A second homotopy group for digital images

Gregory Lupton[1] · Oleg Musin[2] · Nicholas A. Scoville[3] ·
P. Christopher Staecker[4] · Jonathan Treviño-Marroquín[5]

Abstract

We define a second (higher) homotopy group for digital images. Namely, we construct a functor from digital images to abelian groups, which closely resembles the ordinary second homotopy group from algebraic topology. We illustrate that our approach can be effective by computing this (digital) second homotopy group for a digital 2-sphere.

Keywords Digital image · Digital topology · Trivial extension · Subdivision · Digital second homotopy group · Digital sphere

Mathematics Subject Classification Primary 55Q99; Secondary 68U10 · 68R99

1 Introduction

Digital topology refers to the use of notions and methods from (algebraic) topology to study digital images. A digital image in our sense is an idealization of an actual digital image which consists of pixels in the plane, or higher-dimensional analogues of such. Specifically, a *digital image* is a subset of $\mathbb{Z}^n$ together with some chosen "adjacency relation" induced by the integer lattice. The aim of digital topology is to provide useful theoretical background for certain steps of image processing, such as contour filling, border and boundary following, thinning, and feature extraction or recognition (e.g., see p.273 of [12]). There is an extensive literature on digital topology (see, e.g., [4, 9, 12, 18] and the references therein).

A number of authors have studied the fundamental group in the setting of digital topology (see [4, 11, 14, 16] for a sample). In this paper, we begin a development

Extended author information available on the last page of the article

of higher homotopy groups in the digital topology setting, essentially extending the approach of [14, 16] to an initial study of the second homotopy group of a digital image. In subsequent work, we hope to continue the development begun here into a fuller treatment of higher homotopy groups of digital images. These notions do appear in the digital topology literature [17, 20], but these earlier treatments differ from ours, as we describe in Remark 4.4. Higher homotopy groups have also been developed in a graph-theoretic setting that is very closely related to the setting in which we work (see [2, 5, 6] for a sample of work in this area). Our emphasis on the operations of *trivial extension* and *subdivision* of maps between digital images separates our work from this graph-theoretic work because these operations rely on specifics of the orthogonal integer lattice and do not naturally generalize to arbitrary graphs. Also, our present treatment is more computational: the second half of the paper focuses on a very hands-on calculation of a non-trivial second homotopy group.

The paper is organized as follows. In Sect. 2, we give basic definitions and define the fundamental notion of *extension homotopy*. In Sect. 3, we discuss a more general notion of column- or row-doubling, and subdivision of a map on a digital image. In Sect. 4, we define the second homotopy group $\pi_2(X, x_0)$, show that it is an abelian group (Thm. 4.3 and Thm. 4.8), and establish some basic properties such as independence of basepoint (Prop. 4.5) and functoriality (Prop. 4.6). We take the general development far enough to establish behavior with respect to products (Thm. 4.9).

The remainder of the paper is devoted to a calculation of the second homotopy group of a digital 2-sphere. Perhaps unsurprisingly, we find that this group is isomorphic to $\mathbb{Z}$. But the way in which we calculate this to be so involves some interesting combinatorial ingredients. Notably, in Sect. 5 we develop a *triangle-counting function* for a map of the kind that represents an element of the homotopy group. This may be conceived of as the degree of such a map, and it may be determined directly from the map in a very transparent way. We complete the calculation of the second homotopy group of a digital 2-sphere in Sect. 6. A short Sect. 7 ends the paper with some suggestions for future work.

2 Homotopy, trivial extension, and extension-homotopy

Definition 2.1 A *digital image* $X \subset \mathbb{Z}^n$ is a finite subset of the integer lattice, together with a chosen reflexive *adjacency relation* denoted $\sim$.

Typical choices of the adjacency relation are the various adjacencies denoted c_i for $i \in \{1, \ldots, n\}$, in which $x \sim y$ when the coordinates of x and y differ by at most 1 in at most i positions, and are equal in all other positions. These c_i adjacencies follow the lattice structure of $\mathbb{Z}^n$, with different interpretations of diagonal adjacencies. The c_1 relation includes no diagonally adjacent points, while the c_n adjacency counts any diagonal points as adjacent. In $\mathbb{Z}^2$, the c_1 adjacency is referred to as "4-adjacency," because each point is adjacent to 4 points other than itself, while the c_2 adjacency is referred to as "8-adjacency."

In the digital topology literature, the adjacency relation is often taken to be antireflexive, so that a point is not adjacent to itself. In our case, though, we follow [14, 16]

and require our relation to be reflexive, which simplifies some definitions and clarifies the connections to related work in graph theory.

Let $I_n = \{0, \ldots, n\}$ be the *digital interval*, considered with the usual adjacency in which $a \sim b$ if and only if $|a - b| \le 1$.

For a real interval $[n, m]$, define $[n, m]_{\mathbb{Z}} = [n, m] \cap \mathbb{Z}$, so that $I_n = [0, n]_{\mathbb{Z}}$. We will also consider products of the form $I_{m,n} = I_m \times I_n$, which we refer to as *rectangles*. For a rectangle $I_{m,n}$, let $\partial I_{m,n}$ be the *boundary*, defined by:

$$\partial I_{m,n} = (\{0, m\} \times I_n) \cup (I_m \times \{0, n\}).$$

For products of digital images, we always use the categorical product adjacency. That is, if $x, x' \in X$ and $y, y' \in Y$, then $(x, y) \sim (x', y') \in X \times Y$ if and only if $x \sim x'$ and $y \sim y'$. In the case of $I_{m,n} \subset \mathbb{Z}^2$, we have $(a, b) \sim (a', b')$ if and only if $|a - a'| \le 1$ and $|b - b'| \le 1$. This choice of product essentially dictates that we always use the c_2 adjacency (or 8-adjacency) on the rectangle $I_{m,n}$.

Definition 2.2 For two digital images X, Y, a function $f : X \to Y$ is *(digitally) continuous* when $x \sim x'$ implies $f(x) \sim f(x')$ for all $x, x' \in X$.

It is easy to see that the composition of two continuous functions is continuous.

There is a natural interpretation of this setup in the context of graph theory, and in fact a body of very similar work has developed in graph theory, independent from and until recently unknown to the digital topology community. Notably the areas of A-theory (see [1–3]) and ×-homotopy theory (see [5–7]). From the graph theoretic point of view, any digital image X may be regarded as an induced (reflexive) subgraph of the integer lattice $\mathbb{Z}^n$. Then a digitally continuous function $f : X \to Y$ is simply a graph homomorphism, provided that we represent X and Y as reflexive graphs (that is, we must have a looped edge at every vertex in the graph). These looped edges must be present in the codomain to allow the map f to collapse an edge to a vertex and yet to map edges to edges. For example, if $a \sim b \in X$ with $a \ne b \in X$ and $f(a) = f(b) = c$, then there must be a looped edge at c in order for f to carry the edge (a, b) to an edge in Y. These connections with graph theory have often been left implicit in the digital topology literature, but some authors have emphasized the graph-theoretic aspects of their work. See [13] for a recent example.

In abstract terms, the category of digital images and digitally continuous functions is the same as the category of reflexive graphs and graph homomorphisms. Thus in many cases, the constructions used in A-theory and ×-homotopy theory of, e.g., [2, 5, 6] are the same as the constructions in the digital topology literature, which developed independently.

The differences between A-theory and ×-homotopy theory and the digital theory arise in different choices made in the definition of homotopy:

Definition 2.3 Two continuous maps $f, g : X \to Y$ are *(digitally) homotopic* if there is some k with a continuous map $H : X \times I_k \to Y$ with $H(x, 0) = f(x)$ and $H(x, k) = g(x)$ for all x. In this case we write $f \simeq g$.

This notion of homotopy gives an equivalence relation on the set of all maps $X \to Y$ (see Lemma 3.16 of [15], for example). The choice of product in the definition of homotopy has an important effect on developments. As noted above, we use the categorical product, which leads to the homotopy notion typically featured in the $\times$-homotopy theory of [6]. The development in A-theory [2] has traditionally used the "box product," in which $(x, t) \sim (y, s)$ if and only if either $x \sim y$ and $t = s$, or $x = y$ and $t \sim s$, which leads to a weaker notion of homotopy often referred to as the *box homotopy*.

The digital topology literature following Boxer [4] has typically used the word "homotopy" to indicate the box homotopy, though some papers have explored the categorical product homotopy: it is called "strong homotopy" in [19]. The present paper follows the terminology used in [14, 16], with the relation of Definition 2.3 simply called "homotopy."

If $k = 1$ in Definition 2.3, then we refer to the homotopy as a *one-step homotopy*. That is, a one-step homotopy between maps $f, g \colon X \to Y$ is a continuous map $H \colon X \times I_1 \to Y$ with $H(x, 0) = f(x)$ and $H(x, 1) = g(x)$. In this case, we say that f and g are *one-step homotopic*. There is a simple criterion for maps to be one-step homotopic:

Lemma 2.4 *[19, Theorem 2.4] Suppose continuous maps $f, g \colon X \to Y$ satisfy $f(x) \sim g(x')$ in Y whenever $x \sim x'$ in X. Then f and g are one-step homotopic. Indeed, the homotopy $H \colon X \times I_1 \to Y$ defined by $H(x, 0) = f(x)$ and $H(x, 1) = g(x)$ is a one-step homotopy from f to g.*

Proof We need to confirm continuity of H. That is, we require $H(x, t) \sim H(x', t')$ in Y whenever we have $(x, t) \sim (x', t')$ in $X \times I_1$ (recall that we are using the categorical product for adjacencies in $X \times I_1$). If $t' = t \in I_1$, then the adjacencies in Y follow from continuity of f or g. If $t' \neq t \in I_1$, then the hypothesis provides the required adjacencies.

In fact, the hypothesis on f and g in Lemma 2.4 gives a characterization of when maps f and g are one-step homotopic, but we will not need the converse here. We will make repeated use of the following simple kind of a one-step homotopy.

Lemma 2.5 *Let $f \colon X \to Y$ be a continuous map of digital images, with $f(a) = b$ at some particular $a \in X$. Suppose $b' \in Y$ is adjacent to b and also adjacent to $f(a')$ for every $a' \in X$ adjacent to a. Define a map*

$$g(x) = \begin{cases} f(x) & x \neq a \\ b' & x = a \end{cases}$$

Then, g is continuous and one-step homotopic to f via the one-step homotopy $H \colon X \times I_1 \to Y$ with $H(x, 0) = f(x)$ and $H(x, 1) = g(x)$.

Proof Continuity of g follows because (continuous) f and g agree apart from at $a \in X$, where we have $g(a) = b' \sim f(x) = g(x)$ for all $x \sim a$ (but not equal to a) in X. Then the maps $f(x)$ and $g(x)$ satisfy the condition of Lemma 2.4.

Fig. 1 Representation of a typical map $f : (I_{4,4}, \partial I_{4,4}) \to (X, x_0)$ for $x_i \in X$. Each pixel is labeled with its function value, so that, e.g., $f(2, 1) = x_8 \in X$. Values representing the base point x_0 are labeled with a dot
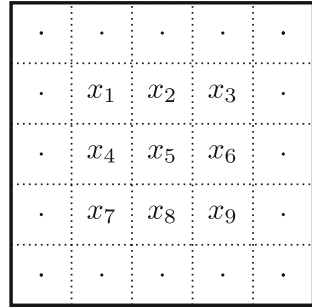
Definition 2.6 The type of one-step homotopy in Lemma 2.5, whereby a map is changed in value at a single point, is called a *spider move*.

The following theorem, which has appeared in both the digital topology literature and the $\times$-homotopy literature, shows that an arbitrary homotopy can be realized by a finite sequence of spider moves.

Theorem 2.7 *[5, Proposition 4.4] [19, Theorem 3.2] For continuous $f, g : X \to Y$, the maps f and g are homotopic if and only if they are homotopic by a finite sequence of spider moves.* □

For positive integers m, n and a pointed digital image (X, x_0), we will consider continuous maps of pairs $f : (I_{m,n}, \partial I_{m,n}) \to (X, x_0)$. That is, continuous maps f with $f(\partial I_{m,n}) = \{x_0\}$.

It is often convenient to visualize a function $f : (I_{m,n}, \partial I_{m,n}) \to (X, x_0)$ as a labeling of the points of the rectangle $I_{m,n}$ with labels taken from the set X. For example, a function $f : (I_{4,4}, \partial I_{4,4}) \to (X, x_0)$ would be represented by the labeled rectangle in Fig. 1. For simplicity in our pictures, we will indicate the label of the basepoint with a dot.

Our second homotopy group is modeled on homotopy classes of maps of pairs $(I_{m,n}, \partial I_{m,n}) \to (X, x_0)$, where the homotopies preserve values at the boundary in the following sense:

Given $f, g : (I_{m,n}, \partial I_{m,n}) \to (X, x_0)$ a homotopy $H : I_{m,n} \times I_k \to X$, we say H is a *homotopy relative to the boundary* when $H(\partial I_{m,n} \times I_k) = \{x_0\}$.

In our development, we often encounter a situation in which we have a "local" homotopy that only involves values of a map in some part of the rectangle $I_{m,n}$. If such a homotopy is stationary on the boundary of a subrectangle, then it may be extended to a homotopy of the whole rectangle in an obvious way.

Lemma 2.8 *Let R be a subrectangle of $I_{m,n}$. Let $f : (I_{m,n}, \partial I_{m,n}) \to (X, x_0)$ be a map for which f_R, the restriction of f to the subrectangle R, is a map of pairs $f_R : (R, \partial R) \to (X, x_0)$. Let $g : (R, \partial R) \to (X, x_0)$ be a map on the subrectangle such that $f_R \simeq g$ by a homotopy that is stationary on the boundary ∂R. Then, the map $A : (I_{m,n}, \partial I_{m,n}) \to (X, x_0)$ defined by:*

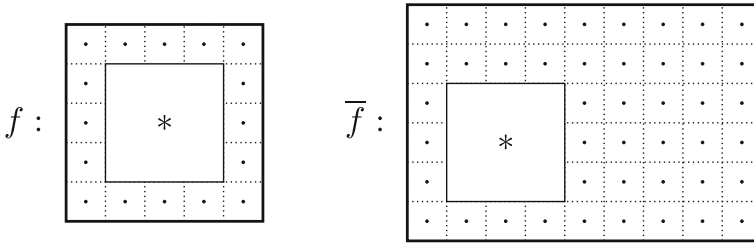$$A(a, b) = \begin{cases} g(a, b) & \text{if } (a, b) \in R \\ f(a, b) & \text{if } (a, b) \notin R \end{cases}$$

Fig. 2 Schematic of a map $f : I_{4,4} \to X$ with its trivial extension $\overline{f} : I_{8,5} \to X$

is continuous and we have $f \simeq A$ by a homotopy relative to the boundary.

Proof Continuity of the map A is assured because the maps f and g agree on the boundary of the rectangle ∂R, which separates $I_{m,n}$ into non-adjacent interior and exterior: A point interior to the rectangle R cannot be adjacent to a point exterior to the rectangle R. Thus, $A(a, b) \sim A(a', b')$ follows from continuity of g for points $(a, b) \sim (a', b')$ in R (including its boundary), and from continuity of f for points $(a, b) \sim (a', b')$ in $R^C \cup \partial R$. (Here, R^C denotes the complement of R in $I_{m,n}$.)

