T$. Suppose $T \to_{C,S} T'$ meaning

$$|T \cap gS| = |T' \cap gS| = |S| - 1 \wedge T \triangle T' \in \mathcal{P}_2(gC).$$

Let $T \setminus T' = \{gc\}$, $T' \setminus T = \{gc'\}$ By the definition of (C, S)-permutivity of X, there are functions $f = f_{g,c} : A^{S \setminus \{c\}} \to A$ and $f' = f_{g,c'} : A^{S \setminus \{c'\}} \to A$, such that for every $g \in G$ we have

$$g^{-1}x|_S = p \implies f(p|_{S \setminus \{c\}}) = p_c$$

and

$$g^{-1}x|_S = p \implies f'(p|_{S \setminus \{c'\}}) = p_{c'}.$$

Let $U = T \cap T'$ and $V = T \cup T'$. Note that $T \setminus U, T' \setminus U$ are singleton sets contained in gC, and similarly for $V \setminus T, V \setminus T'$. Now, fix $x|_U \in A^U$, let $E \subset A$ be the set of symbols $a \in A$ such that $x|_U \sqcup a^{gc} \in X|T$, and $E' \subset A$ the set of symbols $a \in A$ such that $x|_U \sqcup a^{gc'} \in X|T'$. Clearly $E = A$ by the assumption that $T \in \mathrm{TI}(X)$. Our task is precisely to show that $E' = A$.

We give a bijection between these sets. For this, observe that for $x \in X$, the restriction $x|T$ uniquely determines $x|V$ and thus $x|T'$ (using the property of f'), and the restriction $x|T'$ uniquely determines $x|T$. In particular, there is a bijection between $X|T$ and $X|T'$, which can only change symbols in the symmetric difference of T, T'. This is only possible if indeed $E' = A$ (since $E = A$), concluding the proof for the sets of topological independence.

If μ is a probability measure, the proof is similar. Namely, suppose T is a limit of finite patterns Q such that μ is uniform on the finite patterns $X|_Q$. Then fix $Q \supset gS$ and let $\mathcal{P} = X|_Q$. We have that the distribution on the legal patterns $\mathcal{P}' = X|_Q \setminus \{gc\} \cup \{gc'\}$ is also uniform, because as above we have a bijection between $\mathcal{P}$ and $\mathcal{P}'$. Now T' is a limit of such patterns $Q \setminus \{gc\} \cup \{gc'\} = T'$. □

The main result of [7] is that if X is a (C, S)-TEP-subshift on the group G and $C \Subset S$ can be interpreted as a set of corners of S for an abstract convex geometry on G satisfying some additional properties, then the sets of topological independence form a subshift containing every finite convex set, and there is a shift-invariant probability measure μ on X such that all finite convex sets are equidistributed in μ (among the legal patterns).

Remark 4. We note that if μ is a Borel probability measure on a (C, S)-permutive X, then the set of $P \Subset G$ such that μ is equidistributed on the patterns $X|_P$ is also closed under the solitaire process, by a similar proof as above. This observation gives some motivation for studying the solitaire on sets that are are not independence sets for a TEP subshift.

The fact the topological independence subshift is closed under solitaire is the original motivation for the study of the solitaire process. The S-TEP subshifts are an interesting class of subshifts (generalizing the Ledrappier subshift and spacetime subshifts of bipermutive cellular automata), and the sets of topological independence are interesting class of subshifts of the group (which in the Ledrappier case forms a matroid [7]).

8.3 Spanning sets

We now define a kind of dual notion to topological independent sets.

Definition 19. *Let G be a group, $P \subset G$, and $X \subset A^G$ closed. Then $T \subset G$ is a* P-spanning *if there is a function $f : X|_T \to X|_P$ such that for all $x \in X$, $x|_P = f(x|_T)$.*

We use the term "spanning" as an analog of "independence", roughly as the terms are used in linear algebra and matroid theory. However, the G-spanning sets are often called *expansive* or *coding* sets. In the case $G = \mathbb{Z}^d$ an *expansive subspace* $H \leqslant \mathbb{R}^d$ is precisely one such that $H + B_r \cap \mathbb{Z}^d$ is G-spanning [1]. For the cellular automata minded it might also be natural to call this the *determined set*, as we have deterministic (local) rules for it in the following sense:

