HABILITATION THESIS

Martin Tancer

On the interplay of combinatorics, geometry, topology and computational complexity

Department of Applied Mathematics

Prague 2016
# Contents

**Preface**

1 On the interplay of combinatorics, geometry, topology and computational complexity  
1.1 Introduction ................................................. 4  
1.2 Embeddability of simplicial complexes  
   1.2.1 Algorithmic aspects ................................... 6  
   1.2.2 Combinatorial aspects  
1.3 Collapsibility and shellability  
   1.3.1 Collapsibility ............................................. 10  
   1.3.2 Shellability ............................................... 11  
1.4 Curves and Graphs on surfaces  
   1.4.1 Untangling curves on surfaces  
   1.4.2 Shortest paths on surfaces  
   1.4.3 Hanani-Tutte theorem on the projective plane  
1.5 Homology growth of flag complexes  

**Bibliography**

**Reprints of papers**

- On Betti numbers of flag complexes with forbidden induced subgraphs .............................................. 20  
- A direct proof of the strong Hanani-Tutte theorem on the projective plane ............................................. 74  
- On generalized Heawood inequalities for manifolds: a van Kampen–Flores-type nonembeddability result .......... 109  
- Bounding Helly numbers via Betti numbers ............................................................. 124  
- Shortest path embeddings of graphs on surfaces ............................................................. 153  
- Embeddability in the 3-sphere is decidable .................................................................................. 174  
- Untangling two systems of noncrossing curves ........................................................................ 228  
- Shellability of the higher pinched Veronese posets ............................................................. 271  
- Recognition of collapsible complexes is NP-complete .................................................................. 303
This habilitation thesis consists of reprints of nine papers coauthored by Martin Tancer, together with an introductory commentary. The common topic of the papers is that they focus on the interplay among combinatorics, combinatorial geometry, topology and computational complexity. However, the motivation for the results in these papers also partially comes from other areas of mathematics such as algebra or Riemannian geometry.

The main aim of the introductory commentary is to quickly explain main results of each of the individual papers and to provide the links among the papers. We remark that many details regarding the motivation or the ideas for the proofs are left out from the introductory commentary and we refer to the introductions of each of the individual papers. The aim of the introductory commentary is not to repeat these introductions (although some bit of repetition is perhaps unavoidable in order to explain the results).

The nine papers forming the main body of the thesis are listed at the end of preface. Namely, papers [3, 4, 6] relate to the combinatorial and algorithmic properties of embeddability of simplicial complexes (and its applications) and their contents is explained in Section 1.2. Papers [8, 9] focus on collapsibility and shellability, two combinatorial ways how to simplify a topological space (given as a simplicial complex or a poset). Their contents is explained in Section 1.3. Papers [2, 5, 7] study combinatorial and topological properties of graphs and curves drawn on surfaces and they are explained in Section 1.4. Finally, [1] is a result on the growth of homology of certain complexes associated to graphs and it is explained in Section 1.5.


1These papers are in general very recent. For papers that do not currently have a publicly available final journal version yet, we use a preprint from arXiv.


On the interplay of combinatorics, geometry, topology and computational complexity

1.1 Introduction

The fundamental task of combinatorics is to study properties of discrete objects such as their enumeration, extremal properties, interactions or structural properties. Although it is often possible to solve combinatorial problems by intrinsic combinatorial means, in general, combinatorics strongly benefits from interactions with other areas of mathematics (and vice versa).

Combinatorics is strongly linked to the theoretical computer science. Understanding efficient algorithms for recognition of combinatorial objects with certain property or for enumeration of objects is an integral part of combinatorics. For instance, the Kuratowski’s planarity [Kur30] criterion is as central result in graph theory as the existence of a linear time algorithm for recognition of planar graphs by Hopcroft and Tarjan [HT74].

A rich mathematical world appears when we further combine these two subjects with questions in combinatorial geometry and topology. From the point of view of the way how these subjects interact, we may distinguish several areas. They include algorithmic topology, combinatorial topology and graph drawings (on surfaces).

**Algorithmic topology.** The task of algorithmic topology is to design efficient algorithms for topological problems. This usually comes together with some combinatorial model for the topological spaces, maps, etc. in the question so that we may have a finite input for the algorithmic question.

A prominent example in this line of research is the unknot recognition problem which is currently known to belong to \( \text{NP} \cap \text{co-NP} \) [HLP99, Lac16]; but no polynomial time algorithm is known for this problem.

To this area of research we contribute with an algorithm for the 3-embeddability problem [6] (see Section 1.2.1) and with a result on NP-hardness of recognition of collapsible complexes [9] (see Section 1.3.1).

**Combinatorial topology.** By combinatorial topology we mean the area studying the direct interactions of combinatorial and topological objects. This

---

\(^1\)This list is neither complete nor pairwise disjoint. As other areas we could name, for example, topological methods in combinatorics or study of metric embeddings. However, we focus only on the areas covered in this thesis. Regarding disjointness, for example [7] partially belongs to all three mentioned areas.
is usually done via properties of simplicial complexes which can be seen both as purely combinatorial objects (hereditary set systems) as well as topological objects (triangulations of topological spaces).

Such interactions include, for example, combinatorial questions on embeddings of simplicial complexes into $\mathbb{R}^d$ or other topological spaces initiated by van Kampen and Flores [vK32, Flo34], study of clique complexes of graphs with forbidden certain subgraphs as a discrete analogue of bounded curvature [JS06] or study of Morse spectra of triangulated spaces as in [ABL14], for example.

To this area, we contribute by results on the homology growth of clique complexes of graphs with forbidden certain subgraphs [1] (see Section 1.5) and a significant progress towards Kühnel’s conjecture on embeddings of skeleta of simplices into manifolds [3] (see Section 1.2.2). We also provide a very general topological Helly-type theorem [3] (see Section 1.2.2), which can be seen as a result partially belonging to this area and partially to ‘topological methods’ in combinatorics. Similarly, the result [8] on shellability of higher pinched Veronese posets (see Section 1.3.2) can be seen as result partially belonging to topological combinatorics and partially belonging to combinatorial commutative algebra.

**Drawings of graphs on surfaces.** It could be easily argued that studying various aspects of drawings of graphs on surfaces is just a part of combinatorial topology described above. However, the lower-dimensional nature of drawing of graphs on surfaces causes that there are very different interesting questions in this area and it brings the area even closer to combinatorics. That is why we consider this area separately.

Classical questions in this area include to determine which graphs can be drawn on which surface without crossings or what are the other combinatorial properties of graphs drawn on surfaces (such as the chromatic number); see [MT01]. Regarding drawings where we allow crossings, it is very interesting to study various aspects of the crossing number of such graphs [Sch13a].

To this area, we contribute with an alternative proof of the strong Hanani-Tutte theorem on the projective plane [2] (see Section 1.4.3) and with results on drawings of graphs on surfaces with shortest paths [5] (see Section 1.4.2).

Finally, contribution [7] interacts with all three areas. Here we show that two systems of pairwise disjoint curves can be untangled with a self-homeomorphism of a surface applied to one of the systems so that there are not too many intersections among the two systems (see Section 1.4.1). Systems of curves are closely related to drawings of graphs. In addition a part of the main result in [7] serves as a verification of the algorithm in [6].

**On preliminaries.** We assume that the reader is familiar with basic notions from combinatorics, computational complexity and topology. In particular, we frequently use basic properties of simplicial complexes in the text below and we also use the basics of the homology theory. For further reading, we refer, for example, to [Hat01, Mat03].

**Organization of the remaining sections.** In the forthcoming sections, we briefly explain the contents of the individual papers that are part of this thesis. We group the papers together according to the similarity of the topics, as sketched in Preface. It turns out that this is not the same as what we described in this introductory section according to the way how the fields interact.
1.2 Embeddability of simplicial complexes

Let $X$ and $Y$ be two topological spaces, does $X$ embed into $Y$? This is a classical important general question in topology. We cannot expect that there would be a simple criterion that would answer this question. Indeed, it includes, for example, the homeomorphism problem for manifolds which is known to be algorithmically undecidable. Nevertheless there are important classes of spaces $X$ and $Y$ for which the question can be either fully answered or there are important sufficient and/or necessary conditions to be understood.

We will mostly focus on the case where $X$ and $Y$ are topological spaces triangulated as finite simplicial complexes. In this setting, it is possible to represent $X$ and $Y$ in computer and thus we may ask algorithmic questions on embeddability.3 From theoretical point of view, the structure of simplicial complex allows linking topological and combinatorial questions on embeddings which has fruitful consequences as we will argue in 1.2.2.

1.2.1 Algorithmic aspects

From algorithmic point of view one of the most natural settings is the following algorithmic question $\text{Embed}_{k \to d}$, which depends on two positive integers $k$ and $d$, $k \leq d$: Given a simplicial complex $K$ of dimension at most $k$, does $K$ (piecewise linearly) embed into $\mathbb{R}^d$?

The question $\text{Embed}_{k \to d}$ was introduced by Matoušek, the author and Wagner in [MTW11] and based on this paper it was one of the central topics of the author’s PhD thesis. It was previously known that the cases $\text{Embed}_{1 \to 2}$ (graph planarity [HT74]) and $\text{Embed}_{2 \to 2}$ [GR79] are solvable in linear time and that for every $k \geq 3$ fixed, $\text{Embed}_{k \to 2k}$ can be decided in polynomial time (this is based on the work of Van Kampen, Wu, and Shapiro; see [MTW11] for a detailed explanation).

For dimension $d \geq 4$, the currently known understanding of the computational complexity of $\text{Embed}_{k \to d}$ is the following: for all $k$ with $(2d - 2)/3 \leq k \leq d$ it is NP-hard (and even undecidable if $k \geq d - 1 \geq 4$) [MTW11], while for $k < (2d - 2)/3$ it is polynomial-time solvable, assuming $d$ fixed, as was shown in a series of papers on computational homotopy theory [ČKM+14a, ČKM+14b, KMS13, ČKV13].

### Dimension 3.

The new contribution (when compared with [MTW11]), presented as a part of this thesis, is the joint work with J. Matoušek, E. Sedgwick and U. Wagner [6] where we show the following.

**Theorem 1** (Thm 1.1 & Cor. 1.2 in [6]). The problems $\text{Embed}_{2 \to 3}$ and $\text{Embed}_{3 \to 3}$ are algorithmically decidable.

Here we only very briefly sketch the main steps; for a more detailed overview of the idea of the proof we refer to Sections 1, 2 and 3 of [6].

In fact, it is sufficient to show algorithmic decidability of $\text{Embed}_{2 \to 3}$; solution for $\text{Embed}_{3 \to 3}$ then follows by a combinatorial reduction. The first step is to

However, this is not the only possible representation. For example, a representation as a simplicial set might be a more efficient representation of the same space.
show that it is actually sufficient to establish the following variant of the problem for 3-manifolds.

**Theorem 2** (Thm. 1.3 in [6]). *There is an algorithm which decides whether a given triangulated 3-manifold \(X\) with boundary embeds into the 3-sphere \(S^3\).*

Indeed, given a 2-complex \(K\) (that is, an instance of \(\text{EMBED}_{2\to3}\)), we can test all possible *thickenings* of \(K\) to a 3-manifold with boundary (up to a homeomorphism). That is, to a manifold which contains \(K\) and collapses to \(K\). Then \(K\) embeds into \(\mathbb{R}^3\) if and only if at least one thickening of \(K\) embeds into \(S^3\). By an algorithm of Neuwirth, it is possible to generate all possible thickenings [Neu68] (see also [Sko95]).

The bulk of our contribution is to prove the following result.

**Theorem 3** (Thm. 1.4 in [6]). *Let \(X\) be an irreducible 3-manifold, neither a ball nor an \(S^3\), with incompressible boundary and with a 0-efficient triangulation \(T\). If \(X\) embeds in \(S^3\), then there is also an embedding for which \(X\) has a short meridian \(\gamma\), i.e., an essential normal curve \(\gamma \subset \partial X\) bounding a disk in \(S^3 \setminus X\) such that the length of \(\gamma\), measured as the number of intersections of \(\gamma\) with the edges of \(T\), is bounded by a computable function of the number of tetrahedra in \(T\).*

Here an *irreducible manifold* is such that every embedded 2-sphere in \(X\) bounds a ball in \(X\); it has incompressible boundary if any curve in \(\partial X\) that bounds a disc in \(X\) also bounds a disc in \(\partial X\). A 0-efficient triangulation is a technical term that we do not define here (and we refer to the reprint of [6]). A *normal curve* on \(\partial X\) is a closed curve which avoids vertices of the triangulation; it crosses each edge transversally; and it meets each triangle in a finite number of arcs with endpoints on different edges of the triangle. Finally, the *length* of a normal curve is the number of edges of the triangulation it crosses.

Theorem 2 allows to prove Theorem 3 recursively. After standard transformations, we may assume that \(X\) satisfies the assumptions of Theorem 3. Then we may enumerate all normal curves up to the length provided by Theorem 3 as candidate meridians, fill them with a thickened disc and recurse.

For the proof of Theorem 3 we already refer to [6].

### 1.2.2 Combinatorial aspects

From a combinatorial point of view, we plan to present two results regarding embeddability of simplicial complexes.

**Almost embeddings.** First, we need to introduce a certain important notion. Given a simplicial complex \(K\) and a topological space \(Y\), an *almost embedding* is a map \(f: |K| \to Y\) such that \(f(|\sigma|) \cap f(|\tau|) = \emptyset\) whenever \(\sigma\) and \(\tau\) are disjoint simplices of \(K\). Every embedding is an almost embedding but the converse is not true.

The classical results of van Kampen and Flores [vK32, Flo34] state that the following \(k\)-dimensional complexes do not embed into \(\mathbb{R}^{2k}\):

---

3Meaning that \(\gamma\) does not bound a disk in \(\partial X\).

4Unless stated otherwise, we work with abstract simplicial complexes. For a face (simplex) \(\sigma \in K\), the symbol \(|\sigma|\) denotes the geometric simplex corresponding to \(\sigma\) in some fixed geometric realization of \(K\). Finally, \(|K|\) denotes the *underlying space* of \(K\), that is \(\bigcup_{\sigma \in K} |\sigma|\).

---
• $\Delta_{2k+2}^{(k)}$, that is, the $k$-skeleton of the $(2k+2)$-simplex, and
• $D_3^{(k+1)}$, that is, the $(k+1)$-tuple join of the three-element discrete set.

However, the standard proofs provide a stronger conclusion: these complexes do not even almost embed into $\mathbb{R}^{2k}$. In general, almost embeddings are useful tools for understanding embeddings as they are often easier to handle.

(Almost) embeddings on the level of chain maps. For further applications it turned out that it is important to study the (almost) embeddings on the level of chain maps (in $\mathbb{Z}_2$-homology). In a joint work with X. Goaoc, P. Paták, Z. Patáková and U. Wagner [4] we have developed an inductive Ramsey-based approach how to build a certain combinatorially well behaved chain map $C_*(K) \rightarrow C_*(\mathbb{R}^d)$ where $K$ is a simplicial complex. As an application of this approach, we have obtained a Helly-type theorem with very weak topological assumptions (see Theorem 4 below). Subsequently, in a joint work with the same group of coauthors and in addition with I. Mabillard [3], we have utilized a modification of this technique to a different problem regarding embeddability of simplicial complexes into manifolds. (In fact, a simplification of this technique is sufficient in [3] which allows to remove the use of the Ramsey theorem and yields improved quantitative bounds.)

A Helly theorem for collections of convex sets with very weak topological assumptions. Now we explain the statement of the main result of [4].

Helly’s classical theorem [Hel23] states that a finite family of convex subsets of $\mathbb{R}^d$ must have a point in common if any $d+1$ of the sets have a point in common. In the contrapositive, Helly’s theorem asserts that any finite family of convex subsets of $\mathbb{R}^d$ with empty intersection contains a sub-family of size at most $d+1$ that already has empty intersection. This inspired the definition of the 

Helly number of a family $\mathcal{F}$ of arbitrary sets. If $\mathcal{F}$ has empty intersection then its Helly number is defined as the size of the largest sub-family $\mathcal{G} \subseteq \mathcal{F}$ with the following properties: $\mathcal{G}$ has empty intersection and any proper sub-family of $\mathcal{G}$ has nonempty intersection; if $\mathcal{F}$ has nonempty intersection then its Helly number is, by convention, 1. With this terminology, Helly’s theorem simply states that any finite family of convex sets in $\mathbb{R}^d$ has Helly number at most $d+1$.

Helly already realized that bounds on Helly numbers independent of the cardinality of the family are not a privilege of convexity: his topological theorem [Hel30] asserts that a finite family of open subsets of $\mathbb{R}^d$ has Helly number at most $d+1$ if the intersection of any sub-family of at most $d$ members of the family is either empty or a homology cell. Subsequently, several other topological generalizations of the Helly theorem were found. However, as far as we know, all these generalizations require vanishing homology in certain dimension.

Here we offer a generalization that requires only a bounded homology (but possibly non-zero). We consider homology with coefficients in $\mathbb{Z}_2$, and denote by $\tilde{\beta}_i(X)$ the $i$th reduced Betti number (over $\mathbb{Z}_2$) of a space $X$. Furthermore, we use the notation $\bigcap \mathcal{F} := \bigcap_{U \in \mathcal{F}} U$ as a shorthand for the intersection of a family of sets.

---

5By definition, a homology cell is a topological space $X$ all of whose (reduced, singular, integer coefficient) homology groups are trivial, as is the case if $X = \mathbb{R}^d$ or $X$ is a single point.
Theorem 4 (Thm 1. in [4]). For any non-negative integers $b$ and $d$ there exists an integer $h(b,d)$ such that the following holds. If $\mathcal{F}$ is a finite family of subsets of $\mathbb{R}^d$ such that $\beta_i(\bigcap \mathcal{G}) \leq b$ for any $\mathcal{G} \subseteq \mathcal{F}$ and every $0 \leq i \leq \lceil d/2 \rceil - 1$ then $\mathcal{F}$ has Helly number at most $h(b,d)$.

Theorem 4 subsumes many other existential Helly-type theorems as well as it helps to identify new Helly-type theorems for concrete collections of sets. For a detailed overview of consequences and additional background, we refer to Sections 1.1, 1.2 and 1.3 of [4]. For a sketch of a proof, which also explains the relation to (almost) embeddings, we refer the reader to Section 1.4 of [4].

On a conjecture of Kühnel. Now we explain the contents of [3].

The fact that the complete graph $K_5$ does not embed in the plane has been generalized in two independent directions. One generalization is coming from the solution of the classical Heawood problem for graphs on surfaces which implies that the complete graph $K_n$ embeds in a closed surface $M$ (other than the Klein bottle) if and only if $(n-3)(n-4) \leq 6b_1(M)$, where $b_1(M)$ is the first $\mathbb{Z}_2$-Betti number of $M$. Second generalization is the aforementioned van Kampen–Flores theorem saying that the $k$-skeleton of the $n$-dimensional simplex embeds in $\mathbb{R}^{2k}$ if and only if $n \leq 2k+1$.

In [Küh94], Kühnel conjectured the following common generalization.

Conjecture 5 (Kühnel). Let $n, k \geq 1$ be integers. If $\Delta_n^{(k)}$ embeds in a compact, $(k-1)$-connected $2k$-manifold $M$ with $k$th $\mathbb{Z}_2$-Betti number $b_k(M)$, then

$$\binom{n-k-1}{k+1} \leq \binom{2k+1}{k+1} b_k(M). \quad (1.1)$$

In [3], using the aforementioned technique, we obtained the following bound towards the Kühnel conjecture.

Theorem 6 (Thm. 2 in [3]). If $\Delta_n^{(k)}$ almost embeds into a $2k$-manifold $M$ then

$$n \leq 2 \binom{2k+2}{k} b_k(M) + 2k + 4,$$

where $b_k(M)$ is the $k$th $\mathbb{Z}_2$-Betti number of $M$.

The quantitative bound on $n$ from Theorem 6 is much weaker that the bound conjectured by Kühnel. On the other hand Theorem 5 does not require that the manifold is $(k-1)$-connected and also the assumption that $\Delta_n^{(k)}$ almost embeds is weaker. As far as we know, the bound from Theorem 6 is a first finite bound on $n$ of this type. In addition, Theorem 6 further generalizes to the case of mappings not covering a same point $q$-times (where $q$ is a parameter, power of a prime number); see Theorem 3 in [3].

As usual, we refer to the introduction of [3] for more detailed background.

1.3 Collapsibility and shellability

There are various ways how to simplify a simplicial complex step by step while keeping certain topological or combinatorial properties of interest. Two of the most important notions in this respect are collapsibility and shellability of a simplicial complex which we introduce below.
1.3.1 Collapsibility

Let $K$ be a simplicial complex and let $\sigma$ be a nonempty non-maximal face of $K$. We say that $\sigma$ is free if it is contained in only one maximal face $\tau$ of $K$. Let $K'$ be the simplicial complex obtained from $K$ by removing $\sigma$ and all faces above $\sigma$, that is,

$$K' := K \setminus \{ \varnothing \in K : \sigma \subseteq \varnothing \}.$$ 

We say that $K'$ arises from $K$ by an elementary collapse (induced by $\sigma$ and $\tau$). We say that a complex $K$ collapses to a complex $L$ if there exist a sequence of complexes ($K_1 = K, K_2, \ldots, K_{m-1}, K_m = L$), called a sequence of elementary collapses (from $K$ to $L$), such that $K_{i+1}$ arises from $K_i$ by an elementary collapse for any $i \in \{1, \ldots, m-1\}$. A simplicial complex $K$ is collapsible if it collapses to a point.

An important property of elementary collapses is that they preserve homotopy type. Thus, for example, collapsibility of some complex serves as a certificate that the complex is contractible (homotopically trivial). However, even if we start with a complex that is not contractible, it may be very useful to simplify it with collapses to a smaller complex for which we can determine the homotopy type more easily.

From purely theoretical point of view, collapsibility plays an essential role, for example, in PL-topology where it helps to determine properties of regular neighborhoods [RS72] or it is strongly related to the discrete Morse theory [For98] where the Morse functions (roughly) correspond to sequences of collapses. From more practical point of view, an application of collapsibility can be found, for example, in shape reconstruction [AL15].

In [9], we prove the following algorithmic result on collapsibility.

**Theorem 7** (Thm. 1 of [9]). It is NP-complete to decide whether a given 3-dimensional simplicial complex is collapsible.

In the statement above ‘3’ can be replaced with any $d \geq 3$. On the other hand, recognition of collapsible 2-dimensional complexes is polynomial time solvable.

It is easy to see that the problem in Theorem 7 belongs to NP by guessing a right sequence of collapses. Thus, the core is to show that the problem is NP-hard. The proof of the NP-hardness builds on a previous work of Malgouyres and Francès [MP08] showing that it is NP-hard to decide whether a given 3-dimensional complex collapses to 1-complex. The reduction of Malgouyres and Francès uses complexes that are (typically) homotopically non-trivial and therefore the resulting 1-complexes (for positive instances) do not further collapse to a point. The key new step in [9] is to overcome this difficulty by gluing suitable fillings to the ‘holes’ (despite the fact that the exact position of the holes is unknown prior to collapses). This requires introducing several auxiliary triangulated topological spaces including a Bing’s house with three rooms, a modification of famous Bing’s house with two rooms.

For more detailed background on Theorem 7, we refer to Section 1 of [9] and for a sketch of a proof, we refer to Section 3 of [9].
1.3.2 Shellability

Shellability of a simplicial complex is traditionally considered from a dual perspective when compared with collapsibility: this time, we start with an empty complex and we gradually add faces following certain rules until we reach the target complex.

More concretely, we say that a $d$-dimensional simplicial complex is pure if all its maximal faces (called facets) have dimension $d$. For simplicity we restrict ourselves to finite pure complexes when it is easier to grasp the definition. For such a complex $K$, a shelling of $K$ is an ordering $\sigma_1, \ldots, \sigma_t$ of all facets of $K$ such that for all $k \in \{2, \ldots, t\}$ the complex $B_k := \sigma_k \cap \left( \bigcup_{i=1}^{k-1} \sigma_i \right)$ is pure and $(\dim \sigma_k - 1)$-dimensional; here we regard $\sigma_i$ as geometric simplices. A complex is shellable if it admits a shelling.

When compared with collapsibility, shellings do not necessarily preserve the homotopy type of the complex. However, they still may affect homology (or the homotopy type) only in certain ways. In particular, every shellable complex is homotopy equivalent to a wedge of spheres.

Shellability of posets. An important class of simplicial complexes is obtained as order complexes of posets. That is, given a poset $P = (P, \leq)$, the vertices of the order complex $\Delta(P)$ are elements of $P$ and the simplices are the chains in $P$.

We will be interested especially in the following restricted case. Let $P = (P, \leq)$ be a graded poset with rank function $rk$. By $\hat{0}$ we mean the minimum element of $P$ (if it exists) and similarly by $\hat{1}$ we mean the maximum element (if it exists). For $a, b \in P$ we say that $a$ covers $b$, $a \triangleright b$, if $a > b$ and there is no $c$ with $a > c > b$. Equivalently, $a > b$ and $rk(a) = rk(b) + 1$. Pairs of elements $a, b$ with $a \triangleright b$ are also known as edges in the Hasse diagram of $P$. Atoms are elements that cover $\hat{0}$. That is, atoms are elements of rank 1 in a poset that contains the minimum element.

From now on, let us assume that $P$ contains the minimum element. Let $A$ be a set of some atoms in $P$. By $P(A) = (P(A), \leq)$ we mean the induced subposet of $P$ with the ground set

$$P(A) = \{\hat{0}\} \cup \{b \in P: b \geq a \text{ for some } a \in A\}.$$ 

Now we assume that $P$ contains both the minimum and the maximum element. Let $C(P)$ be the set of maximal chains of $P$. A shelling order is an order of chains from $C(P)$ satisfying the following condition.

(Sh) If $c'$ and $c$ are two chains from $C(P)$ such that $c'$ appears before $c$, then there is a chain $c^*$ from $C(P)$ appearing before $c$ such that $c \cap c^* \supseteq c \cap c'$ and also $c$ and $c^*$ differ in one level only (that is, $|c \Delta c^*| = 2$ where $\Delta$ denotes the symmetric difference).

A poset $P$ is shellable if it admits a shelling order. This is equivalent with saying that the order complex of $P$ (which is pure) is shellable as a simplicial complex.

Shellability of a poset serves as a tool how to show that a poset is Cohen-Macaulay. This has further consequences on intrinsic properties of the poset;
see [BGSS2]. On algebraic side, the fact that a certain polynomial ring is Koszul can be verified by checking that all intervals of a certain poset associated to the ring are Cohen-Macaulay; see the results of Peeva, Reiner and Sturmfels [PRS98] (this result is explicitly stated as Proposition 1.2 in [8]).

There are various sufficient criteria how to establish shellability of poset. Such criteria were pioneered by Björner [Bjö80] who proved that a certain edge-lexicographic labelling of the poset implies shellability. This criterion was later on extended by Björner and Wachs [BWS2] to chain-lexicographic labellings. In the next paragraph we describe a new criterion to prove shellability obtained in [8], that we call $A$-shellability. This criterion has been successfully applied to show shellability of so-called pinched Veronese posets where the direct application of the other previously known criteria seems to fail.

$A$-shellability. Now let us assume that $A = (A, \leq^A)$ is a partially ordered set of some atoms in $P$. We say that $P\langle A \rangle$ is $A$-shellable if $P\langle A \rangle$ is shellable with a shelling order respecting the order on $A$. That is, if $c$ and $c'$ are two maximal chains on $P\langle A \rangle$ and the unique atom of $c'$ appears before the unique atom of $c$ in the $\leq^A$ order, then $c'$ appears before $c$ in the shelling.

The strength of this notion is that there are three inductive criteria allowing to prove $A$-shellability inductively for a well behaved class of posets; see Theorems 2.1, 2.2 and 2.3 in [8]. The choice of the partial order on $A$ allows enough freedom not to overlook some important candidate shelling orders. However, on the other hand, if the order on $A$ is non-trivial, it still preserves certain structure that can be useful in induction.

For more details on $A$-shellability we refer to Sections 1 and 4 in [8].

Shellability of pinched Veronese posets. By the $n$-th Veronese poset with spacing on $n$ generators, denoted as $(V_{m,n}, \leq)$ we mean the following poset. Its ground set consists of non-negative integer vectors of length $n$ such the sum of their coordinates is divisible by $m$. The partial order on $V_{m,n}$ is given so that $a \leq b$ if and only if $a$ is less or equal to $b$ in each coordinate. It is not hard to see that every interval in $V_{m,n}$ is shellable and therefore Cohen-Macaulay.

If we set $m = n$, we just speak of the $n$-th Veronese poset $V_n := V_{n,n}$. We can pinch this poset in the following way. We remove the distinguished vector $j$ which contains 1 in each coordinate. We also remove order relations between vectors that differ exactly by $j$ (making them incomparable). In this way we thus obtain the $n$-th pinched Veronese poset $(V^*_n, \leq)$. It is very interesting that removing this single element $j$ (and corresponding order relation) strongly influences understanding the properties of the poset.

By using the properties of $A$-shellability, in [8] we prove the following.

**Theorem 8** (Thm 1.1 in [8]). Let $n \geq 4$. For any $z \in V^*_n$ the interval $[0, z]$ in $V^*_n$ is a shellable poset, where $0$ is the zero vector of length $n$.

Together with the aforementioned result of Peeva, Reiner and Sturmfels [PRS98], Theorem 8 provides a combinatorial proof of the result of Conca, Herzog, Trung and Valla [CHTV97] that the $n$-th pinched Veronese ring is Koszul for $n \geq 4$.

For additional background on the pinched Veronese poset and ring, we refer to Sections 1 and 4 in [8].
1.4 Curves and Graphs on surfaces

In this section we describe the contents of [2], [5] and [7]. The unifying topic of these three results is that they deal with curves and/or graphs on surfaces.

Drawing graphs on surfaces can be seen as a lower-dimensional analogue of the embeddability question into the manifolds. However, in the lower dimension we often encounter different phenomena which often yield different answers or at least different approaches how to reach the goal.

1.4.1 Untangling curves on surfaces

The earliest among the three results [7], obtained in a joint work with J. Matoušek, E. Sedgwick and U. Wagner, considers the following problem: we are given two collections $A = (\alpha_1, \ldots, \alpha_n)$ and $B = (\beta_1, \ldots, \beta_m)$ of simple curves on a surface $M$ with boundary. Each of the curves is either a closed curve avoiding the boundary or an arc meeting the boundary exactly at the two endpoints of the arc. The curves $\alpha_i$ are pairwise disjoint except that they may possibly share endpoints. Similarly $\beta_j$ are pairwise disjoint that they may possibly share endpoints. However, there might possibly be many crossings of the curves $\alpha_i$ with the curves $\beta_j$. Our aim is to untangle the $\beta_j$ from the $\alpha_i$ by some boundary-preserving homeomorphism of $M$; that is we look for a boundary preserving self-homeomorphism $\varphi: M \rightarrow M$ such that the total number of crossings between $\alpha_i$ and $\varphi(\beta_j)$ is as small as possible. We call this minimum number of crossings achievable through any choice of $\varphi$ the entanglement number of the two systems $A$ and $B$.

In the orientable case, let $f_{g,h}(m,n)$ denote the maximum entanglement number of any two systems $A = (\alpha_1, \ldots, \alpha_m)$ and $B = (\beta_1, \ldots, \beta_n)$ of almost-disjoint curves on an orientable surface of genus $g$ with $h$ holes. Analogously, we define $\hat{f}_{g,h}(m,n)$ as the maximum entanglement number of any two systems $A$ and $B$ of $m$ and $n$ curves, respectively, on a nonorientable surface of genus $g$ with $h$ holes.

The main results of [7] are the following; they provide bounds on $f_{g,h}(m,n)$ and $\hat{f}_{g,h}(m,n)$ independent of $g$ and $h$.

**Theorem 9** (Thm. 1.1 of [7]). For planar $M$, we have $f_{0,h}(m,n) = O(mn)$, independent of $h$.

**Theorem 10** (Thm. 1.2 of [7]). (i) For the orientable case, $f_{g,h}(m,n) = O((m+n)^4)$.

(ii) For the nonorientable case, $\hat{f}_{g,h}(m,n) = O((m+n)^4)$.

A small modification of a proof of Theorem 10 provides a bound on $f_{g,h}(m,n)$ and $\hat{f}_{g,h}(m,n)$ which depends on $g$ but is linear in $m$ and $n$ (see Corollary 1.6 in [7]). Such a bound is important for verification of the correctness of the algorithm in [6] and the relation with [6] was our main motivation why we considered this problem.

Independently of us, a similar problem was studied by Geelen, Huynh, and Richter [GHR13], with a rather different and very strong motivation stemming from the theory of graph minors, namely the question of obtaining explicit upper bounds for the graph minor algorithms of Robertson and Seymour. Geelen et
al. Theorem 2.1] show that $f_{g,h}(m,n)$ and $\hat{f}_{g,h}(m,n)$ are both bounded by $n3^m$, but only under the assumption that $\mathcal{M} \setminus (\beta_1 \cup \cdots \cup \beta_n)$ is connected (which is sufficient for their needs).

The proof of Theorem 9 relies on a result of Erten and Kobourov on simultaneous drawings of graphs with bends in the plane.

The proof of Theorem 10 relies on the bound from Theorem 9 via a suitable cut and glue technique. Two other main ingredients are ideas based on the change of the coordinates principle; see and a result on searching for a canonical system of loops in an orientable surface by Lazarus, Pocchiola, Vegter and Verroust. For a more detailed overview of the proof we refer to Table 1 of [7]. As usual, more additional background can be found in the introduction of [7].

1.4.2 Shortest paths on surfaces

A famous theorem of Fáry states that any simple planar graph can be embedded so that edges are represented by straight line segments.