Suppose $G : R \times I_k \to X$ is the homotopy—stationary on ∂R—from f_R to g. Then, the homotopy $H : I_{m,n} \times I_k \to X$ defined by

$$H((a, b), t) = \begin{cases} G((a, b), t) & \text{if } (a, b) \in R \\ f(a, b) & \text{if } (a, b) \notin R \end{cases}$$

starts at f and ends at A. This homotopy is continuous by reasoning similar to that of the first part. Namely, $\partial R \times I_k$ separates $I_{m,n} \times I_k$ into non-adjacent interior and exterior. Then, $H((a, b), t) \sim H((a', b'), t')$ follows from continuity of G for points $((a, b), t) \sim ((a', b'), t')$ in $R \times I_k$ and from continuity of f for points $((a, b), t) \sim ((a', b'), t')$ in $(R^C \cup \partial R) \times I_k$.

Boxer's definition of the fundamental group in [4] uses a construction which he calls *trivial extension* of a loop. We adapt this concept for our higher-dimensional setting by simply repeating values of the base point.

We define *trivial extensions* of maps $(I_{m,n}, \partial I_{m,n}) \to (X, x_0)$ as follows: if $m' \geq m$ and $n' \geq n$, then there is a natural inclusion $I_{m,n} \subseteq I_{m',n'}$. We say $\overline{f} : (I_{m',n'}, \partial I_{m',n'}) \to (X, x_0)$ is a trivial extension of $f : (I_{m,n}, \partial I_{m,n}) \to (X, x_0)$ when

$$\overline{f}(x) = \begin{cases} f(x) & \text{if } x \in I_{m,n}, \\ x_0 & \text{otherwise.} \end{cases}$$

See Fig. 2 for a pictorial representation of a trivial extension.

Whereas homotopy is a relation between maps with the same domain, we will need to compare maps whose domains are differently sized rectangles. We do this through the following device.

Definition 2.9 Given two maps $f : (I_{m,n}, \partial I_{m,n}) \rightarrow (X, x_0)$ and $g : (I_{m',n'}, \partial I_{m',n'}) \rightarrow (X, x_0)$, we write $f \approx g$, and say that f and g are *extension-homotopic*, when there exist $\overline{m} \geq \max(m, m')$ and $\overline{n} \geq \max(n, n')$ and $\overline{f}, \overline{g} : I_{\overline{m},\overline{n}} \rightarrow X$ with $\overline{f}$ a trivial extension of f and $\overline{g}$ a trivial extension of g and $\overline{f}$ homotopic to $\overline{g}$ by a homotopy relative to the boundary.

Theorem 2.10 *Extension homotopy of maps is an equivalence relation on the set of maps $(I_{m,n}, \partial I_{m,n}) \rightarrow (X, x_0)$ for all sizes of rectangles.*

Reflexivity and symmetry follow immediately because homotopy (relative the boundary) of maps is an equivalence relation. Transitivity is a consequence of the following simple lemma:

Lemma 2.11 *Suppose we have maps $f \simeq g : (I_{m,n}, \partial I_{m,n}) \rightarrow (X, x_0)$ homotopic relative to the boundary. Let $f', g' : (I_{m',n'}, \partial I_{m',n'}) \rightarrow (X, x_0)$ be trivial extensions of f and g to the same-sized rectangle, for $m' \geq m$ and $n' \geq n$. Then, we have $f' \simeq g'$.*

Proof The proof is fairly obvious. A homotopy $H : I_{m,n} \times I_k \rightarrow X$ from f to g relative the boundary extends to a homotopy $\overline{H} : I_{m',n'} \times I_k \rightarrow X$ from f' to g' relative the boundary, by setting $\overline{H}$ to be stationary at x_0 on all points of $I_{m',n'}$ not in $I_{m,n}$. This extension $\overline{H}$ is easily seen to be a continuous map on $I_{m',n'} \times I_k$ since a point of $I_{m',n'}$ not in $I_{m,n}$ cannot be adjacent to a point in the interior of $I_{m,n}$, and the original H is already stationary at x_0 on all points of $\partial I_{m,n}$. $\qquad \blacksquare$

Proof of Theorem 2.10 As observed above, we only need to show transitivity. So, suppose we have maps $f_t : (I_{m_t,n_t}, \partial I_{m_t,n_t}) \rightarrow (X, x_0)$ for $t = 1, 2, 3$ and that $f_1 \approx f_2 \approx f_3$. Since $f_1 \approx f_2$ there are $m' \geq \max\{m_1, m_2\}$ and $n' \geq \max\{n_1, n_2\}$ along with trivial extensions $f_1', f_2' : I_{m',n'} \rightarrow X$ of f_1 and f_2 and a homotopy from f_1' to f_2'. Similarly, since $f_2 \approx f_3$ there are $m'' \geq \max\{m_2, m_3\}$ and $n'' \geq \max\{n_2, n_3\}$ along with trivial extensions $f_2'', f_3'' : I_{m'',n''} \rightarrow X$ of f_2 and f_3 and a homotopy from f_2'' to f_3''.

Let $\overline{m} = \max\{m', m''\}$ and $\overline{n} = \max\{n', n''\}$, and let $\overline{f_1}, \overline{f_2}, \overline{f_3} : I_{\overline{m},\overline{n}} \rightarrow X$ be trivial extensions of f_1, f_2 and f_3, respectively. Then, $\overline{f_1}$ and $\overline{f_3}$ are trivial extensions of f_1' and f_3'', respectively, while $\overline{f_2}$ is a common trivial extension of both f_2' and f_2''. Hence, we have $\overline{f_1} \simeq \overline{f_2}$ and $\overline{f_2} \simeq \overline{f_3}$ by Lemma 2.11. Since homotopy is transitive, we have $\overline{f_1} \simeq \overline{f_3}$ and the result follows. $\qquad \blacksquare$

3 Column-Doubling and Row-Doubling

In this section we define column- and row-doubling operators and subdivisions of a digital map, and show their relations to trivial extensions.

Definition 3.1 Given $i \in \{0, \ldots, m\}$ and $j \in \{0, \ldots, n\}$, let $\alpha_i : I_{m+1,n} \rightarrow I_{m,n}$ and $\beta_j : I_{m,n+1} \rightarrow I_{m,n}$ be the maps defined by

$$\alpha_i(a, b) = \begin{cases} (a, b) & \text{if } a \leq i, \\ (a - 1, b) & \text{if } a > i. \end{cases} \qquad \beta_j(a, b) = \begin{cases} (a, b) & \text{if } b \leq j, \\ (a, b - 1) & \text{if } b > j. \end{cases}$$

The map α_i simply omits one column of the domain, so the composition $f \circ \alpha_i : I_{m+1,n} \to X$ is a map which resembles f, but with column i repeated once and all following columns shifted to the right by one position. Namely, viewing f as a labeling of the points of $I_{m,n}$ with values from X, we have *doubled the ith column* of labeled values to result in a similar labeling of $I_{m+1,n}$ for $f \circ \alpha_i$. Likewise, $f \circ \beta_j$ is a map which resembles f, but with row j doubled.

Theorem 3.2 *Let $f : (I_{m,n}, \partial I_{m,n}) \to (X, x_0)$ be continuous. We have homotopies relative the boundary $f \circ \alpha_m \simeq f \circ \alpha_{m-1} \simeq \cdots \simeq f \circ \alpha_0$ and $f \circ \beta_n \simeq f \circ \beta_{n-1} \simeq \cdots \simeq f \circ \beta_0$. Consequently, we have $f \approx f \circ \alpha_i$ for each $i \in \{0, \ldots, m\}$ and $f \approx f \circ \beta_j$ for each $j \in \{0, \ldots, n\}$.*

Proof We will prove the statement for the $f \circ \alpha_i$. The proof for the $f \circ \beta_j$ is similar and we omit it. Let $\overline{f} : I_{m+1,n} \to X$ be the trivial extension of f, which we may write as $\overline{f} = f \circ \alpha_m$. We will show that $f \circ \alpha_i \simeq f \circ \alpha_{i-1}$ for each $i \in \{1, \ldots, m\}$ and it follows that we have $f \approx f \circ \alpha_i$ for each i.

Indeed, we will use the criterion of Lemma 2.4 to show $f \circ \alpha_i$ and $f \circ \alpha_{i-1}$ are one-step homotopic. For $x \sim x' \in I_{m+1,n}$, we must show that $f \circ \alpha_i(x) \sim f \circ \alpha_{i-1}(x')$ in X. Notice that $f \circ \alpha_i$ and $f \circ \alpha_{i-1}$ agree in value except at points $(i, b) \in I_{m+1,n}$: Unless one of x or x' has coordinates (i, b) for some b, the desired conclusion follows from continuity of either map $f \circ \alpha_i$ or $f \circ \alpha_{i-1}$. So, assume $x = (i, b)$ and $x' = (a', b')$ with $i \sim a'$ (and thus $a' \in \{i-1, i, i+1\}$) and $b \sim b'$.

If $a' = i - 1$, we may use continuity of f and the definitions of $f \circ \alpha_i$ and $f \circ \alpha_{i-1}$ to write

$$f \circ \alpha_i(x) = f \circ \alpha_i(i, b) = f(i, b) \sim f(i-1, b') = f \circ \alpha_{i-1}(i-1, b') = f \circ \alpha_{i-1}(x').$$

Similarly, if $a' = i$, we may write

$$f \circ \alpha_i(x) = f \circ \alpha_i(i, b) = f(i, j) \sim f(i-1, b') = f \circ \alpha_{i-1}(i, b') = f \circ \alpha_{i-1}(x').$$

Finally, if $a' = i + 1$, we may write

$$f \circ \alpha_i(x) = f \circ \alpha_i(i, b) = f(i, b) \sim f(i, b') = f \circ \alpha_{i-1}(i+1, b') = f \circ \alpha_{i-1}(x').$$

It follows from Lemma 2.4 that $f \circ \alpha_i$ and $f \circ \alpha_{i-1}$ are one-step homotopic, and the result follows.

Observe that if $f : (I_{m,n}, \partial I_{m,n}) \to (X, x_0)$, then any trivial extension of f may be obtained by repeatedly doubling the mth row and the nth column.

Lemma 3.3 *Let $f : (I_{m,n}, \partial I_{m,n}) \to (X, x_0)$, and let $\overline{f} : (I_{r,s}, \partial I_{r,s}) \to (X, x_0)$ be a trivial extension. Then we have*

$$\overline{f} = f \circ \alpha_m^{r-m} \circ \beta_n^{s-n}.$$

Furthermore, in this expression the $r - m$ iterations of α_m and the $s - n$ iterations of β_n may be shuffled amongst themselves in this expression (any order of these row- or column-doublings achieves the same effect).
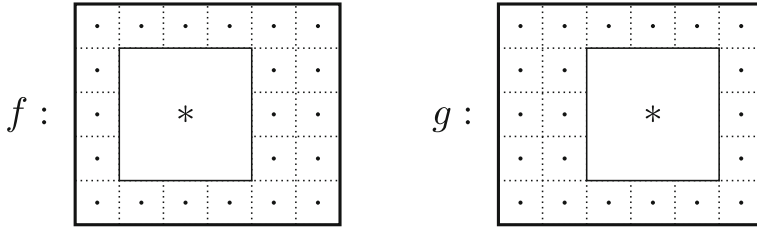
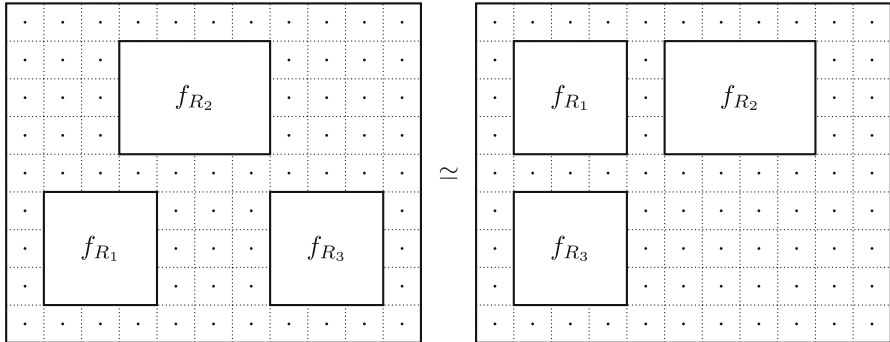**Fig. 3** Schematic of maps f and g from Corollary 3.4: a horizontal shift by one unit

Fig. 4 Translations by homotopy of subrectangles surrounded by basepoint values illustrating Remark 3.5

Proof Each pre-composition with an α_m or a β_n doubles the nth row or mth column of values of f. Since the nth row and mth column of f are constant maps at x_0, this produces the same map as the trivial extension $\overline{f}$.

Corollary 3.4 *Let* $f, g \colon (I_{m,n}, \partial I_{m,n}) \rightarrow (X, x_0)$ *be continuous maps with* $f(m - 1, b) = x_0$ *and* $g(1, b) = x_0$ *for each* $j \in I_n$, *and* $g(a, b) = f(a - 1, b)$ *for each* $a \in \{2, \ldots, m - 1\}$ *and each* $j \in I_n$. *That is, the maps* f *and* g *are* horizontal shifts *of one another, as illustrated in Fig. 3. Then, we have* $f \simeq g$.

Proof The assumption on f means that f is a trivial extension of some continuous map $h \colon I_{m-1,n} \rightarrow X$. In fact we have $f = h \circ \alpha_{m-1}$ and $g = h \circ \alpha_0$. By the proof of Theorem 3.2, we have that $h \circ \alpha_{m-1} \simeq h \circ \alpha_0$, which implies that $f \simeq g$.

Remark 3.5 Repeated application of Corollary 3.4 and Lemma 2.8 allows for shifting horizontally up to homotopy, respectively, extension homotopy, of a subrectangle of a function within a region of $I_{m,n}$ surrounded by basepoints. A similar argument using row-doubling rather than column-doubling in Corollary 3.4 shows that the same is possible for vertical shifts. Combining these, we see that any "translation" of a subrectangle through a region of constant basepoint values will not change the homotopy class of a map. For example, we may achieve up to homotopy any translation of the type appearing in Fig. 4, where three blocks of values from X, each surrounded by basepoints, can be maneuvered into a different configuration within $I_{m,n}$.

The following is a version of Lemma 2.8 for "local extension homotopy."

Lemma 3.6 *Let R be a subrectangle of $I_{m,n}$. Let $f : (I_{m,n}, \partial I_{m,n}) \to (X, x_0)$ be a map for which f_R, the restriction of f to the subrectangle R, is a map of pairs $f_R : (R, \partial R) \to (X, x_0)$. Let $g : (R, \partial R) \to (X, x_0)$ be a map on the subrectangle such that $f_R \approx g$ by an extension homotopy in the sense of Definition 2.9.*

Then, the map $A : (I_{m,n}, \partial I_{m,n}) \to (X, x_0)$ defined by:

$$A(x) = \begin{cases} f(x) & \text{if } x \notin R, \\ g(x) & \text{if } x \in R \end{cases}$$

is continuous and we have $f \approx A$.