Lemma 33. *Let G be a group, and let $X \subset A^G$ be a closed set. Let T be a P-spanning set. Then for all $g \in P$ there exists a finite set $T_g \Subset T$ and a function $f_g : X|_{T_g} \to A$ such that for all $x \in X$, we have $x_g = f_g(x|_{T_g})$.*

Proof. Suppose there is a function $f : X|_T \to X|_P$ such that for all $x \in X$, $x|_P = f(x|_T)$. Note that the graph of this function is just $Y = \{(x|_T, x|_P) \mid x \in X\}$, which is the continuous image of the compact space X, thus Y is compact. A function with compact Hausdorff codomain and closed graph is continuous (this is one of the closed graph theorems from general topology [4]), thus we conclude that f is continuous.

Now the claim is proved as the Curtis-Hedlund-Lyndon theorem: the clopen partition $\{[a]_g \mid a \in A\}$ has clopen preimage in f, and inclusion in a clopen set depends on only finitely many coordinates. ☐

It is obvious that if T is both P-spanning and Q-spanning, then it is $(P \cup Q)$-spanning. It is also clear that the set of sets P for which T is P-spanning is closed in the product topology of $\mathcal{P}(G)$. This gives rise to the following definition.

Definition 20. *Let G be a group, $T \subset G$, and $X \subset A^G$ closed. The* spanned set *$\psi_X(T)$ of T is the maximal subset P of G such that T is P-spanning.*

Lemma 34. *Spanned sets form a closure system.*

Proof. By definition, we need to show $P \subset \psi_X(P)$ (extensivity), $P \subset P' \implies \psi_X(P) \subset \psi_X(P')$ (monotonicity) and $\psi_X(\psi_X(P)) = \psi_X(P)$ (idempotency). All of these properties are obvious from the semantics of the operator (and also easy to verify from the formal definition). ☐

Sets P such that $\psi_X(P) = P$ are called *spanning-closed*. Our particular interest is in spanning sets in TEP subshifts. The following is immediate from the definitions, but we give a detailed proof.

Lemma 35. *Let G be a group, let $C \Subset S \Subset G$ and let $X \subset A^G$ closed and (C, S)-permutive. Let $P \Subset G$. Then P is a $\varphi(P)$-spanning set. In other words, the spanned set always contains the filling closure, or in a formula $\psi_X(P) \supset \varphi(P)$ for all $P \subset G$.*

Proof. By definition of the filling process, we produce $\varphi(P)$ from P by constructing a sequence of sets $P_0 = P, P_1, \ldots$ such that P_{i+1} is obtained from P_i by taking g_i such that $|g_i S \cap P_i| = |S| - 1$ and $|g_i C \cap P_i| = |C| - 1$, and letting $P_{i+1} = P_i \cup \{g_i c_i\}$ where $g_i c_i \notin P_i$.

Now let $x \in X$ be arbitrary. We prove that $x|_{\varphi(P)}$ is uniquely determined by $x|_P$ by following the chain P_i. First, $x|_{P_0} = x|_P$ is trivially determined by $x|_P$. Now suppose $x|_{P_i}$ is determined already.

By the definition of (C, S)-permutivity, there is a function $f_{g_i} : A^{S \setminus \{c_i\}} \to A$ such that

$$g_i^{-1} x|_S = p \implies f_g(p|_S \setminus \{c_i\}) = p_{c_i}.$$

Note that $g_i^{-1}x|_S$ contains the same data as $x|_{g_iS}$, and $p_{c_i} = x|_{g_ic_i}$. so we are precisely saying that the pattern $x|_{P_{i+1}}$ is uniquely determined by $x|_{g_iS}$, in particular by $x|_{g_iS\setminus\{g_ic_i\}}$. The set $g_iS \setminus \{g_ic_i\}$ is contained in P_i since because $|g_iS \cap P_i| = |S| - 1$ and $g_ic_i \notin P_i$. Therefore, $x|_{P_{i+1}}$ is determined by $x|_{P_i}$. □

Lemma 36. *Let G be a group, let $C \Subset S \Subset G$ and let $X \subset A^G$ closed and (C, S)-permutive. Then (C, S)-solitaire preserves the spanned set.*