It is natural to ask whether the following generalization of Fáry’s theorem is possible: Given a surface $S$, is there a metric on $S$ such that every graph embeddable into $S$ can be embedded so that edges are represented by shortest paths? We call such an embedding a shortest path embedding. If such a metric exists, we call it a universal shortest paths metric.

Motivation to study this question comes from various directions. Apart from the relation to Fáry’s theorem, it is also related to a conjecture of Negami which states that there exists a universal constant $c$ such that for any pair of graphs $G_1$ and $G_2$ embedded in a surface $S$, there exists a homeomorphism $h : S \to S$ such that $h(G_1)$ and $G_2$ intersect transversely at their edges and the number of edge crossings satisfies $cr(h(G_1),G_2) \leq c|E(G_1)| \cdot |E(G_2)|$. The connection is that if two graphs are embedded transversally by shortest path embeddings, then indeed no two edges cross more than once, since otherwise one of them could be shortcut.

Similarly, this question is related to untangling curves in [7]. If we had a stronger version of a result of Erten and Kobourov on simultaneous drawings of graphs with bends extended to an arbitrary surface, then we could perhaps improve the bounds in Theorem 10 Answering the question above seems as the first necessary step towards such a result.

We do not know a full answer to the question; however, in a joint work with A. Hubard, V. Kaluža and A. de Mesmay [5] we have reached the following results.

**Theorem 11** (Thm. 1 in [5]). The sphere $S^2$, the projective plane $\mathbb{R}P^2$, the torus $T^2$, and the Klein bottle $K$ can be endowed with a universal shortest path metric.

For surfaces of higher genus, a natural approach would be to look for a universal shortest path metric among hyperbolic metrics. However, we show that most of them are not universal shortest path metrics. For understanding the statement of the theorem below: if we allow that each edge is drawn as a concatenation of $k$ shortest paths, we call such a metric $k$-universal shortest paths metric.

**Theorem 12** (Thm. 3 in [5]). For any $\varepsilon > 0$, with probability tending to 1 as $g$ goes to infinity, a random hyperbolic metric is not a $O(g^{1/3-\varepsilon})$-universal shortest
paths metric. In particular, with probability tending to 1 as \( g \) goes to infinity, a random hyperbolic metric is not a universal shortest path metric.

On the other hand, if we allow \( k \)-universal shortest paths metric for \( k \) linear in \( g \), then there is already such a metric.

**Theorem 13** (Thm. 4 in [5]). For every \( g > 1 \), there exists an \( O(g) \)-universal shortest path hyperbolic metric \( m \) on the orientable surface \( S \) of genus \( g \).

### 1.4.3 Hanani-Tutte theorem on the projective plane

Finally, regarding graphs on surfaces, let us briefly explain the contents of [2], obtained in a joint work with É. Colin de Verdière, V. Kaluža, P. Paták and Z. Patáková.

The strong Hanani-Tutte theorem [CH34, Tut70, PSS07] states that whenever a graph can be drawn in the plane in such a way that every pair of disjoint edges crosses evenly, then the graph is actually planar. Apart from the intrinsic combinatorial beauty of this theorem, it can be also seen as an analogue of completeness of van Kampen obstruction for embedding \( k \)-complexes into \( \mathbb{R}^{2k} \) for \( k \geq 3 \). It can be also seen as a basis for results on various notions of planarity [Sch13b].

It is an open question whether the strong Hanani-Tutte theorem is valid on other surfaces; that is, if a graph can be drawn on a surface \( S \) in such a way that every pair of disjoint edges crosses evenly, then the graph actually embeds into \( S \). Pelsmajer, Schaefer and Stasi [PSS09] proved that the strong Hanani-Tutte theorem is valid on the projective plane via the inspection of the forbidden minors for the projective plane. Unfortunately, this approach cannot be used on other surfaces.

The main aim of [2] is to provide an alternative constructive proof of the strong Hanani-Tutte on the projective plane not relying on forbidden minors. The cost that we pay is that the proof is more complicated. On the other hand, there is a hope that this approach could be extended to other surfaces. (Or it could yield a desired structure for a counterexample if some essential step fails.)

### 1.5 Homology growth of flag complexes

Last but not least, in a joint work with K. Adiprasito and E. Nevo [1] we study the maximal possible growth of homology of clique complexes over graphs with a fixed forbidden induced minor.

More precisely, let \( \mathbb{K} \) be any field, \( H \) be any simple finite graph, and

\[
b_H(n) = b_H(n, \mathbb{K}) = \max_G \left\{ \sum_{i \geq -1} \dim_\mathbb{K} \tilde{H}_i(\text{cl}(G); \mathbb{K}) \right\}
\]

where \( G \) runs over all simple graphs on at most \( n \) vertices without an induced copy of \( H \), \( \text{cl}(G) \) denotes the clique complex of \( G \) and \( \tilde{H}_i(\cdot; \mathbb{K}) \) denotes the \( i \)th reduced homology with coefficients over \( \mathbb{K} \). We are interested in the growth of \( b_H(n) \) as \( n \) tends to infinity.

\(^6\)actually slightly stronger
A similar question was previously considered by Adamaszek [Ada14]. He showed that $b(n) \leq 4^{n/5}$, for

\[ b(n) = \max_G \left\{ \sum_{i \geq -1} \dim_K \tilde{H}_i(\text{cl}(G); K) \right\} \]

where $G$ runs over all graphs on at most $n$ vertices. Adamaszek further showed that for $H = I_3$ (independent set with 3 vertices), the growth is exponential but with a smaller base, at most $\approx 1.2499 < 4^{1/5} \approx 1.3195$. It is also obvious that, if $H = K_d$ is a complete graph on $d$ vertices, then $\text{cl}(G)$ is at most $(d - 2)$-dimensional, and thus $b_{K_d}(n) = O(n^{d-1})$.

Our aim is to provide a systematic approach to this question for a general forbidden graph $H$. A strong part of our motivation also comes from the case $H = C_4$ which can be seen as a discrete analogue of non-positive curvature (for a suitable metric on simplices). It is perhaps a bit surprising that a clique complexes with forbidden induced $C_4$ exist with arbitrary high homology [JS03].

We show that the limit $\lim_{n \to \infty} \sqrt[n]{b_H(n)}$ always exists and that it may attain exactly 5 possible values (four of which we can determine precisely).

**Theorem 14** (Thm. 1.2 in [1]). Let $H$ be any graph. The limit $c_H = \lim_{n \to \infty} \sqrt[n]{b_H(n)}$ exists. In addition:

(i) If $H \not\subseteq K_{5,5,...}$, then $c_H = 4^{1/5} \approx 1.3195$.

(ii) For every $i \in \{1, \ldots, 5\}$ there is a value $c'_i$ with the following property. If $H = K_{i_1, \ldots, i_m}$ with $5 \geq i_1 \geq \cdots \geq i_m \geq 1$, then $c_H = c'_{i_1}$. Moreover, $c'_5 = 3^{1/4} \approx 1.3161, c'_4 = 2^{1/3} \approx 1.2599, c'_3 \in [8^{1/14}, \Gamma_4] \approx [1.1601, 1.2434], \text{ and } c'_2 = c'_1 = 1$.

Here $\Gamma_4 \approx 1.2434$ is a certain fixed constant.

For the interesting case when $H = C_4$, we get the following improved bounds.

**Theorem 15** (Thm. 1.4 in [1]). If $H = K_{2,2} = C_4$ is the 4-cycle, then there are constants $c, C > 0$ such that for any $n$

\[ cn^{3/2} < b_{C_4}(n) < n^{C \sqrt{\log n}}. \]
Bibliography


On Betti numbers of flag complexes with forbidden induced subgraphs

Karim Adiprasito ∗, Eran Nevo† and Martin Tancer‡

May 4, 2016

Abstract

We analyze the asymptotic extremal growth rate of the Betti numbers of clique complexes of graphs on \( n \) vertices not containing a fixed forbidden induced subgraph \( H \).

In particular, we prove a theorem of the alternative: for any \( H \) the growth rate achieves exactly one of five possible exponentials, that is, independent of the field of coefficients, the \( n \)th root of the maximal total Betti number over \( n \)-vertex graphs with no induced copy of \( H \) has a limit, as \( n \) tends to infinity, and, ranging over all \( H \), exactly five different limits are attained.

For the interesting case where \( H \) is the 4-cycle, the above limit is 1, and we prove a slightly superpolynomial upper bound.

∗Einstein Institute of Mathematics, The Hebrew University of Jerusalem, Jerusalem, 91904 Israel. Partially supported by NSF Grant DMS 1128155.
†Einstein Institute of Mathematics, The Hebrew University of Jerusalem. Partially supported by Israel Science Foundation grants ISF-805/11 and ISF-1695/15.
‡Department of Applied Mathematics, Charles University in Prague, Malostranské náměstí 25, 118 00, Praha 1. Partially supported by the GAČR grant 16-01602Y. Part of this work was done when M. T. was affiliated with IST Austria.
1 Introduction

A central subject of extremal graph theory concerns monotone family of graphs without a fixed subgraph, and its extremal properties – starting with Turán’s theorem, the Erdős-Stone theorem and further generalizations and refinements.

The non-monotone family of graphs $G$ without fixed induced subgraphs have also been the subject of extensive research \[CS07\]; for structure (e.g., perfect graphs, chordal graphs, coloring, \[KKTW01\]), enumeration (e.g., \[PS92\]), as well as extremal properties (e.g., Ramsey theory, Erdős-Hajnal conjecture \[EH89, Chu14\]).

One of the most far-reaching aspects, however, is also seemingly the most inaccessible. Following Gromov and subsequent work of Davis, Januszkiewicz and Świątkowski, the Betti numbers of clique complexes without small induced cycles are central to the study of nonpositive curvature in certain groups and manifolds. The extremal properties of cohomology under this condition have been prominently studied by Januszkiewicz and Świątkowski \[JS03\], who created interesting examples of hyperbolic Coxeter groups of very high dimension, something that was previously thought impossible. Here we focus on this fundamental problem from a different perspective, aimed at uniting graph-theoretic and geometric perspectives.

**Question 1.1.** For any simple finite graph $H$, what is the maximal total Betti number over all clique complexes $\text{cl}(G)$ of graphs $G$ with at most $n$ vertices and without an induced copy of $H$?

Let $\mathbb{K}$ be any field, $H$ be any simple finite graph, and

$$b_H(n) = b_H(n, \mathbb{K}) = \max_G \left\{ \sum_{i \geq -1} \dim_{\mathbb{K}} \tilde{H}_i(\text{cl}(G); \mathbb{K}) \right\}$$

where $G$ runs over all simple graph on at most $n$ vertices without an induced copy of $H$, and $\tilde{H}_i(\cdot; \mathbb{K})$ denotes the $i$th reduced homology with coefficients over $\mathbb{K}$. Note that $b_H(0) = 1$ for any $H$, where $G = \emptyset$ is the only graph in the above
We are interested in the growth of $b_H(n)$ as $n$ tends to infinity. The results turn out to be, quite interestingly, independent of the coefficient field.

Adamaszek [Ada14] showed that $b(n) \leq 4^{n/5}$, for

$$b(n) = \max_G \left\{ \sum_{i \geq -1} \dim_{\mathbb{K}} \tilde{H}_i(\text{cl}(G); \mathbb{K}) \right\}$$

where $G$ runs over all graphs on at most $n$ vertices. Moreover the maximum is attained by the complete multipartite graph $K_{5,5,...,5}$ when $n$ is divisible by 5; we deduce that $\lim_{n \to \infty} n^{1/5} b(n) = 4^{1/5}$.

Therefore, if $H$ is not an induced subgraph of the infinite complete multipartite graph $K_{5,5,...}$, then $b_H(n)$ may grow as quickly as $(4^{1/5})^n$ (and again $\lim_{n \to \infty} n^{1/5} b_H(n) = 4^{1/5}$).

Thus, it is only interesting to study the function $b_H(n)$ for induced subgraphs $H$ of $K_{5,5,...}$. The (finite) induced subgraphs of $K_{5,5,...}$ are exactly the complete multipartite graphs $K_{i_1,i_2,...,i_m}$ where (without loss of generality) $5 \geq i_1 \geq \cdots \geq i_m \geq 1$. If $m = 1$, then we get the independent set $I_{i_1}$ on $i_1$ vertices (it should not be confused with $K_{1,1,1,...,1}$, the complete graph on $i_1$ vertices, which is also an induced subgraph of $K_{5,5,...}$).

Adamaszek further showed that for $H = I_3$, the growth is exponential but with a smaller base, at most $\approx 1.2499 < 4^{1/5} \approx 1.3195$. It is also obvious that, if $H = K_d$ is a complete graph on $d$ vertices, then $\text{cl}(G)$ is at most $(d - 2)$-dimensional, and thus $b_{K_d}(n) = O(n^{d-1})$.

We will prove that the limit $\lim_{n \to \infty} \sqrt[n]{b_H(n)}$ exists for any $H$ and we denote this limit by $c_H$. Most strikingly, we will prove a theorem of the alternative: $c_H$, depending on $H$, can attain one of only 5 different values:

**Theorem 1.2.** Let $H$ be any graph. The limit $c_H = \lim_{n \to \infty} \sqrt[n]{b_H(n)}$ exists. In addition:

(i) If $H \not\subseteq K_{5,5,...}$, then $c_H = 4^{1/5} \approx 1.3195$.

(ii) For every $i \in \{1, \ldots, 5\}$ there is a value $c'_i$ with the following property. If $H =$
$K_{i_1, \ldots, i_m}$ with $5 \geq i_1 \geq \cdots \geq i_m \geq 1$, then $c_H = c'_{i_1}$. Moreover, $c'_5 = 3^{1/4} \approx 1.3161$, $c'_4 = 2^{1/3} \approx 1.2599$, $c'_3 \in [8^{1/14}, \Gamma_4] \approx [1.1601, 1.2434]$, and $c'_2 = c'_1 = 1$.

Here $\Gamma_4$ is a certain constant which is precisely defined in the Preliminaries.

We summarize our results (including Adamaszek’s bounds) in Table 1.

<table>
<thead>
<tr>
<th>$H$</th>
<th>$c_H$</th>
<th>lower bound</th>
<th>upper bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H \not\subseteq K_{5,5,\ldots}$</td>
<td>$4^{1/5} \approx 1.3195$</td>
<td>$4^{n/5}$</td>
<td>$4^{n/5}$</td>
</tr>
<tr>
<td>$i_1 = 5$</td>
<td>$I_5$</td>
<td>$K_{5,5,\ldots,5}$ (m parts)</td>
<td>$3^{1/4} \approx 1.3161$</td>
</tr>
<tr>
<td>$i_1 = 4$</td>
<td>$I_4$</td>
<td>$3^{1/3} \approx 1.2599$</td>
<td>$2^{n/3}$</td>
</tr>
<tr>
<td>$i_1 = 3$</td>
<td>$I_3$</td>
<td>$\in [8^{1/14}, \Gamma_4] \approx [1.1601, 1.2434]$</td>
<td>$8^{n/14}$</td>
</tr>
<tr>
<td>$i_1 \leq 2$</td>
<td>$K_{2,2} = C_4$</td>
<td>$\Omega(n^{3/2})$</td>
<td>$n^{O(\sqrt{\log n})}$</td>
</tr>
<tr>
<td>$i_1 \leq 2$</td>
<td>$K_{2,1,\ldots,1}$ (m parts)</td>
<td>$\Theta(n^{m-1})$</td>
<td>$\Theta(n^{m-1})$</td>
</tr>
<tr>
<td>$i_1 \leq 2$</td>
<td>$K_{1,1,\ldots,1}$ (m parts)</td>
<td>$\Theta(n^{m-1})$</td>
<td>$\Theta(n^{m-1})$</td>
</tr>
</tbody>
</table>

Table 1: The value $c_H$ and the upper and lower bounds on $b_H(n)$ for interesting graphs $H$. The lower bounds are valid for infinitely many values of $n$.

Now, let us assume that $H = K_{i_1, \ldots, i_m}$ is an induced subgraph of $K_{5,5,\ldots}$ with $5 \geq i_1 \geq \cdots \geq i_m$. Theorem 1.2 shows that for $i_1 \in \{3, 4, 5\}$, the function $b_H(n)$ grows exponentially. Let $H \subseteq G$ denote that $H$ is an induced subgraph of $G$. The following theorem gives more refined bounds for any $I_5 \leq H \leq K_{5,5,\ldots}$.

**Theorem 1.3.** If $H = K_{5,\ldots,5}$ is $m$-partite, $m \geq 1$, then

$$b_H(n) \leq \left(\frac{4}{3^{5/4}}\right)^{m-1} \cdot 3^{n/4} \approx 1.0131^{m-1}1.3161^n.$$  

This bound is tight if $n - 5(m - 1)$ is divisible by 4 and positive, and is attained by the $(m - 1) + \frac{n-5(m-1)}{4}$-fold join consisting of $m - 1$ copies of $I_5$ and the rest are $I_4$.

The upper bound given in Theorem 1.2 for $H = I_3$ slightly improves the original bound by Adamaszek, but do not believe it to be optimal yet. We present
it mainly for the proof, which sets up a method how to push Adamaszek’s approach further. We believe that by the same method, the obtained value can be further improved, possibly even to the optimal bound, at the cost of a more extensive case analysis.

Regarding $H = I_4$, we show that $c'_4 = c_{I_4} = 2^{1/3}$. The proof requires an extensive case analysis; therefore, we keep it separately in the appendix. (However, some new ideas are needed as well to perform the analysis.) In fact we show exact bound $b_{I_4}(n) \leq 2^{n/3}$; see Theorem A.2 (in complementary setting, explained in the Preliminaries). This bound is tight if $n$ is divisible by 3, which is witnessed by the $n/3$-fold join of $I_3$. In this case, we did not attempt to obtain a more precise bound for $H = K_{4,\ldots,4}$ given the length of the analysis for $I_4$.

We now improve the bounds for graphs where the growth is subexponential, specifically, for certain $H = K_{i_1,\ldots,i_m}$ where $i_j \leq 2$ for any $j$.

**Theorem 1.4.** If $H = K_{2,2} = C_4$ is the 4-cycle, then there are constants $c, C > 0$ such that for any $n$

$$cn^{3/2} < b_{C_4}(n) < n^{C\sqrt{\log n}}.$$ 

**Theorem 1.5.** If $H = K_{i_1,1,\ldots,1}$ where $i_1 \leq 2$, then $b_H(n)$ has a polynomial growth

$$b_H(n) = \Theta(n^{m-1})$$

where $m$ is the number of parts in $H$.

Note that for $C_4 = K_{2,2}$ our upper bound on $b_{C_4}(n)$ is subexponential but superpolynomial. The main problem on the growth of $b_H(n)$ that remains open is the following.

**Question 1.6.** For any $k \geq 2$ and $l \geq 0$ let $H = K_{2,\ldots,2,1,\ldots,1}$ with $k$ parts of size 2 and $l$ parts of size 1. Does $b_H(n)$ have a polynomial growth, namely, is there a function $f(k,l)$ such that $b_H(n) < n^{f(k,l)}$ for any large enough $n$?

A necessary condition for a superpolynomial growth when $H = C_4$ is that for any positive integer $d$ there is a graph $G_d$ with no induces $C_4$ such that $\text{cl}(G_d)$ has
a nonvanishing homology in dimension \( d > d \). As mentioned, such constructions exist: Januszkiewicz and Świątkowski [JŚ03] found \( G_d \) such that \( c_\mathcal{V}(G_d) \) is a \( d \)-dimensional pseudomanifold, for any positive integer \( d \).

In Section 2 we overview relevant results of Adamaszek [Ada14], in Section 3 we prove Theorem 1.4, in Section 4 we prove Theorem 1.5, in Section 5 we prove the existence of the limit \( c_I \), in Section 6 we provide the exponential bounds for \( c_{I_5}, c_{I_3} \) stated in Theorem 1.2 and the refined bounds of Theorem 1.3. Concluding remarks are given in Section 7. Appendix A contains the proof of the optimal bound for \( c_{I_4} \).

2 Preliminaries

We overview some of the results of Adamaszek [Ada14] that will be also useful for us. Following Adamaszek, we present the results in the complementary setting, that is, instead of considering a clique complex over a graph we consider here the independence complex over the complement of the graph. In Section 6 some of the graph theoretical notions that we will meet along the way are more natural in this complementary setting.

Let \( G \) be a graph and we want to estimate the sum of the reduced Betti numbers of the independence complex of \( G \), denoted by \( b(G) \) (computed over some fixed field of coefficients). We will occasionally need the following lemma, which easily follows from the Künneth formula.

**Lemma 2.1** ([Ada14, Lemma 2.1(a)]). Let \( G \) and \( H \) be two graphs. Then \( b(G \sqcup H) = b(G)b(H) \) where \( G \sqcup H \) stands for the disjoint union of \( G \) and \( H \).

Given a graph \( G \), by the symbol \( N[u] = N_G[u] \) we denote the closed neighborhood of a vertex \( u \) in \( G \), that is, the set of neighbors of \( u \) including \( u \). Given a set \( A \) of vertices of \( G \), by \( G - A \) we mean the induced subgraph of \( G \) induced by \( V(G) \setminus A \). We also write \( G - v \) instead of \( G - \{v\} \) for a vertex \( v \) of \( G \). Let us state another lemma by Adamaszek useful for us.
Lemma 2.2 ([Ada14, Lemma 2.1(c)]). For any vertex $v$ of a graph $G$ we have $b(G) \leq b(G - v) + b(G - N[v])$.

The lemma follows from the Mayer-Vietoris long exact sequence for the decomposition of a simplicial complex as the union of a star and anti-star of some vertex.

Now, let us assume that $v$ is a vertex of degree $d$ of $G$ and let $v_1, \ldots, v_d$ be all its neighbors (in arbitrarily chosen order). An iterative application of the previous lemma gives the following recurrent bound; see [Ada14, Eq. (5)]. (Note that Adamaszek states the bound in slightly different notation. He also assumes that $v$ is a vertex of minimum degree. However, this assumption is unimportant in the proof of Eq. (5) in [Ada14]; it is only used in subsequent computations.)

Lemma 2.3. Let $v$ be a vertex of degree $d$ and $v_1, \ldots, v_d$ all its neighbors. Then

$$b(G) \leq \sum_{i=1}^{d} b(G - N[v_i] - \{v_1, \ldots, v_{i-1}\}).$$

From this lemma, Adamaszek deduces bounds on $b(G)$ for arbitrary graph $G$ and for a graph $G$ which is triangle-free. It is very useful for our further approach to describe how to get such bounds from Lemma 2.3.

Given a class $\mathcal{G}'$ of graphs, let $b(\mathcal{G}'; n)$ denote the maximum possible $b(G')$ for a graph $G' \in \mathcal{G}'$ on at most $n$ vertices, assuming that such a graph exists (otherwise $b(\mathcal{G}'; n)$ remains undefined). Let $\mathcal{G}$ denote the class of all graphs and $\mathcal{G}_i$ denote the class of the $K_i$-free graphs, namely graphs with no copy of the complete graph on $i$ vertices.

From now on let us assume that $G$ is a fixed graph with $n$ vertices. We may also assume that $G$ does not contain isolated vertices, otherwise $b(G) = 0$. We also set $n_i$ to be the number of vertices of $G - N[v_i] - \{v_1, \ldots, v_{i-1}\}$ where $v$ and $v_1, \ldots, v_d$ are as above, for $i \in [d]$. In addition, since now we assume that $v$ is a
vertex of minimum degree. Lemma 2.3 implies
\[ b(G) \leq d \cdot b(G; n - d - 1). \] (1)
if \( G \) is arbitrary graph and
\[ b(G) \leq \sum_{i=0}^{d-1} b(G_3; n - i - d). \] (2)
if \( G \) is triangle-free. Indeed, if \( G \) is arbitrary, then \( n_i \) is at most \( n - (d + 1) \) since \(|N[v_{i+1}] \cup \{v_1, \ldots, v_i\}| \geq |N[v_{i+1}]| \geq d + 1\), and if \( G \) is triangle-free, then \( n_i \) is at most \( n - (d + 1 + i - 1) \) since \( N[v_1] \) and \( \{v_1, \ldots, v_{i-1}\} \) are in addition disjoint.

In order to conclude a suitable bound on \( b(G) \), it is sufficient to plug a suitable function into the formulas above and prove the bound inductively. Concretely, we set \( \Theta_d = d^{1/(d+1)} \) and we set \( \Gamma_d \) to be the unique root on \([1, 2]\) of the polynomial equation
\[ x^{2d} - x^{d-1} - x^{d-2} - \cdots - x - 1 = 0. \]
It turns out that the sequence \( \Theta_d \) is increasing on \([1, 4]\) and decreasing on \([4, \infty)\). In particular, it is maximized for \( d = 4 \). Similarly, \( \Gamma_d \) is increasing on \([1, 3]\) and decreasing on \([3, \infty]\), therefore maximized for \( d = 3 \). Later on we will need to know approximative values of \( \Gamma_d \) and \( \Theta_d \) for small \( d \); we provide these values for small \( d \) in Table 2.

Now, if we inductively assume that \( b(G; k) \leq \Theta_d^k \) for \( k < n \), then Equation (1) gives
\[ b(G) \leq d \cdot \Theta_d^{n-d-1} = \Theta_d^n \cdot \frac{d}{\Theta_d^{d+1}} \leq \Theta_d^n \cdot \frac{d}{\Theta_d^{d+1}} = \Theta_d^n. \] (3)
which proves \( b(G, n) \leq \Theta_n^n \). A similar computation yields \( b(G_3, n) \leq \Gamma_n^n \).

The first bound is tight as pointed out by Adamaszek, at least for \( n \) divisible by 5. We will show that the second bound is not tight, and can be improved to \( \Gamma_n^n < \Gamma_3^n \); see Section 6.3.

3 Subexponential growth for \( H = C_4 \)

We now prove Theorem 1.4. We start with the upper bound:

**Theorem 3.1.** \( b_{C_4}(n) < n^{O(\sqrt{\log n})} \).

The proof uses the following simple observation: call a nontrivial homology \( d \)-cycle \( z \) in \( \text{cl}(G) \) \( d \)-minimal for \( G \) if its vertex support, denoted \( z_0 \), satisfies that any strict subset of it is not the vertex support of any nontrivial homology \( d \)-cycle in \( \text{cl}(G) \). For a subset \( A \) of the vertices of \( G \) let \( G[A] \) denote the induced subgraph on \( A \), and \( \text{cl}(A) := \text{cl}(G[A]) \) for short. Let \( N_G(v) \) denote the set of neighbors of \( v \) in \( G \) (excluding \( v \)). For a vertex \( v \) in \( z_0 \) denote by \( \text{lk}_z(v) \) the \((d−1)\)-chain in the complex \( \text{cl}(N_z(v)) \) induced by the link map, where \( N_z(v) \) is the set \( N_G(v) \cap z_0 \); clearly \( \text{lk}_z(v) \) is a cycle. Then:

**Lemma 3.2.** If a cycle \( z \) is \( d \)-minimal for \( G \) and \( v \in z_0 \), then the \((d−1)\)-cycle \( \text{lk}_z(v) \) is homologically nontrivial in \( \text{cl}(N_z(v)) \).

**Proof.** Suppose by contradiction there is a \( d \)-chain \( C \) in \( \text{cl}(N_z(v)) \) with boundary \( \partial C = \text{lk}_z(v) \). Then \( \Gamma = z - z_{\text{star}(v)} + C \) is a \( d \)-cycle whose support is a strict subset of \( z_0 \), where \( z_{\text{star}(v)} \) is the chain obtained by restricting \( z \) to its summands that support \( v \). To reach a contradiction to the minimality of \( z \) it is enough to show that \( \Gamma \) is non-trivial. As \( \text{cl}(G) \) is flag, the cycle \( C - z_{\text{star}(v)} \) is supported on a cone and hence is trivial, so \( C \) and \( z_{\text{star}(v)} \) are homologous chains and thus \( \Gamma \) and \( z \) are homologous. \( \Box \)
Proof of Theorem 3.1. We show that if $G$ has no induced $C_4$ and $\text{cl}(G)$ has non-trivial homology in dimension $d$, then $G$ must have many vertices, specifically at least $2^{\binom{d-1}{2}}$ vertices; this implies the claimed bound.

Define the following three functions:

- $\beta(d)$ is the minimal number of vertices in the support of a nontrivial homological $d$-cycle $z$ in $\text{cl}(G)$, over all graphs $G$ as above; we will show that $\beta(d) \geq 2^{\binom{d-1}{2}}$.

- $\gamma(d)$ is the minimal number of vertices in the support of a nontrivial homological $d$-cycle in $\text{cl}(G)$ that lie outside the star of a vertex $v$ in the support of the cycle, for all graphs $G$ as above.

- $\omega(d)$ is the minimum over all graphs $G$ as above of the maximal size of an independent set in $G$ that is contained in the support of a nontrivial homological $d$-cycle in $\text{cl}(G)$.

The value $\beta(d)$ is clearly realized on graph $G$ and cycle $z$ where $z$ is $d$-minimal for $G$. By considering the link of a vertex $v \in z_0$, Lemma 3.2 immediately gives

$$\beta(d) \geq \beta(d - 1) + \gamma(d) + 1.$$ 

Now if $u$ and $w$ are two vertices in the link $\text{lk}_z(v)$ where $z$ is $d$-minimal for $G$, and $v \in z_0$, and if $u, w$ are non-adjacent in $G$, then by the no induced $C_4$ condition, there is no common neighbor of $u$ and $w$ outside the star of $v$ in $G$. Further, by looking on the link $\text{lk}_z(u)$, the number of vertices outside the star of $v$ in $\text{lk}_z(u)$ is, by definition, at least $\gamma(d - 1)$. Thus, by taking a maximal independent set in $\text{lk}_z(v)$ we get

$$\gamma(d) \geq \gamma(d - 1) \cdot \omega(d - 1).$$ 

First, let us give a simple argument that gives a weaker subexponential upper bound: Clearly $\omega(d) \geq 2$ and $\gamma(1) = 2$, hence $\gamma(d) \geq 2^d$, thus $\beta(d) \geq$
In fact $\beta(d) \geq 2^{d+1}$ for any $d$, so if $G$ has $n$ vertices and no induced $C_4$ then the maximum dimension of a nonzero homology in $\text{cl}(G)$ is $\log n$. Thus, the total Betti number of $\text{cl}(G)$ is upper bounded by the number of subsets of $[n]$ of size at most $1 + \log n$, and the bound $b_{C_4}(n) < n^{C \log n}$ follows.

Now we improve the lower bound on $\omega(d)$, to achieve the bound claimed in the theorem.

Define $\omega'(d)$ to be the minimum over all pairs $(G, a)$ of the size $|T|$ where: (1) $G$ has no induced $C_4$, (2) $\text{cl}(G)$ has nontrivial $d$-th homology, (3) $a$ is a vertex in $G$, and (4) $T$ is a maximal size independent set in $G$ such that $a \in T$ and $T \setminus \{a\}$ is contained in some $d$-minimal cycle for $G$.

As $\omega(d)$ is realized on a $d$-minimal cycle (for appropriate $G$), then $\omega(d) \geq \omega'(d) - 1$. Now we aim to show $\omega'(d) \geq 2\omega'(d-1) - 2$ for any $d \geq 2$. Let us consider a vertex $a \in G$ such that $\omega'(d)$ is realized for the pair $(G, a)$.

(1) If $a$ is not adjacent to any vertex in a $d$-minimal cycle of $G$, let $v$ be any vertex of such cycle $z$.

(2) If $a$ is in some $d$-minimal cycle $z$ of $G$, then let $v$ be any vertex in $z$ adjacent to $a$.

(3) If $a$ is not in any $d$-minimal cycle of $G$, but adjacent to some vertex in such a cycle $z$ then let $v$ be such neighbor of $a$.

Then, in any case, $\text{lk}_{\text{cl}(z \cup \{a\})}(v)$ has nontrivial $(d-1)$-st homology following Lemma 3.2. In case (1) and (2), this is trivial. In case (3), note that if $\text{lk}_{\text{cl}(z \cup \{a\})}(v)$ does not support a homologically nontrivial $(d-1)$-cycle, then $\text{cl}((z \cup \{a\}) \setminus \{v\})$ supports a homology cycle homologous to $z$ from which $a$ cannot be deleted without losing the homology generating cycle, by minimality of $z$, similarly as in the proof of Lemma 3.2. Hence $a$ is in some $d$-minimal cycle of $G$, contradicting the assumption in (3).

Thus, we can consider a maximal size independent set $T$ in $G[z \cup \{a\}]$ such that $a \in T$ and $T \setminus \{a\}$ is contained in some $(d-1)$-minimal cycle for $G[z \cup \{a\}]$. By definition, $|T| \geq \omega'(d-1)$. Let $u$ denote any vertex of $T \setminus \{a\}$.
(it exists as clearly $|T| \geq 2$).

By Lemma 3.2 again, $lk_z(u)$ is a nontrivial $(d-1)$-cycle in $cl(N_z(u))$. As $v \in N_z(u)$, we can consider a maximal size independent set $T'$ such that $v \in T'$ and $T' \setminus \{v\}$ is contained in some $(d-1)$-minimal cycle of $G[N_z(u)]$. Again, $|T'| \geq \omega'(d-1)$ by definition.

First, notice that the intersection $T \cap T'$ is empty, as an element there needs to be both adjacent and nonadjacent to $u$. Next, for $\tilde{T} = T' \cup T \setminus \{u, v\}$ we have $a \in \tilde{T}$, and that $\tilde{T} \setminus \{a\}$ lies in $z$. We claim that $\tilde{T}$ is independent. Indeed, if $x \in T$ and $y \in T'$ form an edge then $(y, u, v, x, y)$ is an induced $C_4$, a contradiction.