Proof Let $R = [r, s]_{\mathbb{Z}} \times [p, q]_{\mathbb{Z}}$. Since $f_R \approx g$, there is a larger rectangle $\overline{R} = [r, s+u]_{\mathbb{Z}} \times [p, q+v]_{\mathbb{Z}}$ with $u, v \geq 0$ and trivial extensions $\overline{f}_R, \overline{g} \colon \overline{R} \to X$ and a homotopy relative to the boundary $H_{\overline{R}} : \overline{R} \times I_k \to X$ from $\overline{f}_R$ to $\overline{g}$. By Lemma 3.3 we have $\overline{f}_R = f_R \circ \alpha_s^u \circ \beta_q^v$ and $\overline{g} = g \circ \alpha_s^u \circ \beta_q^v$.

Now we define a homotopy $H : I_{m+u,n+v} \times I_k \to X$ to be a row- and column-doubled version of f outside of $\overline{R}$, and equal to $H_{\overline{R}}$ inside of R:

$$H(x, t) = \begin{cases} f \circ \alpha_s^u \circ \beta_q^v(x) & \text{if } x \notin \overline{R}, \\ H_{\overline{R}}(x, t) & \text{if } x \in \overline{R}. \end{cases}$$

We will show that H_R is a homotopy relative to the boundary from $f \circ \alpha_s^u \circ \beta_q^v$ to $A \circ \alpha_s^u \circ \beta_q^v$. Since $f \approx f \circ \alpha_s^u \circ \beta_q^v$ and $A \approx A \circ \alpha_s^u \circ \beta_q^v$ by Theorem 3.2 and following, this will demonstrate that $f \approx A$ as desired.

We have:

$$H(x, 0) = \begin{cases} f \circ \alpha_s^u \circ \beta_q^v(x) & \text{if } x \notin \overline{R}, \\ \overline{f}_R(x) & \text{if } x \in \overline{R} \end{cases} = f \circ \alpha_s^u \circ \beta_q^v(x)$$

and

$$H(x, k) = \begin{cases} f \circ \alpha_s^u \circ \beta_q^v(x) & \text{if } x \notin \overline{R}, \\ \overline{g}(x) & \text{if } x \in \overline{R} \end{cases} = A \circ \alpha_s^u \circ \beta_q^v(x)$$

so H begins at $f \circ \alpha_s^u \circ \beta_q^v$ and ends at $A \circ \alpha_s^u \circ \beta_q^v$.

Also it is easy to see that H is a homotopy relative to the boundary: if $x \in \partial I_{m+u,n+v}$ then either $x \notin \overline{R}$ or $x \in \partial \overline{R}$. In either case we have $H(x, t) = x_0$ for all t because $f(\partial I_{m,n}) = x_0$ and H_R is a homotopy relative to the boundary. Thus, it remains only to show that H is continuous.

Let $(x, t) \sim (x', t') \in I_{m+u,n+v} \times I_k$, and we must show that $H(x, t) \sim H(x', t')$. We prove this in simple cases according to whether or not the points x, x' are in $\overline{R}$.

If $x \in \overline{R}$ and $x' \in \overline{R}$, then we have:

$$H(x, t) = H_{\overline{R}}(x, t) \sim H_{\overline{R}}(x', t') = H(x', t')$$

where the middle step is because $H_{\overline{R}}$ is continuous. Thus, $H(x, t) \sim H(x', t')$ as desired.

If $x \notin \overline{R}$ and $x' \notin \overline{R}$, then we have:

$$H(x, t) = f \circ \alpha_s^u \circ \beta_q^v(x) \sim f \circ \alpha_s^u \circ \beta_q^v(x') = H_R(x', t')$$

where the middle step is because f, α_s, and β_q are continuous. Thus, again $H(x, t) \sim H(x', t')$ as desired.

If $x \in \overline{R}$ and $x' \notin \overline{R}$, this is only possible when $x \in \partial \overline{R}$. In that case we must have $f \circ \alpha_s^u \circ \beta_q^v(x) = x_0$ because f maps ∂R to x_0. Since $x \in \partial \overline{R}$ and $H_{\overline{R}}$ maps $\partial \overline{R}$ to x_0, we have:

$$H(x, t) = H_{\overline{R}}(x, t) = x_0 = f \circ \alpha_s^u \circ \beta_q^v(x).$$

Since $x' \notin \overline{R}$, we have $H(x', t') = f \circ \alpha_s^u \circ \beta_q^v(x')$. Since f, α_s, and β_q are continuous and $x \sim x'$, this means

$$H(x', t') = f \circ \alpha_s^u \circ \beta_q^v(x') \sim f \circ \alpha_s^u \circ \beta_q^v(x) = H(x, t)$$

and so $H(x, t) \sim H(x', t')$, which completes the proof that H is continuous. $\square$

The row- and column-doubling operations are closely related to *subdivision* of a digital image, which was defined in [8] to define digitally continuous multivalued maps. The subdivision was used fundamentally in [14] as the basis for the definition of a digital fundamental group. In [14] the authors use a general subdivision of a digital image which they denote $S(X, k)$ for some natural number k, which essentially replaces each point of X by a $k \times k$ block of points. This subdivision comes with a natural projection map $\rho_k : S(X, k) \to X$ which collapses $k \times k$ blocks into single points.

For our purposes, we will only need to subdivide the rectangle $I_{m,n}$, in which case the k-fold subdivision is simply the rectangle $I_{km+k-1,kn+k-1}$. And then the projection map $\rho_k : I_{km+k-1,kn+k-1} \to I_{m,n}$ is obtained by iterated row and column omissions as follows:

$$\rho_k = \alpha_0^{k-1} \circ \alpha_k^{k-1} \circ \cdots \circ \alpha_{mk}^{k-1} \circ \beta_0^{k-1} \circ \cdots \circ \beta_{kn}^{k-1}.$$
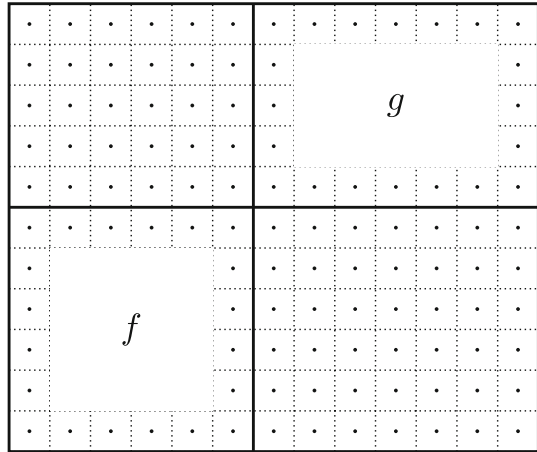
Applying Theorem 3.2 repeatedly to the above gives:

Theorem 3.7 Let $f : (I_{m,n}, \partial I_{m,n}) \to (X, x_0)$ be continuous and $k \geq 1$. Then $f \approx f \circ \rho_k$. $\square$

4 Definition of $\pi_2(X, x_0)$

We now give the definition of our second homotopy group and establish its basic general properties.

Fig. 5 Schematic of the product $f \cdot g$ of two maps f and g

Definition 4.1 Given a based digital image (X, x_0), the *second homotopy group* of (X, x_0), written $\pi_2(X, x_0)$, is the set of equivalence classes of maps $f : (I_{m,n}, \partial I_{m,n}) \to (X, x_0)$, for all rectangles $I_{m,n}$, modulo the equivalence relation of extension homotopy.

The group operation in $\pi_2(X, x_0)$ is induced by the following operation on maps. Let $f \colon (I_{m,n}, \partial I_{m,n}) \to (X, x_0)$ and $g \colon (I_{r,s}, \partial I_{r,s}) \to (X, x_0)$ be maps. Define $f \cdot g \colon (I_{m+r+1,n+s+1}, \partial I_{m+r+1,n+s+1}) \to (X, x_0)$ by

$$(f \cdot g)(a, b) = \begin{cases} f(a, b) & \text{if } (a, b) \in [0, m]_{\mathbb{Z}} \times [0, n]_{\mathbb{Z}} \\ g(a - (m+1), b - (n+1)) & \text{if } (a, b) \in [m+1, m+r+1]_{\mathbb{Z}} \times [n+1, n+s+1]_{\mathbb{Z}} \\ x_0 & \text{otherwise.} \end{cases}$$

See Fig. 5 for an illustration with $I_{m,n} = I_{5,5}$ and $I_{r,s} = I_{6,4}$

Proposition 4.2 *Suppose we have maps*

$$f_1 \colon (I_{m_1,n_1}, \partial I_{m_1,n_1}) \to (X, x_0), \quad f_2 \colon (I_{m_2,n_2}, \partial I_{m_2,n_2}) \to (X, x_0),$$

and

$$g_1 \colon (I_{r_1,s_1}, \partial I_{r_1,s_1}) \to (X, x_0), \quad g_2 \colon (I_{r_2,s_2}, \partial I_{r_2,s_2}) \to (X, x_0)$$

with $f_1 \approx f_2$ and $g_1 \approx g_2$. Then $f_1 \cdot g_1 \approx f_2 \cdot g_2$.

Proof Let f_1, f_2, g_1, and g_2 satisfy the above. Then there are trivial extensions $\overline{f_1}, \overline{f_2} \colon I_{\overline{m},\overline{n}} \to X$ of f_1 and f_2, respectively, that are homotopic relative the boundary. Similarly, there are trivial extensions $\overline{g_1}, \overline{g_2} \colon I_{\overline{r},\overline{s}} \to X$ of g_1 and g_2, respectively, that are homotopic relative the boundary.

In the notation of Definition 3.1, we may write $\overline{f_1}$ as $f_1 \circ \alpha_{m_1}^{\overline{m}-m_1} \circ \beta_{n_1}^{\overline{n}-n_1}$. Since the first $m_1 + 1$ columns and first $n_1 + 1$ rows (aside from extra x_0s to the right and above)

of $f_1 \cdot g_1$ are those of f_1, we may further write $\overline{f_1} \cdot g_1$ as $(f_1 \cdot g_1) \circ \alpha_{m_1}^{\overline{m}-m_1} \circ \beta_{n_1}^{\overline{n}-n_1}$. Continuing to break down the trivial extensions into successive column- and row-doubling, being careful with the indexing of rows and columns in the products, and using Theorem 3.2, we have a sequence of extension homotopies as follows:

$$f_1 \cdot g_1 \approx (f_1 \cdot g_1) \circ \alpha_{m_1}^{\overline{m}-m_1} \circ \beta_{n_1}^{\overline{n}-n_1} = \overline{f_1} \cdot g_1$$
$$\approx (\overline{f_1} \cdot g_1) \circ \alpha_{\overline{m}+r_1+1}^{\overline{r}-r_1} \circ \beta_{\overline{n}+s_1+1}^{\overline{s}-s_1} = \overline{f_1} \cdot \overline{g_1}.$$

Likewise, we may write $f_2 \cdot g_2 \approx \overline{f_2} \cdot \overline{g_2}$ by changing all subscripts from 1 to 2 in the above steps. Now the homotopies $\overline{f_1} \simeq \overline{f_2}$ and $\overline{g_1} \simeq \overline{g_2}$ extend to homotopies $\overline{f_1} \cdot \overline{g_1} \simeq \overline{f_2} \cdot \overline{g_1} \simeq \overline{f_2} \cdot \overline{g_2}$ by Lemma 2.8. It follows that we have $f_1 \cdot g_1 \approx f_2 \cdot g_2$.

Now let $[f], [g] \in \pi_2(X, x_0)$ and define $[f] \cdot [g] = [f \cdot g]$. This operation is well defined by Proposition 4.2.

Theorem 4.3 *With the operation given above, the set of equivalence classes $\pi_2(X, x_0)$ is a group.*

Proof Associativity follows immediately since $(f \cdot g) \cdot h = f \cdot (g \cdot h)$ at the level of maps.

Next, let $c_{x_0} \colon I_{m,n} \to X$ be the constant map at $x_0 \in X$ from any rectangle. Any such map may be viewed as a trivial extension of the constant map $c_{x_0} \colon I_{0,0} \to X$, where $I_{0,0} = \{(0, 0)\}$. We show that $[c_{x_0}]$ acts as a two-sided identity, where—by the preceding remark—we may as well assume the representative c_{x_0} has domain the single point $I_{(0,0)}$. Let $f \colon (I_{m,n}, \partial I_{m,n}) \to (X, x_0)$ be any map. On the right, we see that $f \cdot c_{x_0} \colon I_{m+1,n+1} \to X$ is simply equal to the trivial extension $\overline{f} \colon I_{m+1,n+1} \to X$. Thus $[f] \cdot [c_{x_0}] = [f \cdot c_{x_0}] = [\overline{f}] = [f]$ and so $[c_{x_0}]$ acts on the right as an identity element. On the left, we see that $c_{x_0} \cdot f \colon I_{m+1,n+1} \to X$ may be written as the result of doubling the first row and column of f, in the sense of Sect. 3. From Theorem 3.2, we may write

$$c_{x_0} \cdot f = f \circ \alpha_0 \circ \beta_0 \approx f \circ \alpha_0 \approx f.$$

That is, we have $[c_{x_0}] \cdot [f] = [f]$ and $[c_{x_0}]$ acts as a left identity too.