Proof. Suppose $P \to Q$. We know that solitaire preserves the filling closure, $\varphi(P) = \varphi(Q)$. Then we calculate

$$\psi_X(P) = \psi_X(\psi_X(P)) \supset \psi_X(\varphi(P)) \supset \psi_X(Q).$$

The other inclusion is symmetric. □

Lemma 37. *Let G be a group, let $C \Subset S \Subset G$ and let $X \subset A^G$ closed and (C, S)-permutive. Then every spanning-closed set is filling-closed.*

Proof. Suppose $\psi_X(P) = P$. Then Lemma 35 we have $P \subset \varphi(P) \subset \psi_X(P) = P$ so $\varphi(P) = P$. □

It is tempting to conjecture that in natural situations, the spanned sets of a TEP subshift would in fact correspond to filling closures. We will see in Section 8.5 that this fails in the very classical Ledrappier subshift (which is the spacetime subshift of the two-neighbor XOR CA).

There is an obvious connection between the cardinalities of independent and spanning sets in a general subshift:

Definition 21. *Let G be a group, and let $X \subset A^G$ be a closed set. For $P \Subset G$, write $\mathrm{rank}_{\mathrm{span}}(P)$ for the minimal cardinality of a set $R \subset P$ such that $\psi_X(R) \supset P$. Write $\mathrm{rank}_{\mathrm{indep}}(P)$ for the maximal cardinality of a set $Q \subset P$ such that $X|_Q = A^Q$.*

Proposition 5. *Let G be a group, and let $X \subset A^G$ closed set. Then*

$$\mathrm{rank}_{\mathrm{indep}}(P) \leqslant \mathrm{rank}_{\mathrm{span}}(P)$$

Proof. The inequality

$$\mathrm{rank}_{\mathrm{indep}}(P) \leqslant \mathrm{rank}_{\mathrm{span}}(P)$$

holds in general (Without assuming permutivity). Namely, suppose R is such that $\psi_X(R) \supset P$. This means that $x|_R$ determines $x|_P$ for all $x \in X$, so in particular $|X|_R| \geqslant |X|_P|$. If $X|_Q = A^Q$ and $Q \subset P$, then certainly $|X|_P| \geqslant |A^Q|$, so we must have $|X|_R| \geqslant |A^Q|$ and taking base-$|A|$ logarithms we get $|R| \geqslant |Q|$.

This means every spanning set is larger than any independent set, in particular the minimal cardinality of a spanning set is larger than the maximal cardinality of and independent set, proving the inequality. □

Similarly, there is a connection between the filling rank and the spanning rank.

Proposition 6. *Let G be a group, $C \Subset S \Subset G$, and let $X \subset A^G$ be a (C,S)-permutive closed set. If P is closed under filling, then* $\mathrm{rank}_{\mathrm{indep}}(P) \leqslant \mathrm{rank}(P)$, *where* $\mathrm{rank}(P)$ *is computed with respect to (C,S)-solitaire.*

Proof. Let $R \subset P$ satisfy $\varphi(R) = P$. Then in particular R is a P-spanning set contained in P. Thus its cardinality is at least as large as that of a minimal P-spanning set. □

Alternatively, we can say that the filling process is only one way to find spanning sets, and thus we expect that minimal spanning sets of a pattern can be smaller than the minimal sets that fill it.

8.4 Bases

Definition 22. *Let $P \Subset G$ and let $X \subset A^G$ be closed. We say $T \Subset P$ is a* basis *of P (with respect to X) if it is independent and P-spanning.*

While the bases of a TEP subshift are of great interest, there is not much we can say about them in general. Indeed, even in TEP subshifts with the triangle shape, the family of bases depends on the subshift.

The following related definition is more amenable to analysis through the solitaire, and we will show that for the triangle shape, it does not depend on the subshift:

Definition 23. *Let $P \Subset G$ be a filling closed set, and let $X \subset A^G$ be closed. We say $T \Subset P$ is a* filling basis *of P if it is independent and its filling closure equals P.*

In the conference version [10], we called "filling bases" as "bases", but we feel now that it is better to use the generic term for a concept that is not TEP or solitaire specific.