We conclude that

$$\omega'(d) \geq |\tilde{T}| \geq 2\omega'(d-1) - 2.$$  

We now show that $\omega'(2) \geq 3$. Consider a 2-minimal cycle $z$ for $G$ and a vertex $a \in G$ realizing $\omega'(2)$. As $z$ is nontrivial in $cl(G)$, there exists a vertex $q \in z_0$ not a neighbor of $a$ in $G$. As $lk_z(q)$ is nontrivial in $cl(N_z(q))$, there is a 1-minimal cycle $z'$ for $G[N_z(q)]$ supported on a subset of $lk_z(q)$; then $z'$ is an induced simple cycle of length $\geq 5$ in $G$ (length 3 is excluded as $z'$ is nontrivial, and length 4 is excluded by the no induced $C_4$ condition). If $z'$ has length $\geq 5$ then it contains an independent set of size 3, say with elements $s, t, u$. If $a$ forms no independent set with any two of them, then $a$ is connected w.l.o.g. to $s$ and $t$. Then $(s, a, t, q, s)$ is an induced $C_4$, a contradiction. So we assume $z' = (s, t, u, v, w, s)$, a $C_5$. If $a$ forms no independent set of size 3 with some two of the vertices of $z'$, then $a$ is a neighbor of some 3 consecutive vertices of $z'$, say $s, t, u$. Then $(s, a, u, q, s)$ is an induced $C_4$, a contradiction. Thus, $\omega'(2) \geq 3$.

We conclude $\omega(d) \geq \omega'(d) - 1 \geq 2^{d-2}$, so $\beta(d) \geq \gamma(d) \geq 2^{(d-1)/2}$ and hence $b_{C_4}(n) \leq n^{C\sqrt{\log n}}$. 

We now turn to the lower bound. Note that for any prime $p$, the graph of the projective plane of order $p$ is bipartite, connected, with no $C_4$, it has $v/2 = p^2 + p + 1$ vertices on each side and $e = (p+1)(p^2 + p + 1)$ edges. Thus, its total
Betti number equals the dimension of the first homology group, which equals \( e - v + 1 \). Given \( n \), add some isolated vertices to the above graph where \( p \) is the largest prime for which \( n/2 \geq p^2 + p + 1 \), to obtain a graph \( G \) with \( n \) vertices. As clearly \( n/2 < (2p)^2 + 2p + 1 \), we conclude that \( \text{cl}(G) \) has total Betti number of order \( \Omega(n^{3/2}) \). Thus:

**Corollary 3.3.** There exists a constant \( c \) such that for any \( n \) large enough \( b_{C_4}(n) > cn^{3/2} \).

Combining Theorem 3.1 and Corollary 3.3 gives Theorem 1.4.

4 Polynomial growth for \( H \leq K_{2,1,...,1} \)

We prove Theorem 1.5. Clearly, if \( H = K_d \) is a clique of size \( d \), then all faces of \( \text{cl}(G) \) have size \( \leq d - 1 \) and so \( b_{K_d} = O(n^{d-1}) \). On the other hand, the well known Turán graph on \( n \) vertices and without a \( K_d \), denoted \( T_{d,n} \), satisfies that \( \text{cl}(T_{d,n}) \) has dimension \( d - 2 \) and the \( (d - 2) \)-faces without the least element from each of the \( d - 1 \) color classes form a basis of \( H_{d-1}(\text{cl}(T_{d,n}); k) \); there are \( \Omega(n^{d-1}) \) such faces, thus

\[
b_{K_d} = \Theta(n^{d-1}).
\]

The other case left to consider is \( H = K_{d+1}^- \), the complete graph on \( d + 1 \) vertices minus one edge. Then, any two simplices of \( \text{cl}(G) \) of dimension \( \geq d - 1 \) intersect in a face of dimension at most \( d - 3 \). Thus, an iterated application of the Mayer-Vietoris sequence (or one application of the Mayer-Vietoris spectral sequence) shows that the union of all simplices of dimension \( \geq d - 1 \) in \( \text{cl}(G) \) is a complex with vanishing homology in dimensions \( \geq d - 1 \). Thus, the entire complex \( \text{cl}(G) \) has vanishing homology in dimensions \( \geq d - 1 \), and so \( b_{K_{d+1}^-}(n) = O(n^{d-1}) \). As \( K_d \leq K_{d+1}^- \) the lower bound provided by the Turán graph \( T_{d,n} \) applies, and we conclude

\[
b_{K_{d+1}^-} = \Theta(n^{d-1}).
\]
5 Existence of the limit $\lim_{n \to \infty} \sqrt[n]{b_H(n)}$

From the previous sections, we already know that if $H = C_4$ or $H \leq K_{2,1,\ldots,1}$ then $b_H(n)$ has a subexponential growth, and thus $c_H := \lim_{n \to \infty} \sqrt[n]{b_H(n)} = 1$ is these cases. Further, Adamaszek’s result Theorem 1.2(i) implies that if $H \not\subseteq K_{3,5,\ldots,\ldots}$ then $c_H = \sqrt[4]{3}$. Our goal now is to show that the limit $\lim_{n \to \infty} \sqrt[n]{b_H(n)}$ exists for any $H$.

We will work now in the complementary setting. Recall from the Preliminaries that $b_H(n) := b_{\overline{T}}(n)$, where $\overline{T}$ is the complement of a graph $T$.

First, we consider the case when $H$ is connected.

**Proposition 5.1.** If $H$ is connected, then the limit $\lim_{n \to \infty} \sqrt[n]{b_H(n)}$ exists.

**Proof.** Let, for any positive integer $n$, $G_n$ be a graph on at most $n$ vertices which maximizes $b_H(n)$.

For any two positive integers $m, n$, the graph $G_m \sqcup G_n$ does not contain an induced copy of $H$ as $H$ is connected and $G_m$ and $G_n$ do not contain an induced copy of $H$. (We recall that ‘$\sqcup$’ stands for the disjoint union.) Therefore, by Lemma 2.1, we get $b_H(m+n) \geq b_H(m)b_H(n)$. By the Fekete lemma for superadditive sequences [Fek23] (see also [vLW01, Lem.11.6]), the limit $\lim_{n \to \infty} \sqrt[n]{b_H(n)}$ exists. In addition, this limit is finite since we already know that $b_H(n) \leq \Theta^4_n$ (or we can use the trivial bound $b_H(n) \leq 2^n$).

We now turn to the case where $H$ is disconnected.

Clearly, for a disjoint union $H = H_1 \sqcup H_2$, we have

$$\max_{i=1,2} \liminf_n \sqrt[n]{b_{H_i}(n)} \leq \liminf_n \sqrt[n]{b_H(n)}.$$ 

Denote $C_H := \limsup_n \sqrt[n]{b_H(n)}$. We now show the reverse inequality, which then implies, in the original setting, that for $H = H_1 \ast H_2$, if the limits

$$c_{H_i} = \lim_{n \to \infty} \sqrt[n]{b_{H_i}(n)}$$
exist then the limit of $\sqrt[n]{b_H(n)}$ exists and equals $c_H = \max(c_{H_1}, c_{H_2})$.

**Proposition 5.2.** Let $H_1$ and $H_2$ be any two graphs. Then $C_{H_1 \sqcup H_2} = \max(C_{H_1}, C_{H_2})$.

**Proof.** Obviously, $\max(C_{H_1}, C_{H_2}) \leq C_{H_1 \sqcup H_2}$, since $H_1$ and $H_2$ are induced subgraphs of $H_1 \sqcup H_2$. Therefore, it is sufficient to prove $C_{H_1 \sqcup H_2} \leq \max(C_{H_1}, C_{H_2})$.

For simplicity of subsequent formulas, let $\alpha = C_{H_1 \sqcup H_2}$ and $\alpha_i := C_{H_i}$ for $i \in \{1, 2\}$. Without loss of generality, we will assume that $\alpha_1 \geq \alpha_2$, that is, our task is to show that $\alpha \leq \alpha_1$. We will achieve this task by showing that $\alpha \leq \alpha_1 + \varepsilon$ for any $\varepsilon > 0$.

Form now on, let us fix $\varepsilon > 0$. We also fix a large enough integer parameter $p$ which depends on $\varepsilon$, but we will describe the exact dependency later on. Now let $G_n$ be a graph on at most $n$ vertices which maximizes $b_{H_1 \sqcup H_2}(n)$, in particular, it does not contain an induced copy of $H_1 \sqcup H_2$. By the definition of $\alpha_1$ we get

$$b_{H_1}(n) \leq k(\varepsilon)(\alpha_1 + \varepsilon)^n$$

for every $n$ where $k(\varepsilon)$ is a large enough constant depending only on $\varepsilon$. Since $\alpha_2 \leq \alpha_1$, we can also assume that

$$b_{H_2}(n) \leq k(\varepsilon)(\alpha_1 + \varepsilon)^n$$

eventually by adjusting $k(\varepsilon)$.

Our aim is to show by induction that

$$b(G_n) \leq 2^p k(\varepsilon)(\alpha_1 + \varepsilon)^n.$$  \hspace{1cm} (6)

Note that this inequality is true for $n = 1$ since $\alpha_1 \geq 1$.

It remains to prove Eq. (6) for a fixed $n$ assuming that it is true for every smaller value. Let us distinguish several cases.

In the first case we assume that $G_n$ does not contain an induced copy of $H_2$. Then we get the desired inequality directly from Eq. (5).
In the second case, let us assume that there are at most \( p \) vertices of \( G_n \) such that when we remove these vertices, we get a graph which does not contain an induced copy of \( H_1 \). Our next task is to show that in this case, \( b(G_n) \leq 2^p k(\varepsilon)(\alpha_1 + \varepsilon)^n \) which implies desired Eq. (6). Let \( u_1, \ldots, u_j \) be the removed vertices from \( G \) and let \( G' \) be the resulting graph. Let \( G'' \) be any induced subgraph of \( G \) and let \( i := |V(G'') \cap \{u_1, \ldots, u_j\}| \) be the number of vertices \( u_1, \ldots, u_j \) in \( G'' \). We will prove by induction in \( i \) that

\[
b(G'') \leq 2^i k(\varepsilon)(\alpha_1 + \varepsilon)^n. \tag{7}
\]

When we specify Eq. (7) to \( G \), that is, \( i = p \), we get the desired inequality.

The first induction step for \( i = 0 \) follows from the fact that \( G'' \) is \( H_1 \)-free in this case and from Eq. (4). (We could get a better bound since the number of vertices of \( G'' \) is (typically) less than \( n \), but we do not need such an improvement.)

The second induction step for \( i > 0 \) follows directly from Lemma 2.2 by removing one of the vertices \( u_1, \ldots, u_j \) which is also a vertex of \( G'' \).

Finally, we distinguish a third case when we assume that \( G_n \) contains an induced copy of \( H_2 \) and after removing any \( p \) vertices from \( G_n \) we still get a graph that contains an induced copy of \( H_1 \). Let \( H_2' \) be an induced copy of \( H_2 \) in \( G \) and let \( h_i \) be the number of vertices of \( H_i \) for \( i \in \{1, 2\} \). We will prove that \( H_2' \) contains a vertex of degree at least \( \frac{p - h_2}{h_2} \) in \( G_n \). It is sufficient to show that there are more than \( p - h_2 \) edges connecting \( H_2' \) and the remainder of \( G \). For contradiction, there are at most \( p - h_2 \) such edges. Let \( H' \) be the induced subgraph of \( G \) consisting of \( H_2 \) and all neighbors of vertices of \( H_2 \) inside \( G \). Then \( H' \) has at most \( p \) vertices. Consequently, there is an induced copy \( H_1' \) of \( H_1 \) inside the induced subgraph of \( G_n \) obtained from \( G_n \) by removing the vertices of \( H' \) by our assumption of this distinguished case. By the definition of \( H' \), the two copies \( H_1' \) and \( H_2' \) are connected by no edge and therefore we have found an induced copy of \( H_1 \sqcup H_2 \), this is a contradiction.
$H'_2$ contains a vertex $v$ of degree $d \geq \frac{p-h^2}{h^2}$. Lemma 2.2 gives
\[ b(G_n) \leq b(G_n - v) + b(G_n - N_{G_n}[v]). \]

Note that $G_n - v$ has at most $n - 1$ vertices, $G_n - N_{G_n}[v]$ has at most $n - d - 1$ vertices and both these graphs do not contain an induced copy of $H_1 \sqcup H_1$ since they are induced subgraphs of $G_n$. Therefore, the induction in $n$ gives us
\[ b(G_n) \leq 2^p k(\varepsilon) \left( (\alpha_1 + \varepsilon)^{n-1} + (\alpha_1 + \varepsilon)^{n-d-1} \right). \]

It is easy to check that
\[
(\alpha_1 + \varepsilon)^{-1} + (\alpha_1 + \varepsilon)^{-d-1} \leq 1
\]
if $d$ is large enough, that is, if $p$ is large enough, for fixed $\varepsilon$, since $\alpha_1 \geq 1$. Combining the two above-mentioned inequalities, we get the desired inequality (6). \qed

Combining Propositions 5.1 and 5.2 we conclude the existence of the limit $c_H$ promised in Theorem 1.2, and that for $H = K_{i_1, \ldots, i_m}$ as in Theorem 1.2(ii) indeed $c_H = c_{I_{i_1}}$.

To finish the proof of Theorem 1.2(ii) we need to find or bound the constants $c_{I_k}$ for $k = 3, 4, 5$ (clearly $c_{I_1} = 1 = c_{I_2}$, realized by the empty graph). This we do in the next Section 6.

6 Comparing the exponential growth for graphs

$\quad I_3 \leq H \leq K_{5,3,\ldots}$

In this section we also work in complementary setting, as described in the Preliminaries.
6.1 $K_5$-free graphs

Recall that $G_5$ denotes the class of $K_5$-free graphs, namely, it consists of the graphs with no induced $K_5$. In this case, the upper bound on the homology growth can be improved from $\Theta_4^n$ to $\Theta_3^n$, which is tight.

**Proposition 6.1.** $b(G_5; n) \leq \Theta_3^n$. 

**Proof.** Let $G$ be a $K_5$-free graph with $n$ vertices. The proof is by induction on $n$. The base case $n = 0$ trivially holds as $b(\emptyset) = 1$. We may also assume that $G$ does not contain an isolated vertex otherwise $b(G) = 0$.

Let $d$ be the minimum degree of $G$. If $d \neq 4$, then the same computation as in Equation (3) gives (note that $\Theta_3^n \geq \Theta_d^n$ if $d \neq 4$; see Table 2):

$$b(G) \leq \sum_{i=0}^{d-1} \Theta_3^{n-d-1} = \Theta_3^n \frac{d}{\Theta_3^{d+1}} \leq \Theta_3^n \frac{d}{\Theta_3^{d+1}} = \Theta_3^n.$$

It remains to consider the case $d = 4$. That is, $v$ has neighbors $v_1, \ldots, v_4$. As $G$ is $K_5$ free, there is at least one missing edge among these neighbors. For simplicity, we can assume that this missing edge is $v_1v_2$ since we can choose the order of the neighbors of $v$. We deduce that $n_i \leq n - d - 1 = n - 5$ for $i \in \{1, 2, 3, 4\}$ as usual, where $n_i$ is the number of vertices of $G - N[v_i] - \{v_1, \ldots, v_{i-1}\}$. However, in addition, we can deduce that $n_2 \leq n - 6$ because $N[v_2]$ does not contain $v_1$. Therefore, Lemma 2.3 gives (using that $f(n) = \Theta^n$ is increasing)

$$b(G) \leq 3\Theta_3^{n-5} + \Theta_3^{n-6} = \Theta_3^n \Theta_3^{-6}(3\Theta_3 + 1).$$

Now, the equation $3x + 1 = x^6$ has a root $x_0 \approx 1.3038 < \Theta_3$ (this is the only root on $[1, 2]$), and therefore it is easy to deduce that $3\Theta_3 + 1 \leq \Theta_3^6$ as $\Theta_3 \geq x_0$ (one can also put directly $3\Theta_3 + 1$ and $\Theta_3^6$ into a calculator). This gives the desired bound $b(G) \leq \Theta_3^n$. 

The bound provided by Proposition 6.1 is tight for $n$ divisible by 4, as the disjoint union of $n/4$ copies of $K_4$ shows. For $n$ not divisible by 4 change the sizes
of one or two components such that each of them have size $>1$, to conclude $b(G_5; n) \geq \frac{2}{5} \Theta_n^5$. Thus, we get the following corollary.

**Corollary 6.2.** $c_{I_5}(n) = \Theta_3$.

### 6.2 $mK_5$-free graphs

Let $mK_5$ denote the disjoint union of $m$ copies of $K_5$. By Theorem 1.2(ii) and Proposition 6.1 we already know that for any $I_5 \leq H \leq mK_5$, $c_H = \Theta_3$. Here we refine the upper bound on $b_{mK_5}(n)$, as asserted in Theorem 1.3.

**Proposition 6.3.** Let $G$ be an $mK_5$-free graph (meaning that $mK_5$ does not appear as an induced subgraph of $G$) with $n$ vertices. Then

$$b(G) \leq 4^{m-1} \Theta_3^{n-5(m-1)} = \frac{\Theta_4^{5(m-1)}}{\Theta_3^{5(m-1)}} \Theta_3^n = 1.0131^{m-1} \cdot 1.3161^n.$$ 

**Proof.** We prove the result by a double induction. The outer induction is in $m$, the inner induction is in $n$. The case $m = 1$ was proved in the previous section, thus we can assume $m \geq 2$.

First, let us assume that $G$ contains $k$ isolated copies of $K_5$ for some $k > 0$. Let $G'$ be $G$ without these copies. Note that $k \leq m - 1$ and that $G'$ is $(m - k)K_5$-free. Then

$$b(G) = b(G')b(K_5)^k \leq \left(4^{m-k-1} \Theta_3^{n-5k-5(m-k-1)}\right) \cdot 4^k$$

where the equality follows from Lemma 2.1 and the inequality follows from the induction and from $b(K_5) = 4$. That is, $b(G) \leq 4^{m-1} \Theta_3^{n-5(m-1)}$ as desired.

If $G$ does not contain an isolated copy of $K_5$ then we proceed analogously as in the previous subsection. We let $d$ be the minimum degree of $G$ and we consider a vertex $v$ of degree $d$ and its neighbors.
If $d \neq 4$, then Lemma 2.3 implies
\[ b(G) \leq d \cdot 4^{m-1} \Theta_3^{n-5(m-1)} = 4^{m-1} \Theta_3^{n-5(m-1)} \frac{d}{\Theta_3^{d+1}} \]
\[ \leq 4^{m-1} \Theta_3^{n-5(m-1)} \frac{d}{\Theta_3^{d+1}} = 4^{m-1} \Theta_3^{n-5(m-1)}. \]

Now, let us assume that $d = 4$. Since we assume that $G$ has no isolated $K_5$, we either miss some edge among the neighbors $v_1, \ldots, v_4$, or the degree of some of the vertices $v_1, \ldots, v_4$ is greater than 4. In both cases, Lemma 2.3 provides us with a bound
\[ b(G) \leq 3 \cdot 4^{m-1} \Theta_3^{n-5(m-1)} + 4^{m-1} \Theta_3^{n-6(m-1)} = 4^{m-1} \Theta_3^{n-5(m-1)} \Theta_3^{2} (3\Theta_3 + 1) \]
\[ \leq 4^{m-1} \Theta_3^{n-5(m-1)} \]
as wanted. (Here we again use the inequality $3\Theta_3 + 1 \leq \Theta_3^2$ explained at the end of the proof of Proposition 6.1.)

6.3 $K_3$-free graphs

We recall that it was explained in the Preliminaries how to get Adamaszek’s bound $c_{I_3} \leq \Gamma_3$. We aim to get an improved bound $c_{I_3} \leq \Gamma_4$. The idea behind the improvement is that a more detailed combinatorial analysis of $N[v_i]$, in the setting of Lemma 2.3, reveals one of the following three options. Either $d \neq 3$ and we can use a bound with $\Gamma_4$, or $d = 3$ and $v$ can be chosen so that some of the neighbors of $v$ has degree at least 4 which again improves the bound, or, finally (assuming connectedness), $G$ is a cubic graph which means that $N[v_i]$ is a 2-degenerate graph, which again yields improving the bound.

Before addressing general triangle free graphs, it is useful first to give an upper bound on $b(G)$ for triangle-free graphs $G$ which are 2-degenerate.
6.3.1 2-degenerate triangle free graphs

Let \( D_k \) be the class of \( k \)-degenerate graphs, that is graphs, such that for every \( G \) in \( D_k \) and for every (induced) subgraph \( G' \) of \( G \), the minimum degree of \( G \) is at most \( k \).

**Proposition 6.4.** Let \( G \in D_2 \) be a triangle-free graph on \( n \) vertices. Then
\[
\sqrt{b(G)} \leq \Gamma_2 \approx 1.2207.
\]

The bound \( \Gamma_2 \) is very probably not an optimal one in this case. However, it is sufficient for our purposes.

**Proof.** The proof is essentially the same as the proof of Adamaszek’s bound for triangle free graphs using, in addition, the fact that the minimum degree is at most 2. Assume \( G \) has no isolated vertex, else the assertion is trivial, as \( b(G) = 0 \) in this case.

Let \( v \) be a vertex of minimum degree \( d \) and \( v_1 \) and \( v_2 \) (or just \( v_1 \)) be its neighbors. If \( d = 2 \), Lemma 2.3 yields
\[
b(G) \leq \Gamma_2^{n-3} + \Gamma_2^{n-4} = \Gamma_2^n \Gamma_2^{-4} (\Gamma_2 + 1) = \Gamma_2^n.
\]

In the induction, we crucially use that the subgraphs \( G - N[v_1] \) and \( G - N[v_2] - v_1 \) are also triangle-free graphs in \( D_2 \).

If \( d = 1 \), we even get \( b(G) \leq \Gamma_2^{n-2} < \Gamma_2^{n} \) from Lemma 2.3. \( \square \)

6.3.2 General triangle-free graphs

Here we prove the promised bound, namely,

**Proposition 6.5.** Let \( G \) be a triangle-free graph on \( n \) vertices. Then
\[
b(G) \leq \Gamma_4^n.
\]
Proof. As usual, the proof is by induction on \( n \), again \( d \) is the minimum degree, \( v \) is a vertex of the minimum degree and \( v_1, \ldots, v_d \) are its neighbors.

First, we can assume that \( G \) is connected. Indeed, if \( C_1, \ldots, C_k \) are the components of \( G \) then we can deduce \( b(G) \leq \Gamma^\alpha_n \) from \( b(G) = b(C_1) \cdots b(C_k) \) (see Lemma 2.1) and from the induction.

If \( d \neq 3 \), then we deduce \( b(G) \leq \Gamma^\alpha_n \) from induction analogously to the computations in the proof of Proposition 6.1. Indeed

\[
b(G) \leq \sum_{i=0}^{d-1} \Gamma^{\alpha-i-1}_4 = \Gamma^\alpha_n \sum_{i=0}^{d-1} \Gamma^{\alpha-i-1}_4 \leq \Gamma^3_4 \sum_{i=0}^{d-1} \Gamma^{\alpha-i-1}_d = \Gamma^\alpha_n.
\]

The first inequality follows from the induction analogously to Eq. (2). The last equality follows from the definition of \( \Gamma_d \). Also note that \( \Gamma_4 \) is the largest value among \( \Gamma_d \), with \( d \neq 3 \).

It remains to consider the case \( d = 3 \). We will distinguish two subcases.

In the first subcase, \( G \) is not a cubic graph (3-regular). That means, it contains two vertices, one of them of degree 3 and the second one of degree greater than 3. Thus we can adjust our choice of \( v \) and its neighbors \( v_1, v_2, v_3 \) so that the degree of \( v_1 \) is at least 4. This means that \( n_1 \leq n - 5 \), \( n_2 \leq n - 5 \), and \( n_3 \leq n - 6 \), as there is no edge between \( v_1, v_2, v_3 \). Lemma 2.3 now gives a bound

\[
b(G) \leq \Gamma^\alpha_n (\Gamma^{-5}_4 + \Gamma^{-5}_4 + \Gamma^{-6}_4) = \Gamma^\alpha_n \Gamma^{-6}_4 (2\Gamma_4 + 1).
\]

The equation \( 2x + 1 = x^6 \) has a unique solution \( x_1 \approx 1.2298 \) on \([1, 2]\) and we can deduce that \( b(G) \leq \Gamma^\alpha_n \) since \( \Gamma_4 \geq x_1 \); see Table 2.

In the second subcase we assume that \( G \) is a (connected) cubic graph. In this subcase, we will not save the value on the exponents, but we will save it on the bases. More concretely, in this case we crucially use that the graphs \( G_i - N[v_i] - \{v_1, \ldots, v_{i-1}\} \) belong to \( D_2 \) since they are proper subgraphs of a
connected cubic graph. Therefore, we can use Proposition 6.4 and together with Lemma 2.3, for \( n \geq 7 \), we deduce

\[
b(G) \leq \Gamma_2^n (\Gamma_2^{-4} + \Gamma_2^{-5} + \Gamma_2^{-6}) \leq \Gamma_4^n.
\]

It is easy to check that for \( n \leq 6 \), the only possible cubic triangle-free graph is \( K_{3,3} \). In this case \( b(K_{3,3}) = 1 \) and the required inequality is satisfied as well. \( \square \)

7 Concluding remarks

As mentioned in the Introduction, we still do not know whether there exists a graph \( H \) for which \( b_H(n) \) grows subexponentially and superpolynomially. See Question 1.6 for the candidates for such \( H \).

The computation of \( b_H(n) \) reduces to graphs with exactly \( n \) vertices:

Monotonicity. By definition, for any graph \( H \) clearly \( b_H(n) \) is weakly increasing. Let \( b_H(n) \) be the maximum total Betti number among all graphs with no induced copy of \( H \) and with exactly \( n \) vertices. In fact,

Observation 7.1. For any graph \( H \), the function \( b_H(n) \) is weakly increasing in \( n \).

Proof. First note that when adding to \( G \) an isolated vertex \( v \), the total Betti number of \( \text{cl}(G \sqcup v) \) is one more than of \( \text{cl}(G) \), where the 0th Betti number is increased by one. Thus, the result holds for \( H \) with no isolated vertex. Next, for \( H = H' \sqcup u \), \( G \) a maximizer of \( b_H(n) \) and a vertex \( w \in G \), let \( G' \) be obtained from \( G \) by adding a new vertex \( v \) and connecting it to \( w \) and all neighbors of \( w \). Then \( \text{cl}(G') \) deformation retracts to \( \text{cl}(G) \), so they have the same total Betti number. If \( H \leq G' \) then any induced copy of \( H \) in \( G' \) must contain \( v \) and \( w \); but then for \( G'' = G \sqcup v \) we get \( H \not\leq G'' \), and again the total Betti number of \( \text{cl}(G'') \) is one more than of \( \text{cl}(G) \). \( \square \)
References


A \( K_4 \)-free graphs

Preliminaries. We keep the notational standards introduced in Section 2. We recall, that \( N[v] = N_G[v] \) denotes the closed neighborhood of vertex \( v \) in a graph \( G \), that is, the set of neighbors of \( v \) together with \( v \). We, however, modify the definition of the open neighborhood from Section 3. Throughout the appendix we assume that \( N(v) = N_G(v) \) is the subgraph of \( G \) induced by neighbors of \( v \). That is, it is not only the set of neighbors as in Section 3. (For further considerations of the closed neighborhoods, it is not important whether we consider the subgraph or just the set of vertices.)

Since we plan to use Lemma 2.3 quite heavily, it pays of to set up certain additional notational conventions. Once we fix \( v \) and the order of the neighbors, \( v_1, \ldots, v_d \) we define \( G^i = G - N[v_i] - \{v_1, \ldots, v_{i-1}\} \) for \( i \in [d] \). That is, the inequality in Lemma 2.3 can be rewritten as

\[
\text{b}(G) \leq \sum_{i=1}^{d} \text{b}(G^i). \tag{8}
\]

We also denote by \( k_i \) the size of the set \( V(N[v_i]) \cup \{v_1, \ldots, v_{i-1}\} \), that is \( G^i \) has \( n_i = n - k_i \) vertices.

Lemma A.1. Let \( v_1, \ldots, v_d \) be vertices forming a cut in \( G \) and let \( C \) be one of the components of \( G - \{v_1, \ldots, v_d\} \) and \( G' \) be the union of the remaining components. Then

\[
\text{b}(G) \leq \text{b}(C)\text{b}(G') + \sum_{i=1}^{d} \text{b}(G^i).
\]

The proof of this lemma is essentially the same as the proof of Lemma 2.3 in Adamaszek’s paper [Ada14].
Proof. This lemma is obtained by an iterative application of Lemma 2.2. We
remove all the vertices \( v_1, \ldots, v_d \) one by one in the given order. Finally, we use
that \( b(G - \{v_1, \ldots, v_d\}) = b(C)b(G') \) by Lemma 2.1.

The main bound. We prove the following bound for \( K_4 \)-free graphs.

**Theorem A.2.** Let \( G \) be a graph with \( n \) vertices and without an induced copy of \( K_4 \). Then
\[
b(G) \leq \Theta_2^n = 2^{n/3} \approx 1.2599^n.
\]

If, in addition, \( G \) contains a vertex of degree at most 3 which is not in a component consisting of a single triangle, then
\[
b(G) \leq (\Theta_2^{-4} + \Theta_2^{-5} + \Theta_2^{-6})\Theta_2^n.
\]

This first bound is asymptotically optimal as witnessed by the disjoint union of triangles.

Given that \( \Theta_2^{-4} + \Theta_2^{-5} + \Theta_2^{-6} \approx 0.9618 \), the improvement from the second bound is very minor. However, it will be our crucial tool for ruling out 4-regular graphs.

**Minimal counterexample approach.** The proof is in principle given by induction in the spirit of previous proofs; however, some new ingredients are needed. From practical point of view, it is better to reformulate the induction in this case as the minimal counterexample approach. That is, we will assume that \( G \) is a counterexample to Theorem A.2 with the least number of vertices and we will gradually narrow the set of possible counterexamples until we show that such \( G \) cannot exist. It is easy to check that the theorem is valid for \( n = 1 \) or \( n = 2 \).
A.1 Roots of suitable polynomials

As our approach in previous sections suggest, we will need to know the roots of several suitable polynomials. Here we extend the considerations from Section 2. Given an ordered $t$-tuple of positive integers $(a_1, \ldots, a_t)$, we will consider the equation

$$1 = x^{-a_1} + x^{-a_2} + \cdots + x^{-a_t}. \quad (9)$$

This can be understood as a polynomial equation after multiplying with a suitable power of $x$. We are interested in a solution of this equation for $x \in [1, \infty)$. Note that the right hand side is at least 1 for $x = 1$ and it is a decreasing function in $x$ tending to 0. Therefore, there is a unique solution, which we denote by $r_{a_1, \ldots, a_t}$. In our previous terminology, $\Gamma_d = r_{d+1, \ldots, 2d}$ and $\Theta_d = r_{d+1, \ldots, d+1}$ where there are $d$ arguments. We will frequently use the following simple observation.

**Lemma A.3.** Whenever $\Omega$ is a real number such that $\Omega \geq r_{a_1, \ldots, a_t}$, then $\Omega^n \geq \Omega^{n-a_1} + \cdots + \Omega^{n-a_t}$, for any positive integer $n$.

**Proof.** It is sufficient to prove $1 \geq \Omega^{-a_1} + \cdots + \Omega^{-a_t}$. This immediately follows from the definition of $r_{a_1, \ldots, a_t}$. \qed

We will need to know the approximative numerical values of $r_{a_1, \ldots, a_t}$ for various $t$-tuples $(a_1, \ldots, a_t)$, so that we can mutually compare them. We present the values important for this section in Table 3; we also include some of the important values that we met previously.

We will also often use monotonicity, that is, if $(b_1, \ldots, b_t) \geq (a_1, \ldots, a_t)$ entry-by-entry, then $r_{b_1, \ldots, b_t} \leq r_{a_1, \ldots, a_t}$. This allows us to skip computing precise values for many sequences $(a_1, \ldots, a_t)$.

A.2 Initial observations about the minimal counterexample

**Lemma A.4.** Let $G$ be a disconnected graph. Then $G$ is not a minimal counterexample to Theorem A.2.
\[(a_1, \ldots, a_t) \text{ the root of the root of } \Theta - a_1^2 + \cdots + \Theta - a_t^2 \]

<table>
<thead>
<tr>
<th>(a_1, \ldots, a_t)</th>
<th>approx. value of the root</th>
<th>approx. value of (\Theta^{-a_1} + \cdots + \Theta^{-a_t})</th>
</tr>
</thead>
<tbody>
<tr>
<td>(3, 3)</td>
<td>(\Theta_2 = 2^{1/3})</td>
<td>1.2599</td>
</tr>
<tr>
<td>(6, 6, 6, 6)</td>
<td>(\Theta_2 = 2^{1/3})</td>
<td>1.2599</td>
</tr>
<tr>
<td>(5, 7, 10, 11, 11, 12, 12)</td>
<td>(r_{5,7,10,11,12})</td>
<td>1.2590</td>
</tr>
<tr>
<td>(6, 6, 9, 10, 11, 11, 12, 13)</td>
<td>(r_{6,6,9,10,11,12,13})</td>
<td>1.2590</td>
</tr>
<tr>
<td>(6, 6, 7, 8, 9)</td>
<td>(r_{6,6,7,8,9})</td>
<td>1.2564</td>
</tr>
<tr>
<td>(1, 7)</td>
<td>(r_{1,7})</td>
<td>1.2554</td>
</tr>
<tr>
<td>(5, 6, 6, 8)</td>
<td>(r_{5,6,6,8})</td>
<td>1.2541</td>
</tr>
<tr>
<td>(5, 6, 7, 7)</td>
<td>(r_{5,6,7,7})</td>
<td>1.2519</td>
</tr>
<tr>
<td>(4, 5, 6)</td>
<td>(\Gamma_4)</td>
<td>1.24985</td>
</tr>
<tr>
<td>(5, 5, 5)</td>
<td>(3^{1/5})</td>
<td>1.2457</td>
</tr>
<tr>
<td>(3, 4)</td>
<td>(\Gamma_2)</td>
<td>1.2207</td>
</tr>
</tbody>
</table>

Table 3: Solutions of Equation (9) for suitable \(t\)-tuples and values \(\Theta^{-a_1} + \cdots + \Theta^{-a_t}\) for some of them.