Finally, we consider inverses. For a map $f \colon I_{m,n} \to X$, define $f^{-1} \colon I_{m,n} \to X$ by $f^{-1}(a, b) = f(m - a, b)$. As a pre-processing step, we show that $f \cdot f^{-1}$ has the same extension homotopy type of the map that we denote by $(f \mid f^{-1}) \colon (I_{2m+1,n}, \partial I_{2m+1,n}) \to (X, x_0)$ and define as

$$(f \mid f^{-1})(a, b) = \begin{cases} f(a, b) & \text{if } 0 \le a \le m \\ f^{-1}(a - (m + 1), b) & \text{if } m + 1 \le a \le 2m + 1. \end{cases}$$

To see this, note that on the sub-rectangle $R = [m+1, 2m+1]_{\mathbb{Z}} \times [n+1, 2n+1]_{\mathbb{Z}} \subseteq I_{2m+1,2n+1}$, the map $f \cdot f^{-1}$ restricts to the map $f^{-1} \circ \beta_0^{n+1}$ in the notation of Theorem 3.2. By using translations of the type described in Remark 3.5 and which

flow from Corollary 3.4 and Lemma 2.8, $f^{-1} \circ \beta_0^{n+1}$ and $f^{-1} \circ \beta_{n+1}^{n+1}$ on the right-hand half of the rectangle are homotopic via a homotopy that extends to one of $f \cdot f^{-1}$ on the whole rectangle $I_{2m+1,2n+1}$ and leaves the left-hand half fixed. This map may now be written as $(f \mid f^{-1}) \circ \beta_{n+1}^{n+1}$ and repeated application of Theorem 3.2 yields an extension homotopy to $(f \mid f^{-1})$. This pre-processing step may be summarized pictorially as a combination of translation and collapsing of repeated rows as follows:

$$f \cdot g = \begin{array}{|c|c|} \hline x_0 & f^{-1} \\ \hline f & x_0 \\ \hline \end{array} \simeq \begin{array}{|c|c|} \hline x_0 & x_0 \\ \hline f & f^{-1} \\ \hline \end{array} \approx \begin{array}{|c|c|} \hline f & f^{-1} \\ \hline \end{array} = (f \mid f^{-1})$$

Now display the values of $(f \mid f^{-1})$ on the rectangle $I_{2m+1,n}$ in column-wise form as $(f \mid f^{-1}) = [\mathbf{v}_0 \mid \mathbf{v}_1 \mid \cdots \mid \mathbf{v}_{m-1} \mid \mathbf{v}_m \mid \mathbf{v}_m \mid \mathbf{v}_{m-1} \mid \cdots \mid \mathbf{v}_0]$, where each $\mathbf{v}_i$ is a column vector of entries from X given by

$$\mathbf{v}_i = \begin{bmatrix} f(a, n) \\ f(a, n-1) \\ \vdots \\ f(a, 1) \\ f(a, 0) \end{bmatrix}.$$

Since the middle pair of columns repeat, we may write $(f \mid f^{-1}) = g_m \circ \alpha_m$, in the notation of Theorem 3.2, where the values of g_m on the rectangle $I_{2m,n}$ in column-wise form are

$$g_m = [\mathbf{v}_0 \mid \mathbf{v}_1 \mid \cdots \mid \mathbf{v}_{m-1} \mid \mathbf{v}_m \mid \mathbf{v}_{m-1} \mid \cdots \mid \mathbf{v}_0].$$

Namely, we have "collapsed" the repeated $\mathbf{v}_m$ column into a single column and we have $(f \mid f^{-1}) \approx g_m$ by Theorem 3.2. Now define, for each $k = 0, \dots, m$, a map $g_k : (I_{2k,n}, \partial I_{2k,n}) \to (X, x_0)$ by

$$g_k(a, b) = \begin{cases} f(a, b) & a = 0, \dots, k \\ f(2k - a, b) & a = k+1, \dots, 2k. \end{cases}$$

The map g_m we arrived at above is the case in which $k = m$, and the general g_k may be pictured column-wise as a reduced form of g_m, with

$$g_k = [\mathbf{v}_0 \mid \cdots \mid \mathbf{v}_{k-1} \mid \mathbf{v}_k \mid \mathbf{v}_{k-1} \mid \cdots \mid \mathbf{v}_0].$$

We note in this case that:

$$g_{k-1} \circ \alpha_{k-1}^2 = [\mathbf{v}_0 \mid \cdots \mid \mathbf{v}_{k-1} \mid \mathbf{v}_{k-1} \mid \mathbf{v}_{k-1} \mid \cdots \mid \mathbf{v}_0].$$

Claim. For each $k \in \{m, m-1, \ldots, 1\}$, we have $g_k \simeq g_{k-1} \circ \alpha_{k-1}^2 \approx g_{k-1}$, where the first homotopy is relative to the boundary.

Proof of Claim. Repeated application of Theorem 3.2 gives an extension homotopy $g_{k-1} \circ \alpha_{k-1}^2 \approx g_{k-1}$, so we need only show that $g_k \simeq g_{k-1} \circ \alpha_{k-1}^2$. In fact we will prove that $g_k \simeq g_{k-1} \circ \alpha_{k-1}^2$ by a one-step homotopy. By Lemma 2.4, take $(a, b) \sim (a', b')$, and we must show that $g_k(a, b) \sim g_{k-1} \circ \alpha_{k-1}^2(a', b')$.

Since g_k and $g_{k-1} \circ \alpha_{k-1}^2$ differ only in column k, we need only consider the cases where $\{a, a'\} = \{k-1, k\}$. (The cases where $\{a, a'\} = \{k, k+1\}$ are similar.)

In the case where $a = k-1$ and $a' = k$, we have:

$$g_{k-1} \circ \alpha_{k-1}^2(a', b') = g_{k-1} \circ \alpha_{k-1}^2(k, b') = g_k(k-1, b') \sim g_k(k-1, b) = g_k(a, b)$$

so $g_{k-1} \circ \alpha_{k-1}^2(a', b') \sim g_k(a, b)$ as desired.

In the case where $a = k$ and $a' = k-1$, we have:

$$g_{k-1} \circ \alpha_{k-1}^2(a', b') = g_{k-1} \circ \alpha_{k-1}^2(k-1, b') = g_k(k, b') \sim g_k(k-1, b') = g_k(a', b')$$

so again $g_{k-1} \circ \alpha_{k-1}^2(a', b') \sim g_k(a, b)$ as desired. *End of Proof of Claim.*

The preceding arguments give a chain of extension homotopies as follows:

$$f \cdot f^{-1} \approx (f \mid f^{-1}) \approx g_m \approx g_{m-1} \approx \cdots \approx g_1 \approx g_0,$$

where the final map is a constant map at x_0. This shows that, for any map f, we have a right inverse $[f^{-1}]$ for $[f]$, which is sufficient for $[f^{-1}]$ to be a two-sided inverse for $[f]$.

Remark 4.4 The earlier papers [17, 20] define a higher digital homotopy group in a way that superficially uses the same ingredients that we do here. Some of the deductions in that work appear to be logically flawed, and a number of proofs are omitted. But the main difference between that work and ours stems from subtle but vital differences in the basic approach. In [17, 20] a rectangle $I_{m,m}$ is assumed to have only 4-adjacencies, which means that many more maps from a rectangle are admitted as continuous than are in our work. Furthermore, the "box" homotopy is used, which means that maps are more easily homotopic there than in our work here. These differences result in totally different invariants. For instance, the 2-sphere that we use in this paper is easily shown to be contractible if the box homotopy is used in place of ours. All the higher homotopy groups of [17, 20] would thus be trivial for our 2-sphere and in fact no non-trivial example of a higher homotopy group is given in that work.

The next result shows that the second homotopy group is independent of choice of basepoint.

Proposition 4.5 *If x and x' are in the same (path-connected) component of X, then* $\pi_2(X, x) \cong \pi_2(X, x')$.

Proof Since x and x' are in the same component of X, there is a sequence of adjacencies $x = x_0 \sim x_1 \sim \cdots \sim x_n = x'$. By induction, it suffices to show that $\pi_2(X, x_0)$ is isomorphic to $\pi_2(X, x_1)$. For a map $f : (I_{m,n}, \partial I_{m,n}) \to (X, x_0)$, define a map $f_{x_1} : (I_{m+2,n+2}, \partial I_{m+2,n+2}) \to (X, x_1)$ by

$$f_{x_1}(a, b) = \begin{cases} f(a - 1, b - 1) & \text{if } (a, b) \in [1, m + 1]_{\mathbb{Z}} \times [1, n + 1]_{\mathbb{Z}} \\ x_1 & \text{otherwise .} \end{cases}$$

If we picture maps from a rectangle as a labeling of points by their values in X, as we have before, this simply takes f and surrounds it by a border of x_1. We claim that this induces a well-defined map $\Phi_{x_1} : \pi_2(X, x_0) \to \pi_2(X, x_1)$. For suppose we have $[f] = [g] \in \pi_2(X, x_0)$, with $f : (I_{m,n}, \partial I_{m,n}) \to (X, x_0)$ and $g : (I_{r,s}, \partial I_{r,s}) \to (X, x_0)$. Then, there are trivial extensions $\overline{f}, \overline{g} : (I_{\overline{m},\overline{n}}, \partial I_{\overline{m},\overline{n}}) \to (X, x_0)$ of each and a homotopy relative the boundary $H : I_{\overline{m},\overline{n}} \to X$ from $\overline{f}$ to $\overline{g}$. By Lemma 2.8, this homotopy extends to a homotopy relative the boundary $H : I_{\overline{m}+2,\overline{n}+2} \to X$ from $(\overline{f})_{x_1}$ to $(\overline{g})_{x_1}$. In the following sequence of extension homotopies, identifications and homotopies

$$f_{x_1} \approx f_{x_1} \circ \alpha_{m+2}^{\overline{m}-m} \circ \beta_{n+2}^{\overline{n}-n} = (\overline{f})_{x_1} \simeq (\overline{g})_{x_1} = g_{x_1} \circ \alpha_{r+2}^{\overline{m}-r} \circ \beta_{s+2}^{\overline{n}-s} \approx g_{x_1},$$

the first and last extension homotopies follow from repeated application of Theorem 3.2, the middle homotopy is the one we just observed, and the identifications follow from the definitions of the maps involved. Hence, we may define a map $\Phi_{x_1} : \pi_2(X, x_0) \to \pi_2(X, x_1)$ by setting $\Phi_{x_1}([f]) = [f_{x_1}]$. We show that Φ_{x_1} is an isomorphism.

To show that Φ_{x_1} is a homomorphism, we must show that $[(f \cdot g)_{x_1}] = [f_{x_1}] \cdot [g_{x_1}]$ in $\pi_2(X, x_1)$, for $[f], [g] \in \pi_2(X, x_0)$. Firstly, if $f : (I_{m,n}, \partial I_{m,n}) \to (X, x_0)$ and $g : (I_{r,s}, \partial I_{r,s}) \to (X, x_0)$, then we have an extension homotopy

$$(f \cdot g)_{x_1} \approx (f \cdot g)_{x_1} \circ \alpha_{m+1}^2 \circ \beta_{n+1}^2$$

from repeated application of Theorem 3.2. For brevity, let $h = (f \cdot g)_{x_1} \circ \alpha_{m+1}^2 \circ \beta_{n+1}^2$.

It suffices now to show that $h \simeq f_{x_1} \cdot g_{x_1}$. These maps are pictured in Fig. 6. We see that h and $f_{x_1} \cdot g_{x_1}$ differ only in certain points whose value under h is x_0, while the value under $f_{x_1} \cdot g_{x_1}$ is x_1. But all of these points have adjacent values only of x_0 or x_1. Thus, we may perform repeated spider moves on h which change all of these values of x_0 to x_1. In this way, we obtain a homotopy $h \simeq f_{x_1} \cdot g_{x_1}$ as required.

Next we show that given a continuous map of based digital images $\phi : (X, x_0) \to (Y, y_0)$, there is a natural induced homomorphism on the fundamental group. This induced homomorphism is invariant under the following natural type of homotopy: We say that two based maps $\phi, \psi : (X, x_0) \to (Y, y_0)$ are *homotopic relative to the basepoint* or *based homotopic* when there is some continuous $H : X \times I_k \to Y$ such that H is a homotopy from ϕ to ψ and $H(x_0, t) = y_0$ for all $t \in I_k$.

$(f \cdot g)_{x_1}$

$h = (f \cdot g)_{x_1} \circ \alpha_{m+1}^2 \circ \beta_{n+1}^2$
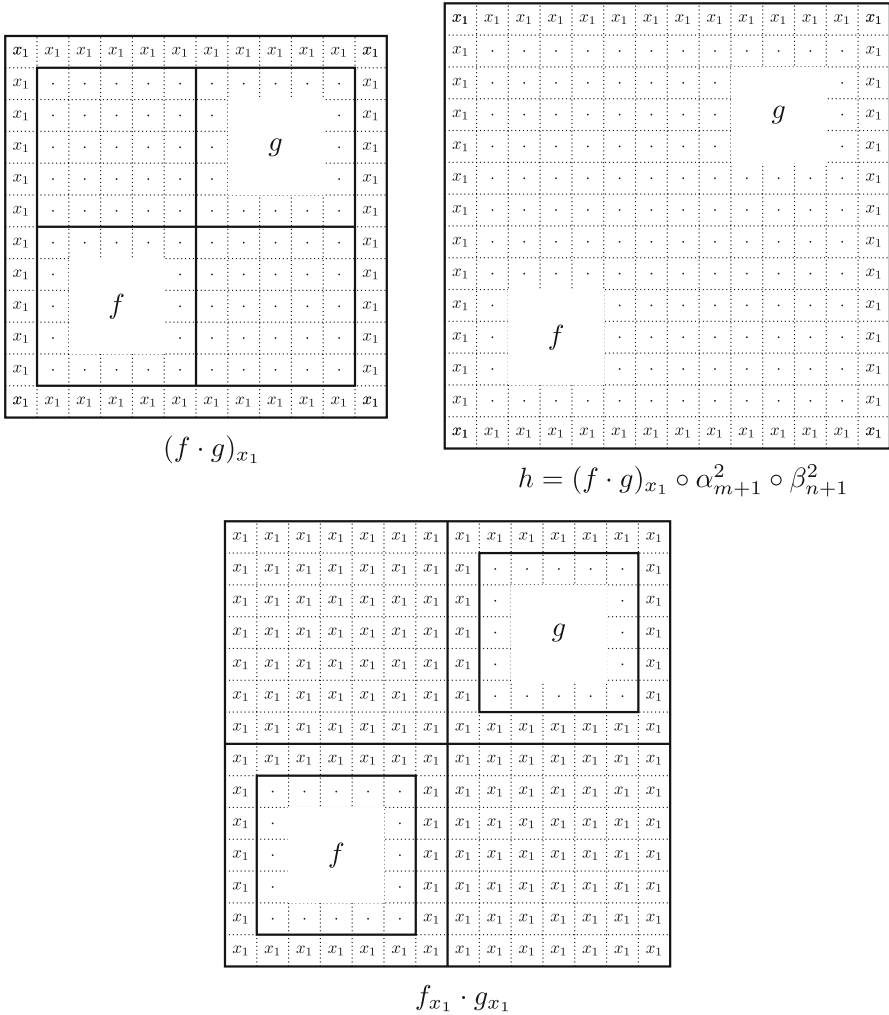
$f_{x_1} \cdot g_{x_1}$

Fig. 6 Various maps used in the proof of Proposition 4.5, which shows that $(f \cdot g)_{x_1} \approx (f \cdot g)_{x_1} \circ \alpha_{m+1}^2 \circ \beta_{n+1}^2 \simeq f_{x_1} \cdot g_{x_1}$. (Dots represent the base point x_0)

Proposition 4.6 *If $\phi \colon (X, x_0) \to (Y, y_0)$ is a based digital map between based digital images, there is an induced homomorphism $\phi_* \colon \pi_2(X, x_0) \to \pi_2(Y, y_0)$ given by $\phi_*([f]) = [\phi \circ f]$.*

This $*$ operator is functorial in the sense that $(\phi \circ \psi)_* = \phi_* \circ \psi_*$ for any based maps ϕ and ψ, and also $(\mathrm{id}_X)_* = \mathrm{id}_{\pi_2(X, x_0)}$, where id_X denotes the identity function of X.

Furthermore, if ϕ and ψ are based homotopic, then $\phi_* = \psi_*$.

Proof To show that ϕ_* is a group homomorphism, we observe:

$$\phi_*([f] \cdot [g]) = \phi_*([f \cdot g]) = [\phi \circ (f \cdot g)] = [(\phi \circ f) \cdot (\phi \circ g)] = [\phi \circ f] \cdot [\phi \circ g] = \phi_*([f]) \cdot \phi_*([g]).$$

It is clear from our definitions that $(\psi \circ \phi)_*([f]) = [\psi \circ \phi \circ f] = \psi_*[\phi \circ f] = \psi_* \circ \phi_*([f])$ and $(\mathrm{id}_X)_*([f]) = [\mathrm{id}_X \circ f] = [f] = \mathrm{id}_{\pi_2(X,x_0)}[f]$, proving functoriality.

For the last statement, let $\phi, \psi : X \to Y$ be based homotopic, and we will show $\phi_* = \psi_*$. Since ϕ and ψ are homotopic, there exists a based homotopy $H : X \times I_k \to Y$. Let $[f] \in \pi_2(X, x_0)$. To see that $[\phi \circ f] = [\psi \circ f]$, observe that $I_{m,n} \times I_k \to X \times I_k \to Y$ is a homotopy between $\phi \circ f$ and $\psi \circ g$. Because H is a based homotopy, this homotopy of $\phi \circ f$ and $\psi \circ g$ is a homotopy relative to the boundary, and so $[\phi \circ f] = [\psi \circ f]$.