Now, the following is obvious:

Lemma 38. *Any filling basis for a filling closed set is a basis for it.*

Lemma 39. *Let P be a (filling) basis for $F = \varphi(P)$. Then every pattern in its solitaire orbit is also a (filling) basis for F.*

Proof. Independent sets are closed under solitaire, filling closure is preserved under solitaire, and F-spanning is preserved under solitaire. □

Definition 24. *Let $N = \{0, \dots, n-1\}$ and let A be an alphabet. A* simple k-permutation *of A^N is a permutation π that only reads and modifies at most k*

elements of N. More precisely, it is one for which there exists $\pi' : A^M \to A^M$ with $M \subset N$ and $|M| \leqslant k$, such that

$$\pi(x)_i = \begin{cases} x_i & \text{if } i \notin M \\ \pi'(x|_M)_i & \text{if } i \in M. \end{cases}$$

If $|A| \geqslant 3$, then a simple permutation *refers to a simple 2-permutation. If $|A| = 2$, then it refers to a simple 3-permutation.*

If P, Q are bases for the same set, then there is a natural bijection $\pi : X|_P \to X|_Q$, namely the one with graph $\{(x|_P, x|_Q) \mid x \in X\}$.

Lemma 40. *Let $P \to^n Q$, then for any identifications of P and Q with $N = \{0, \ldots, |P|-1\}$, the permutation of A^N corresponding to the natural bijection $\pi : X|_P \to X|_Q$ can be computed with $O(n + |P| \log |P|)$ many simple permutations.*

Proof. We first show that this is true for simple $|S|$-permutations.

We start by showing that this is true for some identifications between P and N, and between Q and N. Let $P_0 = P \to P_1 \to \cdots \to P_n = Q$ be a sequence of solitaire moves.

Start with any identification between P_0 and N, and when applying a solitaire move, copy the identification for all cells except the ones in the symmetric difference. When we use this identification of P_{i+1} and N, it is clear that the operation of determining the contents of P_{i+1} from those of P_i is a k-simple permutation.

Finally, for any other identification between Q and N, it suffices to sort this final result, for which $|Q| \log |Q|$ simple 2-permutations suffice.

To show that simple $|S|$-permutations can be turned into simple permutations, one can apply Lemma 3.2 of [8]. ☐

8.5 Bipermutive cellular automata and the Ledrappier subshift

A standard source of TEP subshifts with the triangle shape are bipermutive cellular automata. Recall that a *cellular automaton* (here, one-dimensional) is a function $f : A^{\mathbb{Z}} \to A^{\mathbb{Z}}$ (for finite alphabet A) that is continuous and shift-commuting.

A cellular automaton always has a *local rule* $F : A^{\{\ell,\ldots,r\}} \to A$ such that $f(x)_i = F(x|_{\{i+\ell,\ldots,i+r\}})$ for all $x \in A^{\mathbb{Z}}, i \in \mathbb{Z}$. We say a cellular automaton is *bipermutive* if, taking ℓ maximal and r minimal, we have $\ell < r$ and F satisfies that $F(au) \neq F(bu)$ and $F(ua) \neq F(ub)$ whenever $u \in A^{r-\ell}$ and $a, b \in A$ are distinct letters. (Here, we identify words and patterns with an interval shape in the obvious way.)

The following is easy to show, the main observations in the proof being that bipermutivity implies that a is determined by the pair $(u, F(ua))$ and by the pair $(u, F(bu))$; and that for each $u \in A^{r-\ell}$, $F(ua)$ and $F(au)$ can take any values (since an injective function between finite sets of equal cardinality is surjective).

Lemma 41. *Let $f : A^{\mathbb{Z}} \to A^{\mathbb{Z}}$ be a bipermutive cellular automaton and let $F : A^{\{\ell,\dots,r\}} \to A$ be a local rule for it with ℓ maximal and r minimal. Then*

$$X = \{x \in A^{\mathbb{Z}^2} \mid x_{(a,b+1)} = F(x_{(a+\ell,b)}, \dots x_{(a+r,b)})\}$$

is a (C, S)-TEP subshift, where

$$S = (\{\ell, \dots, r\} \times \{0\}) \cup \{(0, 1)\}$$

and

$$C = \{(\ell, 0), (r, 0), (0, 1)\}.$$

Remark 5. The subshift X can be defined for any cellular automaton, and is called the *spacetime subshift*. It follows from the theory in [7] that every element of $A^{\mathbb{Z}}$ appears on rows of this subshift in the TEP case. More generally, this holds for surjective cellular automata. For general cellular automata, the set of configurations appearing on the rows is precisely the limit set of the cellular automaton.

The theory developed in Section 5 applies to all bipermutive cellular automata with $r - \ell = 1$, as in these cases (up to shearing) the shape S above is precisely the triangle, and $C = S$.