**Proof.** The proof follows directly from Lemma 2.1. Indeed, let \(H_1, \ldots, H_m\) be the components of \(G\), where \(m \geq 2\). Let \(n_i\) be the size of \(H_i\). For contradiction, let us assume that \(G\) is a minimal counterexample to Theorem A.2. Then \(b(H_i) \leq \Theta_2^{n_i}\). If in addition, \(H_i\) is not a triangle and it contains a vertex of degree at most 3, then \(b(H_i) \leq (\Theta_2^{-4} + \Theta_2^{-5} + \Theta_2^{-6})\Theta_2^{n_i}\). Therefore, Lemma 2.1 gives \(b(G) \leq \Theta_2^{n_1 + \cdots + n_m}\). If at least one \(H_i\) is not a triangle and it contains a vertex of degree at most 3. This contradicts that \(G\) is a counterexample to Theorem A.2.

A.3 Vertices of degree at most 2.

We begin by excluding vertices of degree at most 2.

**Lemma A.5.** Let \(G\) be a minimal counterexample to Theorem A.2. Then the minimum degree of \(G\) is at least 3.

**Proof.** For contradiction assume the minimum degree of \(G\) is less than 3.
Trivially, the minimum degree of \( G \) cannot be zero (otherwise \( b(G) = 0 \)).

If the minimum degree of \( G \) equals 1, let \( v \) be a vertex in \( G \) of degree 1. Let \( v_1 \) be the neighbor of \( v \). Lemma 2.3 gives

\[
b(G) \leq b(G - N[v_1]).
\]

This immediately gives that \( G - N[v_1] \) is a smaller counterexample.

It remains to consider the case when the minimum degree of \( G \) equals 2. Let \( v \) be a vertex of degree 2 and let \( v_1 \) and \( v_2 \) be its neighbors. If possible, we pick \( v \) so that \( v, v_1 \) and \( v_2 \) do not induce a component consisting of a single triangle. Lemma 2.3 gives

\[
b(G) \leq b(G^1) + b(G^2).
\]

Note that the size of \( G^1 \), as well as of \( G^2 \), is at most \( n - 3 \).

If \( G \) is a minimal counterexample to Theorem A.2, then

\[
b(G) \leq \Theta_2^{n-3} + \Theta_2^{n-3} = \Theta_2^n
\]

which gives the required contradiction for the first bound in Theorem A.2.

If, in addition, \( v, v_1 \) and \( v_2 \) do not induce a component consisting of a single triangle, then the size of \( G^1 \) or of \( G^2 \) is at most \( n - 4 \).

Since \( G \) is a minimal counterexample to Theorem A.2, we get

\[
b(G) \leq (\Theta_2^{n-3} + \Theta_2^{n-4}) = (\Theta_2^{n-3} + \Theta_2^{n-4})\Theta_2^n \leq (\Theta_2^{-4} + \Theta_2^{-5} + \Theta_2^{-6})\Theta_2^n
\]

where the last inequality \( \Theta_2^{-3} + \Theta_2^{-4} < \Theta_2^{-4} + \Theta_2^{-5} + \Theta_2^{-6} \) can be checked in Table 3.

Therefore \( G \) is not a counterexample to Theorem A.2.

A.4 Vertices of degree 3

We continue our analysis by excluding vertices of degree 3.

29
Proposition A.6. Let \( G \) be a minimal counterexample to Theorem A.2. Then the minimum degree of \( G \) is at least 4.

We need a number of lemmas ruling out various cases.

Lemma A.7. Let \( G \) be a minimal counterexample to Theorem A.2. Then \( G \) does not contain a vertex \( v \) of degree 3 such that the open neighborhood \( N(v) \) consists of three isolated points.

Proof. For contradiction, let us assume that \( G \) contains such a vertex \( v \) and let \( v_1, v_2, \) and \( v_3 \) be its neighbors. As usual, Lemma 2.3 gives
\[
b(G) \leq b(G^1) + b(G^2) + b(G^3).
\]
We already know that the minimum degree of \( G \) is at least 3 by Lemma A.5. Since \( v_1, v_2, \) and \( v_3 \) are three isolated points we get that the sizes of the three graphs on the right-hand side are at least \( n - 4, n - 5 \) and \( n - 6 \).

If \( G \) is a minimal counterexample to Theorem A.2, by Lemma A.3 we get
\[
b(G) \leq \Theta_2^{n-4} + \Theta_2^{n-5} + \Theta_2^{n-6} = (\Theta_2^{-5} + \Theta_2^{-6})\Theta_2^n.
\]
This is the required contradiction. (Note that we have assumed only the weaker bound in Theorem A.2 for the graphs \( G^1, G^2, \) and \( G^3 \), but we still could derive the stronger bound for \( G \).) \( \square \)

By the previous lemma, we have ruled out a case when a vertex of degree three sees three isolated vertices. Now we will focus on the case when it sees an isolated vertex and an edge. At first we do not rule it out completely but set up some necessary condition.

Lemma A.8. Let \( G \) be a minimal counterexample to Theorem A.2. If \( G \) contains a vertex \( v \) of degree 3 such that the open neighborhood \( N(v) \) consists of an edge and an isolated vertex, then all neighbors of \( v \) have degree 3.
Proof. We know that the minimum degree of $G$ is at least 3 by Lemma A.5. For contradiction, let us assume that $G$ contains a vertex $v$ such that it has three neighbors $v_1$, $v_2$ and $v_3$; $\deg v_1 \geq 4$, $\deg v_2, v_3 \geq 3$, and the induced subgraph of $G$ on $\{v_1, v_2, v_3\}$ consists of an edge and an isolated vertex. Without loss of generality we assume that $v_1$ and $v_2$ are not connected with an edge (otherwise we swap $v_2$ and $v_3$). As usual, Lemma 2.3 gives

$$b(G) \leq b(G^1) + b(G^2) + b(G^3).$$

The size of all three graphs on the right-side is at most $n - 5$.

Therefore

$$b(G) \leq 3\Theta_2^{n-5} \leq (\Theta_2^{-4} + \Theta_2^{-5} + \Theta_2^{-6})\Theta_2^{n}$$

since $3\Theta_2^{-5} < \Theta_2^{-4} + \Theta_2^{-5} + \Theta_2^{-6}$ which follows from Table 3 (or from the convexity of the function $\Theta_2^x$).

Now we may rule out the case of a vertex of degree 3 which sees two edges in its neighborhood.

Lemma A.9. Let $G$ be a minimal counterexample to Theorem A.2. Then $G$ does not contain a vertex $v$ of degree 3 such that the open neighborhood $N(v)$ consists of the path of length 2.

Proof. For contradiction, let $v$ be a vertex in $G$ contradicting the statement of the lemma and let $v_1$, $v_2$ and $v_3$ be its neighbors. Without loss of generality, $v_3$ is adjacent to $v_1$ and to $v_2$ but $\{v_1, v_2\}$ is not an edge.

We need to distinguish some cases and subcases.

(i) First we assume that $\deg v_3 = 3$.

(a) Now we consider a subcase $\deg v_2 = 3$. See Figure 1.

In this subcase let $w_2$ be the unique neighbor of $v_2$ different from $v$ and $v_3$. Let $C$ be the edge $vv_3$ and $G' = G - \{v, v_1, v_2\}$. Then Lemma A.1
Figure 1: Subcase (ia), $G$ and $G - N[v_1]$. 

gives 

$$b(G) \leq b(C)b(G') + b(G - N[v_1]) + b(G - N[v_2] - v_1).$$

We observe that in the graph $G - N[v_1]$, the vertex $v_2$ has degree 1. Thus we further get $b(G - N[v_1]) \leq b(G - N[v_1] - N[w_2])$ by Lemma 2.3. Note that $b(C) = 1$ and the size of $G'$ is $n - 4$. We also know that the size of $G - N[v_2] - v_1$ is at most $n - 5$. Finally, the size of $G - N[v_1] - N[w_2]$ is at most $n - 6$, even if $w_2$ and $v_1$ are neighbors. As usual, if $G$ is a minimal counterexample to Theorem A.2, we get 

$$b(G) \leq 1 \cdot \Theta_2^{n-4} + \Theta_2^{n-6} + \Theta_2^{n-5}$$

as required.

(b) If we consider a subcase $\deg v_1 = 3$, it can be solved analogously to the previous subcase by swapping $v_1$ and $v_2$.

(c) Finally, we consider the subcase $\deg v_1 \geq 4$ and $\deg v_2 \geq 4$. Here we use the usual bound via Lemma 2.3 which gives 

$$b(G) \leq b(G^1) + b(G^2) + b(G^3).$$

The size of $G^1$ is at most $n - 5$, the size of $G^2$ is at most $n - 6$, and the size of $G^3$ is at most $n - 4$. Therefore, we get a contradiction as above.

(ii) Now we consider the case $\deg v_3 \geq 4$. 

32
If at least one of the vertices $v_1$ or $v_2$ has degree 4, or if $\deg v_3 \geq 5$, we use again the bound
\[ b(G) \leq b(G^1) + b(G^2) + b(G^3). \]

The sizes of the three graphs on the right-hand side are either at least $n - 5$ or they are at least $n - 4$, $n - 5$ and $n - 6$ respectively. This yields the required contradiction eventually using that $3\Theta_2^{-5} < \Theta_2^{-4} + \Theta_2^{-5} + \Theta_2^{-6}$.

Finally, we know that $\deg v_1 = \deg v_2 = 3$ and $\deg v_3 = 4$. In such case either $v_3$ and $v_1$ have a single common neighbor (namely $v$), or $v_3$ and $v_2$ have a single common neighbor (again $v$), or we get the graph on Figure 2. (Indeed, if the rightmost vertex in Figure 2 has degree at least 4 then repeating the analysis above in (ii) for $v_1$ instead of $v$ gives the desired contradiction.) In the first case, we get a contradiction with Lemma A.8 for $v_1$. The second case is symmetric. In the last case, the independence complex of $G$ consists of two edges and an isolated vertex; therefore $b(G) = 2 \leq (\Theta_2^{-4} + \Theta_2^{-5} + \Theta_2^{-6})\Theta_2^5$. A contradiction.

Now we may rule out the only remaining case of minimum degree 3 when we have 3-regular graph where the open neighborhood $N(v)$ of every vertex $v$ consists of an edge and an isolated vertex.

**Lemma A.10.** Let $G$ be a minimal counterexample to Theorem A.2. Then $G$ is not a cubic (3-regular) graph such that the open neighborhood $N(v)$ of every vertex $v$ consists
of an edge and isolated vertex.

Proof. For contradiction assume $G$ is a minimal counterexample to Theorem A.2 and $G$ satisfies the condition (C) that the open neighborhood $N(v)$ of every vertex $v$ consists of an edge and isolated vertex. Equivalently, the condition (C) can be reformulated so that $G$ is a cubic graph where every vertex is incident to exactly one triangle. By contracting each triangle to a point, graphs satisfying (C) are in one to one correspondence with 3-regular multigraphs. (We allow multiple edges but we disallow loops.) See some examples on Figure 3. Let $G'$ be the multigraph obtained from $G$ by contracting the triangles of $G$. First, we show that $G'$ is actually a graph. Indeed, if $G'$ contains a triple edge, then $G'$ must be the graph on the left part of Figure 3, as $G$ is connected. In such case $b(G) = 1$ since the independence complex of $G$ is the 6-cycle. If $G'$ contains a double edge, then $G$ contains a subgraph as on Figure 4. The vertices $a$ and $v_1$ may or may not be neighbors. Let $C$ be the subgraph of $G$ induced by the vertices $u, w, x, y, z$. We get $b(C) = 0$ since the independence complex of $C$ is the path of length 4. Consequently, Lemma A.1 gives

$$b(G) \leq b(G - N[v_1]) + b(G - N[v_2] - v_1) \leq \Theta_2^{n-4} + \Theta_2^{n-5} < (\Theta_2^{-4} + \Theta_2^{-5} + \Theta_2^{-6})\Theta_2^n.$$ 

This yields the required contradiction.

Now we know that $G'$ is a graph. We distinguish two cases: either $G'$ contains an induced path of length 2 or not.
If $G'$ does not contain an induced path of length 2, let us consider any vertex $s$ of $G'$. We get that any pair of neighbors of $s$ is adjacent. Therefore $G'$ is the graph $K_4$ (as $G'$ is cubic and connected as well). We get that $G$ is the graph on the right part of Figure 3 and it remains to bound $b(G)$ for this particular graph. We need to show that $b(G) \leq \Theta \frac{12}{4}(\Theta \frac{4}{4} + \Theta \frac{5}{4} + \Theta \frac{6}{4}) \approx 15.3893$. Given that $b(G)$ is an integer, our aim is in fact to show $b(G) \leq 15$.

We choose a vertex $v$ of $G$ arbitrarily, and we choose its neighbors $v_1, v_2, v_3$ so that $v_2$ and $v_3$ are adjacent. We use the usual bound via Lemma 2.3 which gives $b(G) \leq b(G^1) + b(G^2) + b(G^3)$. This bound would not be in general sufficient for a vertex $v$ with such a neighborhood; however, we will show that three summands on the right hand-side are small enough integers for the graph at hand. The size of $G^1$ is 8, the sizes of $G^2$ and $G^3$ are 7. Since $G$ is a minimal counterexample, the first summand may be bounded by $\Theta \frac{8}{4} \approx 6.3496$. Given that this is an integer, we bound the first summand by 6. Similarly, we can bound the remaining two summands, but this time we use the stronger conclusion of Theorem A.2 which allows to bound each of the summands by $\Theta \frac{7}{4}(\Theta \frac{4}{4} + \Theta \frac{5}{4} + \Theta \frac{6}{4}) \approx 4.8473$, that is, we may bound these summands by 4. Altogether, we get $b(G) \leq 14$ which gives the required contradiction.

Finally, it remains to consider the case when $G'$ contains an induced path of length 2. In this case, $G$ contains the subgraph on Figure 5. (Some of the pairs of vertices $w_i$ and $w_j$ may be adjacent.)
Figure 5: Part of $G$ corresponding to an induced path of length 2.

We use the usual bound via Lemma 2.3 which gives

$$b(G) \leq b(G^1) + b(G^2) + b(G^3).$$

For the required contradiction, it would not be sufficient to check the orders of the graphs on the right hand-side. However, we may get a better bound for $b(G^1)$ by applying Lemma 2.3 again to this graph.

For the cut $\{v_4, v_3\}$ in $G^1$ we get

$$b(G - N[v_1]) \leq b(G - N[v_1] - N[v_4]) + b(G - N[v_1] - N[v_3] - v_4).$$

The orders of the two graphs on the right hand-side of this inequality are $n - 8$. The orders of the graphs $G - N[v_2] - v_1$ and $G - N[v_3] - \{v_1, v_2\}$ are $n - 5$. Since $G$ is a minimal counterexample, we get

$$b(G) \leq \Theta^2_2(2\Theta^{-5}_2 + 2\Theta^{-8}_2) = \Theta^5_2(3\Theta^{-5}_2) \leq \Theta^4_2(\Theta^{-4}_2 + \Theta^{-5}_2 + \Theta^{-6}_2)$$

as required. (The equality in the middle follows since $\Theta^2_2 = 2^{1/3}$; the last inequality follows from Table 3.) This gives the required contradiction.

Now we conclude everything to a proof of Proposition A.6.

Proof of Proposition A.6. Let $G$ be a minimal counterexample to Theorem A.2. By Lemma A.5 we know that the minimum degree of $G$ is at least 3. It is, therefore, sufficient to show that $G$ does not contain a vertex of degree 3.
Since $G$ is $K_4$-free the open neighborhood of any vertex must be triangle-free. That is, the open neighborhood of any vertex are either three isolated points; an edge and a point; or a path of length 2. Any of these options is ruled out by Lemmas A.7; A.9; and A.8 and A.10, respectively. 

A.5 Vertices of degree at least 6

Now we bound the maximum degree of a possible minimal counterexample.

**Lemma A.11.** Let $G$ a minimal counterexample to Theorem A.2, then the degree of every vertex of $G$ is at most 5.

**Proof.** For contradiction, let $v$ be a vertex of degree $d \geq 6$ in $G$. By Lemma 2.2, we have

$$b(G) \leq b(G - v) + b(G - N[v]).$$

Since $G$ is a minimal counterexample to Theorem A.2, we get that the right hand side of the inequality above is at most $\Theta_2^{n-1} + \Theta_2^{n-(d+1)} \leq \Theta_2^{n-1} + \Theta_2^{n-7}$. Since $r_{1,7} < \Theta_2$ (see Table 3), we get $b(G) \leq \Theta_2^{n-1} + \Theta_2^{n-7} \leq \Theta_2^n$. Together with Proposition A.6, this contradicts the fact that $G$ is a counterexample to Theorem A.2. 

A.6 Vertices of degree 4

We continue our analysis by excluding vertices of degree 4. As above, we let $G$ to be a minimal counterexample on $n$ vertices. By Proposition A.6 we know that the minimum degree of $G$ is at least 4, and by Lemma A.11, the maximum degree of $G$ is at most 5. These are already quite restrictive conditions. On the other hand, the treatment of vertices of degree 4 is perhaps the most complicated part of the proof of Theorem A.2.
We consider a vertex $v$ of degree 4 (if it exists). We check its open neighborhood $N(v)$, and depending on $N(v)$ and on the degrees of vertices of $N(v)$ in $G$, we rule out many cases how may $N(v)$ look like. Once we rule out these cases, we get graphs with certain structure; and this structure helps us to estimate $b(G)$ more precisely. This will rule out the remaining cases.

Let $v_1, \ldots, v_4$ denote the vertices of $N(v)$, and recall the bound (8)

$$b(G) \leq b(G^1) + b(G^2) + b(G^3) + b(G^4),$$

where $G^i$ stands for $G - N[v_i] - \{v_1, \ldots, v_{i-1}\}$. We will often alternate the order of the vertices $v_1, \ldots, v_4$ in order to get the best bound.

We also recall that $k_i$ is set up so that $G^i$ has $n - k_i$ vertices. If we show that $r_{k_1, \ldots, k_4} \leq \Theta_2$, then we are done, since we obtain

$$b(G) \leq \Theta_2^n$$

by Lemma A.3. This is the required contradiction. In particular, we achieve this task, if $(k_1, \ldots, k_4) \geq (6, 6, 6, 6)$ or $(k_1, \ldots, k_4) \geq (5, 6, 7, 7)$, up to possibly permuting $k_1, \ldots, k_4$; see Table 3. (This is not the same as permuting $v_1, \ldots, v_4$; permuting the vertices may yield an essentially different values of $k_1, \ldots, k_4$.) On the other hand, it is insufficient to achieve that $(k_1, \ldots, k_4)$ is $(5, 5, 7, 7)$ or $(5, 6, 6, 7)$, since $r_{k_1, \ldots, k_4} > \Theta_2$ in these cases (very tightly). This will complicate our analysis.

Now, let us inspect the possible neighborhoods $N(v)$. Since $G$ is $K_4$-free, we get that $N(v)$ is triangle-free. There are 7 options for the isomorphism class of $N(v)$ depicted on Figure 6.

The discussion above immediately gives that the last two options cannot occur for a minimal counterexample.

**Lemma A.12.** Let $G$ be a minimal counterexample to Theorem A.2. Then $G$ does not contain vertex $v$ such that $N(v)$ is isomorphic to $I_4$ or $P_2 + I_2$; see Figure 6.
**Figure 6:** The possible isomorphism classes of $N(v)$. The number at a vertex $v_i$ denotes $k_i$ under the condition that the degree of $v_i$ in $G$ is 4.

**Proof.** For contradiction, there is such $v$ in the minimal counterexample. We already know that we may assume that the minimum degree of $G$ is at least 4. Therefore, from the discussion above we get $(k_1, \ldots, k_4) \geq (5, 6, 7, 7)$, which yields the required contradiction.

Our next task is to show that if $G$ is a minimal counterexample which contains a vertex of degree 4, then $G$ is actually 4-regular. We do this in two steps. First we significantly restrict the possible isomorphism classes of $N(v)$ where $v$ is a vertex of degree 4 incident to a vertex of degree 5. Next, we analyze the remaining options in more details so that we may rule them out as well.

**Lemma A.13.** Let $G$ be a minimal counterexample to Theorem A.2. Let $v$ be a vertex of degree 4 in $G$, which is incident to a vertex of degree greater or equal to 5. Then one of the following options hold.

(a) $N(v)$ is isomorphic to $C_4$, one vertex of $N(v)$ has degree 5 in $G$ and the three remaining vertices have degrees 4 in $G$.

(b) $N(v)$ is isomorphic to $C_4$, two opposite vertices of $N(v)$ have degrees 5 in $G$ and the two remaining vertices have degrees 4 in $G$. 

39
Figure 7: Excluding the case that \( N(v) \) is isomorphic to \( P_4 \). The label at \( v_i \) denotes \( k_i \) under the condition that all bullet vertices have degree 4. (In general, it is a lower bound for \( k_i \).)

\[ (c) \text{ } N(v) \text{ is isomorphic to } K_{1,3}, \text{ one vertex of } N(v) \text{ has degree 5 in } G \text{ and the three remaining vertices have degrees 4 in } G. \]

Proof. Let \( v \) be the vertex from the statement. We gradually exclude all remaining cases. By Lemma A.12, we already know that \( N(v) \) is not isomorphic to \( P_2 + I_2 \) or \( I_4 \).

First let us consider the case that \( N(v) \) is isomorphic to \( 2P_2 \) or \( P_3 + I_1 \). Let us choose \( v_1, \ldots, v_4 \) according to Figure 6. Since one of the vertices \( v_1, \ldots, v_4 \) has degree 5, we get that \((k_1, \ldots, k_4) \geq (6, 6, 6, 7)\) or \((k_1, \ldots, k_4) \geq (5, 6, 7, 7)\) or \((k_1, \ldots, k_4) \geq (5, 6, 8)\). Therefore this option is excluded since the three roots \( r_{6,6,6,6}, r_{5,6,7,7}, \) and \( r_{5,6,6,8} \) are less than \( \Theta_2 \); see Table 3.

Now let us consider the case that \( N(v) \) is isomorphic to \( P_4 \). At least one vertex of \( N(v) \) has degree 5 in \( G \). Up to isomorphism, there are two (non-exclusive) options depicted at Figure 7. Depending on these options we label the vertices of \( N(v) \) by \( v_1, \ldots, v_4 \) according to Figure 7. In both cases, we get \((k_1, \ldots, k_4) \geq (6, 6, 6, 6)\) which contradicts that \( G \) is a minimal counterexample.

Let us continue with the case that \( N(v) \) is isomorphic to \( K_{1,3} \). If there is only one vertex in \( N(v) \) of degree 5 in \( G \), we get the case \( (c) \) of the statement of this lemma. Therefore, we may assume that \( N(v) \) contains at least two vertices of degree 5 in \( G \) and we want to exclude this case. We label the vertices of \( N(v) \) by \( v_1, \ldots, v_4 \) according to Figure 6. Up to a self-isomorphism of \( K_{1,3} \) we may assume that \( v_2 \) has degree 5 and also \( v_1 \) or \( v_3 \) has degree 5. Therefore, we get \((k_1, \ldots, k_4) \geq (6, 6, 6, 7)\) or \((k_1, \ldots, k_4) \geq (5, 6, 7, 7)\) which gives the required
Finally, it remains to consider the case that $N(v)$ is isomorphic to $C_4$. In this case, it is sufficient to exclude the case that $N(v)$ contains two vertices of degree 5 in $G$ which are neighbors. If there are two such vertices, we label the vertices of $N(v)$ by $v_1, \ldots, v_4$ so that $v_1$ and $v_2$ have degrees 5. Then $(k_1, \ldots, k_4) \geq (6, \ldots, 6)$ which is the required contradiction.

**Lemma A.14.** Let $G$ be a minimal counterexample to Theorem A.2. Then $G$ does not contain the graph on 5 vertices from Figure 8, left, as an induced subgraph, where $\deg(v_1) = \deg(v_4) = 4$, $\deg(w_4) = 5$ and $\deg(w_1), \deg(v) \in \{4, 5\}$.

**Proof.** Assume by contradiction $G$ contains such a subgraph, and observe that the neighborhood $N(v_1)$ is isomorphic to $K_{1,3}$. Therefore, Lemma A.13 gives that $\deg(v) = \deg(w_1) = 4$.

Now, let us focus on $N(v)$. Since we know all neighbors of $v_1$ and $v_4$, we get that $N(v)$ is isomorphic to one of the graphs $2P_2$ or $P_2 + I_2$. But Lemma A.12 excludes the latter case. In addition, Lemma A.13 implies that all neighbors of $v$ have degree 4 in $G$. Let $v_2$ and $v_3$ be the two remaining neighbors of $v$; see Figure 8, right.

Now, we want to use Lemma 2.3 on $v$. In this case $(k_1, \ldots, k_4) = (5, 6, 6, 7)$ which is not sufficient but we may gain a slight improvement if we inspect the
graphs on the right hand-side of Lemma 2.3 in this case:

\[ b(G) \leq b(G^1) + b(G^2) + b(G^3) + b(G^4). \] (10)

We have that the size of \( G' \) is \( n - k_i \). However, we may also check that \( G^2 = G - N[v_2] - v_1 \) contains a vertex of degree at most 3 which is not contained in a component of \( G^2 \) consisting of a single triangle. Indeed, \( v_4 \) is such a vertex. (Note that \( w_1 \) or \( w_4 \) may or may not belong to \( G^2 \)). Similarly, \( G^3 \) contains a vertex of degree at most 3 which is not contained in a component of \( G^3 \) consisting of a single triangle, which is again witnessed by \( v_4 \).

Therefore, since \( G \) is a minimal counterexample to Theorem A.2, we get

\[ b(G^2), b(G^3) \leq \Theta_2^{n-6}(\Theta_2^{-4} + \Theta_2^{-5} + \Theta_2^{-6}). \]

Hence (10) gives

\[ b(G) \leq \Theta_2^n(\Theta_2^{-5} + (\Theta_2^{-6} + \Theta_2^{-6})(\Theta_2^{-4} + \Theta_2^{-5} + \Theta_2^{-6}) + \Theta_2^{-7}). \]

We get a contradiction to the assumption that \( G \) is a counterexample to Theorem A.2 as \( r_{5,7,10,10,11,11,12,12} \leq \Theta_2 \); see Table 3.

Now we have enough tools to exclude the remaining cases of Lemma A.13.

**Lemma A.15.** Let \( G \) be a minimal counterexample to Theorem A.2. If \( G \) contains a vertex of degree 4, then \( G \) is 4-regular.

**Proof.** We know that \( G \) is connected by Lemma A.4, and has minimal degree 4 by Proposition A.6. For contradiction, let us suppose that \( G \) contains a vertex of degree 4 but \( G \) is not 4-regular. In particular, \( G \) contains a vertex \( v \) of degree 4 which is incident to a vertex of degree 5; thus one of the three options (a,b,c) in Lemma A.13 must hold.

First, let us consider the case (c), that is, \( N(v) \) is isomorphic to \( K_{1,3} \) and \( v \) is
incident to exactly one vertex of degree 5. Let us label the vertices of \( N(v) \) as in Figure 6. Now, there are two subcases, either \( v_1 \) is the vertex of degree 5, or, without loss of generality, \( v_2 \) is the vertex of degree 5.

In the first subcase, \( v_1 \) has a single neighbor \( w \) different from \( v, v_2, v_3 \) and \( v_4 \). Now let us describe \( N(v_2) \). By Lemma A.13, \( N(v_2) \) is isomorphic to \( C_4 \) or \( K_{1,3} \) (\( v_1 \) is a neighbor of \( v_2 \) of degree 5). In addition, \( v \) and \( v_1 \) belong to \( V(N(v_2)) \). We also have \( \deg_{N(v_2)}(v) = 1 \) since \( \deg_G(v) = 4 \) and \( v_2 \) and \( v_3 \) are not neighbors of \( v_2 \). Similarly, we deduce \( \deg_{N(v_2)}(v_1) \leq 2 \). This rules out both options, \( C_4 \) and \( K_{1,3} \) for the isomorphism class of \( N(v_2) \). A contradiction.

In the second subcase, we suppose that \( v_2 \) has degree 5 in \( G \). Therefore, \( v_1 \) has degree 4 in \( G \) and consequently \( N(v_1) \) is isomorphic to \( K_{1,3} \). Therefore, up to relabeling of the vertices, \( G \) contains the induced subgraph from Lemma A.14. This gives the required contradiction.

This way, we have ruled out the case (c) of Lemma A.13. Therefore, we may assume that any vertex \( v \) of degree 4 of \( G \), incident to a vertex of degree 5, falls into the case (a) or (b) of Lemma A.13.

Now, let us consider an arbitrary vertex \( w \) of degree 5, incident to a vertex \( v \) of degree 4. Our aim is to show that \( N(w) \) is isomorphic either to \( C_5 \), or to \( C_4 + I_1 \); see Figure 9.

By inspecting \( N(v) \), we get that \( v \) and \( w \) have two common neighbors, say \( v_1 \) and \( v_2 \), which are not incident. Now, we analogously inspect the 4-vertex graphs \( N(v_1), N(v_2) \) and so on (for the other vertices of degree 4 incident to \( w \)), and we arrive at one of the two cases in Figure 9 (keeping the degree notation of Figure 8).

Therefore, it is sufficient to distinguish two subcases according to the isomorphism type of \( N(w) \).

First we suppose that \( N(w) \) is isomorphic to \( C_5 \). Then all vertices of \( N(w) \) have degree 4 in \( G \). Let us label the vertices of \( N(w) \) according to Figure 9 and let \( x \) be the neighbor of \( v \) different from \( w, v_1 \) and \( v_2 \). By checking \( N(v) \) again,
we see that \( v_1 \) and \( v_2 \) are neighbors of \( x \) as well. Then by checking \( N(v_1) \) and \( N(v_2) \) we get that all vertices of \( N(w) \) are incident to \( x \), and we get that \( G \) is the graph on Figure 10. In this case, we easily observe that the independence complex of \( G \) consists of an edge and a cycle. Therefore \( b(G) = 2 \leq \Theta^2 \) which is the required contradiction.

Now, we suppose that \( N(w) \) is isomorphic to \( C_4 + I_1 \). Let us again label the vertices of \( N(w) \) according to Figure 9. In this case, the four vertices of \( C_4 \) have degree 4 in \( G \). (The last vertex \( y \) has degree 5 in \( G \), but we do not need this information.) Analogously to the previous case, we deduce that there is another vertex \( x \) incident to the vertices of \( C_4 \) in \( N(w) \). The degree of \( x \) in \( G \) may be 4 or 5. See Figure 11. Let us apply Lemma A.1 to the cut formed by the vertices \( w \) and \( x \) (in this order, which is relevant for the lemma). We obtain

\[
b(G) \leq 1 \cdot \Theta^{2-6} + \Theta^{n-6} + \Theta^{2-6} = \frac{3}{4} \Theta^2.
\]

As usual, this contradicts that \( G \) is a counterexample to Theorem A.2. \( \Box \)
4-regular graphs. Now we know that if a minimal counterexample $G$ contains a vertex of degree 4, then it must be 4-regular. Our next step is to rule out this case.

We will often need to check that a certain graph satisfies the stronger condition in the statement of Theorem A.2. Here is a useful sufficient condition which allows us to avoid distinguishing various special cases.

**Lemma A.16.** Let $G$ be a connected 4-regular graph and let $H$ be a proper subgraph of $G$ such that the number of vertices of $H$ is not divisible by 3. Then $H$ contains a vertex of degree at most 3 in $H$ which is not in a component consisting of a single triangle.

**Proof.** Let us consider a component $C$ of $H$ which has the number of vertices not divisible by 3. In particular $C$ is not a triangle. Since $H$ is a proper subgraph of a connected 4-regular graph, $C$ must contain a vertex of degree at most 3. □

In Lemma A.12 we have ruled out certain options for the neighborhood of a vertex of degree 4. Now, we may rule out further options.

**Lemma A.17.** Let $G$ be a minimal counterexample to Theorem A.2. Then $G$ does not contain vertex $v$ such that $N(v)$ is isomorphic to $2P_2$ or $P_3 + I_1$; see Figure 6.

**Proof.** By Lemma A.15, we know that $G$ is 4-regular. For contradiction, let us assume that there is a vertex $v$ such that $N(v)$ is isomorphic to $2P_2$ or $P_3 + I_1$. Let us label the neighbors of $v$ according to Figure 6. In our usual notation, this gives $(k_1, k_2, k_3, k_4) = (5, 6, 6, 7)$. This is insufficient to rule out these cases.
directly, but it will help us to focus on the stronger conclusion of Theorem A.2. Lemma 2.3 gives

\[ b(G) \leq b(G^1) + b(G^2) + b(G^3) + b(G^4) \]

where \( G^i = G - N[v_i] - \{v_1, \ldots, v_{i-1}\} \) as usual. Now, let us consider two cases depending on whether the number of vertices of \( G \) is divisible by 3. If it is divisible by 3, Lemma A.16, together with the fact that \( G \) is a minimal counterexample, gives

\[ b(G) \leq \Theta_2 n - 5 (\Theta_2^{-4} + \Theta_2^{-5} + \Theta_2^{-6}) + \Theta_2^{n-6} + \Theta_2^{n-6} + \Theta_2^{n-7} (\Theta_2^{-4} + \Theta_2^{-5} + \Theta_2^{-6}). \]

If the number of vertices of \( G \) is not divisible by 3, we analogously get

\[ b(G) \leq \Theta_2^{n-5} + 2 \Theta_2^{n-6} (\Theta_2^{-4} + \Theta_2^{-5} + \Theta_2^{-6}) + \Theta_2^{n-7}. \]

In both cases, we get the required contradiction, since \( r_{6,6,9,10,11,12,13} \leq \Theta_2 \) as well as \( r_{5,7,10,11,12,12} \leq \Theta_2 \). See Table 3.