Stated more abstractly, the above means that the second homotopy group π_2 is a functor from the category of based digital images and homotopy classes of digitally continuous maps to the category of abelian groups and group homomorphisms.

Two based digital images (X, x_0) and (Y, y_0) are *based homotopy equivalent* when there are based maps $\phi : (X, x_0) \to (Y, y_0)$ and $\psi : (Y, y_0) \to (X, x_0)$ with $\phi \circ \psi$ and $\psi \circ \phi$ each based homotopic to identity maps on (X, x_0) and (Y, y_0), respectively.

Theorem 4.7 *Let (X, x_0) and (Y, y_0) be based homotopy equivalent. Then, $\pi_2(X, x_0)$ and $\pi_2(Y, y_0)$ are isomorphic.*

Proof Suppose we have based maps $\phi : (X, x_0) \to (Y, y_0)$ and $\psi : (Y, y_0) \to (X, x_0)$ with $\phi \circ \psi$ and $\psi \circ \phi$ each based homotopic to identity maps on (X, x_0) and (Y, y_0), respectively. Then, it follows from Proposition 4.6 that we have $\phi_* \circ \psi_* = (\phi \circ \psi)_* = (\mathrm{id}_X)_* = \mathrm{id}_{\pi_2(X,x_0)}$ and likewise $\psi_* \circ \phi_* = \mathrm{id}_{\pi_2(Y,y_0)}$. Hence, each of ϕ_* and ψ_* must be an isomorphism. ∎

Theorem 4.8 *Given any pointed digital image (X, x_0), the group $\pi_2(X, x_0)$ is abelian.*

Proof The result follows using translations of the type described in Remark 3.5 and which flow from Corollary 3.4 and Lemma 2.8. We have $f \cdot g \simeq g \cdot f$ by homotopies indicated as follows:

Recall that both f and g map their boundaries to the base point x_0. This means that, in the diagrams above, we may slide these blocks alongside each other without breaking continuity of the intermediate maps.

The last general result that we give shows that our second homotopy group behaves with respect to products like the second homotopy group in the ordinary topological

setting. That is, our second homotopy group preserves products in the functorial sense. Recall that, for based digital images (X, x_0) and (Y, y_0), their product $X \times Y$ is the categorical product and its basepoint is the point $(x_0, y_0) \in X \times Y$. The product of two (abelian) groups, denoted here by '$\times$,' means their direct product. This result uses the induced homomorphisms just discussed.

Theorem 4.9 *Let (X, x_0) and (Y, y_0) be any based digital images. Let $p_1 \colon X \times Y \to X$ and $p_2 \colon X \times Y \to Y$ denote the projections onto either factor. Define a map*

$$\Psi \colon \pi_2\big(X \times Y; (x_0, y_0)\big) \to \pi_2(X; x_0) \times \pi_2(Y; y_0),$$

by setting $\Psi([\alpha]) := \big((p_1)_*([\alpha]), (p_2)_*([\alpha])\big)$ for each $[\alpha] \in \pi_2\big(X \times Y; (x_0, y_0)\big)$. Then Ψ is an isomorphism.

Proof Because $(p_1)_*$ and $(p_2)_*$ are both well defined and homomorphisms, it follows that so too is Ψ well-defined and a homomorphism. We show that Ψ is both surjective and injective, and thus an isomorphism.

For surjectivity, suppose that we have $([\alpha], [\beta]) \in \pi_2(X; x_0) \times \pi_2(Y; y_0)$, with $\alpha \colon (I_{m,n}, \partial I_{m,n}) \to (X, x_0)$ and $\beta \colon (I_{m',n'}, \partial I_{m',n'}) \to (Y, y_0)$. Then the maps $(\alpha, c_{y_0}) \colon I_{m,n} \to X \times Y$ and $(c_{x_0}, \beta) \colon I_{m',n'} \to X \times Y$ represent elements of $\pi_2\big(X \times Y; (x_0, y_0)\big)$. We have

$$\Psi\big([(\alpha, c_{y_0})] \cdot [(c_{x_0}, \beta)]\big) = \Psi\big([(\alpha, c_{y_0})]\big)\Psi\big([(c_{x_0}, \beta)]\big) = ([\alpha], [c_{y_0}])([c_{x_0}], [\beta]) = ([\alpha], [\beta]).$$

It follows that Ψ is surjective.

For injectivity, suppose that we have $[\alpha] \in \pi_2\big(X \times Y; (x_0, y_0)\big)$ represented by a map $\alpha \colon I_{m,n} \to X \times Y$, such that $\alpha \in \ker \Psi$. That is, $\Psi([\alpha]) = ([c_{x_0}], [c_{y_0}]) \in \pi_2(X; x_0) \times \pi_2(Y; y_0)$. Then, $p_1 \circ \alpha$ and $p_2 \circ \alpha$ are extension-homotopic to constant maps. Suppose we have $\overline{p_1 \circ \alpha} \simeq c_{x_0}$ via a homotopy relative to the boundary $H \colon I_{m',n'} \times I_T \to X$ and $\overline{p_2 \circ \alpha} \simeq c_{y_0}$ via a homotopy relative to the boundary $G \colon I_{m'',n''} \times I_S \to Y$. The idea is simply to "same size" the domains of these homotopies using a process akin to a "3D trivial extension.". Firstly, we trivially extend $\overline{p_1 \circ \alpha}$ and $\overline{p_2 \circ \alpha}$ to a larger common domain $I_{m''',n'''}$. We continue to write these new trivial extensions as $\overline{p_1 \circ \alpha}$ and $\overline{p_2 \circ \alpha}$. On any point (i, j) of $I_{m''',n'''}$ not in $I_{m',n'}$, extend H by setting $H(i, j, t) = x_0$ for all $t \in I_T$. Likewise, extend G to the stationary homotopy at y_0 on points of $I_{m''',n'''}$ not in $I_{m'',n''}$. Since H and G were originally homotopies relative to the boundary (stationary at their respective basepoints), these extensions of the homotopies to the larger domains are evidently continuous. Secondly, if $S \neq T$, then we may lengthen the shorter one by adding stationary steps at the basepoint. Note that this depends on either homotopy ending at the constant map. Thus, we have trivially extended our homotopies to maps $H \colon I_{m''',n'''} \times I_R \to X$ and $G \colon I_{m''',n'''} \times I_R \to Y$, on some common domain. Then, we have a homotopy relative to the boundary

$$\big(H, G\big) \colon I_{m''',n'''} \times I_R \to X \times Y$$

from $\big(\overline{p_1 \circ \alpha}, \overline{p_2 \circ \alpha}\big)$ to $\big(c_{x_0}, c_{y_0}\big)$. Now may write $\alpha = (p_1 \circ \alpha, p_2 \circ \alpha)$ and hence $\big(\overline{p_1 \circ \alpha}, \overline{p_2 \circ \alpha}\big)$ as a trivial extension $\overline{\alpha}$ of α, and also $c_{(x_0, y_0)} = \big(c_{x_0}, c_{y_0}\big)$. it follows

that we have $[\alpha] = [c_{(x_0, y_0)}] \in \pi_2\big(X \times Y; (x_0, y_0)\big)$. Hence Ψ is injective and the result follows.

## 5 A triangle-counting function on $\pi_2(S^2)$

We now turn to the computation of the second homotopy group of a digital two-sphere. Various models have appeared in the literature of sphere-like digital images (see [10] for instance). We will define the digital sphere S^n as in classical geometry as the set of points in $(n + 1)$-dimensional space at unit distance from the origin. In the digital setting, namely with points from $\mathbb{Z}^n$, this includes only points with a single nonzero coordinate of magnitude 1.

Specifically let $\mathbf{e}_i = (0, \ldots, 0, 1, 0, \ldots, 0)$ be the i-th standard basis vector in $\mathbb{Z}^{n+1}$, and define *the digital n-sphere*

$$S^n := \{\pm\mathbf{e}_1, \ldots, \pm\mathbf{e}_{n+1}\}.$$

By S^n we always mean this digital image of $2n$ points, rather than the classical manifold S^n. We always consider S^n as a digital image with c_n adjacency, that is, two points are adjacent when they differ by at most 1 in each of their coordinates.

From now on, we deal exclusively with S^2, the set of 6 points whose (non-self-) adjacencies form the octahedral graph:
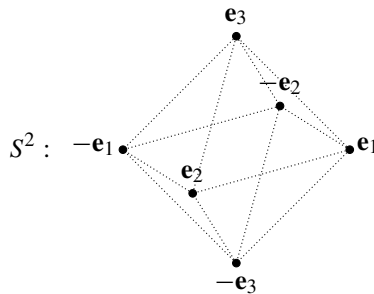
For $a, b \in S^2$ with our chosen adjacency, we will have $a \sim b$ if and only if $a \neq -b$. As discussed between Definitions 2.2 and 2.3, we should really add loops at each vertex in the figure above to represent our S^2 strictly from the graph-theoretic point of view. But we typically suppress these self-adjacencies from our figures.

The rest of the paper is concerned with showing that we have a group isomorphism $\pi_2(S^2, -\mathbf{e}_1) \cong \mathbb{Z}$. The main tool we use for showing this is the following *triangle-counting function* for maps from the rectangle into S^2.

We may view a map $f : (I_{m,n}, \partial I_{m,n}) \to (S^2, -\mathbf{e}_1)$ as a labeling of the points of the rectangle $I_{m,n}$ with values from $S^2 = \{\pm\mathbf{e}_1, \pm\mathbf{e}_2, \pm\mathbf{e}_3\}$. Furthermore, we choose the triangulation of the rectangle $I_{m,n}$ that uses horizontal and vertical edges (adjacencies) between points, together with all positively sloped diagonal edges (adjacencies). (Here,

Fig. 7 Triangulation of $I_{m,n}$ shown with $m = 5$ and $n = 4$. Labeling with points from S^2 shows a map with $d(f) = 0$
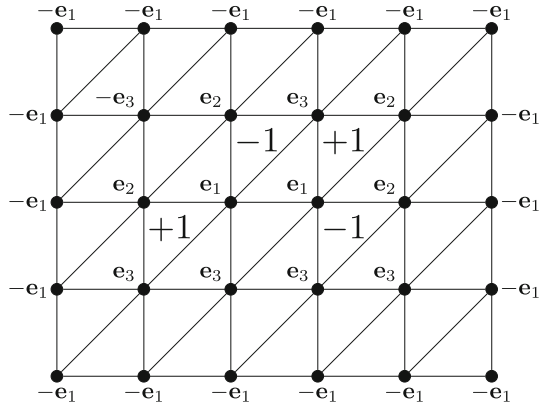
we are setting aside all negatively sloped diagonal edges (adjacencies) that also exist in the adjacency relation of $\mathbb{Z}^2$.)

Definition 5.1 (Triangle-counting function) With the conventions above, define an integer $d(f)$ as the signed sum of the number of oriented triangles labeled $\langle \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3 \rangle$ in this triangulation. By oriented triangles, we mean that a triangle labeled $\langle \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3 \rangle$ in a counter-clockwise sense counts as $+1$ and a triangle labeled $\langle \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3 \rangle$ in a clockwise sense counts as -1.

Example 5.2 Take the map $f : (I_{5,4}, \partial I_{5,4}) \to (S^2, -\mathbf{e}_1)$ whose values on the points of $I_{5,4}$ are as specified in Fig. 7. We have triangulated $I_{5,4}$ in the way described above. Note that although we have not included diagonal edges of slope -1 in our triangulation, pairs of points that would be connected by such must be labeled with adjacent values from S^2 as well, to preserve continuity of the map f. We find that there are two triangles labeled $\langle \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3 \rangle$ in a counter-clockwise sense and two in a clockwise sense, leading to a signed sum of 0. We will see that d provides a function $d : \pi_2(S^2, -\mathbf{e}_1) \to \mathbb{Z}$, which we will eventually prove is a group isomorphism. Thus, this map represents the trivial element in $\pi_2(S^2, -\mathbf{e}_1)$, namely it must be extension-homotopic to the constant map $c_{-\mathbf{e}_1} \in S^2$.

First we show that the triangle-counting function is preserved by extension-homotopy.

Lemma 5.3 *If $f \approx g$, then we have $d(f) = d(g)$. Thus, the integer $d(f)$ described above induces a well-defined induced triangle-counting function $d : \pi_2(S^2, -\mathbf{e}_1) \to \mathbb{Z}$, given by setting $d([f]) = d(f)$.*

Proof Let $f : (I_{m,n}, \partial I_{m,n}) \to (S^2, -\mathbf{e}_1)$ and $g : (I_{m',n'}, \partial I_{m',n'}) \to (S^2, -\mathbf{e}_1)$ be maps with $f \approx g$. Recall that this means there are trivial extensions $\overline{f}$ of f and $\overline{g}$ of g such that $\overline{f}$ and $\overline{g}$ are defined on the same-sized rectangle as each other and are homotopic via a homotopy relative to the boundary. Firstly, it is clear that we have $d(f) = d(\overline{f})$ and $d(g) = d(\overline{g})$, since a trivial extension preserves the labels of the original rectangle and simply labels additional points in the larger containing
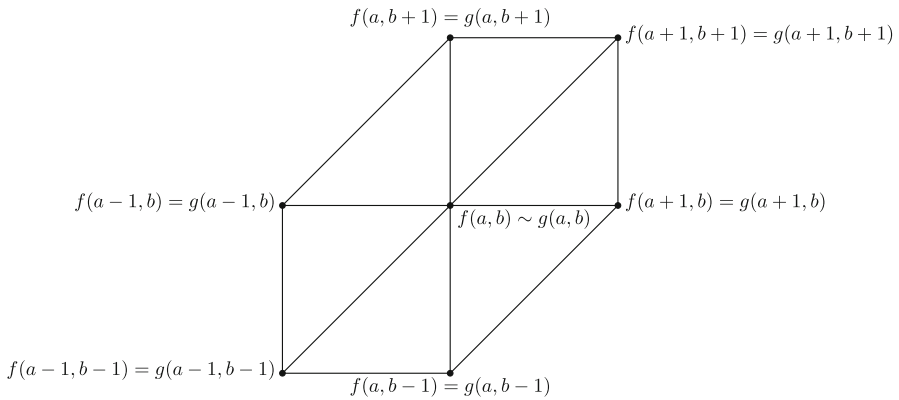
Fig. 8 The hexagon of the 6 triangles in $I_{m,n}$ surrounding (a, b) labeled with points from S^2

rectangle with $-\mathbf{e}_1$, thereby preserving all the original triangles labeled $\langle \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3 \rangle$ and not introducing any additional ones. So, it is sufficient to show that we have $d(f) = d(g)$ when f and g are homotopic (and defined on the same-sized rectangle as each other).

By Theorem 2.7, any homotopy can be effected by a sequence of spider moves which change only one point at a time. Thus, it suffices to show that $d(f) = d(g)$ when f and g are homotopic by a spider move that changes the values of f at only one point. So, we consider the situation in which $f, g \colon (I_{m,n}, \partial I_{m,n}) \to (S^2, -\mathbf{e}_1)$ differ in value only at the point $(a, b) \in I_{m,n}$, for some a and b with $0 < a < m$ and $0 < b < n$ (recall that we do not change the value of boundary points through our homotopy).