A TEP subshift of special interest, and having the triangle shape, is the *Ledrappier subshift*, which is the spacetime subshift of the XOR cellular automaton on alphabet $A = \{0, 1\}$, defined by $f(x)_i = x_i \oplus x_{i+1}$ where $\oplus$ is addition modulo 2. Equivalently, this is the TEP subshift where the allowed T-patterns are $p \in A^T$ such that $\sum p = 0$ where addition is again modulo 2.

This example shows that spanned sets can be larger than filled sets.

Proposition 7. *There exists a triangular TEP subshift X such that we can find n, and $P \subset T_n$ such that $\psi_X(P) = T_n$ but $|T_n| - |\varphi(P)|$ is arbitrarily large. Indeed, the Ledrappier subshift is such.*

Proof. Next, observe that the restriction of the Ledrappier subshift to the subgroup $2\mathbb{Z}^2$ of index 4 is $2T$-permutive. This can be verified by a direct computation. (Abstractly, we may deduce it from the fact the subshift is algebraic: we can see configurations as infinite series in polynomials in two commuting variables with coefficients in $\mathbb{Z}_2$, and T corresponds to an annihilating polynomial p. Then $p(x, y)^2 = p(x^2, y^2)$ because the field has characteristic 2, see e.g. [Nivat].)

It follows that $P = L_{2n} \cup \{(2n + 1, 0), (1, 2n)\}$ spans T_{2n+2}, since L_{2n} spans T_{2n}. Since $2L_n \subset L_{2n}$ and $2L_{n+1} = 2L_n \cup \{(2n + 1, 0)\}$, using the shape $2P$ we can span $2T_{n+2}$, in particular we span every second coordinate in the top right edge of the triangle T_{2n+2}.

In particular, the spanned set contains $(0, 2n + 1)$. Since P contains $(1, 2n)$, the spanned set contains $(0, 2n)$, and we conclude that the spanned set contains the left edge of T_{2n+2}, thus all of T_{2n+2}.

On the other hand, for sufficiently large n, the filling closure of P is $T_{2n} \cup \{(2n + 1, 0), (1, 2n)\}$, as this set is closed under T-filling. The difference in cardinality is of course arbitrarily large.

8.6 Filling bases for triangular TEP subshifts

We continue with TEP subshifts with the triangle shape, and show that the filling bases of a triangle (with respect to a TEP subshift with the triangle shape) correspond to the line orbit.

Let X_R be a TEP subshift for some $R \subset \Sigma^T$ and fix some $n \in \mathbb{N}$. Recall from [7] that the pattern $q \in \Sigma^{T_n}$ appears in X_R (i.e. is equal to the restriction of some $x \in X_R$ to T_n) if and only if it does not explicitly contain a translate of a pattern in $\Sigma^T \setminus R$.

Theorem 10. *The following are equivalent for $P \subset T_n$, and a TEP subshift $X \subset \Sigma^{\mathbb{Z}^2}$ with the triangle shape:*

1. *P is a filling basis (with respect to X),*
2. *P fills (i.e. has filling closure T_n and no excess),*
3. *P is in the line orbit.*

Proof. The equivalence of the last two items is a special case of Corollary 1. We show that (1) implies (2) and (3) implies (1).

Suppose thus first that P is a filling basis. Then by definition $\varphi(P) = T_n$. Since the line is a filling basis, it is T_n-spanning, and thus $X|_{T_n}$ cannot have cardinality larger than $|\Sigma|^n$. Thus we cannot have more than n elements in P since P is independent. Now, we know that any filling set with at most n element is in the line orbit.

The line is an independent set by Remark 5, and of course fills, thus it is a filling basis. The property of being a filling basis is preserved under solitaire (Lemma 39), thus every pattern in the line orbit is a filling basis. □

Note that the first item talks about a specific (but arbitrary) TEP subshift, since in the definition of the filling basis we require that $X|_T = A^T$ for the specific subshift. The other two items only talk about the abstract solitaire process, i.e. the set of bases is independent of Σ and R.

The solitaire process allows us to translate patterns on one basis to ones on another, more space efficiently than the direct method suggests, indeed we give a polynomial time in-place algorithm for this.