Now, we may also rule out an open neighborhood isomorphic to \( K_{1,3} \).

**Lemma A.18.** Let \( G \) be a minimal counterexample to Theorem A.2. Then \( G \) does not contain a vertex \( v \) such that \( N(v) \) is isomorphic to \( K_{1,3} \); see Figure 6.

**Proof.** For contradiction, there is such a vertex \( v \). Let us label the neighbors of \( v \) according to Figure 6. By Lemma A.15, we know that \( G \) is 4-regular. Therefore, the only common neighbor of \( v_2 \) and \( v_1 \) is \( v \). Similarly, the only common neighbor of \( v_2 \) and \( v \) is \( v_1 \). Therefore \( N(v_2) \) must be isomorphic to \( 2P_2 \) or to \( P_2 + I_2 \). However this is already ruled out by Lemmas A.12 and A.17.

Now let us establish two graph classes that will help us to work with 4-regular graphs such that the open neighborhood of every vertex is isomorphic either to the cycle \( C_4 \) or to the path on 4 vertices \( P_4 \). The **triangular path** on \( n \)
vertices is the graph $TP_n$ such that $V(TP_n) := [n]$ and

$$E(TP_n) := \left\{ ij \in \binom{[n]}{2} : |i - j| \leq 2 \right\}.$$ 

Similarly, we define triangular cycle so that we consider the distance cyclically. That is, we get a graph $TC_n$ such that $V(TC_n) := V(TP_n) = [n]$ and

$$E(TC_n) := \left\{ ij \in \binom{[n]}{2} : i - j \pmod{n} \in \{n - 2, n - 1, 1, 2\} \right\}.$$ 

See Figure 12.

If we consider the clique complex $\text{cl}(TC_n)$, then we get a triangulation of an annulus for $n \geq 8$ even, whereas we get a triangulation of the Möbius band for $n \geq 7$ odd. We establish the following structural result for graphs with the remaining two options for open neighborhoods.

**Lemma A.19.** Let $G$ be a connected 4-regular graph such that the open neighborhood of every vertex is isomorphic either to $C_4$ or to $P_4$. Then $G$ is isomorphic to $TC_n$ for some $n \geq 6$.

**Proof.** Let us consider the clique complex $\text{cl}(G)$. By the condition on the neighborhoods, we get that $\text{cl}(G)$ is a triangulated surface, possibly with boundary.

Let $k$ be the number of vertices of $G$ such that their neighborhood is isomorphic to $C_4$ and $\ell$ be the number of remaining vertices. Therefore, by double counting, we get that $\text{cl}(G)$ has $k + \ell$ vertices, $2(k + \ell)$ edges, and $\frac{1}{3}k + \ell$ triangles. That is, the Euler characteristic $\chi(\text{cl}(G))$ equals $k + \ell - 2(k + \ell) + \frac{1}{3}k + \ell = \frac{1}{3}k$. 

Figure 12: The graphs $TC_8$ and $TC_7$ (after the identification of the vertices labeled 1 and 2).
However; surfaces with nonnegative Euler characteristic are rare, which will help us to rule out many options.

It is easy to check that $G$ has at least 6 vertices because the closed neighborhood of a single vertex has already 5 vertices.

If $\ell = 0$, then $k \geq 6$ and therefore $\chi(\text{cl}(G)) \geq 2$. This leaves an only option that $\text{cl}(G)$ is a sphere, $\chi(\text{cl}(G)) = 2$ and $k = 6$. Consequently (by checking how to extend a neighborhood of arbitrary vertex), we get that $G = K_{2,2,2} = TC_6$.

If $\ell > 0$, then $\text{cl}(G)$ must be a surface with boundary and the only options are the disc (with Euler characteristic 1), the annulus and the Möbius band (the latter two have Euler characteristic 0).

In the case of a disc, we get $k = 3$. We say, that a vertex $v$ of $G$ is a $C_4$-vertex, if $N(v)$ is isomorphic to $C_4$. The interior vertices of the disc are precisely the $c_4$-vertices. By a local check of the neighborhoods, we see that every $C_4$ vertex is adjacent to at least two $C_4$-vertices, and therefore, the $C_4$-vertices form a triangle in $G$. We now count the edges according to the number of $C_4$-vertices they contain. Note that no edge of the disc connects two boundary vertices, as such edge $e$ would separate the disc into two regions, and in the region with no interior vertex there would be a boundary vertex, not in $e$, of degree at most 2; a contradiction. Thus, each boundary vertex has exactly two neighbors in the boundary and two in the interior. The number of boundary edges is clearly $l$. To summarize, the total number of edges is $3 + l + 2l$, but it is also $2(3 + l)$, thus $l = 3$. This is a contradiction as then the triangle on the 3 boundary vertices is in $\text{cl}(G)$, eliminating the boundary of the disc.

Finally, it remains to consider the case of the annulus and the Möbius band. In this case, $k = 3 \chi(\text{cl}(G)) = 0$; so all the vertices are on the boundary of $\text{cl}(G)$. Now a simple local inspection gives that $G$ is isomorphic to $TC_n$ for $n \geq 7$. (We consider an arbitrary vertex $v$ and its neighborhood $N(v)$, then we check the neighborhoods of the vertices of $N(v)$ which locally determines the graph uniquely. We continue this inspection, until we reach a vertex from ‘two direc-
Now we need to bound $b(TC_n)$ for $n \geq 6$ in order to finish the case of graphs of minimum degree 4. First, we provide a bound for $b(TP_n)$ which will be useful for bounding $b(TC_n)$.

**Lemma A.20.** For $n \neq 3$, we have $b(TP_n) \leq 2^{n/4}$. Furthermore $b(TP_3) = 2$.

**Proof.** It is easy to determine the first few initial values by checking the corresponding independence complexes. We obtain $b(TP_0) = 1$, $b(TP_1) = 0$, $b(TP_2) = 1$, $b(TP_3) = 2$, and $b(TP_4) = 2$, where $TP_0$ stands for the empty graph.

Next, we use Lemma 2.3 to vertex $n$ and its neighbors $n-1$ and $n-2$. We get

$$b(TP_n) \leq b(TP_{n-4}) + b(TP_{n-5}).$$

This further gives $b(TP_5) \leq 1$, $b(TP_6) \leq 1$, $b(TP_7) \leq 3$, and $b(TP_8) \leq 4$. Therefore, $b(TP_n) \leq 2^{n/4}$ for $n \in [8] \setminus \{3\}$. (Note that $2^{7/4} \approx 3.3636$.) Furthermore, it is trivial to show that $b(TP_n) \leq 2^{n/4}$ for $n \geq 9$ by induction using (11). □

Now we bound $b(TC_n)$.

**Lemma A.21.** For $n \geq 9$ we have $b(TC_n) \leq 2^{n/4}(2^{-1/2} + 2^{-1/4}) \approx 1.5480 \cdot 2^{n/4}$.

**Proof.** First remove the vertex $n$ and then the vertex $n-1$ from $TC_n$. Lemma 2.2 then gives

$$b(TC_n) \leq b(TC_n - n) + b(TC_n - N[n])$$
$$\leq b(TC_n - \{n, n-1\}) + b(TC_n - n - N[n-1]) + b(TC_n - N[n])$$
$$\leq b(TP_{n-2}) + 2b(TP_{n-5}).$$

Therefore, Lemma A.20 gives

$$b(TC_n) \leq 2^{(n-2)/4} + 2 \cdot 2^{(n-5)/4} = 2^{n/4}(2^{-1/2} + 2^{-1/4}).$$
Now we may rule out 4-regular graphs.

**Lemma A.22.** Let $G$ be a minimal counterexample to Theorem A.2. Then $G$ is not a 4-regular graph.

**Proof.** For contradiction, let us assume that there is such $G$. By Lemma A.4 we know that $G$ is connected. By Lemmas A.12, A.17 and A.18 we know that the open neighborhood of every vertex in $G$ is isomorphic either to $C_4$ or to $P_4$. Lemma A.19 implies that $G$ is isomorphic to $TC_n$ for $n \geq 6$. Therefore, in order to obtain a contradiction, it is sufficient to show that $b(TC_n) \leq \Theta_2^n = 2^{n/3}$.

We treat separately the cases $n \in \{6, 7, 8\}$. The independence complex of $TC_6$ consists of three edges and therefore $b(TC_6) = 2$. The independence complex of $TC_7$ is the cycle $C_7$ which gives $b(TC_7) = 1$. Finally, the independence complex of $TC_8$ is a connected 3-regular graph (triangle-free) with 8 vertices, thus with 12 edges. Therefore $b(TC_8) = 5$. In all three cases, we easily see that $b(TC_n) \leq 2^{n/3}$.

Now we consider $n \geq 9$. Lemma A.21 gives $b(TC_n) \leq 2^{n/4}(2^{-1/2} + 2^{-1/4})$. Therefore, we need to check the inequality $2^{-1/2} + 2^{-1/4} \leq 2^{n/12}$. This inequality holds for $n \geq 8$ since $2^{8/12} \approx 1.5874$ while $2^{-1/2} + 2^{-1/4} \approx 1.5480$.

Proposition A.6 and Lemmas A.11, A.15 and A.22 together imply the following corollary.

**Proposition A.23.** Let $G$ be a minimal counterexample to Theorem A.2. Then $G$ is a 5-regular graph.

**A.7 5-regular graphs**

It remains to rule out 5-regular graphs. We use an analogous approach as in the case of 4-regular graphs. Given a minimal counterexample $G$, which is 5-regular
Figure 13: Triangle-free graphs with 5 vertices and at least 4 edges.

by Proposition A.23, and a vertex \( v \) of \( G \), we consider all possible isomorphism classes of \( N(v) \). Those are triangle free graphs on 5 vertices. All triangle free graphs on 5 vertices with at least 4 edges are depicted on Figure 13. All other triangle-free graphs on 5 vertices are subgraphs of \( P_5 \) or \( K_{1,4} \). (It is easy to check both claims from the well known list of graphs on 5 vertices.)

We use the standard approach via Lemma 2.3 to rule out the cases when \( N(v) \) does not have many edges.

**Lemma A.24.** Let \( G \) be a minimal counterexample to Theorem A.2 and let \( v \) be any vertex of \( G \). Then \( N(v) \) contains at least 5 edges.

**Proof.** For contradiction, \( G \) is a minimal counterexample and \( v \) is a vertex of \( G \) such that \( N(v) \) contains at most 4 edges. We know that \( G \) is 5-regular. Therefore, \( N(v) \) is one of the four graphs with 4 edges on Figure 13, or their subgraph. Let us label the vertices of \( N(v) \) according to Figure 13 (we fix one choice of a subgraph if \( G \) has less than 4 edges). We use Lemma 2.3 to \( v \). In our standard notation, we get \( (k_1, \ldots, k_5) \geq (6, 6, 7, 8, 9) \) (if \( G = C_4 + I_1 \), we have to permute last two coordinates). Therefore, \( G \) cannot be a counterexample since \( r_{6,6,7,8,9} < \Theta_2 \); see Table 3.

Therefore, it remains to consider the connected 5-regular graphs such that the open neighborhood of every vertex is isomorphic to \( K_{2,3}, C_5 \) or \( K_{2,3} - e \); see Figure 13. Fortunately, such graphs are very rare. In fact, we will show that there are only two such graphs. One of them is the graph of the icosahedron, which we denote by \( G_{ico} \). The second graph is the join (as a graph, not as a
simplicial complex) of $C_5$ and $I_3$ which we denote by $C_5 \star I_3$. That is, $C_5 \star I_3$ is the graph with the vertex set $V(C_5 \star I_3) = V(C_5) \cup V(I_3)$, assuming that $V(C_5)$ and $V(I_3)$ are disjoint, and with the set of edges

$$E(C_5 \star I_3) = E(C_5) \cup \{uv: u \in C_5, v \in I_3\}.$$ 

**Lemma A.25.** Let $G$ be a 5-regular graph such that the open neighborhood of every vertex is isomorphic to $K_{2,3}$, $C_5$ or $K_{2,3} - e$. Then $G$ is isomorphic to $G_{ico}$ or to $C_5 \star I_3$.

**Proof.** Let us first consider the case that $G$ contains a vertex $v$ such that $N(v)$ is isomorphic to $K_{2,3} - e$. Let us label the vertices of $N(v)$ according to Figure 13. Now, let us focus on $N(w)$. It contains $x$ and $v$, which are neighbors. Moreover, $\deg_{N(w)} v = 1$ since $v$ has degree 5 in $G$ and $v$ is also incident to $y_1$, $y_2$, and $z$ which are not incident with $w$. Therefore $N(w)$ must be isomorphic to $K_{2,3} - e$ as well, which implies that $\deg_{N(w)} x = 3$. But this is a contradiction, since $x$ has too many neighbors, namely $v$, $w$, $y_1$, $y_2$, and two other neighbors which are incident to $w$. Altogether, $G$ cannot contain a vertex such that its open neighborhood is isomorphic to $K_{2,3} - e$.

Now, let us consider the case that $G$ contains a vertex $v$ such that $N(v)$ is isomorphic to $K_{2,3}$. Let us label the vertices of $N(v)$ according to Figure 13. Now let us focus on $N(w_1)$. It contains a subgraph formed by the vertices $v$, $x_1$, $x_2$, and $x_3$ isomorphic to $K_{1,3}$. Therefore $N(w_1)$ cannot be isomorphic to $C_5$, so it must be isomorphic to $K_{2,3}$. Now we focus on $N(x_1)$. By checking $N(v)$, we get that $\deg_{N(x_1)} v = 2$. By an analogous argument, $\deg_{N(x_1)} w_1 = 2$ since we already know that $N(w_1)$ is isomorphic to $K_{2,3}$, and $w_1 v$ is an edge in $N(x_1)$. Therefore $N(x_1)$ cannot be isomorphic to $K_{2,3}$ which implies that it is isomorphic to $C_5$. Analogously, we deduce that $N(x_2)$ and $N(x_3)$ are isomorphic to $C_5$.

As $G$ is a connected 5-regular graph, in order to show that $G$ is isomorphic to $C_5 \star I_3$, it is sufficient to show that $N(x_1) = N(x_2) = N(x_3)$. We will show $N(x_1) = N(x_2)$ and the other equality $N(x_1) = N(x_3)$ will be analogous. Let us again focus on $N(v)$. The two edges $vw_1$ and $vw_2$ belong simultaneously
to $N(x_1)$ and $N(x_2)$. Now, if we refocus to $N(w_1)$, we see that the two edges of $N(x_1)$ incident with $w_1$ belong also to $N(x_2)$, as $w_1$ has degree 5 in $G$. By repeating this argument for $w_2$, we get that $N(x_1)$ and $N(x_2)$ share a 4-path on 5 vertices. As both $N(x_1)$ and $N(x_2)$ are isomorphic to $C_5$ we conclude $N(x_1) = N(x_2)$.

Finally, it remains to consider the case that the open neighborhood of every vertex of $G$ is isomorphic to $C_5$. In this case, the clique complex $\text{cl}(G)$ is a closed triangulated surface without boundary. Let $n$ be the number of vertices of $G$. By double-counting, $\text{cl}(G)$ contains $\frac{5}{2}n$ edges and $\frac{5}{3}n$ triangles. Therefore, the Euler characteristic $\chi(\text{cl}(G))$ equals $n - \frac{5}{2}n + \frac{5}{3}n = \frac{n}{6}$. In particular, $\chi(\text{cl}(G))$ is positive; therefore $\text{cl}(G)$ must be the sphere or the projective plane. The case of projective plane cannot occur, because in such case, we would have $n = 6\chi(\text{cl}(G)) = 6$, forcing $G = K_6$ in the unique 6-vertex triangulation of the projective plane; but then $\text{cl}(G)$ is the 5-simplex, a contradiction. Hence we know that $\text{cl}(G)$ is the sphere and $n = 6\chi(\text{cl}(G)) = 12$. However, it is well known that the only 5-regular graph that triangulates the sphere is the graph of the icosahedron. □

It remains to rule out the two cases from the previous lemma as minimal counterexamples.

**Lemma A.26.** Neither $C_5 \ast I_3$ nor $G_{ico}$ is a counterexample to Theorem A.2.

**Proof.** It is easy to compute that $b(C_5 \ast I_3) = 2$ since the corresponding independence complex is the union of a 5-cycle and a triangle. Since $2 < 2^{8/3} = \Theta_2^6$, we get that $C_5 \ast I_3$ is not a counterexample to Theorem A.2.

It is a bit harder to determine $b(G_{ico})$ precisely since the corresponding independence complex is 2-dimensional. Let $K$ be the independence complex of $G_{ico}$. Let $v$ be an arbitrary vertex of $G_{ico}$. Then $v$ is not incident to a subgraph of $G_{ico}$ forming the wheel graph $W_5$; see Figure 14. This gives that the link $\text{lk}_K v$ is the independence complex of $W_5$, that is, the disjoint union $C_5 + I_1$ of $C_5$ and a vertex. Let $K'$ be the complex obtained from $K$ by removing all edges which are
not incident to any triangle. This means removing 6 edges. Then the link of every vertex of $K'$ is isomorphic to $C_5$. Now, the same reasoning as in the last part of the proof of Lemma A.25 gives that $K'$ is the boundary of the icosahedron (but the vertices are significantly permuted when compared to the icosahedron for $G_{ico}$). This implies that $K$ is homotopy equivalent to the wedge of one 2-sphere and six 1-spheres. We obtain $b(G_{ico}) = 7$. Given that $7 < 8 < 2^{12/3} = \Theta_2^{12}$, we get that $G_{ico}$ is not a counterexample to Theorem A.2. \hfill $\square$

Now we conclude everything and obtain the final result.

**Proof of Theorem A.2.** For contradiction, there is a minimal counterexample $G$ to Theorem A.2. By Proposition A.23, $G$ is 5-regular. By Lemmas A.24 and A.25, $G$ must be isomorphic to $G_{ico}$ or to $C_5 \ast I_3$. However, Lemma A.26 excludes these two options. \hfill $\square$
A Direct Proof of the Strong Hanani–Tutte Theorem on the Projective Plane

Éric Colin de Verdière1 Vojtěch Kaluža2 Pavel Paták3 Zuzana Patáková3
Martin Tancer2

1Département d’informatique, École normale supérieure, Paris and CNRS, France
2Department of Applied Mathematics, Charles University in Prague, Czech Republic
3Einstein Institute of Mathematics, The Hebrew University of Jerusalem, Israel

Abstract
We reprove the strong Hanani–Tutte theorem on the projective plane. In contrast to the previous proof by Pelsmajer, Schaefer and Stasi, our method is constructive and does not rely on the characterization of forbidden minors, which gives hope to extend it to other surfaces. Moreover, our approach can be used to provide an efficient algorithm turning a Hanani–Tutte drawing on the projective plane into an embedding.

1 Introduction
A drawing of a graph on a surface is a Hanani–Tutte drawing if no two vertex-disjoint edges cross an odd number of times. We call vertex-disjoint edges independent.
Pelsmajer, Schaefer and Stasi [PSS09] proved the following theorem via consideration of the forbidden minors for the projective plane.

Theorem 1 (Strong Hanani–Tutte for the projective plane, [PSS09]). A graph $G$ can be embedded into the projective plane if and only if it admits a Hanani–Tutte drawing on the projective plane.\(^1\)

Our main result is a constructive proof of Theorem 1. The need for a constructive proof is motivated by the strong Hanani–Tutte conjecture, which states that an analogous result is valid on an arbitrary (closed) surface. This conjecture is known to be valid only on the sphere (plane) and on the projective plane. The approach via forbidden minors is relatively simple on the projective plane; however, this approach does not seem applicable to other surfaces, because there is no reasonable characterization of forbidden minors for them. (Already for the torus or the Klein bottle, the exact list is not known.)

On the other hand, our approach reveals a number of difficulties that have to be overcome in order to obtain a constructive proof. If the conjecture is true, our approach may serve as a
basis for its proof on a general surface. If the conjecture is not true, then our approach may perhaps help to reveal appropriate structure needed for a construction of a counterexample.

Unfortunately, our approach needs to build an appropriate toolbox for manipulating with Hanani–Tutte drawings on the projective plane (many tools are actually applicable to a general surface). This significantly prolongs the paper. Therefore, we present the main ideas of our approach in the first four sections of the paper while postponing the technical details to the later sections.

The Hanani–Tutte theorem on the plane and related results. Let us now briefly describe the history of the problem; for complete history and relevant results we refer to a nice survey by Schaefer [Sch13a]. Following the work of Hanani [Cho34], Tutte [Tut70] made a remarkable observation now known as the (strong) Hanani–Tutte theorem: a graph is planar if and only if it admits a Hanani–Tutte drawing in the plane. The theorem has also a parallel history in algebraic topology, where it follows from the ideas of van Kampen, Flores, Shapiro and Wu [vK33, Wu55, Sha57, Lev72].

It is a natural question whether the strong Hanani–Tutte theorem can be extended to graphs on other surfaces; as we already said before, it has been confirmed only for the projective plane [PSS09] so far. On general surfaces, only the weak version [CN00, PSS07b] of the theorem is known to be true: if a graph is drawn on a surface so that every pair of edges crosses an even number of times \(^2\), then the graph can be embedded into the surface while preserving the cyclic order of the edges at all vertices. \(^3\) Note that in the strong version we require that only independent edges cross even number of times, while in the weak version this condition has to hold for all pairs of edges.

We remark that other variants of the Hanani–Tutte theorem generalizing the notion of embedding in the plane have also been considered. For instance, the strong Hanani–Tutte theorem was proved for partially embedded graphs [Sch13b] and both weak and strong Hanani–Tutte theorem were proved also for 2-clustered graphs [FKMP15].

The strong Hanani–Tutte theorem is important from the algorithmic point of view, since it implies the Trémaux crossing theorem, which is used to prove de Fraysseix-Rosenstiehl’s planarity criterion [dFR85]. This criterion has been used to justify the linear time planarity algorithms including the Hopcroft-Tarjan [HT74] and the Left-Right [dFOdM12] algorithms. For more details we again refer to [Sch13a].

One of the reasons why the strong Hanani–Tutte theorem is so important is that it turns planarity question into a system of linear equations. For general surfaces, the question whether there exists a Hanani–Tutte drawing of \( G \) leads to a system of quadratic equations [Lev72] over \( \mathbb{Z}_2 \). If the strong Hanani–Tutte theorem is true for the surface, any solution to the system then serves as a certificate that \( G \) is embeddable. Moreover, if the proof of the Hanani–Tutte theorem is constructive, it gives a recipe how to turn the solution into an actual embedding. Unfortunately, solving systems of quadratic equations is NP-complete.

For completeness we mention that for each surface there exists a polynomial time algorithm that decides whether a graph can be embedded into that surface [Moh99, KMR08]; however, the hidden constant depends exponentially on the genus.

\(^2\)including 0 times
\(^3\)In fact, the embedding preserves the embedding scheme of the graph, where the notion of embedding scheme is a generalization of the rotation systems to arbitrary (even non-orientable) surfaces. For more details on this topic, we refer to [GT87, Chap. 3.2.3], where embedding schemes are called rotation systems and our rotation systems are called pure.
The original proofs of the strong Hanani–Tutte theorem in the plane used Kuratowski’s theorem [Kur30], and therefore are non-constructive. In 2007, Pelsmajer, Schaefer and Štefankovič [PSS07a] published a constructive proof. They showed a sequence of moves that change a Hanani–Tutte drawing into an embedding.

A key step in their proof is their Theorem 2.1. We say that an edge is \textit{even} if it crosses every other edge an even number of times (including the adjacent edges).

**Theorem 2** (Theorem 2.1 of [PSŠS07a]). \textit{If $D$ is a drawing of a graph $G$ in the plane, and $E_0$ is the set of even edges in $D$, then $G$ can be drawn in the plane so that no edge in $E_0$ is involved in an intersection and there are no new pairs of edges that intersect an odd number of times.}

Unfortunately, an analogous result is simply not true on other surfaces, as is shown in [PSŠS07b]. In particular, this is an obstacle for a constructive proof of Theorem 1.

**Our approach—replacement of Theorem 2.1 in [PSS07a].** The key step of our approach is to provide a suitable replacement of Theorem 2.1 in [PSS07a] (Theorem 2); see also Lemma 3 in [FPSŠ12]. For a description of this replacement, let us focus on the following simplified setting.

Let us consider the case that we have a graph $G$ with a Hanani–Tutte drawing $D$ on the sphere $S^2$. Let $Z$ be a cycle of $G$ which is \textit{simple}, that is, drawn without self-intersections, and such that every edge of $Z$ is even. Theorem 2 then implies that $G$ can be redrawn so that $Z$ is free of crossings without introducing new pairs of edges crossing oddly.

Actually, a detailed inspection of the proof in [PSS07a] reveals something slightly stronger in this setting. The drawing of $Z$ splits the plane into two parts that we call the \textit{inside} and the \textit{outside}. This in turn splits $G$ into two parts. The inside part consists of vertices that are inside $Z$ and of the edges that have either at least one endpoint inside $Z$, or they have both endpoints on $Z$ and they enter the inside of $Z$ next to both endpoints. The outside part is defined analogously. Because we have started with a Hanani–Tutte drawing, it is easy to check that every vertex and every edge is on $Z$ or inside or outside. The proof of Theorem 2 in [PSS07a] then implies that the inside and the outside may be fully separated.
Figure 2: Projective-planar drawing of $K_5$ where the outside and the inside cannot be separated by a continuous motion (right) and a solution by duplicating the crosscap (middle) and removing one of them (right).

in the drawing; see Fig. 1. Actually, this can be done even by a continuous motion—if the drawing is considered on the sphere (instead of the plane).

The trouble on $\mathbb{R}P^2$ is that it may not be possible to separate the outside and the inside by a continuous motion (of each of the parts separately). This is demonstrated by a projective-planar drawing of $K_5$ in Fig. 2, left. (The symbol ‘⊗’ stands for the crosscap in the picture.)

It would actually help significantly if we were allowed to duplicate the crosscap as in Fig. 2, middle. However, the problem is that we cannot afford raising the genus. On the other hand, if we give up on a continuous motion, we may observe that the inside vertices and edges in Fig. 2, middle, may be actually redrawn in a planar way if we remove the ‘inside’ crosscap. This step changes the homotopy/homology type of many cycles in the drawing.

Our main technical contribution is to show that it is not a coincidence that this simplification of the drawing in Fig. 2 was possible. We will show that it is always possible to redraw one of the sides without using the ‘duplicated’ crosscap. The precise statement is given by Theorem 10.

The remainder of the proof. As we mentioned above, Theorem 2 is a key ingredient in the proof of the strong Hanani–Tutte theorem in the plane. The rough idea is to find a suitable order on some of the cycles of the graph so that Theorem 2 can be used repeatedly on these cycles eventually obtaining a planar drawing. A detailed proof of Pelsmajer, Schaefer and Štefankovič uses an induction based on this idea.

Similarly, we use Theorem 10 in an inductive proof of Theorem 1. The details in our setting are more complicated, because we have to take care of two types of cycles in the graph based on their homological triviality. We also need to put more effort to set up the induction in a suitable way for using Theorem 10, because our setting for Theorem 10 is slightly more restrictive than the setting of Theorem 2.

Organization of the paper. In Sect. 2 we describe Hanani–Tutte drawings on the projective plane and their properties. There we also set up several tools for modifications of the drawings. In particular, we describe how to transform the Hanani–Tutte drawings on $\mathbb{R}P^2$ into drawings on the sphere satisfying a certain additional condition. This helps significantly in several cases with manipulating these drawings. In Sect. 3 we describe the precise statement of Theorem 10. We also provide a proof of this theorem in that section, however, we postpone the proofs of many auxiliary results to later sections. In Sect. 4 we prove Theorem 1.
using Theorem 10 and some of the auxiliary results from Sect. 3. The remaining sections are devoted to the missing proofs of auxiliary results.

2 Hanani–Tutte Drawings

In this section, we consider Hanani–Tutte drawings of graphs on the sphere and on the projective plane. We use the standard notation from graph theory. Namely, if $G$ is a graph, then $V(G)$ and $E(G)$ denote the set of vertices and the set of edges of $G$, respectively. Given a vertex $v$ or an edge $e$, by $G - v$ or $G - e$ we denote the graph obtained from $G$ by removing $v$ or $e$, respectively.

Regarding drawings of graphs, first, let us recall a few standard definitions considered on an arbitrary surface. We put the standard general position assumptions on the drawings. That is, we consider only drawings of graphs on a surface such that no edge contains a vertex in its interior and every pair of edges meets only in a finite number of points, where they cross transversally. However, we allow three or more edges meeting in a single point (we do not mind them because we study the pairwise interactions of the edges only). Let us also mention that, in all this paper, we can assume that in every drawing, every edge is free of self-crossings. Indeed, we can remove any self-crossing without changing the image of the edge, except in a small neighborhood of the self-crossing.

We recall from the introduction that two edges are independent if they do not share a vertex. Given a surface $S$ and a graph $G$, a (strong) Hanani–Tutte drawing of $G$ on $S$ is a drawing of $G$ on $S$ such that every pair of independent edges crosses an even number of times. We will often abbreviate the term (strong) Hanani–Tutte drawing to HT-drawing.

Crossing numbers. Let $D$ be a drawing of a graph $G$ on a surface $S$. Given two distinct edges $e$ and $f$ of $G$ by $\text{cr}(e, f) = \text{cr}_D(e, f)$ we denote the number of crossings between $e$ and $f$ in $D$ modulo 2. We say that an edge $e$ of $G$ is even if $\text{cr}(e, f) = 0$ for any $f \in E(G)$ distinct from $e$. We emphasize that we consider the crossing number as an element of $\mathbb{Z}_2$ and all computations throughout the paper involving it are done in $\mathbb{Z}_2$.

HT-drawings on $\mathbb{R}P^2$. It is convenient for us to set up some conventions for working with the HT-drawings on the (real) projective plane, $\mathbb{R}P^2$. There are various ways to represent $\mathbb{R}P^2$. Our convention will be the following: we consider the sphere $S^2$ and a disk (2-ball) $B$ in it. We remove the interior of $B$ and identify the opposite points on the boundary $\partial B$. This way, we obtain a representation of $\mathbb{R}P^2$. Let $\gamma$ be the curve coming from $\partial B$ after the identification. We call this curve a crosscap. It is a homologically (homotopically) non-trivial simple cycle (loop) in $\mathbb{R}P^2$, and conversely, any homologically (homotopically) nontrivial simple cycle (loop) may serve as a crosscap up to a self-homeomorphism of $\mathbb{R}P^2$. In drawings, we use the symbol $\otimes$ for the crosscap coming from the removal of the disk ‘inside’ this symbol. We also use this symbol for ends of proofs.

Given an HT-drawing of a graph on $\mathbb{R}P^2$, it can be slightly shifted so that it meets the crosscap in a finite number of points and only transversally, still keeping the property that we have an HT-drawing. Therefore, we may add to our conventions that this is the case for our HT-drawings on $\mathbb{R}P^2$.

Now, we consider a map $\lambda: E(G) \to \mathbb{Z}_2$. For an edge $e$, we let $\lambda(e)$ be the number of crossings of $e$ and the crosscap $\gamma$ modulo 2. We emphasize that $\lambda$ depends on the choice of
Figure 3: Transformations between HT-drawings on $\mathbb{R}P^2$ and projective HT-drawings on $S^2$. The crosscap. Afterwards, it will be useful to alter $\lambda$ via so-called vertex-crosscap switches, which we will explain a bit later.

Given a (graph-theoretic) cycle $Z$ in $G$, we can distinguish whether $Z$ is drawn as a homologically nontrivial cycle by checking the value $\lambda(Z) := \sum \lambda(e) \in \mathbb{Z}_2$ where the sum is over all edges of $Z$. The cycle $Z$ is homologically nontrivial if and only if $\lambda(Z) = 1$. In particular, it follows that $\lambda(Z)$ does not depend on the choice of the crosscap.

**Projective HT-drawings on $S^2$.** Let $D$ be an HT-drawing of a graph $G$ on $\mathbb{R}P^2$. It is not hard to deduce a drawing $D'$ of the same graph on $S^2$ such that every pair $(e,f)$ of independent edges satisfies $\text{cr}(e,f) = \lambda(e)\lambda(f)$. Indeed, it is sufficient to ‘undo’ the crosscap, glue back the disk $B$ and then let the edges intersect on $B$. See the two leftmost pictures in Fig. 3. This motivates the following definition.

**Definition 3.** Let $D$ be a drawing of a graph $G$ on $S^2$ and $\lambda: E(G) \to \mathbb{Z}_2$ be a function. Then the pair $(D,\lambda)$ is a projective HT-drawing of $G$ on $S^2$ if $\text{cr}(e,f) = \lambda(e)\lambda(f)$ for any pair of independent edges $e$ and $f$ of $G$. (If $\lambda$ is sufficiently clear from the context, we say that $D$ is a projective HT-drawing of $G$ on $S^2$.)

It turns out that a projective HT-drawing on $S^2$ can also be transformed to an HT-drawing on $\mathbb{R}P^2$.

**Lemma 4.** Let $(D,\lambda)$ be a projective HT-drawing of a graph $G$ on $S^2$. Then there is an HT-drawing $D_\circ$ of $G$ on $\mathbb{R}P^2$ such that $\text{cr}_{D_\circ}(e,f) = \text{cr}_D(e,f) + \lambda(e)\lambda(f)$ for any pair of distinct edges of $G$, possibly adjacent. In addition, if $e$ and $f$ are arbitrary two edges such that $\lambda(e) = \lambda(f) = 0$ and $D(e)$ and $D(f)$ are disjoint; then $D_\circ(e)$ and $D_\circ(f)$ are disjoint as well.