Since the spider move only changes the map at the point (a, b), with everything else remaining unchanged, the signed counts of triangles $d(f)$ and $d(g)$ may only differ according as the counts of triangles labeled $\langle \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3 \rangle$ differ in what we will call *the hexagon* of labeled points in $I_{m,n}$, namely the points (labeled in their respective positions) illustrated in Fig. 8.

If either $f(a, b)$ or $g(a, b)$ has a value from $\{-\mathbf{e}_1, -\mathbf{e}_2, -\mathbf{e}_3\}$ then the hexagon does not contribute any triangles labeled $\langle \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3 \rangle$ either before or after the spider move. This follows because both $f(a, b)$ and $g(a, b)$ must be adjacent to all values taken by points of the hexagon—including both $f(a, b)$ and $g(a, b)$. But if one of $f(a, b)$ or $g(a, b)$ takes a negative value $-\mathbf{e}_i$, then its positive counterpart $\mathbf{e}_i$ must be absent from the hexagon—again including both values $f(a, b)$ and $g(a, b)$. However, we need all three values $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ to be taken on the hexagon by f or g in order to have any triangles to count at all. Thus, in this case the signed sums $d(f)$ and $d(g)$ are determined entirely by triangles in $I_{m,n}$ not involving the point (a, b), and on which f and g agree. That is, we have $d(f) = d(g)$.

It remains to consider the cases in which both of $f(a, b)$ and $g(a, b)$ have (different) values from $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$. Say $f(a, b) = \mathbf{e}_i$ and $g(a, b) = \mathbf{e}_j$. Let k be the third value so that no two of $\{\mathbf{e}_i, \mathbf{e}_j, \mathbf{e}_k\}$ are equal. Without loss of generality, assume that $j = i + 1$ mod 3, which means that $k = i + 2 \mod 3$. Then the oriented count of $\langle \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3 \rangle$

Fig. 9 Subgraph of S^2 on which we have the circuit (1)
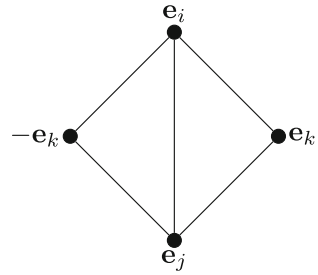
triangles will equal the oriented count of $\langle \mathbf{e}_i, \mathbf{e}_j, \mathbf{e}_k \rangle$ triangles, so we can compute $d(f)$ and $d(g)$ by counting $\langle \mathbf{e}_i, \mathbf{e}_j, \mathbf{e}_k \rangle$ triangles.

Consider the values assigned to the 6 points of the hexagon that surround (a, b). Since each of these points must be labeled with values in S^2 adjacent to both $\mathbf{e}_i$ and $\mathbf{e}_j$, all 6 points must be labeled from amongst the points $\{\mathbf{e}_i, \mathbf{e}_j, \pm\mathbf{e}_k\} \subseteq S^2$. If none of these 6 points is labeled $\mathbf{e}_k$, then the hexagon does not display any triangles labeled $\langle \mathbf{e}_i, \mathbf{e}_j, \mathbf{e}_k \rangle$ either before or after this spider move. So, further assume that at least one of the 6 points of the hexagon that surround (a, b) is labeled $\mathbf{e}_k$. Starting at one such point, travel counter-clockwise in a loop around the six vertices, listing the values with which they are labeled. The result is a circuit γ of length at most 6

$$\gamma = (\mathbf{e}_k, v_1, v_2, v_3, v_4, v_5, \mathbf{e}_k) \tag{1}$$

in the subgraph $\{\mathbf{e}_i, \mathbf{e}_j, \pm\mathbf{e}_k\}$ of S^2, pictured with its adjacencies in Fig. 9.

Let $N(f)$ be the oriented count of $\langle \mathbf{e}_i, \mathbf{e}_j, \mathbf{e}_k \rangle$ triangles of f occurring in the hexagon of Fig. 8. Let $N(g)$ be the same count in the hexagon for g. Since f and g agree outside of the hexagon, we need only show that $N(f) = N(g)$.

Since f maps the center point to $\mathbf{e}_i$, and γ is oriented clockwise around the hexagon, the count $N(f)$ will equal the oriented count of the number of edges $(\mathbf{e}_j, \mathbf{e}_k)$ in γ. Similarly, the count $N(g)$ equals the oriented count of the number of edges $(\mathbf{e}_k, \mathbf{e}_i)$ in γ.

For some edge given as a vertex-pair (v, v'), let $K_\gamma(v, v')$ be the number of occurrences of the edge (v, v') in γ. Then, the above means that:

$$N(f) = K_\gamma(\mathbf{e}_j, \mathbf{e}_k) - K_\gamma(\mathbf{e}_k, \mathbf{e}_j),$$
$$N(g) = K_\gamma(\mathbf{e}_k, \mathbf{e}_i) - K_\gamma(\mathbf{e}_i, \mathbf{e}_k).$$

Since γ is a cycle based at $\mathbf{e}_k$, its in-degree at $\mathbf{e}_k$ must match its out-degree at $\mathbf{e}_k$. That is, $K_\gamma(\mathbf{e}_i, \mathbf{e}_k) + K_\gamma(\mathbf{e}_j, \mathbf{e}_k) = K_\gamma(\mathbf{e}_k, \mathbf{e}_i) + K_\gamma(\mathbf{e}_k, \mathbf{e}_j)$. Combining this with the equations just displayed gives $N(f) - N(g) = 0$ and so $N(f) = N(g)$ as desired.

Proposition 5.4 *The induced triangle-counting function* $d : \pi_2(S^2, -\mathbf{e}_1) \to \mathbb{Z}$ *is a group homomorphism. In particular, we have* $d\left([c_{-\mathbf{e}_1}]\right) = 0$ *and* $d\left([f]^{-1}\right) = d\left([f^{-1}]\right) = -d\left([f]\right)$ *for any* $[f] \in \pi_2(S^2, -\mathbf{e}_1)$.

Fig. 10 The map
$T : (I_{4,4}, \partial I_{4,4}) \to (S^2, -\mathbf{e}_1)$
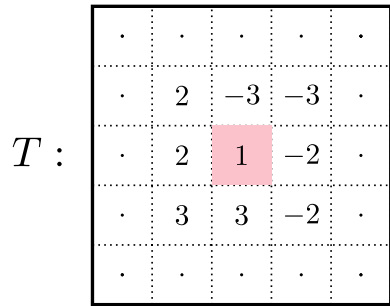
Proof Let $f, g : (I_{m,n}, \partial I_{m,n}) \to (S^2, -\mathbf{e}_1)$ be two maps representing elements of $\pi_2(S^2, -\mathbf{e}_1)$. It is sufficient to show that $d(f \cdot g) = d(f) + d(g)$.

Recall that $f \cdot g$ simply juxtaposes the grids defining f and g into a larger grid. Since f and g each individually map their boundary rectangles to the base point, there is no opportunity for the formation of new triangles involving $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ where the grids of f and g meet. Thus, the total oriented count of $\langle \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3 \rangle$ triangles of $f \cdot g$ will equal the sum of the oriented counts for each of f and g, which is to say $d(f \cdot g) = d(f) + d(g)$.

The last two assertions follow formally for any group homomorphism $d : G \to \mathbb{Z}$: with G any group, we have $d(e) = 0$ and $d(g^{-1}) = -d(g)$ for any $g \in G$ and $e \in G$ the identity element. The first of these is also easy to see directly: a constant map $c_{-\mathbf{e}_1}$ contains no triangles labeled $\langle \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3 \rangle$. The second is not so easy to see directly, as passing from f to f^{-1} involves re-labeling the points of a rectangle while preserving the triangulation—see the examples of $T, T^{-1} : (I_{4,4}, \partial I_{4,4}) \to (S^2, -\mathbf{e}_1)$ given below.

We can now state the main result of this part of the paper:

Theorem 5.5 *The induced triangle-counting function $d : \pi_2(S^2, -\mathbf{e}_1) \to \mathbb{Z}$ is a group isomorphism.*

The proof of Theorem 5.5 occupies the remainder of the paper. We will prove surjectivity immediately; the proof of injectivity requires some preparation.

First, we introduce a specific map that will end up playing a prominent role in our calculation as a generator of $\pi_2(S^2, -\mathbf{e}_1)$.

Definition 5.6 Let $T : (I_{4,4}, \partial I_{4,4}) \to (S^2, -\mathbf{e}_1)$ be the map given by the labeling of points of $I_{4,4}$ with values from S^2 as in Fig. 10. In this diagram, we have used the style of earlier sections and indicated a label of the basepoint $-\mathbf{e}_1$ with a dot. Furthermore, we have indicated labels of the three standard basis vectors with their subscript, and labels of $-\mathbf{e}_2$, respectively, $-\mathbf{e}_3$, by -2, respectively, -3. We will adopt this style of diagram going forward.

Lemma 5.7 *(Surjective part of Theorem 5.5) The map $T : (I_{4,4}, \partial I_{4,4}) \to (S^2, -\mathbf{e}_1)$ given in Definition 5.6 satisfies $d(T) = 1$. Hence, the triangle-counting function $d : \pi_2(S^2, -\mathbf{e}_1) \to \mathbb{Z}$ is surjective.*

Fig. 11 The map
$T : (I_{4,4}, \partial I_{4,4}) \to (S^2, -\mathbf{e}_1)$
has $d(T) = 1$
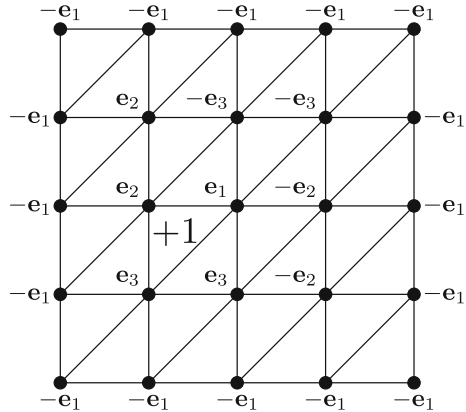
Proof Figure 11 shows the map T with the triangulation we use to define the value of d. In the figure, we see exactly one $\langle \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3 \rangle$ triangle, oriented in the positive direction. Thus we have $d([T]) = 1$. The consequence for surjectivity of d follows immediately, since we already have shown that d is a homomorphism. Note that, *per* Proposition 5.4, we have $d([T^{-1}]) = -1$ and $d\left([c_{-\mathbf{e}_1}]\right) = 0$.

We now break off from the proof of Theorem 5.5 to prepare for showing injectivity of the induced triangle-counting function. The proof of Theorem 5.5 is completed at the end of the next section.

6 Islands, Flooding, and Injectivity of d

We show several lemmas that will be used in showing injectivity of d. These lemmas are summarized in Theorem 6.5 below. The first lemma is a simple application of spider moves.

Lemma 6.1 *Let* $f : (I_{4,4}, \partial I_{4,4}) \to (S^2, -\mathbf{e}_1)$ *be continuous with* $f(x) = \mathbf{e}_1$ *for exactly one* $x \in I_{4,4}$. *Then* f *is homotopic to a map of the form:*
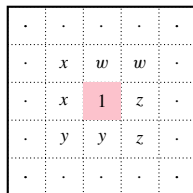
for $x, y, z, w \in \{\pm \mathbf{e}_2, \pm \mathbf{e}_3\}$.

Furthermore, if $\{x, y, z, w\} \neq \{\pm \mathbf{e}_2, \pm \mathbf{e}_3\}$, *then* f *is homotopic to the constant map with constant value* $-\mathbf{e}_1$.

Proof Because f maps the boundary of $I_{4,4}$ to $-\mathbf{e}_1$, the only point which can map to $\mathbf{e}_1$ is the center point $(2, 2) \in I_{4,4}$. Thus our map f must take the following form:

$$f = \begin{array}{|c|c|c|c|c|} \hline \cdot & \cdot & \cdot & \cdot & \cdot \\ \hline \cdot & x_1 & x_2 & x_3 & \cdot \\ \hline \cdot & x_4 & 1 & x_5 & \cdot \\ \hline \cdot & x_6 & x_7 & x_8 & \cdot \\ \hline \cdot & \cdot & \cdot & \cdot & \cdot \\ \hline \end{array}$$

for some $x_1, \ldots, x_8 \in \{\pm\mathbf{e}_2, \pm\mathbf{e}_3\}$. We may do several spider move homotopies as in Lemma 2.6. For example the value of $f(1, 3) = x_1$ may be changed to x_4 because all neighbors of $(1, 3)$ have labels which are already adjacent to x_4. Performing 4 similar spider moves results in the following map homotopic to f:

$$f \simeq \begin{array}{|c|c|c|c|c|} \hline \cdot & \cdot & \cdot & \cdot & \cdot \\ \hline \cdot & x_4 & x_2 & x_2 & \cdot \\ \hline \cdot & x_4 & 1 & x_5 & \cdot \\ \hline \cdot & x_7 & x_7 & x_5 & \cdot \\ \hline \cdot & \cdot & \cdot & \cdot & \cdot \\ \hline \end{array}$$

which has the desired format, proving the first statement of the lemma.

For the second statement, assume that there is some $v \in \{\pm\mathbf{e}_2, \pm\mathbf{e}_3\}$ which is different from all values of f. Then all interior values of f can be changed by spider moves to $-v$, and then to $-\mathbf{e}_1$:

$$f \simeq \begin{array}{|c|c|c|c|c|} \hline \cdot & \cdot & \cdot & \cdot & \cdot \\ \hline \cdot & -v & -v & -v & \cdot \\ \hline \cdot & -v & -v & -v & \cdot \\ \hline \cdot & -v & -v & -v & \cdot \\ \hline \cdot & \cdot & \cdot & \cdot & \cdot \\ \hline \end{array} \simeq \begin{array}{|c|c|c|c|c|} \hline \cdot & \cdot & \cdot & \cdot & \cdot \\ \hline \cdot & \cdot & \cdot & \cdot & \cdot \\ \hline \cdot & \cdot & \cdot & \cdot & \cdot \\ \hline \cdot & \cdot & \cdot & \cdot & \cdot \\ \hline \cdot & \cdot & \cdot & \cdot & \cdot \\ \hline \end{array}$$

Now the map T introduced in Definition 5.6 has $T(x) = \mathbf{e}_1$ at exactly one point x and satisfies $d(T) = 1$. The next lemma shows that these properties effectively characterize T up to extension homotopy.

Lemma 6.2 *Let $f : (I_{4,4}, \partial I_{4,4}) \to (S^2, -\mathbf{e}_1)$ be continuous with $f(x) = \mathbf{e}_1$ for exactly one $x \in I_{4,4}$ and $d(f) = 1$. Then, $f \approx T$.*

Proof By Lemma 6.1 we may assume that f takes the form:

.	.	.	.	.
.	x	w	w	.
.	x	1	z	.
.	y	y	z	.
.	.	.	.	.

where the values x, y, z, w are taken from the set of possible values $\{\pm \mathbf{e}_2, \pm \mathbf{e}_3\}$. Since we assume that $d(f) = 1$, this means that f is not homotopic to a constant map, and so $\{x, y, z, w\} = \{\pm \mathbf{e}_2, \pm \mathbf{e}_3\}$ by the second part of Lemma 6.1.