We now specialize Lemma 40 to the triangle case. If $P, Q \subset T_n$ are bases, any pattern $p \in \Sigma^P$ uniquely determines a pattern in $q \in \Sigma^Q$ in the natural way, by deducing the unique extension of one pattern to T_n and then restricting to the domain of the other. If we biject P and Q with $N = \{1, \ldots, n\}$ we obtain a bijection $\varphi : \Sigma^n \to \Sigma^n$.

Recall that when $|\Sigma| \geqslant 3$, a simple permutation of Σ^n is one that ignores all but two cells. If the cells are $1, 2$, this means that for some $\hat{\pi} \in \mathrm{Sym}(\Sigma^2)$ we have $\pi(a_1 a_2 a_3 a_4 \cdots a_n)_i = \hat{\pi}(a_1 a_2) a_3 a_4 \cdots a_n$ for all $a_1 a_2 \cdots a_n \in \Sigma^n$. In general, one conjugates by a reordering the cells. If $|\Sigma| = 2$, a simple permutation may look at three cells.

Theorem 11. *The bijection φ can be computed with $O(n^3)$ simple permutations.*

Proof. This is a special case of Lemma 40, observing that $n \log n \leqslant n^3$. □

9 Prospects for future work

While we know the diameter of the line orbit (for the triangle shape), and some bounds on its cardinality, we do not have a good understanding of its shape.

Problem 1. Describe the shape of the line orbit as a graph, with solitaire moves as edges.

Question 1. What does a random element of the line orbit look like? What is its expected distance from the line? Can the uniform distribution on the line orbit be sampled efficiently?

Although we have no particularly strong evidence one way or another, we conjecture a nice formula for the size of the line orbit:

Conjecture 3. There are constants $\frac{2}{e} \leqslant c \leqslant e$ and d such that $|\gamma(L_n)| = \Theta\left(\left(\frac{n}{c}\right)^{n+d}\right)$.

Question 2. For which groups G, and subsets $C \Subset S \Subset G$, do we have a nice solitaire theory?

Question 3. Does $G = \mathbb{Z}^2$ (or more generally $\mathbb{Z}^d$) have nice solitaire theory for all $C = S$?

An interesting class of shapes are the simplices $\{\vec{v} \geq 0 \mid \sum \vec{v}_i \leq 1\}$ in $\mathbb{Z}^d$. We do not even know whether $G = \mathbb{Z}^3$ has a nice solitaire theory for the pyramid $\{(0,0,0),(1,0,0),(0,1,0),(0,0,1)\}$.

Question 4. Which groups G have the finite filling (resp. solitaire) property?

We introduced in Section 4 the notion of an excess set. What can be said about the family of excess sets as a set system? How can we determine the maximum cardinality of an excess set? We do not know the answers even for the triangle case.

We also mention the following problem:

Problem 2. Describe the bases of triangles for the Ledrappier subshift, in the sense of Section 8.4. Can they be described by a variant of the solitaire process?

References

1. Boyle, M., Lind, D.: Expansive subdynamics. Trans. Amer. Math. Soc. **349**(1), 55–102 (1997). https://doi.org/10.1090/S0002-9947-97-01634-6
2. Franks, J., Kra, B.: Polygonal $\mathbb{Z}^2$-subshifts. Proc. Lond. Math. Soc. (3) **121**(2), 252–286 (2020). https://doi.org/10.1112/plms.12313
3. Hanna, P.D.: The On-Line Encyclopedia of Integer Sequences, published electronically at https://oeis.org/A101481, 2022, Sequence A101481
4. Kelley, J.L.: General topology. Courier Dover Publications (2017)
5. Kirchner, G.: The On-Line Encyclopedia of Integer Sequences, published electronically at http://oeis.org/A295928, 2022, Sequence A295928

6. Salo, V.: Soliatire (2020), https://villesalo.com/notes/solitaire.html [Accessed: (2024-4-18)]
7. Salo, V.: Cutting corners. J. Comput. System Sci. **128**, 35–70 (2022). https://doi.org/10.1016/j.jcss.2022.03.001
8. Salo, V.: Universal groups of cellular automata. In: Colloquium Mathematicum. vol. 169, pp. 39–77. Instytut Matematyczny Polskiej Akademii Nauk (2022)
9. Salo, V., Schabanel, J.: Triangle solitaire solver. https://github.com/vosalo/trianglesolitaire (2022)
10. Salo, V., Schabanel, J.: Triangle solitaire. In: International Workshop on Cellular Automata and Discrete Complex Systems. pp. 123–136. Springer (2023)