**Proof.** It is sufficient to consider a small disk $B$ which does not intersect $D(G)$, replace it with a crosscap and redraw the edges $e$ with $\lambda(e) = 1$ appropriately as described below. (Follow the two pictures on the right in Fig. 3.) From each edge $e$ with $\lambda(e) = 1$, we pull a thin ‘finger-move’ towards the crosscap which intersects every other edge in pairs of intersection points. Then we redraw the edge in a close neighbourhood of the crosscap as indicated in Fig. 4. After this redrawing, each edge $e$ such that $\lambda(e) = 1$ passes over the crosscap once and each edge $e$ with $\lambda(e) = 0$ does not pass over it. This agrees with our original definition of $\lambda$ for HT-drawings on $\mathbb{R}P^2$. In addition, we indeed obtain an HT-drawing on $\mathbb{R}P^2$ with $\text{cr}_{D_\circ}(e,f) = \text{cr}_D(e,f) + \lambda(e)\lambda(f)$, because in the last step we introduce one more crossing among pairs of edges $e, f$ such that $\lambda(e) = \lambda(f) = 1$. $\otimes$
In summary, Lemma 4 together with the previous discussion provide us with two viewpoints on the Hanani–Tutte drawings.

**Corollary 5.** A graph $G$ admits a projective HT-drawing on $S^2$ (with respect to some function $\lambda: E(G) \to \mathbb{Z}_4$) if and only if it admits an HT-drawing on $\mathbb{R}P^2$.

The main strength of Corollary 5 relies in the fact that in projective HT-drawings on $S^2$, we can ignore the actual geometric position of the crosscap and work in $S^2$ instead, which is simpler. This is especially helpful when we need to merge two drawings. On the other hand, it turns out that for our approach it will be easier to perform certain parity counts in the language of HT-drawings on $\mathbb{R}P^2$.

In order to distinguish the usual HT-drawings on $S^2$ from the projective HT-drawings, we will sometimes refer to the former as to the *ordinary* HT-drawings on $S^2$.

**Nontrivial walks.** Let $(D, \lambda)$ be a projective HT-drawing of a graph $G$ and $\omega$ be a walk in $G$. We define $\lambda(\omega) := \sum_{e \in E(\omega)} \lambda(e)$ where $E(\omega)$ is the multiset of edges appearing in $\omega$. Equivalently, it is sufficient to consider only the edges appearing an odd number of times in $\omega$, because $2\lambda(e) = 0$ for any edge $e$. We say that $\omega$ is *trivial* if $\lambda(\omega) = 0$ and *nontrivial* otherwise.

We often use this terminology in special cases when $\omega$ is an edge, a path, or a cycle. In particular, a cycle $Z$ is trivial if and only if it is drawn as a homologically trivial cycle in the corresponding drawing $D_\circ$ of $G$ on $\mathbb{R}P^2$ from Lemma 4.

Given two homologically nontrivial cycles on $\mathbb{R}P^2$, it is well known that they must cross an odd number of times (assuming they cross at every intersection). This fact is substantiated by Lemma 30 later on. However, we first present a weaker version of this statement in the setting of projective HT-drawings, which we need sooner.

**Lemma 6.** Let $(D, \lambda)$ be a projective HT-drawing of a graph $G$ on $S^2$. Then $G$ does not contain two vertex-disjoint nontrivial cycles.

**Proof.** For contradiction, let $Z_1$ and $Z_2$ be two vertex-disjoint nontrivial cycles in $G$. That is, $Z_1$ as well as $Z_2$ contains an odd number of nontrivial edges. Therefore, there is an odd number of pairs $(e_1, e_2)$ of nontrivial edges where $e_1 \in Z_1$ and $e_2 \in Z_2$. According to Definition 3, $Z_1$ and $Z_2$ must have an odd number of crossings. But this is impossible for two cycles in the plane which cross at every intersection (in $D$). \(\Box\)
**Vertex-edge and vertex-crosscap switches.** Let $D$ be a drawing of a graph $G$ on $S^2$. Let us consider a vertex $v$ and an edge $e$ of $G$ such that $v$ is not incident to $e$. We modify the drawing $D$ into drawing $D'$ so that we pull a thin finger from the interior of $e$ towards $v$ and we let this finger pass over $v$. We say that $D'$ is obtained from $D$ by the *vertex-edge switch* $(v,e)$.ootnote{Another name for the *vertex-edge switch* is the *finger-move* common mainly in topological context in higher dimensions.} If we have an edge $f$ incident to $v$, then the crossing number $cr(e,f)$ of this pair changes (from 0 to 1 or vice versa), but it does not change for any other pair, because the ‘finger’ intersects the other edges in pairs.

Now, let $(D,\lambda)$ be a projective HT-drawing of $G$ on $S^2$. It is very useful to alter $\lambda$ at the cost of redrawing $G$. Given a vertex $v$, we perform the vertex-edge switches $(v,e)$ for all edges $e$ not incident to $v$ such that $\lambda(e) = 1$ obtaining a drawing $D'$. We also introduce a new function $\lambda': E(G) \rightarrow \mathbb{Z}_2$ derived from $\lambda$ by switching the value of $\lambda$ on all edges of $G$ incident to $v$. In this case, we say that $D'$ (and $\lambda'$) is obtained by the *vertex-crosscap switch* over $v$.ootnote{In the case of drawings on $\mathbb{R}P^2$, a vertex-crosscap switch corresponds to passing the crosscap over $v$, which motivates our name. On the other hand, it is beyond our needs to describe this correspondence exactly.} It yields again an HT-drawing.

**Lemma 7.** Let $(D,\lambda)$ be a projective HT-drawing of $G$ on $S^2$. Let $D'$ and $\lambda'$ be obtained from $D$ and $\lambda$ by a vertex-crosscap switch. Then $(D',\lambda')$ is a projective HT-drawing of $G$ on $S^2$.

**Proof.** It is routine to check that $cr_{D'}(e,f) = \lambda'(e)\lambda'(f)$ for any pair of independent edges $e$ and $f$.

Indeed, let $v$ be the vertex inducing the switch. If neither $e$ nor $f$ is incident to $v$, then

$$cr_{D'}(e,f) = cr_D(e,f) = \lambda(e)\lambda(f) = \lambda'(e)\lambda'(f).$$

It remains to consider the case that one of the edges, say $e$, is incident to $v$. Note that $\lambda(e) = 1 - \lambda'(e)$ and $\lambda(f) = \lambda'(f)$ in this case.

If $\lambda(f) = 0$, then

$$cr_{D'}(e,f) = cr_D(e,f) = \lambda(e)\lambda(f) = 0 = \lambda'(e)\lambda'(f).$$

Finally, if $\lambda(f) = 1$, then

$$cr_{D'}(e,f) = 1 - cr_D(e,f) = 1 - \lambda(e)\lambda(f) = \lambda(f) - \lambda(e)\lambda(f) = \lambda'(e)\lambda'(f).$$

We also remark that a vertex-crosscap switch keeps the triviality or nontriviality of cycles. Indeed, let $Z$ be a cycle. If $Z$ avoids $v$, then $\lambda(Z) = \lambda'(Z)$ since $\lambda(e) = \lambda(e')$ for any edge $e$ of $Z$. If $Z$ contains $v$, then $\lambda(Z) = \lambda'(Z)$ as well since $\lambda(e) \neq \lambda'(e)$ for exactly two edges of $Z$.

**Planarization.** As usual, let $(D,\lambda)$ be a projective HT-drawing of $G$ on $S^2$. Now let us consider a subgraph $P$ of $G$ such that every cycle in $P$ is trivial. Then $P$ essentially behaves as a planar subgraph of $G$, which we make more precise by the following lemma.
Lemma 8. Let \((D, \lambda)\) be a projective HT-drawing of \(G\) on \(S^2\) and let \(P\) be a subgraph of \(G\) such that every cycle in \(P\) is trivial. Then there is a set \(U \subseteq V(P)\) with the following property.

Let \((D_U, \lambda_U)\) be obtained from \((D, \lambda)\) by the vertex-crosscap switches over all vertices of \(U\) (in any order). Then \((D_U, \lambda_U)\) is a projective HT-drawing of \(G\) on \(S^2\) and \(\lambda_U(e) = 0\) for any edge \(e\) of \(E(P)\).

Proof. The drawing \((D_U, \lambda_U)\) is a projective HT-drawing by Lemma 7. Let \(F\) be a spanning forest of \(P\), the union of spanning trees of each connected component of \(P\), rooted arbitrarily. We first make \(\lambda(e) = 0\) for each edge of \(F\), as follows: do a breadth-first search on each tree in \(F\); when an edge \(e \in F\) with \(\lambda(e) = 1\) is encountered, perform a vertex-crosscap switch on the vertex of \(e\) farther from the root of the tree. Let \(\lambda_U\) be the resulting map, which is zero on the edges of \(F\). Each edge \(e\) in \(E(P) \setminus E(F)\) belongs to a cycle \(Z\) such that \(Z - e \subseteq F\). Since \(\lambda_U(Z) = \lambda(Z) = 0\), we have \(\lambda_U(e) = 0\) as well.

3 Separation Theorem

In this section, we state the separation theorem announced in the introduction.

As it was explained in the introduction, a simple cycle \(Z\) such that every edge of \(Z\) is even (in a drawing) splits the graph into the outside and the inside. We first introduce a notation for this splitting.

Definition 9. Let \(G\) be a graph and \(D\) be a drawing of \(G\) on \(S^2\). Let us assume that \(Z\) is a cycle of \(G\) such that every edge of \(Z\) is even and it is drawn as a simple cycle in \(D\). Let \(S^+\) and \(S^-\) be the two components of \(S^2 \setminus D(Z)\). We call a vertex \(v \in V(G) \setminus V(Z)\) an inside vertex if it belongs to \(S^+\) and an outside vertex otherwise. Given an edge \(e = uv \in E(G) \setminus E(Z)\), we say that \(e\) is an inside edge if either \(u\) is an inside vertex or if \(u \in V(Z)\) and \(D(e)\) points locally to \(S^+\) next to \(D(u)\). Analogously we define an outside edge.\(^6\) We let \(V^+\) and \(E^+\) be the sets of the inside vertices and the inside edges, respectively. Analogously, we define \(V^-\) and \(E^-\). We also define the graphs \(G^{+0} := (V^+ \cup V(Z), E^+ \cup E(Z))\) and \(G^{-0} := (V^- \cup V(Z), E^- \cup E(Z))\).

Now, we may formulate our main technical tool—the separation theorem for projective HT-drawings.

Theorem 10. Let \((D, \lambda)\) be a projective HT-drawing of a 2-connected graph \(G\) on \(S^2\) and \(Z\) a cycle of \(G\) that is simple in \(D\) and such that every edge of \(Z\) is even. Moreover, we assume that every edge \(e\) of \(Z\) is trivial, that is, \(\lambda(e) = 0\). Then there is a projective HT-drawing \((D', \lambda')\) of \(G\) on \(S^2\) satisfying the following properties.

- The drawings \(D\) and \(D'\) coincide on \(Z\);
- the cycle \(Z\) is completely free of crossings and all of its edges are trivial in \(D'\);
- \(D'(G^{+0})\) is contained in \(S^+ \cup D'(Z)\);
- \(D'(G^{-0})\) is contained in \(S^- \cup D'(Z)\); and
- either all edges of \(G^{+0}\) or all edges of \(G^{-0}\) are trivial (according to \(\lambda'\)); that is, at least one of the drawings \(D'(G^{+0})\) or \(D'(G^{-0})\) is an ordinary HT-drawing on \(S^2\).

\(^6\)It turns out that every edge \(e \in E(G) \setminus E(Z)\) is either an outside edge or an inside edge, because every edge of \(Z\) is even.
The assumption that $G$ is 2-connected is not essential for the proof of Theorem 10, but it will slightly simplify some of the steps. (For our application, it will be sufficient to prove the 2-connected case.)

In the remainder of this section, we describe the main ingredients of the proof of Theorem 10 and we also derive this theorem from the ingredients. We will often encounter the setting when $G$, $(D, \lambda)$ and $Z$ satisfy the assumptions of Theorem 10. Therefore, we say that $G$, $(D, \lambda)$ and $Z$ satisfy the separation assumptions if (1) $G$ is a 2-connected graph; (2) $(D, \lambda)$ is a projective HT-drawing of $G$; (3) $Z$ is a cycle in $G$ drawn as a simple cycle in $D$; (4) every edge of $Z$ is even in $D$ and trivial.

**Arrow graph.** From now on, let us fix $G$, $(D, \lambda)$ and $Z$ satisfying the separation assumptions. This also fixes the distinction between the outside and the inside.

**Definition 11.** A bridge $B$ of $G$ (with respect to $Z$) is a subgraph of $G$ that is either an edge not in $Z$ but with both endpoints in $Z$ (and its endpoints also belong to $B$), or a connected component of $G - V(Z)$ together with all edges (and their endpoints in $Z$) with one endpoint in that component and the other endpoint in $Z$. (This is a standard definition; see, e.g., Mohar and Thomassen [MT01, p. 7].)

We say that $B$ is an inside bridge if it is a subgraph of $G^+0$, and an outside bridge if it is a subgraph of $G^-0$ (every bridge is thus either an inside bridge or an outside bridge).

A walk $\omega$ in $G$ is a proper walk if no vertex in $\omega$ belongs to $V(Z)$, except possibly its endpoints, and no edge of $\omega$ belongs to $E(Z)$. In particular, each proper walk belongs to a single bridge.

Since we assume that $G$ is 2-connected, every inside bridge contains at least two vertices of $Z$. The bridges induce partitions of $E(G) \setminus E(Z)$ and of $V(G) \setminus V(Z)$. See Fig. 5.

We want to record which pairs of vertices on $V(Z)$ are connected with a nontrivial and proper walk inside or outside.\footnote{We recall that nontrivial walks are defined in Sect. 2, a bit below Corollary 5.} For this purpose, we create two new graphs $A^+$ and $A^-$, possibly with loops but without multiple edges. In order to distinguish these graphs from $G$, we draw their edges with double arrows and we call these graphs an inside arrow graph and an outside arrow graph, respectively. The edges of these graphs are called the inside/outside arrows. We set $V(A^+) = V(A^-) = V(Z)$.

Now we describe the arrows, that is, $E(A^+)$ and $E(A^-)$. Let $u$ and $v$ be two vertices of $V(Z)$, not necessarily distinct. By $W^+_{uv}$ we denote the set of all proper nontrivial walks in $G^+0$.
with endpoints $u$ and $v$. We have an inside arrow connecting $u$ and $v$ in $E(A^+)$ if and only if $W^+_uv$ is nonempty. In order to distinguish the edges of $G$ from the arrows, we denote an arrow by $\overrightarrow{uv} = \overleftarrow{vu}$. An arrow which is a loop at a vertex $v$ is denoted by $\overrightarrow{vv}$. (This convention will allow us to work with arrows $\overrightarrow{uv}$ without a distinction whether $u = v$ or $u \neq v$.) Analogously, we define the set $W^-uv$ and the outside arrows.

See Fig. 6 for the arrow graph(s) of the drawing of $K_5$ depicted in Fig. 2, left.

It follows from the definition of the inside bridges that any walk $\omega \in W^+uv$ stays in one inside bridge. Given an inside bridge $B$, we let $W^+_{uv,B}$ be the set of all walks $w \in W^+uv$ which belong to $B$. In particular, $W^+_uv$ decomposes into the disjoint union of the sets $W^+_{uv,B_1}, \ldots, W^+_{uv,B_k}$ where $B_1, \ldots, B_k$ are all inside bridges. Given an inside arrow $\overrightarrow{uv}$ and an inside bridge $B$, we say that $B$ induces $\overrightarrow{uv}$ if $W^+_{uv,B}$ is nonempty. An inside bridge $B$ is nontrivial if it induces at least one arrow. Given two inside arrows $\overrightarrow{uv}$ and $\overrightarrow{xy}$ we say that $\overrightarrow{uv}$ and $\overrightarrow{xy}$ are induced by different bridges if there are two different inside bridges $B$ and $B'$ such that $B$ induces $\overrightarrow{uv}$ and $B'$ induces $\overrightarrow{xy}$. As usual, we define analogous notions for the outside as well. Note that it may happen that there is an inside bridge inducing both $\overrightarrow{uv}$ and $\overrightarrow{xy}$ even if $\overrightarrow{uv}$ and $\overrightarrow{xy}$ are induced by different bridges.

Possible configurations of arrows. We plan to utilize the arrow graph in the following way. On one hand, we will show that certain configurations of arrows are not possible; see Fig. 7. On the other hand, we will show that, since the arrow graph does not contain any of the forbidden configurations, it must contain one of the configurations in Fig. 8 inside or outside. (These configurations are precisely defined in Definition 15.) We will also show that the configurations in Fig. 8 are redrawable, that is, they may be appropriately redrawn without the crosscap. The precise statement for redrawings is given by Proposition 17 below.

More concretely, we prove the following three lemmas forbidding the configurations of arrows from Fig. 7. We emphasize that in all three lemmas we assume that the notions used there correspond to a fixed $G$, $(D, \lambda)$ and $Z$ satisfying the separation assumptions.

Lemma 12. Every inside arrow shares a vertex with every outside arrow.

Lemma 13. Let $\overrightarrow{ab}$ and $\overrightarrow{xy}$ be two arrows induced by different inside bridges of $G^{+0}$. If the two arrows do not share an endpoint, their endpoints have to interleave along $Z$.

Lemma 14. There are no three vertices $a$, $b$, $c$ on $Z$, an inside bridge $B^+$, and an outside bridge $B^-$ such that $B^+$ induces the arrows $\overrightarrow{ab}$ and $\overrightarrow{ac}$ (and no other arrows) and $B^-$ induces the arrows $\overrightarrow{ab}$ and $\overrightarrow{bc}$ (and no other arrows).
Figure 7: Forbidden configurations of arrows. The cyclic order in (a) may be arbitrary whereas it is important in (b) that the arrows there do not interleave. Different dashing of lines in (b) correspond to arrows induced by different inside bridges. The arrows of the same colour in (c) are induced by the same bridge.

We prove these three lemmas in Sect. 6. By symmetry, Lemmas 13 and 14 are also valid if we swap the inside and the outside (Lemma 12 as well, but here already the statement of the lemma is symmetric).

Now we describe the redrawable configurations.

**Definition 15.** We say that $G$ forms

(a) an **inside fan** if there is a vertex common to all inside arrows. (The arrows may come from various inside bridges.)

(b) an **inside square** if it contains four vertices $a$, $b$, $c$ and $d$ ordered in this cyclic order along $Z$ and the inside arrows are precisely $\overline{ab}$, $\overline{bc}$, $\overline{cd}$ and $\overline{ad}$. In addition, we require that the inside graph $G^{+0}$ has only one nontrivial inside bridge.

(c) an **inside split triangle** if there exist three vertices $a$, $b$ and $c$ such that the arrows of $G$ are $\overline{ab}$, $\overline{ac}$ and $\overline{bc}$. In addition, we require that every nontrivial inside bridge induces either the two arrows $\overline{ab}$ and $\overline{ac}$, or just a single arrow.

See Fig. 8. We have analogous definitions for an **outside fan**, **outside square** and **outside split triangle**.

More precisely the notions in Definition 15 depend on $G$, $(D, \lambda)$ and $Z$ satisfying the separation assumptions.

A relatively direct case analysis, using Lemmas 12, 13 and 14, reveals the following fact.

**Proposition 16.** Let $(D, \lambda)$ be a projective HT-drawing on $S^2$ of a graph $G$ and let $Z$ be a cycle in $G$ satisfying the separation assumptions. Then $G$ forms an (inside or outside) fan, square, or split triangle.

On the other hand, any configuration from Definition 15 can be redrawn without using the crosscap:

**Proposition 17.** Let $(D, \lambda)$ be a projective HT-drawing of $G^{+0}$ on $S^2$ and $Z$ be a cycle satisfying the separation assumptions. Moreover, let us assume that $D(G^{+0}) \cap S^- = \emptyset$ (that is, $G^{+0}$ is fully drawn on $S^+ \cup D(Z)$). Let us also assume that $G^{+0}$ forms an inside fan, an inside square or an inside split triangle. Then there is an ordinary HT-drawing $D'$ of $G^{+0}$ on $S^2$ such that $D$ coincides with $D'$ on $Z$ and $D'(G^{+0}) \cap S^- = \emptyset$. 
Figure 8: Schematic drawings of the redrawable configurations of arrows from Definition 15. Different dashing of lines correspond to different inside bridges. The loop in the right drawing (a) is an inside loop (drawn outside due to lack of space). The drawing (c) is only one instance of an inside split triangle.

Proposition 16 is proved in Sect. 5 (assuming there the validity of Lemmas 12, 13 and 14). Proposition 17 is proved in Sect. 7.

Now we are missing only one tool to finish the proof of Theorem 10. This tool is the “redrawing procedure” of Pelsmajer, Schaefer and Štefankovič [PSŠ07a]. More concretely, we need the following variant of Theorem 2. (Note that the theorem below is not in the setting of projective HT-drawings. However, the notions used in the statement are still well defined according to Definition 9.)

**Theorem 18.** Let $D$ be a drawing of a graph $G$ on the sphere $S^2$. Let $Z$ be a cycle in $G$ such that every edge of $Z$ is even and $Z$ is drawn as a simple cycle. Then there is a drawing $D''$ of $G$ such that

- $D''$ coincides with $D$ on $Z$;
- $D''(G^{+0})$ belongs to $S^+ \cup D(Z)$ and $D''(G^{-0})$ belongs to $S^- \cup D(Z)$;
- whenever $(e,f)$ is a pair of edges such that both $e$ and $f$ are inside edges or both $e$ and $f$ are outside edges, then $cr_{D''}(e,f) = cr_D(e,f)$.

It is easy to check that the proof of Theorem 2 in [PSŠ07a] proves Theorem 18 as well. Additionally, we note that an alternative proof of Theorem 2 in [FPSŠ12, Lemma 3] can also be extended to yield Theorem 18. Nevertheless, for completeness, we provide its proof in Sect. 8.

Finally, we prove Theorem 10, assuming the validity of the aforementioned auxiliary results.

**Proof of Theorem 10.** Let $G$ be the graph, $(D,\lambda)$ be the drawing and $Z$ be the cycle from the statement.

We use Theorem 18 to $G$ and $D$ to obtain a drawing $D''$ keeping in mind that all edges of $Z$ are even. See Fig. 9; follow this picture also in the next steps of the proof. We get that $Z$ is drawn on $D''$ as a simple cycle free of crossings. We also get that $D''(G^{+0})$ is contained in $S^+ \cup D''(Z)$ and $D''(G^{-0})$ is contained in $S^- \cup D''(Z)$. However, there may be no $\lambda''$ such that $(D'',\lambda'')$ is a projective HT-drawing; we still may need to modify it to obtain such a drawing.

By Proposition 16, $G$ forms one of the redrawable configurations on one of the sides; that is, an inside/outside fan, square or split triangle. Without loss of generality, it appears inside.
It means that $D''$ restricted to $G^+$ satisfies the assumptions of Proposition 17. Therefore, there is an ordinary HT-drawing $D^+$ of $G^+$ satisfying the conclusions of Proposition 17. Finally, we let $D'$ be the drawing of $G$ on $S^2$ which coincides with $D^+$ on $G^+$ and with $D''$ on $G^{-}$. Both $D''$ and $D^+$ coincide with $D$ on $Z$; therefore, $D'$ is well defined. We set $\lambda'$ so that $\lambda'(e) := \lambda(e)$ for an edge $e \in E^-$ and $\lambda'(e) := 0$ for any other edge. Now, we can easily verify that $(D', \lambda')$ is the required projective HT-drawing.

4 Proof of the Strong Hanani–Tutte Theorem on $\mathbb{R}P^2$

In this section, we prove Theorem 1 assuming validity of Theorem 10 as well as few other auxiliary results from the previous section, which will be proved only in the later sections.

Given a graph $G$ that admits an HT-drawing on the projective plane, we need to show that $G$ is actually projective-planar. By Corollary 5, we may assume that $G$ admits a projective HT-drawing $(D, \lambda)$ on $S^2$. We head for using Theorem 10. For this, we need that $G$ is 2-connected and contains a suitable trivial cycle $Z$ that may be redrawn so that it satisfies the assumptions of Theorem 10. Therefore, we start with auxiliary claims that will bring us to this setting. Many of them are similar to auxiliary steps in [PS07] (sometimes they are almost identical, adapted to a new setting).

Before we state the next lemma, we recall the well known fact that any graph admits a (unique) decomposition into blocks of 2-connectivity [Die10, Ch. 3]. Here, we also allow the case that $G$ is disconnected. Each block in this decomposition is either a vertex (this happens only if it is an isolated vertex of $G$), an edge or a 2-connected graph with at least three vertices. The intersection of two blocks is either empty or it contains a single vertex (which is a cut in the graph). The blocks of the decomposition cover all vertices and edges (a vertex may occur in several blocks whereas any edge belongs to a unique block).

Lemma 19. If $G$ admits a projective HT-drawing on $S^2$, then at most one block of 2-connectivity in $G$ is non-planar. Moreover, if all blocks are planar, $G$ is planar as well.
We note that in [SˇS13] it was proved that a minimal counterexample to the strong Hanani–Tutte theorem on any surface is vertex 2-connected. However, for the projective plane the same result can be obtained by much simpler means; therefore, we include its proof here.

**Proof.** First, for contradiction, let us assume that \( G \) contains two distinct non-planar blocks \( B_1 \) and \( B_2 \). If \( B_1 \) and \( B_2 \) are disjoint, then Lemma 6 implies that at least one of these blocks, say \( B_2 \), does not contain any non-trivial cycle. However, it means that \( B_2 \) admits an ordinary HT-drawing on \( S^2 \) by Lemma 8. Therefore, \( B_2 \) is planar by the strong Hanani–Tutte theorem in the plane [Cho34, Tut70, PSˇS07a]. This contradicts our original assumption. It remains to consider the case when \( B_1 \) and \( B_2 \) share a vertex \( v \) (it must be a cut vertex). Let us set \( H := B_1 \cup B_2 \). Let \( P \) be a spanning tree of \( H \) with just two edges \( e_1, e_2 \) incident to \( v \) and such that \( e_1 \in B_1 \) and \( e_2 \in B_2 \). Note that such a tree always exists, because \( B_1 \) and \( B_2 \) are connected after removing \( v \). By Lemma 8 we may assume that all the edges of \( P \) are trivial (after a possible alteration of \( \lambda \)).

Any nontrivial edge \( e \) from \( E(H) \setminus E(P) \) creates a nontrivial cycle in the corresponding block. If \( e \) is not incident to \( v \), then the cycle avoids \( v \) by the choice of \( P \). Using Lemma 6 again, we see that at least one of the blocks, say \( B_2 \), satisfies that all its nontrivial edges are incident with \( v \). This already implies that \( B_2 \) is a planar graph, because \( D \) is an HT-drawing of \( B_2 \) on \( S^2 \) (there are no pairs of nontrivial independent edges in \( G \)). This is again a contradiction.

The last item in the statement of this lemma is a well known property of planar graphs. It is sufficient to observe that a disjoint union of two planar graphs is a planar graph, and moreover, that if a graph \( G \) contains a cut vertex \( v \) and all the components after cutting (and reattaching \( v \)) are planar, then \( G \) is planar as well.

**Observation 20.** Let \((D, \lambda)\) be a drawing of a 2-connected graph. If \( D \) does not contain any trivial cycle, then \( G \) is planar.

**Proof.** As \( G \) is 2-connected, it is either a cycle or it contains three disjoint paths sharing their endpoints. A cycle is a planar graph as we need. In the latter case, two of the paths are both trivial or both nontrivial. Together, they induce a trivial cycle, therefore this case cannot occur.

**Lemma 21.** Let \((D, \lambda)\) be a projective HT-drawing on \( S^2 \) of a graph \( G \) and let \( Z \) be a cycle in \( G \). Then \( G \) can be redrawn only by local changes next to the vertices of \( Z \) to a projective HT-drawing \( D' \) on \( S^2 \) so that \( \lambda \) remains unchanged and \( cr_D(e,f) = \lambda(e)\lambda(f) \), for any pair \((e,f) \in E(Z) \times E(G)\) of distinct (not necessarily independent) edges. In particular, if \( \lambda(e) = 0 \) for every edge \( e \) of \( Z \), then every edge of \( Z \) becomes even in \( D' \).

**Proof.** Since we have a projective HT-drawing, \( cr_D(e,f) = \lambda(e)\lambda(f) \) for every pair of independent edges. To prove the claim it remains to show that local changes allow to change the parity of \( cr_D(e,f) \) whenever \( e \) is an edge of \( Z \) and \( e \) and \( f \) share a vertex.

This can be done in two steps. First we use local move c) from Fig. 10 to obtain the desired parity of \( cr_D(e,f) \), for all pairs of consecutive edges \((e,f)\) on \( Z \). This move may change the parity of crossings between edges on \( Z \) and dependent edges not on \( Z \).

Next we use local moves a) and b) from Fig. 10 to obtain the desired parity of crossings between edges on \( Z \) and dependent edges not on \( Z \). If \( v \) is the vertex common to \( h \), \( e \) and \( f \), where \( e \) and \( f \) are edges on \( Z \), move a) is used when we need to change the parity of \( cr_D(e,h) \).
and its symmetric version to change the parity of $cr_D(f, h)$. Move b) is used when we need to change the parity for both $cr_D(e, h)$ and $cr_D(f, h)$. Since these moves do not change the parity of $cr_D(e, h')$ or $cr_D(f, h')$ for any other edge $h'$, the claim follows.

Once we know that the edges of a cycle can be made even we also need to know that such a cycle can be made simple.

**Lemma 22.** Let $(D, \lambda)$ be a projective HT-drawing on $S^2$ of a graph $G$ and let $Z$ be a cycle in $G$ such that each of its edges is even. Then $G$ can be redrawn so that $Z$ becomes a simple cycle, its edges remain even and the resulting drawing is still a projective HT-drawing (with $\lambda$ unchanged).

**Proof.** First, we want to get a drawing such that there is only one edge of $Z$ which may be intersected by other edges. Let us consider three consecutive vertices $u, v$ and $w$ on $Z$, with $v \not\in \{u, w\}$. We almost-contract $uv$ so that we move the vertex $v$ towards $u$ until we remove all intersections between $uv$ and other edges. Note that the image of the cycle $Z$ is not changed; we only slide $v$ towards $u$ along $Z$. This way, $uw$ is now free of crossings and these crossings appear on $vw$. See the two leftmost pictures in Fig. 11. (The right picture will be used in the proof of Theorem 18.)

Since $uw$ as well as $vw$ were even edges in the initial drawing, $vw$ remains even after the redrawing. If $uv$ and $vw$ intersected, then this step introduces self-intersections of $vw$.

After performing such redrawing repeatedly, we get there is only one edge of $Z$ which may be intersected by other edges, as required. We remove self-crossings of this edge, as described in Sect. 2, and we are done.

Apart from lemmas tailored to set up the separation assumptions, we also need one more lemma that will be useful in the inductive proof of Theorem 1.

**Lemma 23.** Let $(D, \lambda)$ be a Hanani–Tutte drawing of $G$ and let $Z$ be a cycle satisfying the separation assumptions. Let $B$ be an inside bridge such that any path with both endpoints on $V(B) \cap V(Z)$ is nontrivial. Then $|V(B) \cap V(Z)| = 2$ and $B$ induces a single arrow and no loop.
Proof. First, we show that there is no nontrivial cycle in $B$. For contradiction, there is a nontrivial cycle $N$ in $B$. By the 2-connectivity of $G$ there exist two vertex disjoint paths $p_1$ and $p_2$ (possibly of length zero) that connect $Z$ to $N$. We consider the shortest such paths; thus, each of the paths shares only one vertex with $Z$ and one vertex with $N$. Let $y_1$ and $y_2$ be the endpoints of $p_1$ and $p_2$ on $N$, respectively. Let $p_3$, $p_4$ be the arcs of $N$ between $y_1$ and $y_2$. We consider two paths $q_1$ and $q_2$ where $q_1$ is obtained from the concatenation of $p_1$, $p_3$ and $p_2$, while $q_2$ is obtained from the concatenation of $p_1$, $p_4$ and $p_2$. Since $N$ is non-trivial, one of these paths is trivial, which provides the required contradiction.

Next, we observe that $B$ does not induce any loop in the inside arrow graph. For contradiction, it induces a loop at a vertex $x$ of $Z$. This means that there is a proper nontrivial walk $\kappa$ in $B$ with both endpoints $x$. We set up $\kappa$ so that it is the shortest such walk. We already know that $\kappa$ cannot be a cycle, thus it contains a closed nonempty subwalk $\kappa'$ and we set up $\kappa'$ so that it is the shortest such subwalk. Therefore, it must be a cycle; by the previous part of this proof, it is trivial. However, it means that $\kappa$ can be shortened by leaving out $\kappa'$, which is the required contradiction.

Now, we show that $|V(B) \cap V(Z)| = 2$. By the 2-connectedness of $G$, we have that $|V(B) \cap V(Z)| \geq 2$. Thus, for contradiction, let $a$, $b$, $c$ be three distinct vertices of $V(B) \cap V(Z)$. Let $v$ be one of the inner vertices of $B$ (there must be such a vertex since $B$ cannot be a single edge in this case). By the definition of inside/outside bridges, there exist proper walks $p_a$, $p_b$ and $p_c$ connecting $v$ to $a$, $b$ and $c$, respectively. By the pigeonhole principle, two of the walks have the same value of $\lambda$; without loss of generality, let them be $p_a$ and $p_b$. It follows that the proper walk obtained from the concatenation of $p_a$ and $p_b$ is trivial. Since $B$ does not contain any non-trivial cycle, this walk can be shortened to a trivial proper path between $a$ and $b$ by an analogous argument as in the previous paragraph. A contradiction.