Since (x, y, z, w) are all distinct, and we must have $x \sim y \sim z \sim w \sim x$, these values in order must be some cyclic permutation of $(\mathbf{e}_2, \mathbf{e}_3, -\mathbf{e}_2, -\mathbf{e}_3)$ or of $(\mathbf{e}_2, -\mathbf{e}_3, -\mathbf{e}_2, \mathbf{e}_3)$. In fact, only cyclic permutations of the first type will result in $d(f) = 1$. The others will result in $d(f) = -1$, and so we do not consider them. Thus, f is homotopic to one of the following four maps:

$$
\begin{array}{cccc}
\begin{array}{|ccccc|}
\hline
. & . & . & . & . \\
. & 2 & -3 & -3 & . \\
. & 2 & 1 & -2 & . \\
. & 3 & 3 & -2 & . \\
. & . & . & . & . \\
\hline
\end{array}
&
\begin{array}{|ccccc|}
\hline
. & . & . & . & . \\
. & -3 & -2 & -2 & . \\
. & -3 & 1 & 3 & . \\
. & 2 & 2 & 3 & . \\
. & . & . & . & . \\
\hline
\end{array}
&
\begin{array}{|ccccc|}
\hline
. & . & . & . & . \\
. & -2 & 3 & 3 & . \\
. & -2 & 1 & 2 & . \\
. & -3 & -3 & 2 & . \\
. & . & . & . & . \\
\hline
\end{array}
&
\begin{array}{|ccccc|}
\hline
. & . & . & . & . \\
. & 3 & 2 & 2 & . \\
. & 3 & 1 & -3 & . \\
. & -2 & -2 & -3 & . \\
. & . & . & . & . \\
\hline
\end{array}
\end{array}
\tag{2}
$$

The first of these maps is T, so it will suffice to show that any of these maps is extension-homotopic to the others.

We do this by demonstrating a "rotation of values" by extension homotopies. Our demonstration is pictorial. In the sequence of diagrams below, each step is either a pair of consecutive spider moves or a row doubling:
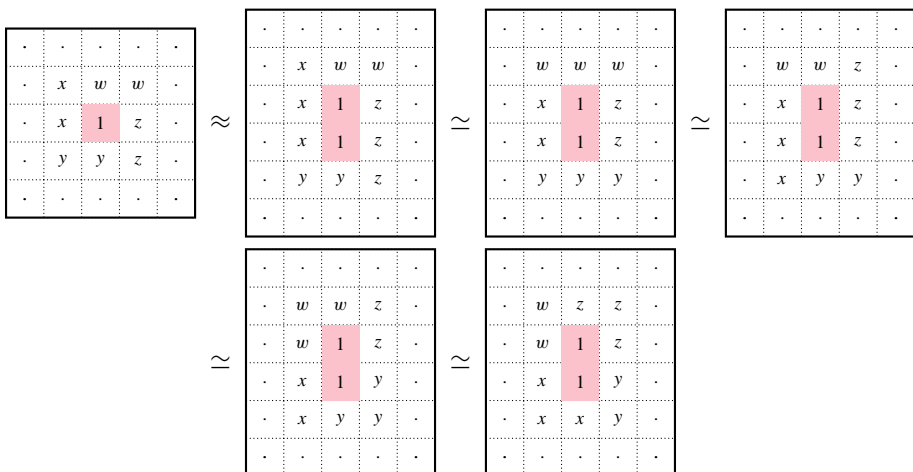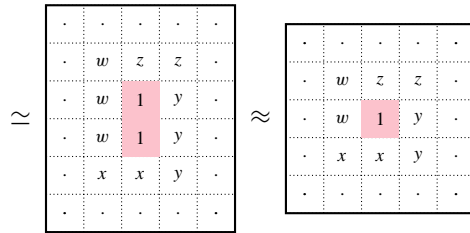
By repeatedly applying this rotation, we see that all four maps of (2) are extension-homotopic. Since f is homotopic to one of these four maps, and the first one is T, we have $f \approx T$.

Let $T^{-1}: (I_{4,4}, \partial I_{4,4}) \to (S^2, -\mathbf{e}_1)$ be the inverse of T, as we defined inverses in the proof of Theorem 4.3. In diagrammatic terms, this is given as follows:


The following result corresponds to Lemma 6.2 for maps with $d(f) = -1$ proved using the same arguments. We omit the details of its proof.

Lemma 6.3 *Let $f : (I_{4,4}, \partial I_{4,4}) \to (S^2, -\mathbf{e}_1)$ be continuous with $f(x) = \mathbf{e}_1$ for exactly one $x \in I_{4,4}$ and $d(f) = -1$. Then $f \approx T^{-1}$.* □

We also consider such maps having $d(f) = 0$.

Lemma 6.4 *Let $f : (I_{4,4}, \partial I_{4,4}) \to (S^2, -\mathbf{e}_1)$ be continuous with $f(x) = \mathbf{e}_1$ for exactly one $x \in I_{4,4}$ and $d(f) = 0$. Then, f is homotopic to a constant map.*

Proof By Lemma 6.1 we may assume that f takes the form:
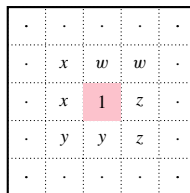
where the values x, y, z, w are taken from this set of 4 possible values: $\{\pm\mathbf{e}_2, \pm\mathbf{e}_3\}$. By the second part of Lemma 6.1, if $\{x, y, z, w\} \neq \{\pm\mathbf{e}_2, \pm\mathbf{e}_3\}$ then f is homotopic to a constant map and we are done. So it remains only to consider the case when $\{x, y, z, w\} = \{\pm\mathbf{e}_2, \pm\mathbf{e}_3\}$, and we will show that this case leads to a contradiction.

The proof of Lemma 6.2 shows how the values x, y, z, w can be "rotated" by extension homotopy. Thus we may assume without loss of generality that $x = \mathbf{e}_2$.

Since x appears adjacent to both y and w, neither y or w can be $-\mathbf{e}_2$, and so we must have $z = -\mathbf{e}_2$. Thus, either $y = \mathbf{e}_3$ and $w = -\mathbf{e}_3$, or $y = -\mathbf{e}_3$ and $w = \mathbf{e}_3$.

If $y = \mathbf{e}_3$ and $w = -\mathbf{e}_3$, then $f = T$ which contradicts our assumption that $d(f) = 0$, since $d(T) = 1$. In the other case, we have $w = -\mathbf{e}_3$ and $y = \mathbf{e}_3$, and so f is a rotation of T^{-1}, which again contradicts our assumption that $d(f) = 0$.

We summarize the three lemmas above as follows:

Theorem 6.5 *Let* $f : (I_{4,4}, \partial I_{4,4}) \to (S^2, -\mathbf{e}_1)$ *be continuous with* $f(x) = \mathbf{e}_1$ *for exactly one* $x \in I_{4,4}$, *and let:*

$$
T = \begin{array}{|c|c|c|c|c|}
\hline
\cdot & \cdot & \cdot & \cdot & \cdot \\
\hline
\cdot & 2 & -3 & -3 & \cdot \\
\hline
\cdot & 2 & 1 & -2 & \cdot \\
\hline
\cdot & 3 & 3 & -2 & \cdot \\
\hline
\cdot & \cdot & \cdot & \cdot & \cdot \\
\hline
\end{array} \,, \qquad
T^{-1} = \begin{array}{|c|c|c|c|c|}
\hline
\cdot & \cdot & \cdot & \cdot & \cdot \\
\hline
\cdot & -3 & -3 & 2 & \cdot \\
\hline
\cdot & -2 & 1 & 2 & \cdot \\
\hline
\cdot & -2 & 3 & 3 & \cdot \\
\hline
\cdot & \cdot & \cdot & \cdot & \cdot \\
\hline
\end{array}
$$

If $d(f) = 1$, *then* $f \approx T$. *If* $d(f) = -1$, *then* $f \approx T^{-1}$. *If* $d(f) = 0$, *then* f *is homotopic to a constant map.*

Going forwards, we will refer to the 3×3 blocks of values with a value of $\mathbf{e}_1$ in the center as *islands*. We will eventually see that any map may be reduced, up to extension homotopy, to one whose values are represented by a number of these islands isolated from one another in a surrounding "sea" of values of $-\mathbf{e}_1$.

Next, we introduce two more ingredients used to prove injectivity of d: a "flooding" homotopy; and a pre-processing step that involves subdivision.

Given any map $f : (I_{m,n}, \partial I_{m,n}) \to (S^2, -\mathbf{e}_1)$ and some particular value $b \in S^2$, we define a map $f_b : (I_{m,n}, \partial I_{m,n}) \to (S^2, -\mathbf{e}_1)$ which we call *the flood of f with b* or *the b-flood of f*, and which is a map that agrees in values with f except that its values away from the boundary have been changed whenever possible to equal b. Specifically, we define:

$$
f_b(x) = \begin{cases} b & \text{if } x \notin \partial I_{m,n} \text{ and if } f(z) \neq -b \text{ for all } z \sim x, \\ f(x) & \text{otherwise.} \end{cases}
$$

Figure 12 shows an example of a map f with its flood by $-\mathbf{e}_1$.

Lemma 6.6 *For* $f : (I_{m,n}, \partial I_{m,n}) \to (S^2, -\mathbf{e}_1)$ *and* $b \in S^2$, *the flood of f with b*, $f_b : (I_{m,n}, \partial I_{m,n}) \to (S^2, -\mathbf{e}_1)$ *is continuous and homotopic to f by a homotopy relative to the boundary.*

Proof First we show that f_b is continuous. Take $x \sim y \in I_{m,n}$, and we will show that $f_b(x) \sim f_b(y)$. If $f(x) = -b$ or $f(y) = -b$, then by the definition of f_b we will have $f_b(x) = f(x)$ and $f_b(y) = f(y)$ and so $f_b(x) \sim f_b(y)$ as desired. In the case where neither of $f(x)$ and $f(y)$ are $-b$, then $f_b(x) \in \{f(x), b\}$ and $f_b(y) \in \{f(y), b\}$. Since neither of $f(x)$ or $f(y)$ is $-b$, we have $f(x) \sim b$ and $f(y) \sim b$, and $f_b(x) \sim f_b(y)$ as desired.
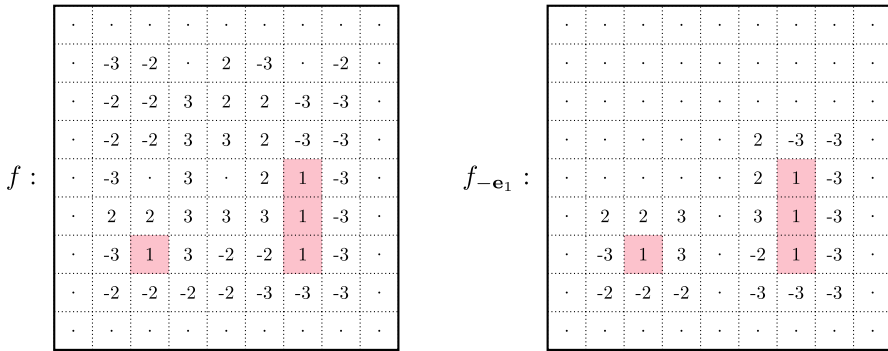
Fig. 12 An example of a map $f : I_{8,8} \to S^2$ and the map $f_{-\mathbf{e}_1}$, the $-\mathbf{e}_1$-flood of f

Now we show that f_b is homotopic to f in a single step. Take $x \sim y \in I_{m,n}$. Since each of f and f_b are continuous, it is sufficient to show that $f_b(x) \sim f(y)$, by Lemma 2.4. We investigate the various cases appearing in the definition of $f_b(x)$.

If $x \in \partial(I_{m,n})$ then $f_b(x) = -\mathbf{e}_1$, and also $f(x) = -\mathbf{e}_1$ and so $f_b(x) = f(x) \sim f(y)$ as desired.

If $f(z) \neq -b$ for all $z \sim x$, then in particular $f(y) \neq -b$, and $f_b(x) = b$. Since $f_b(x) = b$ and $f(y) \neq -b$, we have $f_b(x) \sim f(y)$ as desired.

The final case is when x is outside the boundary and $f(z) = -b$ for some $z \sim x$. In this case $f_b(x) = f(x) \sim f(y)$ as in the first case.

Next, we show how any $f : (I_{m,n}, \partial I_{m,n}) \to (S^2, -\mathbf{e}_1)$ can be changed up to extension homotopy into a map in which the value e_1 occurs only at isolated points.

Lemma 6.7 *Let $f : (I_{m,n}, \partial I_{m,n}) \to (S^2, -\mathbf{e}_1)$. Then there is a map $g \approx f$ having the property that no two adjacent points of the domain both map to $\mathbf{e}_1$. Furthermore, we may construct g so that any two points x, y in the domain of g with $g(x) = g(y) = \mathbf{e}_1$ will be arbitrarily far apart.*

Proof Choose some $k \geq 5$. We construct g in three steps, beginning with $\hat{f} = f \circ \rho_k :$ $S(I_{m,n}, k) \to S^2$. This is the k-fold subdivision of f; we view it as a continuous function $\hat{f} : (I_{km+k-1,kn+k-1}, \partial I_{km+k-1,kn+k-1}) \to (S^2, -\mathbf{e}_1)$. By Theorem 3.7 we have $\hat{f} \approx f$.

For the second step we modify certain values of $\hat{f}$ in specific ways based on the specific locations of values with $\hat{f}(x) = \mathbf{e}_1$. Recalling that all values of $\hat{f}$ occur in constant $k \times k$ blocks, we make two adjustments described in Figs. 13 and 14. Making these two adjustments produces a new map $\hat{f}'$.

The first adjustment will change values of $\hat{f}$ in a small region surrounding two diagonally adjacent points $x \sim y$ with value $\hat{f}(x) = \hat{f}(y) = \mathbf{e}_1$, in which x and y are mutually adjacent to a pair of diagonal points with value different from $\mathbf{e}_1$. We adjust the values of $\hat{f}$ by making all points adjacent to either x or y have value $\mathbf{e}_1$, as in Fig. 13.