Finally, we know that there are two vertices in $V(B) \cap V(Z)$. Let $x$ and $y$ be these two vertices. Since any path connecting $x$ and $y$ is nontrivial, $B$ induces the arrow $xy$ in $A^+$. No other arrow in $A^+$ is possible since there are no loops.

Proposition 24 below is our main tool for deriving Theorem 1 from Theorem 10. It is set up in such a way that it can be inductively proved from Theorem 10. Then it implies Theorem 1, using the auxiliary lemmas from the beginning of this section, relatively easily.

**Proposition 24.** Let $(D, \lambda)$ be a projective HT-drawing of a 2-connected graph $G$ on $S^2$ and $Z$ a cycle in $G$ that is completely free of crossings in $D$ and such that each of its edges is trivial in $D$. Assume that $(V^+, E^+)$ or $(V^-, E^-)$ is empty. Then $G$ can be embedded into $\mathbb{R}P^2$ so that $Z$ bounds a face of the resulting embedding homeomorphic to a disk. If, in addition, $D$ is an ordinary HT-drawing on $S^2$, then $G$ can be embedded into $S^2$ so that $Z$ bounds a face of the resulting embedding (this face is again homeomorphic to a disk—there is in fact no other option on $S^2$).\(^8\)

**Proof.** The proof proceeds by induction on the number of edges of $G$. The base case is when $G$ is a cycle.

Without loss of generality, we assume that $(V^-, E^-)$ is empty. That is, $G = G^{+0}$. If $(V^+, E^+)$ is also empty, $G$ consists only of $Z$ and such a graph can easily be embedded into the plane or projective plane as required. Therefore, we assume that $(V^+, E^+)$ is nonempty.

We find a path $\gamma$ in $(V(G^{+0}), E(G^{+0}) \setminus E(Z))$ connecting two points $x$ and $y$ lying on $Z$. We may choose $x$, $y$ so that $x \neq y$ since $G$ is 2-connected.

\(^8\)We need to consider the case of ordinary HT-drawings in this proposition for a well working induction.
Case 1: There exists a trivial $\gamma$. First we solve the case that at least one such path $\gamma$ is trivial. We show that all edges of $\gamma$ can be made even and simple in the drawing while preserving simplicity of $Z$, the fact that $Z$ is free of crossings and the projective Hanani–Tutte condition on the whole drawing of $G^{+0}$.

As the first step, we use Lemma 8 in order to achieve that $\lambda(e) = 0$ for any edge $e$ of $Z$ and $\gamma$ simultaneously. By inspecting the proof of Lemma 8 we see that we can achieve this by vertex-crosscap switches only over the inner vertices of $\gamma$ (for this, we set up the root in the proof to be one of the endpoints of $\gamma$). In particular we can perform these vertex-crosscap switches inside $Z$ without affecting $Z$.

Now, we want to make the edges of $\gamma$ even, again without affecting $Z$. First, for any pair $(e, f)$ of adjacent edges of $\gamma$ which intersect oddly, we locally perform the move c) from Fig. 10 similarly as in Lemma 21. Next, we consider any edge $e \notin E(\gamma)$ adjacent to a vertex $u \in V(\gamma) \setminus V(Z)$. For such an edge we eventually perform one of the moves a) or b) from Fig. 10 so that we achieve that $e$ intersects evenly each of the two edges of $\gamma$ incident with $u$. Finally, we consider any edge $e \notin E(\gamma) \cup E(Z)$ adjacent to $u \in \{x, y\}$, one of the endpoints of $\gamma$ on $Z$. Let $f$ be the edge of $\gamma$ incident with $u$. If $e$ and $f$ intersect oddly, we perform the move from Fig. 12. This is possible since $Z$ is free of crossings. This way we achieve that every edge of $\gamma$ is even.

As the last step of the redrawing of $\gamma$, we want to make $\gamma$ simple (again without affecting $Z$). This can be done in the same way as in Lemma 22. We almost-contract all edges of $\gamma$ but one so that there is only one edge of $\gamma$ that intersects with other edges. Then we remove eventual self-intersections.

The rest of the argument is easier to explain if we switch inside and outside (this is easily doable by a homeomorphism of $S^2$) and treat drawings on $S^2$ as drawings in the plane.

We may assume that after the homeomorphism $Z$ is drawn in the plane as a circle with the inner region empty and with $x$ and $y$ antipodal. The vertices $x$ and $y$ split $Z$ into two paths; we denote by $p_1$ the ‘upper’ one and by $p_2$ the ‘lower’ one. We may also assume that $\gamma$ is ‘above’ $p_1$ by eventually adapting the initial choice of the correspondence between $S^2$ and the plane.

Now we continuously deform the plane so that $Z$ becomes flatter and flatter until it coincides with the line segment connecting $x$ to $y$, as depicted in Fig. 13 a). We may further require that no inner vertex of $p_1$ was identified with any inner vertex of $p_2$.

This way, we get a projective HT-drawing $(\bar{D}, \bar{\lambda})$ of a new graph $\bar{G}$: all the vertices of $G$ remain present in $\bar{G}$, that is, $V(\bar{G}) = V(G)$. Also the edges of $G$ which are not on $Z$ are present in $\bar{G}$. Only some of the edges of $Z$ may disappear and they are replaced with edges forming a path $\bar{p}$ between $x$ and $y$. Note that we did not introduce any multiple edges, because there is no edge in $G$ connecting an inner vertex of $p_1$ with an inner vertex of $p_2$. It also turns out that $\bar{G}$ has one edge less than $G$. Regarding $\bar{\lambda}$, we have $\lambda(e) = \bar{\lambda}(e)$ if $e$ is an
Figure 13: The deformation of the plane that changes $G$ into $\bar{G}$, the redrawing of $\bar{G}$ and the resulting embeddings of $\bar{G}$ and $G$.

edge of $E(G) \setminus E(Z)$ and we have $\bar{\lambda}(e) = 0$ if $e$ belongs to $p$.

Now consider the cycle $\bar{Z}$ in $\bar{G}$ formed by $\gamma$ and $p$. It is trivial and simple. In particular, we distinguish the inside and the outside according to Definition 9. For example, $\bar{G}^{+0}$ corresponds to the part of $G$ in between $\gamma$ and $p_1$ before the flattening; see Fig. 13 a) and b).

Now, we apply Theorem 10 and we get a drawing $D'$ of $\bar{G}$. When we look at the two sides of $\bar{G}$ separately, we get that the drawing of one of the sides, say the drawing of $\bar{G}^{+0}$, is a projective HT-drawing, while there is an ordinary HT-drawing on $S^2$ on the other side. If, in addition, $D$ were already an ordinary HT-drawing, we get an ordinary HT-drawing on both sides by Theorem 18.

Note also that since $G$ was 2-connected, both parts of $\bar{G}$ are 2-connected as well. Subsequently, we examine each of these two parts separately and use the inductive hypothesis; we obtain an embedding of $\bar{G}^{+0}$ into $\mathbb{R}P^2$ such that $\bar{Z}$ bounds a face homeomorphic to a disk as well as an embedding of $\bar{G}^{-0}$ into $S^2$ such that $\bar{Z}$ bounds a face homeomorphic to a disk. If, in addition, $D$ were already an ordinary HT-drawing, we get also the required embedding of $\bar{G}^{+0}$ into $S^2$. We merge these two embeddings along $\bar{Z}$ obtaining an embedding of $\bar{G}$ into $\mathbb{R}P^2$ (or $S^2$ if $D$ were an ordinary HT-drawing). See Fig. 13 c) and d).

Finally, we need to undo the identification of $p_1$ and $p_2$ into $p$. Whenever we consider a vertex $v$ on $p$ different from $x$ and $y$, it is uniquely determined whether it comes from $p_1$ or $p_2$. In addition, if $v$ comes from $p_1$, then any edge $e \in E(G) \setminus E(Z)$ incident with $v$ must belong to $\bar{G}^{+0}$. Similarly, if $v$ comes from $p_1$, then any edge $e \in E(G) \setminus E(Z)$ incident with $v$ must belong to $\bar{G}^{-0}$. Therefore, it is possible to undo the identification and we get the required embedding of $G$. See Fig. 13 e).
Case 2: All choices of \( \gamma \) are nontrivial. Now we deal with the situation when all possible choices of \( \gamma \) are nontrivial. We will first analyse which situations allow such configuration. Later we will show how to draw each of these situations.

Let us consider the inside arrow graph \( A^+ \). Since all choices of \( \gamma \) are nontrivial, Lemma 23 shows that every inside bridge induces a single inside arrow. This allows us to redraw inside bridges separately as is provided by the following claim.

Claim 24.1. For any inside bridge \( B \) there exists a planar drawing of \( Z \cup B \) in which \( Z \) is the outer face.

Proof. Since we know that \( B \) induces only a single arrow, we get that \( Z \cup B \) forms an inside fan, according to Definition 15. It follows from Proposition 17 that \( Z \cup B \) admits an ordinary HT-drawing such that \( Z \) is an outer cycle. However, the setting of ordinary HT-drawings is already fully resolved in Case 1. That is, we may already use Proposition 24 for this drawing and we get the required conclusion.

We consider the graph \( A^{+0} \) obtained from \( A^+ \) by adding the edges of \( Z \) to it, where \( A^+ \) is the inside arrow graph. (Note that \( V(A^+) = V(Z) \) according to our definition of the arrow graph.)

Our main aim will be to find an embedding of \( A^{+0} \) to \( \mathbb{R}P^2 \) such that \( Z \) bounds a face. As soon as we reach this task, then we can replace an embedding of each arrow by the embedding of inside bridges inducing this arrow via Claim 24.1 in a close neighbourhood of the arrow. If there are, possibly, more inside bridges inducing the arrow, then they are embedded in parallel.

Finally, we show that it is possible to embed \( A^{+0} \) in the required way. By Lemma 13, any two disjoint arrows interleave.

Let us consider two concentric closed disks \( E_1 \) and \( E_2 \) such that \( E_1 \) belongs to the interior of \( E_2 \). Let us draw \( Z \) to the boundary of \( E_1 \). Let \( a \) be the number of arrows of \( A^+ \) and let us consider \( 2a \) points on the boundary of \( E_1 \) making the vertices of regular \( 2a \)-gon. These points will marked by ordered pairs \((x, y)\) where \( xy \) is an inside arrow. We mark the points so that the cyclic order of the points respect the cyclic order as on \( Z \) in the first coordinate (in particular pairs with the same first coordinate are consecutive). However, for a fixed \( x \), the pairs \((x, y_1), \ldots, (x, y_k)\) corresponding to all arrows emanating from \( x \) are ordered in the reverted order when compared with the order of \( y_1, \ldots, y_k \) on \( Z \). See Fig. 14.

We show that it follows that the points marked \((x, y)\) and \((y, x)\) are directly opposite on \( E_1 \) for every inside arrow \( xy \). For contradiction, let us assume that \((x, y)\) and \((y, x)\) are not directly opposite for some \( xy \). Then there is another arrow \( uv \) such that \((x, y)\) and \((y, x)\) do not interleave with \((u, v)\) and \((v, u)\). Indeed, such an arrow must exist because the arrows induce a matching on the points, and \((x, y)\) and \((y, x)\) do not split the points equally. However, if \( xy \) and \( uv \) do not share an endpoint, we get a contradiction with the fact that disjoint arrows interleave. If \( xy \) and \( uv \) share an endpoint, we get a contradiction that we have reverted the order on the second coordinate.

Now, we get the required drawing in the following way. For any arrow \( xy \) we connect \( x \) with the point \((x, y)\) and \( y \) with \((y, x)\). We can do all the connections simultaneously for all arrows without introducing any crossing since we have respected the cyclic order on the first coordinate. We remove the interior of \( E_1 \) and we identify the boundary. This way we introduce a crosscap. Finally, we glue another disk along its boundary to \( Z \) and we get the required drawing on \( \mathbb{R}P^2 \).
Finally, we prove Theorem 1.

Proof of Theorem 1. We prove the result by induction in the number of vertices of G. We can trivially assume that G has at least three vertices.

If G has at least two blocks of 2-connectivity, G can be written as G₁∪G₂, where G₁∩G₂ is a minimal cut of G and, therefore, has at most one vertex. By Lemma 19 we may assume that G₁ is planar and G₂ non-planar. By induction, there exists an embedding D₂ of G₂ into ℜP^2. So G₁ is planar, G₂ is embeddable into ℜP^2 and G₁∩G₂ has at most one vertex. From these two embeddings, we easily derive an embedding of G = G₁∪G₂ in ℜP^2.

We are left with the case when G is 2-connected. By Observation 20, we may assume that there is at least one trivial cycle Z in (D,λ). We can also make each of its edges trivial by Lemma 8 and even by Lemma 21. Then we make Z, in addition, simple using Lemma 22. Hence G, Z and the current projective HT-drawing satisfy the separation assumptions.

Then we use Z to redraw G as follows. At first, we apply Theorem 10 to get a projective HT-drawing (D',λ') that separates G⁺₀ and G⁻₀. We define D⁺ := D'(G⁺₀) and D⁻ := D'(G⁻₀)—without loss of generality, D⁻ is an ordinary HT-drawing on S², while D⁺ is a projective HT-drawing on S².

Finally, we apply Proposition 24 above to D⁺ and D⁻ separately. Thus, we get embeddings of G⁺₀ and G⁻₀—one of them in S², the other one in ℜP^2. In addition, Z bounds a face in both of them; hence, we can easily glue them to get an embedding of the whole graph G into ℜP^2.

5 Labellings of Inside/Outside Bridges and the Proof of Proposition 16

In this section, given an inside (or outside) bridge B, we first describe what are possible combinations of arrows induced by B. Then we use the obtained findings for a proof of Proposition 16, assuming validity of Lemmas 12, 13 and 14 which will be proved in Sect. 6.

Labelling the vertices of the inside/outside bridges. We start with the first step. As usual, we only describe the ‘inside’ case; the ‘outside’ case will be analogous. We introduce certain labellings of V(B) ∩ V(Z) which will help us to determine arrows.
Definition 25 (Labelling of $V(B) \cap V(Z)$). A valid labelling $L = L_B$ for $B$ is a mapping $L: V(B) \cap V(Z) \to \{(0), \{1\}, \{0,1\}\}$ obtained in the following way.

If $V(B) \setminus V(Z) \neq \emptyset$ we pick a reference vertex $v_B \in V(B) \setminus V(Z)$ for $L$. Then we fix a labelling parameter $\alpha_B \in \mathbb{Z}_2$ for $L$. Finally, for any $u \in V(B) \cap V(Z)$ and for any proper walk $\omega$ with endpoints $u$ and $v_B$, the vertex $u$ receives the label $\alpha_B + \lambda(\omega) \in \mathbb{Z}_2$. Note that $u$ may receive two labels after considering all such walks. On the other hand, each vertex of $V(B) \cap V(Z)$ obtains at least one label, which follows from the definition of bridges (Definition 11).

If $V(B) \subseteq V(Z)$, then $B$ comprises only of one edge $e = uv$ connecting two vertices of $V(Z)$. In such case, there are two valid labellings for $B$. We set $L(u) = \{\alpha_B\}$ and $L(v) = \{\lambda(e) + \alpha_B\}$ for a chosen labelling parameter $\alpha_B \in \mathbb{Z}_2$.

If the bridge $B$ is understood from the context we may write just $v$ instead of $v_B$ for the reference vertex and $\alpha$ instead of $\alpha_B$ for the labelling parameter. By alternating the choice of $\alpha$ in the definition we may swap all labels. This means that there are always at least two valid labellings for a given inside bridge. On the other hand, a different choice of the reference vertex either does not influence the resulting labelling, or has the same effect as swapping the value of the labelling parameter $\alpha$. In other words, there are always exactly two valid labellings of the given inside/outside bridge $B$ corresponding to two possible choices of the labelling parameter $\alpha$, as is explained below.

To see this, consider a vertex $u \in V(B) \setminus V(Z)$ different from $v = v_B$. By Definition 11, there is a proper $uv$-walk $\gamma$ in $B$ not using any vertex of $Z$. Now, for any $x \in V(B) \cap V(Z)$ and for any proper $xv$-walk $\omega_{vu}$ in $B$, the concatenation of the walks $\omega_{vu}$ and $\gamma$ is a proper $xu$-walk in $B$ of type $\lambda(\omega_{vu}) + \lambda(\gamma)$. Also, for any proper $xu$-walk $\omega_{ux}$ in $B$, the concatenation of the walks $\omega_{ux}$ and $\gamma$ is a proper $xv$-walk in $B$ of type $\lambda(\omega_{ux}) + \lambda(\gamma)$. As a result, choosing $u$ as the reference vertex with $\alpha + \lambda(\gamma)$ as the labelling parameter leads to the same labelling as the choice of $v$ as the reference vertex with the labelling parameter $\alpha$.

The idea presented above can be used to establish the following simple observation, which we later use several times in the proofs.

Observation 26. Let $B$ be an inside or an outside bridge containing at least one inside/outside vertex. Moreover, let $L$ be a valid labelling for $B$ and $v$ the reference vertex for $L$. Let $x, y \in V(B)$ and let $\omega$ be a proper $xy$-walk in $B$. Then there is a proper $xy$-walk $\omega'$ in $B$ containing the reference vertex $v$ such that $\lambda(\omega) = \lambda(\omega')$.

Proof. If $\omega$ contains inside/outside vertices, we choose one of them and denote it by $u$. If it does not contain any such vertex, then $x \in V(Z)$ and $x = y$, since $B$ cannot consist of just one edge. In this case we choose $u = x$.

Now we find a proper $uv$-walk $\gamma$ in $B$ and use it as a detour. More precisely, $\omega'$ starts at $x$ and follows $\omega$ to the first occurrence of $u$ in $\omega$. Then it goes to $v$ and back along $\gamma$. Finally, it continues to $y$ along $\omega$. It is clear that $\lambda(\omega) = \lambda(\omega')$. By the choice of $u$, the walk $\omega'$ is also proper. \hfill \square

Now, whenever $u$ and $w$ are two vertices from $V(B) \cap V(Z)$, there is an arrow $\overrightarrow{uw}$ arising from $B$ if and only if the vertices $u$ and $w$ were assigned different labels by $L_B$—this is proved in Proposition 27 below.

Proposition 27. Let $B$ be an inside bridge and $L$ be a valid labelling for $B$. Let $x, y \in V(B) \cap V(Z)$ (possibly $x = y$). Then the inside arrow graph $A^+$ contains an arrow $\overrightarrow{xy}$ arising from $B$ if and only if $L(x) \cup L(y) = \{0,1\}$. 22
Proof. It is straightforward to check the claim if $B$ is just an edge $e$. Indeed, if $x \neq y$, then $e = xy$, and it defines the arrow $\overline{xy}$ arising from $B$ if and only if $\lambda(e) = 1$, which in turn happens if and only if $L(x) \cup L(y) = \{0, 1\}$ according to Definition 25. If $x = y$, then $\overline{xy}$ is not induced by $B$ and $|L(x) \cup L(x)| = 1$.

If $V(B) \setminus V(Z) \neq \emptyset$, let $v = v_B$ be the reference vertex for $L$. First, let us assume that $L(x) \cup L(y) = \{0, 1\}$. Let us consider a proper $xv$-walk $\omega_{xv}$ and a proper $vy$-walk $\omega_{vy}$ in $B$ such that $\lambda(\omega_{xv}) \neq \lambda(\omega_{vy})$. Such walks exist by Definition 25, since $L(x) \cup L(y) = \{0, 1\}$. Then the concatenation of these two walks is a nontrivial walk which belongs to $W_{xy,B}^+$. Therefore, $\overline{xy}$ is induced by $B$.

On the other hand, let us assume that there is a nontrivial walk $\omega$ in $W_{xy,B}^+$ defining the arrow $\overline{xy}$. We can assume that $\omega$ is not just an edge, because it would mean that $B$ consists only of that edge. By Observation 26, we may assume that $\omega$ contains the reference vertex $v$. This vertex splits $\omega$ into two proper walks $\omega_1$ and $\omega_2$ so that each of them has at least one edge. Since $\lambda(\omega) = 1$, we have $\lambda(\omega_1) \neq \lambda(\omega_2)$. Consequently, $L(x) \cup L(y) = \{0, 1\}$. \(\Box\)

The argument from the last two paragraphs of the proof above can also be used to establish the following lemma.

Lemma 28. Let $B$ be an inside or an outside bridge, let $L$ be a valid labelling for $B$, and let $x, y \in V(B) \cap V(Z)$ be two distinct vertices. Moreover, we assume that $|L(x)| = |L(y)| = 1$. Then for any proper $xy$-walks $\omega_1$, $\omega_2$ in $B$ we have $\lambda(\omega_1) = \lambda(\omega_2)$.

Proof. If $B$ contains just the edge $xy$, the observation is trivially true. Therefore, we assume that there is the inside/outside reference vertex $v \in V(B)$ for $L$. By the assumption, every two proper $xy$-walks in $B$ have the same $\lambda$-value. The same holds also for proper $vy$-walks in $B$. By Observation 26, we can assume that both $\omega_1$ and $\omega_2$ contain $v$. Then the lemma follows. \(\Box\)

We will also need the following description of inside arrows induced by an inside bridge which does not induce any loop.

Lemma 29. Let $B$ be an inside bridge which does not induce any loop. Then the inside arrows induced by $B$ form a complete bipartite graph. (One of the parts is empty if $B$ does not induce any arrow.)

Proof. Let us consider a valid labelling $L$ for $B$. By Proposition 27, $|L(x)| = 1$ for any $x \in V(B) \cap V(Z)$, since $B$ does not induce any loop. By Proposition 27 again, the inside arrows induced by $B$ form a complete bipartite graph, in which one part corresponds to the vertices labelled $0$ and the second part corresponds to the vertices labelled $1$. \(\Box\)

We conclude this section a by a proof of Proposition 16.

Proof of Proposition 16. We need to distinguish few cases.

First, we consider the case when we have two disjoint inside arrows, but at least one of them is a loop. In this case, it is easy to see that Lemma 12 implies that $G$ forms the outside fan and we are done.

Second, let us consider the case that we have two disjoint inside arrows $\overline{ab}$ and $\overline{cd}$ which are not loops. Lemma 12 implies that the only possible outside arrows are $\overline{ac}$, $\overline{ad}$, $\overline{bc}$, $\overline{bd}$. (In particular, there are no loops outside.) If there are not two disjoint arrows outside, then
G forms an outside fan and we are done. Therefore, we may assume that there are two disjoint arrows outside, without loss of generality, \( \overrightarrow{ac} \) and \( \overrightarrow{bd} \) (otherwise we swap \( a \) and \( b \)). By swapping outside and inside in the previous argument, we get that only further possible arrows inside are \( \overrightarrow{ad} \) and \( \overrightarrow{bc} \).

Now we distinguish a subcase when there is an inside bridge inducing the inside arrows \( \overrightarrow{ab} \) and \( \overrightarrow{cd} \). In this case, \( \overrightarrow{ad} \) and \( \overrightarrow{bc} \) must be inside arrows as well by Lemma 29. By Lemma 12, we know that \( \overrightarrow{ac} \) and \( \overrightarrow{bd} \) are the only outside arrows and we get that they must alternate by Lemma 13. That is, up to relabelling of the vertices, we get the right cyclic order for an inside square. In order to check that \( G \) indeed forms an inside square, it remains to verify that \( G \) has only one nontrivial inside bridge. The inside arrows are \( \overrightarrow{ab}, \overrightarrow{bc}, \overrightarrow{cd} \) and \( \overrightarrow{ad} \). If any of these arrows, for example \( \overrightarrow{ab} \), is induced by two bridges, then we get a contradiction with Lemma 13, in this case on arrows \( \overrightarrow{ab} \) and \( \overrightarrow{cd} \).

By swapping inside and outside we solve the subcase when there is an outside bridge inducing the outside arrows \( \overrightarrow{ac} \) and \( \overrightarrow{bd} \); we get that \( G \) forms an outside square.

It remains to consider the subcase when \( \overrightarrow{ab} \) and \( \overrightarrow{cd} \) arise from different inside bridges and \( \overrightarrow{ac} \) and \( \overrightarrow{bd} \) arise from different outside bridges. However, Lemma 13 applied to the inside and then to the outside reveals that these two events cannot happen simultaneously. Consequently, we have proved Proposition 16 in case there are two disjoint inside arrows. Analogously, we resolve the case when we have two disjoint arrows outside.

Finally, we consider the case when every pair of inside arrows shares a vertex and every pair of outside arrows shares a vertex. If there is a vertex \( v \) common to all the inside arrows, then we get an inside fan and we are done.

It remains to consider the last subcase when there is no vertex common to all inside arrows while every pair of inside arrows shares a vertex. This leaves the only option that there are three distinct vertices \( a, b \) and \( c \) on \( Z \) and all three inside arrows \( \overrightarrow{ab}, \overrightarrow{ac} \) and \( \overrightarrow{bc} \) are present. Then, the only possible outside arrows are \( \overrightarrow{ab}, \overrightarrow{ac} \) and \( \overrightarrow{bc} \) as well due to Lemma 12. In addition, all three outside arrows \( \overrightarrow{ab}, \overrightarrow{ac} \) and \( \overrightarrow{bc} \) must be present, otherwise we have an outside fan and we are done.

In the present case, an inside bridge can induce at most two arrows by Lemma 29. Let us consider the three pairs of arrows \( \{ \overrightarrow{ab}, \overrightarrow{ac} \}, \{ \overrightarrow{ab}, \overrightarrow{bc} \}, \text{ and } \{ \overrightarrow{ac}, \overrightarrow{bc} \} \). If at most one of these pairs is induced by an inside bridge, then \( G \) forms an inside split triangle and we are done. Analogously, we are done, if at most one of these pairs is induced by an outside bridge. Therefore, it remains to consider the case that at least two such pairs are induced by inside bridges and at least two such pairs are induced by outside bridges. However, this yields a contradiction to Lemma 14.

\( \otimes \)

6 Forbidden Configurations of Arrows

In this section we show that certain combinations of arrows are not possible. That is, we prove Lemmas 12, 13 and 14. As before, we have a fixed graph \( G \), its drawing \((D, \lambda)\) on \( S^2 \) and a cycle \( Z \) in \( G \). Again, we assume that \( G, (D, \lambda) \) and \( Z \) satisfy the separation assumptions.

Homology and intersection forms. We start with a brief explanation of intersection forms that will help us to prove the required lemmas.

We assume that the reader is familiar with basics of homology theory, otherwise we refer to the introductory books by Hatcher [Hat02] or Munkres [Mun00]. We always work with
homology over \( \mathbb{Z}_2 \) and, unless stated otherwise, we work with singular homology. Let \( S \) be a surface. We will mainly work with the first homology group and we denote by \( B_1(S), Z_1(S) \) and \( H_1(S) := Z_1(S)/B_1(S) \) the group of 1-boundaries, of 1-cycles and the first homology group, respectively. Given a 1-cycle \( z \in Z_1(S) \), if there is no risk of confusion, we also consider it as an element of \( H_1(S) \), although, formally speaking, we should consider its homology class \([z]\). Similarly, if there is no risk of confusion, we do not distinguish a 1-cycle and its support. Namely, by an intersection of two 1-cycles we actually mean an intersection of their images.

We use the same convention for crossings, that is, transversal intersections.

Let \( S \) be a surface. The intersection form on \( S \) is a unique bilinear map \( \Omega_S : H_1(S) \times H_1(S) \to \mathbb{Z}_2 \) with the following property. Whenever \( z_1, z_2 \in Z_1(S) \) are two 1-cycles intersecting in finite number of points and crossing in every such point (i.e., intersecting transversally), then \( \Omega_S(z_1, z_2) \) is the number of crossings of \( z_1 \) and \( z_2 \) modulo 2; we refer to \[FV04, \text{Sect. 8.4}\] for the existence of \( \Omega_S \). In particular, \( \Omega_{S^2} \) is the trivial map since \( H_1(S^2) \) is trivial. On the other hand, \( \Omega_{R^2P^2} \) is already nontrivial:

**Lemma 30** (Intersection form on \( \mathbb{R}P^2 \)). Let \( z_1 \) and \( z_2 \) be two homologically nontrivial 1-cycles in \( \mathbb{R}P^2 \). Then \( \Omega_{\mathbb{R}P^2}(z_1, z_2) = 1 \). In particular, if \( z_1 \) and \( z_2 \) have a finite number of intersections and they cross at every intersection, then they have to cross an odd number of times.

**Proof.** Since the intersection form \( \Omega_{\mathbb{R}P^2} \) depends only on the homology class, and since \( H_1(\mathbb{R}P^2) = \mathbb{Z}_2 \), it is sufficient to exhibit any two nontrivial 1-cycles that intersect an odd number of times on \( \mathbb{R}P^2 \). This is an easy task.

**From sphere to the projective plane.** Although it is overall simpler to do the proof of Theorem 1 in the setting of projective HT-drawings on \( S^2 \), it is easier to prove Lemmas 12, 13 and 14 in the setting of HT-drawings on \( \mathbb{R}P^2 \). A small drawback is that we need to check that splitting of \( S^2 \) to the inside and outside part works analogously on \( \mathbb{R}P^2 \) as well.

**Lemma 31.** Let \((D, \lambda)\) be a projective HT-drawing of a graph \( G \) on \( S^2 \) and let \( Z \) be a cycle satisfying the separation assumptions. Let \( D_\varnothing \) be the HT-drawing of \( G \) on \( \mathbb{R}P^2 \) coming from the proof of Lemma 4. Then \( D_\varnothing(Z) \) is a simple cycle such that each of its edges is even, which separates \( \mathbb{R}P^2 \) into two parts, \((\mathbb{R}P^2)^+\) and \((\mathbb{R}P^2)^-\). In addition, every inside edge (with respect to \( D \)) which is incident to a vertex of \( Z \) points locally into \((\mathbb{R}P^2)^+\) in \( D_\varnothing \) as well as every outside edge (with respect to \( D \)) which is incident to a vertex of \( Z \) points locally into \((\mathbb{R}P^2)^-\).

**Proof.** By statement of Lemma 4 we already know that \( D_\varnothing(Z) \) is a simple cycle and that each of its edges is even. For the rest, we need to inspect the construction of \( D_\varnothing \) in the proof of Lemma 4. However, we get all the required conclusions directly from this construction.

**Drawings of walks.** We also need to set up a convention regrading drawings of walks in a graph \( G \). Let \( D \) be a drawing of a graph \( G \) on a surface \( S \). Let \( \omega \) be a walk in \( G \). Then \( D \) induces a continuous map \( D(\omega) : [0, 1] \to S \); it is given by the concatenation of drawings of edges of \( \omega \). Here we also allow that \( \omega \) is a walk of length 0 consisting of a single vertex \( v \). Then \( D(\omega) \) is a constant map whose image is \( D(v) \). If \( \omega \) is a closed walk, then we may regard it as an element of \( H_1(S) \).

25
Proofs of the lemmas. Now we have introduced enough tools to prove the required lemmas. In all three proofs, $D_\otimes$ stands for the HT-drawing on $\mathbb{R}P^2$ from Lemma 31. First, we prove Lemma 13 which has a very simple proof. In fact, we prove slightly stronger statement which we plan to reuse later on. Lemma 13 follows directly from Lemma 32 below.

Lemma 32. Let $a$, $b$, $x$ and $y$ be four distinct vertices of $Z$ such that $x$ and $y$ are on the same arc of $Z$ when split by $a$ and $b$. Then any two walks $\omega_{ab}^+ \in W_{ab}^+$ and $\omega_{xy}^+ \in W_{xy}^+$ must share a vertex.

Proof. We consider a closed walk $\kappa_{ab}^+$ arising from a concatenation of the walk $\omega_{ab}^+$ and the arc of $Z$ connecting $a$ and $b$ not containing $x, y$. We also consider the closed walk $\kappa_{xy}^+$ obtained analogously. See Fig. 15. The homological 1-cycles corresponding to $D_\otimes(\kappa_{ab}^+)$ and $D_\otimes(\kappa_{xy}^+)$ are both non-trivial; therefore, by Lemma 30, $D_\otimes(\kappa_{ab}^+)$ and $D_\otimes(\kappa_{xy}^+)$ must have an odd number of crossings. (Note that, for example, $D_\otimes(\kappa_{ab}^+)$ may have self-intersections or self-touchings, but there is a finite number of intersections between $D_\otimes(\kappa_{ab}^+)$ and $D_\otimes(\kappa_{xy}^+)$ which are necessarily crossings.) However, if $\omega_{ab}^+ \in W_{ab}^+$ and $\omega_{xy}^+ \in W_{xy}^+$ did not have a vertex in common, then $D_\otimes(\kappa_{ab}^+)$ and $D_\otimes(\kappa_{xy}^+)$ would have an even number of crossings, because $D_\otimes$ is an HT-drawing by Lemma 4.

We have proved Lemma 13 and we continue with the proofs of the next two lemmas.

Proof of Lemma 12. To the contrary, we assume that we have an inside arrow $xy$ and an outside arrow $uv$ which do not share any endpoint. However, we allow $x = y$ or $u = v$, that is, we allow loops. As before, we consider a closed walk $\kappa_{xy}^+$ obtained from the concatenation of a walk from $\omega_{xy}^+ \in W_{xy}^+$ and any of the two arcs of $Z$ connecting $x$ and $y$. If $x = y$, then we do not add the arc from $Z$. Analogously, we have a closed walk $\kappa_{uv}^+$ coming from a walk in $W_{uv}^-$ and an arc of $Z$ connecting $u$ and $v$. Both of these walks are non-trivial and we aim to get a contradiction with Lemma 30.