The second adjustment will change values of $\hat{f}$ in a small region near two orthogonally adjacent points $x \sim y$ with value $\hat{f}(x) = \hat{f}(y) = \mathbf{e}_1$, in which x and y arise
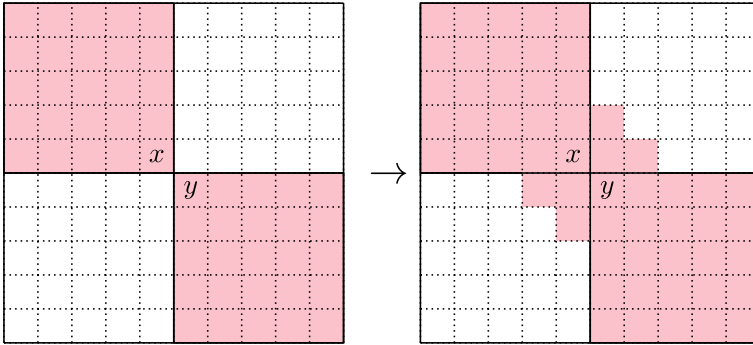
Fig. 13 Adjustment #1, in which all values adjacent to x and y become $\mathbf{e}_1$. In the picture, the shaded squares are pixels with value $\mathbf{e}_1$ under $\hat{f}$, and the white squares have value different from $\mathbf{e}_1$ (but not equal to $-\mathbf{e}_1$). In the picture, we use $k = 5$
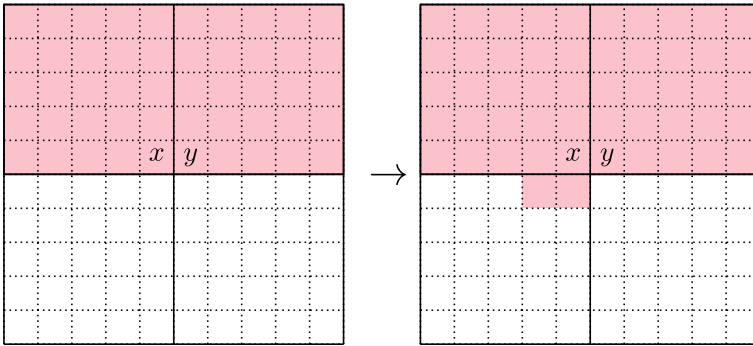

Fig. 14 Adjustment #2, in which 2 values adjacent to x become $\mathbf{e}_1$. In the picture, the shaded squares are pixels with value $\mathbf{e}_1$ under $\hat{f}$, and the white squares have value different from $\mathbf{e}_1$ (but not equal to $-\mathbf{e}_1$). In the picture, we use $k = 5$

from different pixels of $I_{m,n}$ before the subdivision, and each is adjacent to points with value different from $\mathbf{e}_1$. We adjust the values of $\hat{f}$ to make 2 other points adjacent to x have value $\mathbf{e}_1$, as in Fig. 14.

Applying these two adjustments in all applicable locations produces a map $\hat{f}' : (I_{km+k-1,kn+k-1}, \partial I_{km,kn}) \to (S^2, -\mathbf{e}_1)$. Since each of these adjustments may evidently be effected by a (short) sequence of spider moves, we have $\hat{f}' \simeq \hat{f}$.

The third step of our construction is to consider $g = \hat{f}'_{\mathbf{e}_2,\mathbf{e}_3,-\mathbf{e}_1}$. This is the iterated flood of $\hat{f}'$ with $\mathbf{e}_2$ followed by $\mathbf{e}_3$ followed by $-\mathbf{e}_1$. By Lemma 6.6 we will have $\hat{f}' \approx \hat{f}'_{\mathbf{e}_2,\mathbf{e}_3,-\mathbf{e}_1}$, and thus $g \approx f$ as desired. It remains to show that the values x where $g(x) = \mathbf{e}_1$ are separated from one another.

The set S of points x with $\hat{f}'(x) = \mathbf{e}_1$ consists of a (possibly disconnected) union of $k \times k$ blocks of points, together with some extra points added by the two adjustments. This set looks something like Fig. 15, in which $k = 5$.

This set S includes two types of special points: the "exterior corner points," lying on corners of the $k \times k$ blocks which are not adjacent to other $k \times k$ blocks in S, and
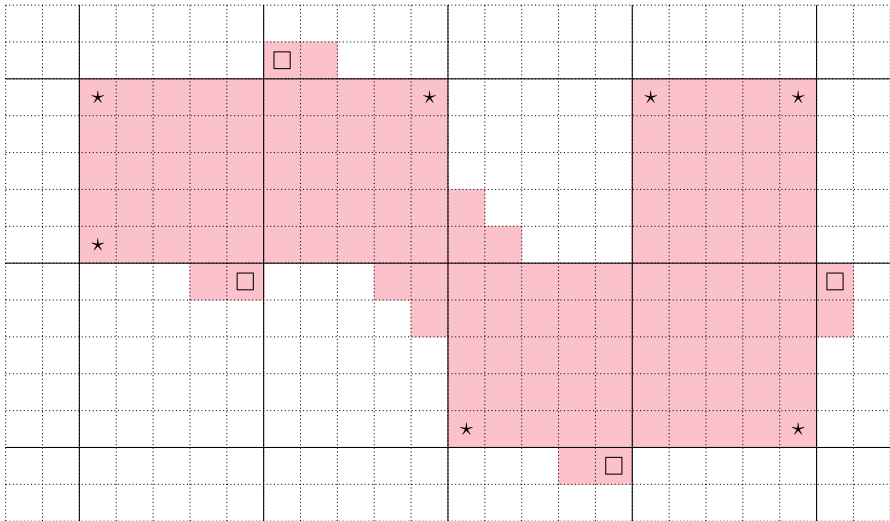
Fig. 15 Possible arrangement of points with $\hat{f}'(x) = \mathbf{e}_1$. Points marked with □ are "adjusted corner" points, which may be adjacent to other points with two different labels. Points marked with ⋆ are "exterior corner" points, which may be adjacent to other points with 3 different labels. All other points are adjacent to 0 or 1 labels other than $\mathbf{e}_1$. In the picture we use $k = 5$

the "adjusted corner points", the points created by Adjustment #2 which are adjacent to the neighboring $k \times k$ block. We label these points with ⋆ and □, respectively, in Fig. 15. We will show that these exterior and adjusted corner points are the only points which can be labeled $\mathbf{e}_1$ by g. To do this, we show that all other points $x \in S$ have $g(x) \neq \mathbf{e}_1$. Since the exterior and adjusted corner points are never adjacent, and will be separated from each other by at least $k - 2$ points in a horizontal or vertical direction, this will complete the proof.

Let $x \in S$ be some point which is not an exterior or adjusted corner of S. We consider cases according to the number of different values which can occur for points $\hat{f}'(y)$ with $y \sim x$. By our construction, since x is not an exterior or adjusted corner, it will be adjacent to at most 1 other label different from $\mathbf{e}_1$.

If all points $y \sim x$ have the same value $\hat{f}'(y) = \mathbf{e}_1$, then $\hat{f}'_{\mathbf{e}_2}(x) = \mathbf{e}_2$ by the definition of the flood. Since no subsequent flood by $\mathbf{e}_3$ or $-\mathbf{e}_1$ can cause the label of x to become $\mathbf{e}_1$, we see that $g(x) = \hat{f}'_{\mathbf{e}_2,\mathbf{e}_3,-\mathbf{e}_1}(x) \neq \mathbf{e}_1$ as desired.

Our second case is when x is adjacent to only 1 label other than $\mathbf{e}_1$. That is, there is some $a \in S^2$ such that all points $y \sim x$ have either the value $\hat{f}'(y) = \mathbf{e}_1$ or $\hat{f}'(y) = a$. If $a \neq -\mathbf{e}_2$, then $\hat{f}'_{\mathbf{e}_2}(x) = \mathbf{e}_2$ and so as above we will have $g(x) \neq \mathbf{e}_1$ as desired. If $a = -\mathbf{e}_2$, then the flood by $\mathbf{e}_2$ will not change the label on x, so $\hat{f}'_{\mathbf{e}_2}(x) = \mathbf{e}_1$. But the subsequent flood by $\mathbf{e}_3$ will change the value to $\mathbf{e}_3$, giving $\hat{f}'_{\mathbf{e}_2,\mathbf{e}_3}(x) = \mathbf{e}_3$. Again as above this means that $g(x) \neq \mathbf{e}_1$ as desired.

Finally, we are ready to complete the proof that $d : \pi_2(S^2, -\mathbf{e}_1) \to \mathbb{Z}$ is an isomorphism.

Proof of Theorem 5.5 In Proposition 5.4 and Lemma 5.7, we have already shown that $d : \pi_2(S^2, -\mathbf{e}_1) \to \mathbb{Z}$ is a surjective homomorphism. Here, we show that $\ker d$ is trivial. Let $f : (I_{m,n}, \partial I_{m,n}) \to (S^2, -\mathbf{e}_1)$ with $d(f) = 0$; we will show that $f \in \pi_2(S^2, -\mathbf{e}_1)$ is the trivial element.

By Lemma 6.7, we may assume—up to extension homotopy—that f has only isolated values of $\mathbf{e}_1$ occurring at the center of 3×3 blocks of points, and outside of these 3×3 "islands," f is constant with value $-\mathbf{e}_1$. (The last flood by $-\mathbf{e}_1$ in the proof of Lemma 6.7 will achieve the latter.) Furthermore we may assume that these islands are separated from each other by any distance we wish.

Using Lemma 3.6 to apply Theorem 6.5 to each island locally, we may replace each island by the maps T, or T^{-1} according to their triangle-count, or replace the island entirely by constant values $-\mathbf{e}_1$ if the triangle count for that island is zero.

Since these islands are separated by arbitrarily large regions of constant values of $-\mathbf{e}_1$, by Remark 3.5 we may translate the remaining islands into any configuration we wish. Since $d(f) = 0$, the number of islands of type T must equal the number of islands of type T^{-1}. Say that there are k islands of each type. Then we may arrange them all to be stacked diagonally as in Fig. 5 (extending the domain if necessary), so that $f \approx kT \cdot kT^{-1}$, and this is the trivial element because T and T^{-1} are inverses.

7 Future work

An obvious direction in which to continue is to define, for each n, a homotopy group $\pi_n(X, x_0)$ for X a digital image. We believe this should be a straightforward generalization of the approach taken here, with the group consisting of (suitably defined) extension-homotopy equivalence classes of maps $(I, \partial I) \to (X, x_0)$ with $I = I_{m_1} \times \cdots \times I_{m_n}$ an n-fold product of intervals. We believe that most, if not all, of the results through Sect. 4 should have direct generalizations. One main issue in proceeding with this is simply an expositional one, dealing with the increasingly burdensome notational complexity. It also seems reasonable to extend and generalize the development here to include suitable relative homotopy groups and, if possible, develop the long exact sequence in homotopy groups in this digital context.

In [16] it was shown that the fundamental group (as defined there) of a 2D digital image is a free group. A reasonable question to ask here is: must $\pi_2(X, x_0)$ for X a 3D digital image be a free abelian group? On a related note, we may ask about torsion in $\pi_2(X, x_0)$. For instance, what is an example of a digital image X that has $\pi_2(X, x_0)$ with non-trivial torsion subgroup? Questions of this sort about $\pi_1(X, x_0)$ are effectively resolved in [16] by the identification of the digital fundamental group with the ordinary fundamental group of the spatial realization of the clique complex of the digital image considered as a graph. Here, then, we can ask whether $\pi_2(X, x_0)$ may be identified with the ordinary second homotopy group of the spatial realization of the clique complex of X?

In a different direction, the calculation of $\pi_2(S^2, -\mathbf{e}_1)$ given here suggests many points of contact with classical topology. Although our maps $I_{m,n} \to S^2$ are really just graph homomorphisms, the line of argument is strongly suggestive of topological

ingredients such as the degree of a map, the classical homotopy group(s) of a topological space, triangulations, simplicial complexes, polyhedra and so-on. Is it possible to somehow make these connections more precise?

References

1. Babson, E., Barcelo, H., de Longueville, M., Laubenbacher, R.: Homotopy theory of graphs. J. Algebraic Combin. **24**(1), 31–44 (2006)
2. Barcelo, H., Kramer, X., Laubenbacher, R., Weaver, C.: Foundations of a connectivity theory for simplicial complexes. Adv. Appl. Math. **26**(2), 97–128 (2001)
3. Barcelo, H., Laubenbacher, R.: Perspectives on *A*-homotopy theory and its applications. Discrete Math. **298**(1–3), 39–61 (2005)
4. Boxer, L.: A classical construction for the digital fundamental group. J. Math. Imaging Vision **10**(1), 51–62 (1999)
5. Chih, T., Scull, L.: A homotopy category for graphs. J. Algebraic Combin. **53**(4), 1231–1251 (2021)
6. Dochtermann, A.: Hom complexes and homotopy in the category of graphs. The International Conference on Topological and Geometric Graph Theory, Electron. Notes Discrete Math., vol. 31, Elsevier Sci. B.V., Amsterdam, 2008, pp. 131–136
7. Dochtermann, A.: Homotopy groups of Hom complexes of graphs. J. Combin. Theory Ser. A **116**(1), 180–194 (2009)
8. Escribano, C., Giraldo, A., Sastre, M.A.: Digitally continuous multivalued functions, morphological operations and thinning algorithms. J. Math. Imaging Vision **42**(1), 76–91 (2012)
9. Evako, A.V.: Topological properties of closed digital spaces: One method of constructing digital models of closed continuous surfaces by using covers. Comput. Vision Image Underst. **102**, 134–144 (2006)
10. Evako, A.V.: Classification of digital *n*-manifolds. Discrete Appl. Math. **181**, 289–296 (2015)
11. Kong, T.Y.: A digital fundamental group. Comput. Graph. **13**, 159–166 (1989)
12. Kong, T.Y., Rosenfeld, A.: Digital topology: a comparison of the graph-based and topological approaches, topology and category theory in computer science. Oxford Sci. Publ., Oxford Univ. Press, Oxford, NY, 1991, pp. 273–289
13. Lee, D.-W., Staecker, P.C.: Digital topological groups. Topology Appl. **338**(20), 108644 (2023)
14. Lupton, G., Oprea, J., Scoville, N.: A fundamental group for digital images. J. Appl. Comput. Topol. **5**(2), 249–311 (2021)
15. Lupton, G., Oprea, J., Scoville, N.: Homotopy theory in digital topology. Discrete Comput. Geom. **67**(1), 112–165 (2022)
16. Lupton, G., Scoville, N.: Digital fundamental groups and edge groups of clique complexes. J. Appl. Comput. Topol. **6**(4), 529–558 (2022)
17. Meric, E.T., Vergili, T., Karaca, I.: Computing higher dimensional digital homotopy groups. Appl. Math. Inf. Sci. **8**(5), 2417–2425 (2014)
18. Rosenfeld, A.: 'Continuous' functions on digital pictures. Pattern Recognition Letters **4**, 177–184 (1986)
19. Staecker, P.C.: Digital homotopy relations and digital homology theories. Appl. Gen. Topol. **22**(2), 223–250 (2021)

20. Vergili, T., Karaca, I.: Some properties of higher dimensional homotopy groups for digital images. Appl. Math. Inf. Sci. **9**(6), 3027–3035 (2015)

Authors and Affiliations

Gregory Lupton[1] [iD] **· Oleg Musin[2] · Nicholas A. Scoville[3] · P. Christopher Staecker[4] · Jonathan Treviño-Marroquín[5]**

✉ Gregory Lupton
g.lupton@csuohio.edu

Oleg Musin
oleg.musin@utrgv.edu

Nicholas A. Scoville
nscoville@ursinus.edu

P. Christopher Staecker
cstaecker@fairfield.edu

Jonathan Treviño-Marroquín
jonathan.trevino@cimat.mx

[1] Department of Mathematics and Statistics, Cleveland State University, Cleveland, OH 44115, USA

[2] School of Mathematical and Statistical Sciences, The University of Texas Rio Grande Valley, Edinburg, TX 78539, USA

[3] Department of Mathematics, Computer Science, and Statistics, Ursinus College, Collegeville, PA 19426, USA

[4] Department of Mathematics, Fairfield University, Fairfield, CT 06824, USA

[5] Centro de Investigación en Matemáticas, A.C., 36023 Guanajuato, GTO, Mexico