Unlike the previous proof, this time $D_\otimes(\kappa_{xy}^+)$ and $D_\otimes(\kappa_{uv}^+)$ may not cross at every intersection. Namely, $\kappa_{xy}^+$ and $\kappa_{uv}^+$ may share some subpath of $Z$, but apart from this subpath the intersections are crossings. We slightly modify these drawings in the following way. Let us recall that $D_\otimes(Z)$ splits $\mathbb{R}P^2$ into two parts $(\mathbb{R}P^2)^+$ and $(\mathbb{R}P^2)^-$ according to Lemma 31. We slightly push into $(\mathbb{R}P^2)^+$ the subpath of $\kappa_{xy}^+$ shared with $Z$ (possibly consisting of a single vertex). This way, we obtain a drawing $D_\otimes^+$ of $\kappa_{xy}^+$. Similarly, we slightly push the subpath of $\kappa_{uv}^+$ shared with $Z$ into $(\mathbb{R}P^2)^-$, obtaining a drawing $D_\otimes^-$ of $\kappa_{uv}^+$. See Fig. 16. Now, $D_\otimes^+(\kappa_{xy}^+)$ and $D_\otimes^-(\kappa_{uv}^+)$ cross at every intersection and the crossings of $D_\otimes^+(\kappa_{xy}^+)$ and $D_\otimes^-(\kappa_{uv}^+)$ correspond to the crossings of $D_\otimes(\kappa_{xy}^+)$ and $D_\otimes(\kappa_{uv}^+)$. 

![Figure 15: Walks in Lemma 32.](image)
We now consider the crossings of $D(\kappa_{xy}^+)$ and $D(\kappa_{uv}^-)$. Whenever $e$ is an edge of $\kappa_{xy}^+$ and $f$ is an edge of $\kappa_{uv}^-$ such that $e$ and $f$ are independent, then $D(e)$ and $D(f)$ have an even number of crossings, because $D$ is an HT-drawing. However, if $e$ and $f$ are adjacent, then they still cross evenly since one of these edges must belong to $Z$. Here we crucially use that $\overline{xy}$ and $\overline{uv}$ do not share any endpoint. Therefore, $D(\kappa_{xy}^+)$ and $D(\kappa_{uv}^-)$ have an even number of crossings, and consequently, $D^+(\kappa_{xy}^+)$ and $D^-(\kappa_{uv}^-)$ as well. This is a contradiction to Lemma 30.

Proof of Lemma 14. For contradiction, there is such a configuration.

Let $e_a^+$ be any edge of $E(B^+)$ incident to $a$. Analogously, we define edges $e_a^-$, $e_b^+$, $e_b^-$, $e_c^+$, and $e_c^-$. We observe that there is a walk $\omega_{ab}^+ \in W_{ab}^+$ which uses the edges $e_a^+$ and $e_b^+$. Indeed, it is sufficient to consider arbitrary proper walk using $e_a^+$ and $e_b^+$ in $B^+$. This walk is nontrivial by Lemma 28. (The assumptions of the lemma are satisfied by Proposition 27 since $B^+$ does not induce any inside loops.) We also let $\kappa_{ab}^+$ be the closed walk obtained from the concatenation of $\omega_{ab}^+$ and the arc of $Z$ connecting $a$ and $b$ and avoiding $c$. Analogously, we define $\omega_{ac}^+$, $\omega_{bc}$, and closed walks $\kappa_{ac}^+$, $\kappa_{bc}^-$ and $\kappa_{bc}^-$. When defining the closed walks, we always use the arc of $Z$ which avoids the third point among $a$, $b$, and $c$. All these eight walks are nontrivial.

Now, we aim to show that $e_a^+$ and $e_c^-$ cross oddly in the drawing $D$. We consider the closed walks $\kappa_{ab}^+$ and $\kappa_{ac}^+$ and their drawings $D(\kappa_{ab}^+)$ and $D(\kappa_{ac}^+)$. The walks $\kappa_{ab}^+$ and $\kappa_{ac}^+$ share only the point $a$; therefore, $D(\kappa_{ab}^+)$ and $D(\kappa_{ac}^+)$ cross at every intersection possibly except $D(a)$. By Lemma 31 we know that $e_a^+$ and $e_c^-$ point to different sides of $Z$ (in $D$); thus, $D(\kappa_{ab}^-)$ and $D(\kappa_{ac}^-)$ actually touch in $D(a)$. This touching can be removed by a slight perturbation of these cycles, analogously as in the proof of Lemma 12, without affecting other intersections. By Lemma 30 we therefore get that $D(\kappa_{ab}^+)$ and $D(\kappa_{ac}^+)$ have an odd number of crossings. However, if we consider any pair of edges $(e, f)$ where $e$ is an edge of $\kappa_{ab}^+$ and $f$ is an edge of $\kappa_{ac}^+$ different from $(e_a^+, e_c^+)$, we get that $e$ and $f$ cross an even number of times. Indeed, if we have such $(e, f) \neq (e_a^+, e_c^+)$, then either $e$ or $f$ belongs to $Z$, or they are independent. Consequently, the odd number of crossings of $D(\kappa_{ab}^-)$ and $D(\kappa_{ac}^-)$ has to be realized on $e_a^+$ and $e_c^-$.

Analogously, we show that $e_b^+$ and $e_b^-$ must cross oddly by considering the walks $\kappa_{ab}^+$ and $\kappa_{bc}$.

Now let us consider the closed walk $\kappa_{ab}^+$ and a closed walk $\mu_{ab}^+$ obtained from the concatenation of $\omega_{ab}^+$ and the arc of $Z$ connecting $a$ and $b$ which contains $c$. By analogous ideas...
as before, we get that \( D \otimes (\kappa_{ab}^+) \) and \( D \otimes (\mu_{ab}^-) \) touch in \( D \otimes (a) \) and \( D \otimes (b) \); if they intersect anywhere else, they cross there. Using a small perturbation as before, they must have an odd number of crossings by Lemma 30. On the other hand, the pairs of edges \((e_a^+, e_a^-)\) and \((e_b^+, e_b^-)\) cross oddly, as we have already observed. Any other pair \((e, f)\) of edges where \( e \) is an edge of \( \kappa_{ab}^+ \) and \( f \) is an edge of \( \mu_{ab}^- \) must cross evenly since they are either independent or one of them belongs to \( Z \). This means that \( D \otimes (\kappa_{ab}^+) \) and \( D \otimes (\mu_{ab}^-) \) intersect evenly, which is a contradiction.

**Intersection of trivial interleaving walks.** We conclude this section by a proof of a lemma similar in spirit to Lemma 32. We will need this Lemma in Sect. 7, but we keep the lemma here due to its similarity to previous statements.

**Lemma 33.** Let \( a, b, x \) and \( y \) be four distinct vertices of \( Z \) such that \( x \) and \( y \) are on different arcs of \( Z \) when split by \( a \) and \( b \). Let \( \omega_{ab}^+ \) and \( \omega_{xy}^+ \) be a proper ab-walk and a proper xy-walk in \( G^+ \), respectively, such that \( \lambda(\omega_{ab}^+) = \lambda(\omega_{xy}^+) = 0 \). Then \( \omega_{ab}^+ \) and \( \omega_{xy}^+ \) must share a vertex.

**Proof.** We proceed by contradiction. As usual, we consider closed walks \( \kappa_{ab}^+ \) and \( \kappa_{xy}^+ \) defined as follows. The walks \( \kappa_{ab}^+ \) consists of \( \omega_{ab}^+ \) and an arc of \( Z \) connecting \( a \) and \( b \), while the walk \( \kappa_{xy}^+ \) is formed by \( \omega_{xy}^+ \) and an arc of \( Z \) connecting \( x \) and \( y \). This time, \( \omega_{ab}^+ \) and \( \omega_{xy}^+ \) are trivial.

We push \( D_\otimes(\kappa_{ab}^+) \) a bit inside and \( D_\otimes(\kappa_{xy}^+) \) a bit outside of \( Z \), similarly as in the proof of Lemma 12. This time, however, we introduce one more crossing, because both \( \kappa_{ab}^+ \) and \( \kappa_{xy}^+ \) are walks in \( G^+ \). Since the intersection form of trivial cycles corresponding to the drawings of \( \kappa_{ab}^+ \) and \( \kappa_{xy}^+ \) is trivial, we get that these drawings have to cross an even number of times. This in turn means that the drawings of \( \omega_{ab}^+ \) and \( \omega_{xy}^+ \) cross an odd number of times. Since \( D_\otimes \) is an HT-drawing, this yields a contradiction to the assumption that \( \omega_{ab}^+ \) and \( \omega_{xy}^+ \) do not share a vertex.

\( \otimes \)

## 7 Redrawings

We will prove Proposition 17 in this section separately for each case. That is, we show that if \( G^+ \) forms any of the configurations depicted in Fig. 8, then \( G^+ \) admits an ordinary HT-drawing on \( S^2 \). However, we start with a general redrawing result that we will use in all cases.

**Lemma 34.** Let \((D, \lambda)\) be a projective HT-drawing of \( G^+ \) on \( S^2 \) and \( Z \) a cycle satisfying the separation assumptions. Let us also assume that that \( D(G^+) \cap S^- = \emptyset \). Let \( B \) be one of the inside bridges different from an edge and let \( L \) be a valid labelling of \( B \). Let us assume that there is at least one vertex \( x \in V(B) \cap V(Z) \) such that \( |L(x)| = 1 \). Then there is a projective HT-drawing \((D', \lambda')\) of \( G^+ \) on \( S^2 \) such that

(a) \( D \) coincides with \( D' \) on \( Z \) and \( D'(G^+) \cap S^- = \emptyset \);
(b) every edge \( e \in E(G^+) \setminus E(B) \) satisfies \( \lambda(e) = \lambda'(e) \);
(c) every edge \( e \in E(B) \) that is not incident to \( Z \) satisfies \( \lambda'(e) = 0 \); and
(d) for every edge \( uv = e \in E(B) \) such that \( u \in V(Z) \), we have \( \lambda'(e) \in L(u) \).
Note that the condition (b) allows that the edges in inside bridges other than B may be redrawn, but only under the condition, that their triviality/nontriviality is not affected.

**Proof.** Let $B^+$ be the subgraph of $B$ induced by the vertices of $V(B) \setminus V(Z)$. By the definition of the inside bridge, the graph $B^+$ is connected; it is also nonempty since we assume that $B$ is not an edge.

Every cycle of the graph $B^+$ must be trivial. Indeed, if $B^+$ contained a nontrivial cycle, then this cycle could be used to obtain a nontrivial proper walk from $x$ to $x$. This would contradict the fact that $|L(x)| = 1$ via Proposition 27. That is, $B^+$ satisfies the assumptions of Lemma 8. Let $U \subseteq V(B^+)$. In addition, we have $U \subseteq V(B^+)$ be the set of vertices obtained from Lemma 8. That is, if we perform the vertex-crosscap switches on $U$, we obtain a projective HT-drawing $(D_U, \lambda_U)$ such that $\lambda_U(e) = 0$ for any edge $e \in E(B^+)$. Let us recall that every vertex-crosscap switch over a vertex $xu$ is obtained from vertex-edge switches of nontrivial edges over $y$ and then from swapping the value of $\lambda$ on all edges incident to $y$. The vertex-edge switches do not affect the value of $\lambda$. Overall, we get that $D_U$ coincides with $D$ on $Z$. We also require that all vertex-edge switches are performed in $S^+$; therefore, $D_U$ does not reach $S^-$. Altogether, $D_U$ and $\lambda_U$ satisfy (a), (b) and (c), but we do not know yet whether (d) is satisfied.

In fact, (d) may not be satisfied and we still may need to modify $D_U$ and $\lambda_U$. Let $e_0$ be any edge incident with $x$. If $L(x) = \{\lambda_U(e_0)\}$, we set $D' := D_U$ and $\lambda' := \lambda_U$. If $L(x) \neq \{\lambda_U(e_0)\}$, we further perform vertex-crosscap switches over all vertices in $V(B^+)$, obtaining $D'$ and $\lambda'$.

We want to check (a) to (d) for $D'$ and $\lambda'$. It is sufficient to check (a), (b) and (c) only in the latter case. Regarding (a), we again change the drawing only by vertex-edge switches over edges $e$ with $\lambda_U(e) = 1$ inside $S^+$. Validity of (b) is obvious from the fact that $\lambda_U$ may be changed only on edges incident with $V(B^+)$. Regarding (c), for any edge $e \in E(B^+)$ we perform the vertex-crosscap switch for both endpoints of $e$. Therefore, $\lambda'(e) = \lambda_U(e) = 0$. It remains to check (d).

First, we realize that we have set up $D'$ and $\lambda'$ in such a way that $L(x) = \{\lambda'(e_0)\}$. Indeed, if $L(x) \neq \{\lambda_U(e_0)\}$, then we have made a vertex-crosscap switch over exactly one endpoint of $e_0$. In particular, we have just checked (d) if $e = e_0$.

Now, let $e = uv \neq e_0$ be an edge from (d). We need to check that $\lambda'(e) \subseteq L(u)$. If $L(u) = \{0, 1\}$, then we are done; therefore, we may assume that $|L(u)| = 1$. Let $\omega$ be any proper $xu$-walk in $B$ containing $e_0$ and $e$. Such a walk exists from the definition of an inside bridge (see Definition 11). We have $\lambda(\omega) = \lambda'(\omega)$ because the vertex-crosscap switches over the inner vertices of $\omega$ do not affect the triviality of $\omega$. But we also have $\lambda'(\omega) = \lambda'(e_0) + \lambda'(e)$ because $\lambda'(f) = 0$ for any edge $f \in E(B^+)$. Since $L(x) = \{\lambda'(e_0)\}$ and $|L(u)| = 1$, it follows that $L(u) = \{\lambda'(e)\}$ by Proposition 27 and Lemma 28 applied to $x$ and $u$.

**Inside fan.** Now we may prove Proposition 17 for inside fans, which is the simplest case.

**Proof of Proposition 17 for inside fans.** We assume that $G^{+0}$ forms an inside fan; see Fig. 8. Let $x \in V(Z)$ be the endpoint common to all inside arrows. Let us consider any inside bridge $B$, possibly trivial. Let $L = L_B$ be a valid labelling of $B$. It follows from Proposition 27 that $|L(u)| = 1$ for any $u \in V(B) \cap V(Z)$ different from $x$. (Actually, there is at least one such $u$, because we assume that $G$ is 2-connected; this is contained in the separation assumptions.) In addition, all $u \in V(B) \cap V(Z)$ different from $x$ have to have the same labels, because there
are no arrows among them. Since we may switch all labels in a valid labelling by changing
the value of the labelling parameter, we may assume that \( L(u) = \{0\} \) for any such \( u \).

Now, we consider all inside bridges \( B_1, \ldots, B_\ell \) (possibly trivial) and the corresponding
labellings \( L_{B_1}, \ldots, L_{B_\ell} \) as above. We apply Lemma 34 to each of these bridges which is not
an edge one by one. This way we get a projective HT-drawing \((D_1, \lambda_1)\) which satisfies:

(i) \( D \) coincides with \( D_1 \) on \( Z \) and \( D_1(G^{+0}) \cap S^- = \emptyset \);

(ii) every edge \( e \in E(G^{+0}) \) which is not incident with \( Z \) satisfies \( \lambda_1(e) = 0 \);

(iii) every edge \( e \in E(G^{+0}) \) such that \( \lambda_1(e) = 1 \) is incident with \( x \).

Indeed, property (i) follows from the iterative application of property (a) of Lemma 34. Property (ii)
follows from the iterative application of properties (b) and (c) of Lemma 34. Finally, property (iii)
follows from (ii), from the iterative application of properties (b) and (d) of Lemma 34 and from the fact that any nontrivial inside bridge which is a single edge must contain \( x \).

Finally, we set \( D' := D_1 \) and let \( \lambda' \colon E(G^{+0}) \to \{0, 1\} \) be the constantly zero function.
We observe that from (ii) and (iii), it follows that \( \lambda(e)\lambda'(f) = \lambda_1(e)\lambda_1(f) \) for any pair of
independent edges of \( G^{+0} \). Therefore \((D', \lambda')\) is a projective HT-drawing as well. But, since
\( \lambda' \) is identically zero function, \( D' \) is also just an ordinary HT-drawing on \( S^2 \).

**Inside square.** Now we prove Proposition 17 for an inside square. Let \( B \) be the inside
bridge inducing the inside square and let \( a, b, c \) and \( d \) be the vertices of \( V(B) \cap V(Z) \) labelled
according to Definition 15. The main ingredient for our proof of Proposition 17 is the following
lemma, which shows that \( B \) must have a suitable cut vertex.

**Lemma 35.** The inside bridge \( B \), inducing the inside square, contains a vertex \( v \) such that the
graph \( B - v \) is disconnected and the vertices \( a, b, c \) and \( d \) belong to four different components
of \( B - v \).

We first show how Proposition 17 for inside squares follows from Lemma 35. The proof is
analogous to the previous proof.

**Proof of Proposition 17 for inside squares.** We assume that \( B \) is the unique inside bridge in-
ducing the inside square and \( a, b, c \) and \( d \) are vertices of \( V(B) \cap V(Z) \) as above. In addition,
let \( v \) be the vertex from Lemma 35.

First we consider valid labellings of trivial inside bridges. After possibly switching the
value of the labelling parameter, we may achieve that all labels of a trivial inside bridge are
0. We apply Lemma 34 to all trivial inside bridges (which are not an edge) and we get a
projective HT-drawing \((D_1, \lambda_1)\) such that \( \lambda_1(e) = 0 \) for any edge of \( G^{+0} \) which does not
belong to the nontrivial inside bridge \( B \). Also, we did not affect \( \lambda \) on edges of \( B, D_1 \) coincides with
\( D \) on \( Z \) and we still have \( D_1(G^{+0}) \cap S^- = \emptyset \).

Now, we consider a valid labelling \( L \) of \( B \). It is easy to check that, up to switching all
labels, we have \( L(a) = L(c) = \{1\} \) and \( L(b) = L(d) = \{0\} \). We apply Lemma 34 to \( B \)
according to this labelling and we get a projective HT-drawing \((D_2, \lambda_2)\) such that the only
edges \( e \) of \( G^{+0} \) with \( \lambda_2(e) = 1 \) are the edges of \( B \) incident to \( a \) or \( c \).

Next, let \( C_a \) and \( C_c \) be the components of \( B - v \) which contains \( a \) and \( c \), respectively.
We perform vertex-crosscap switches over all vertices of \( C_a \) and \( C_c \) except \( a, c \) and \( v \). We

30
perform the switches inside $S^+$ as usual. This way we get a projective HT-drawing $(D_3, \lambda_3)$ such that only edges $e$ of $G^+0$ such that $\lambda_3(e) = 1$ are the edges of $B$ incident to $v$.

Finally, we let $D' = D_3$ and we set $\lambda'(e) = 0$ for any edge $e$ of $G^+0$. Analogously as in the previous proof, $\lambda_3(e)\lambda_3(f) = \lambda'(e)\lambda'(f)$ for any pair of independent edges of $G^+0$. Therefore, $(D', \lambda')$ is a projective HT-drawing on $S^2$ and $D'$ is also an ordinary HT-drawing on $S^2$, as required.

It remains to prove Lemma 35 to conclude the case of inside squares.

We start with a certain separation lemma in a general graph and then we conclude the proof by verification that the assumptions of this lemma are satisfied.

**Lemma 36.** Let $G'$ be an arbitrary connected graph and $A = \{a_1, \ldots, a_4\} \subseteq V(G')$ be a set of four distinct vertices. Let us assume that any $a_ia_j$-path has a common point in $V(G') \setminus A$ with any $a_ka_l$-path whenever \{i, j, k, l\} = \{1, 2, 3, 4\}. Then there is a cut vertex $v$ of $G'$ such that $a_1, \ldots, a_4$ are in four distinct components of $G' - v$.

**Proof.** Let us consider an auxiliary graph $G''$ which is obtained from $G'$ by adding two new vertices $x$, $y$ and attaching $x$ to $a_1, a_2$ and $y$ to $a_3, a_4$. By the assumptions, $G''$ is connected and moreover, there are no two vertex-disjoint paths connecting $x$ and $y$. By Menger’s theorem (see, e.g., [Die10, Corollary 3.3.5]), there is a cut-vertex $v \in V(G'') \setminus \{x, y\} = V(G')$ disconnecting $x$ and $y$. Let $C_1$ be the connected component of $G'' - v$ containing $x$ and $C_2$ be the component containing $y$. Let $C_i'$, for $i = 1, 2$, be the subgraph of $G'$ induced by $v$ and the vertices of $C_i \cap G'$. Note that, since $G'$ is connected, both $C_1'$ and $C_2'$ are connected. We show that $v$ is the desired cut vertex.

Let $p_1$ be an $a_1a_2$-path in $C_1'$ and $p_2$ an $a_3a_4$-path in $C_2'$. Since $C_1'$ and $C_2'$ are connected, such paths $p_1$ and $p_2$ exist. Moreover, $p_1$ and $p_2$ may intersect only in $v$; however, according to the assumptions, they have to intersect in a vertex outside $A$. Therefore, they must intersect in $v$ and $v \notin A$. Overall, we have verified that any $a_ia_j$-path passes through $v$, for $1 \leq i < j \leq 4$, which shows that $v$ is the desired cut vertex.

**Proof of Lemma 35.** We apply Lemma 36 to $B$ and to $A = \{a, b, c, d\}$. Let us consider a valid labelling $L$ of $B$. Up to swapping the labels, we may assume that $L(a) = L(c) = \{1\}$ and $L(b) = L(d) = \{0\}$. Then Proposition 27 together with Lemma 28 imply that any proper $ab$, $bc$, $cd$, or $ad$-walk is nontrivial, whereas any proper $ac$ or $bd$-walk is trivial. Then, the assumptions of Lemma 36 are satisfied due to Lemmas 32 and 33.

**Inside split triangle.** Finally, we prove Proposition 17 for an inside split triangle.

**Proof of Proposition 17 for an inside split triangle.** Let $a$, $b$, $c$ be the three vertices of $Z$ from the definition of the inside split triangle; see Definition 15 or Fig. 8.

First, similarly as in the proof for inside squares, we take care of trivial inside bridges via suitable labellings and Lemma 34. We reach a projective HT-drawing $(D_1, \lambda_1)$ still satisfying the assumptions of Proposition 17, which in addition satisfies $\lambda_1(e) = 0$ for any edge $e$ of $G^+0$ that does not belong to a nontrivial bridge.

Now, let us consider nontrivial inside bridges. By the assumptions, each such bridge is either an $a$-bridge, that is, a nontrivial inside bridge which contains $a$ (and $b$ or $c$ or both), or a $bc$-bridge which contains $b$ and $c$, but not $a$. We consider valid labellings of these bridges. As usual, we may swap all labels in a valid labelling when needed. This way, it is easy to check
that every $a$-bridge $B$ admits a valid labelling $L_B$ such that $L_B(a) = \{1\}$, whereas all other labels are 0. Similarly, each $bc$-bridge $B$ admits a valid labelling $L_B$ such that $L_B(b) = \{1\}$ and $L_B(c) = \{0\}$. We apply Lemma 34 and we reach a projective HT-drawing $(D_2, \lambda_2)$ still satisfying the assumptions of Proposition 17, which in addition satisfies the following property. The edges $e$ of $G^{+0}$ with $\lambda_2(e) = 1$ are exactly the edges of an $a$-bridge which are incident to $a$ or edges of a $bc$-bridge incident to $b$.

If we do not have any $bc$-bridge, then all nontrivial edges are incident to $a$ and we finish the proof by setting $D' = D_2$ and letting $\lambda'$ be identically 0, similarly as in the cases of an inside fan or an inside square. However, if we have $bc$-bridge(s), we need to be more careful.

Let $E_a^x$ and $E_{bc}^x$ be the sets of edges incident to a vertex $x$ in an $a$-bridge and the set of edges incident to $x$ in a $bc$-bridge, respectively. Because $D_2$ is a projective HT-drawing, we have $\lambda_2(e)\lambda_2(f) = \text{cr}_{D_2}(e, f)$ for any pair of independent edges $e$ and $f$. In particular, $\text{cr}_{D_2}(e, f) = 1$ for a pair of independent edges if and only if one of the edges belongs to $E_a^x$ and the second one to $E_{bc}^x$.

Now, for every edge $e \in E_{bc}^x$, we perform the vertex-edge switch over each vertex different from $a$, $b$, $c$ of each $a$-bridge obtaining a drawing $D_3$. We perform the switches inside $S^+$. This way, we change the crossing number of such $e$ with edges from $E_a^x$, $E_b^x$ and $E_c^x$. In particular, after this redrawing, we get $\text{cr}_{D_3}(e, f) = 1$ for a pair of independent edges if and only if one of the edges belongs to $E_a^x$ and the second one to $E_{bc}^x$. See Fig. 17.

Finally, for every edge $e \in E_{bc}^x$, we perform the vertex-edge switch over each vertex different from $b$ and $c$ of each $bc$-bridge obtaining the final drawing $D'$. Again, we perform the switches inside $S^+$. This way, we change the crossing number of such $e$ with edges from $E_{bc}^x$ and $E_{bc}^x$. However, it means that $\text{cr}_{D'}(e, f) = 0$ for any pair of independent edges. That is, $D'$ is the required ordinary HT-drawing on $S^2$. See Fig. 17.

Figure 17: An example of redrawing an inside split triangle with one $a$-bridge and one $bc$-bridge. The edges participating in independent pairs crossing oddly are thick. For simplicity of the picture, the drawings $D_3$ and $D'$ are actually simplified. For example, the vertex-edge switches used to obtain $D_3$ from $D_2$ introduce many pairs of independent edges crossing evenly and some pairs of adjacent edges crossing oddly. These intersections are removed in the picture as they do not play any role in the argument. (In particular, the drawing $D'$ is, in fact, typically not a plane drawing.)
8 Redrawing by Pelsmajer, Schaefer and Štefankovič

It remains to prove Theorem 18. As mentioned above, our proof is almost identical to the proof of Theorem 2.1 in [PSŠ07a]. The only notable difference is that we avoid contractions.\footnote{Our reason why we avoid contractions is mainly for readability issues. Contractions yield multigraphs and, formally speaking, we would have to redo several notions for multigraphs. Introducing multigraphs in the previous sections would be disturbing and it is not convenient to repeat all the definitions in such setting now.}

As noted before, the proof of Lemma 3 in [FPSŠ12] can also be extended to yield the desired result.

Proof. First, we want to get a drawing such that there is only one edge of $Z$ which may be intersected by other edges. Here, part of the argument is almost the same as the analogous argument in the proof of Lemma 22.

Let us consider an edge $e = uv \in E(Z)$ intersected by some other edges and let $f = vw \in E(Z)$ be a neighbouring edge of $e$. We again almost-contract $e$ so that we move the vertex $v$ towards $u$ until we remove all intersection of $e$ with other edges. This way, $e$ is now free of crossings and these crossings appear on $f$. Since both $e$ and $f$ were even edges in the initial drawing, $f$ remains even after the redrawing as well. Finally, since we want to keep the position of $Z$, we consider a self-homeomorphism of $S^2$ which sends $v$ back to its original position. See Fig. 11.

By such redrawings, it can be achieved that only one edge $e_0 = u_0v_0$ of $Z$ may be intersected by other edges while keeping $Z$ fixed and $e_0$ even. Without loss of generality, we may assume that the original drawing $D$ satisfies these assumptions.

Let $p$ be the path in $Z$ connecting $u_0$ and $v_0$ avoiding $e_0$. Let us also consider an arc $\gamma$ connecting $u_0$ and $v_0$ outside (that is in $S^-$) close to $D(p)$ such that it does not cross any inside edge. The closed arc obtained from $\gamma$ and $D(p)$ bounds two disks (2-balls). Let $B$ be the open disk which contains $S^+$. Finally, we consider a self-homeomorphism $h$ of $S^2$ that keeps $D(p)$ fixed and maps $B$ to $S^+$. Considering the drawing $h \circ D$ on $G^{+0} - e_0$, it turns...
out that $G^{+0} - e_0$ is now drawn in $S^+$, up to $p$, which stays fixed. For the edge $e_0$, we also keep its original position, that is, we do not apply $h$ to this edge. See Fig. 18.

Since the redrawing is done by a self-homeomorphism, we do not change the number of crossings among pairs of edges in $G^{+0}$. Analogously, we map $G^{-0}$ to $S^-$ and we get the required drawing.

Acknowledgment

We would like to thank Alfredo Hubard for fruitful discussions and valuable comments.

References


On Generalized Heawood Inequalities for Manifolds: a van Kampen–Flores-type Nonembeddability Result∗ †

Xavier Goaoc†, Isaac Mabillard‡, Pavel Paták§, Zuzana Patáková¶, Martin Tancer¶, and Uli Wagner‡

1Université Paris-Est, LIGM (UMR 8049), CNRS, ENPC, ESIEE, UPEM, F-77454, Marne-la-Vallée, France
2IST Austria, Am Campus 1, 3400 Klosterneuburg, Austria
3Einstein Institute of Mathematics, Hebrew University of Jerusalem, Givat Ram, Jerusalem, 9190401, Israel
4Department of Applied Mathematics and Institute for Theoretical Computer Science, Charles University,
Malostranské nám. 25, 118 00 Praha 1, Czech Republic

Abstract

The fact that the complete graph $K_2$ does not embed in the plane has been generalized in two independent directions. On the one hand, the solution of the classical Heawood problem for graphs on surfaces established that the complete graph $K_n$ embeds in a closed surface $M$ (other than the Klein bottle) if and only if $(n - 3)(n - 4) \leq 6b_1(M)$, where $b_1(M)$ is the first $\mathbb{Z}$-Betti number of $M$. On the other hand, van Kampen and Flores proved that the $k$-skeleton of the $n$-dimensional simplex (the higher-dimensional analogue of $K_{n+1}$) embeds in $\mathbb{R}^{2k}$ if and only if $n \leq 2k + 1$.

Two decades ago, Kühnel conjectured that the $k$-skeleton of the $n$-simplex embeds in a compact, $(k - 1)$-connected $2k$-manifold with $4k+2$-th $\mathbb{Z}$-Betti number $b_k$ only if the following generalized Heawood inequality holds: $(\frac{n-k}{k+1}) \leq (\frac{2k+1}{k+1})b_k$. This is a common generalization of the case of graphs on surfaces as well as the van Kampen–Flores theorem.

In the spirit of Kühnel’s conjecture, we prove that if the $k$-skeleton of the $n$-simplex embeds in a $2k$-manifold with $4k+2$-th $\mathbb{Z}$-Betti number $b_k$, then $n \leq 2b_k(2k+2) + 2k + 4$. This bound is weaker than the generalized Heawood inequality, but does not require the assumption that $M$ is $(k - 1)$-connected.

Our results generalize to maps without $q$-covered points, in the spirit of Tverberg’s theorem, for $q$ a prime power. Our proof uses a result of Volovikov about maps that satisfy a certain homological triviality condition.

1 Introduction

Given a closed surface $M$, a natural question is to determine the maximum integer $n$ such that the complete graph $K_n$ can be embedded (drawn without crossings) into $M$ (e.g., $n = 4$ if $M = S^2$ is the 2-sphere, and $n = 7$ if $M$ is a torus). This classical problem was raised in the late 19th century by Heawood [Hea90] and Heffter [Heff91] and completely settled in the 1950–60’s through a sequence of works by Gustin, Guy, Mayer, Ringel, Terry, Welch, and Youngs (see [Rin74, Ch. 1] for a discussion of the history of the problem and detailed references). Heawood already observed that if $K_n$ embeds into $M$ then

$$(n - 3)(n - 4) \leq 6b_1(M) = 12 - 6\chi(M),$$

where $\chi(M)$ is the Euler characteristic of $M$ and $b_1(M) = 2 - \chi(M)$ is the first $\mathbb{Z}_2$-Betti number of $M$, i.e., the dimension of the first homology group $H_1(M; \mathbb{Z}_2)$ (here and throughout the paper, we work with

∗†The work by Z.P. was partially supported by the Israel Science Foundation grant ISF-768/12. The work by Z.P. and M.T. was partially supported by the project CE-ITI (GACR P202/12/G061) of the Czech Science Foundation and by the ERC Advanced Grant No. 267165. Part of the research work of M.T. was conducted at IST Austria, supported by an IST Fellowship. The research of P.P. was supported by the ERC Advanced grant no. 320924. The work by I.M. and U.W. was supported by the Swiss National Science Foundation (grants SNSF-200020-138230 and SNSF-PP00P2-138948).

‡An extended abstract of this work was presented at the 31st International Symposium on Computational Geometry (GMP+15).
homology with $\mathbb{Z}_2$-coefficients). Conversely, for surfaces $M$ other than the Klein bottle, the inequality is tight, i.e., $K_n$ embeds into $M$ if and only if (1) holds; this is a hard result, the bulk of the monograph [Rin74] is devoted to its proof. (The exceptional case, the Klein bottle, has $b_1 = 2$, but does not admit an embedding of $K_7$, only of $K_6$.)

The question naturally generalizes to higher dimension: Let $\Delta_n^{(k)}$ denote the $k$-skeleton of the $n$-simplex, the natural higher-dimensional generalization of $K_{n+1} = \Delta_n^{(1)}$ (by definition $\Delta_n^{(k)}$ has $n + 1$ vertices and every subset of at most $k + 1$ vertices forms a face). Given a $2k$-dimensional manifold $M$, what is the largest $n$ such that $\Delta_n^{(k)}$ embeds (topologically) into $M$? This line of enquiry started in the 1930’s when van Kampen [vK32] and Flores [Flo33] showed that $\Delta_{2k+2}^{(k)}$ does not embed into $\mathbb{R}^{2k}$ (the case $k = 1$ corresponding to the non-planarity of $K_5$). Somewhat surprisingly, little else seems to be known, and the following conjecture of Kühnel [Küh94, Conjecture B] regarding a generalized Heawood inequality remains unresolved:

**Conjecture 1** (Kühnel). Let $n, k \geq 1$ be integers. If $\Delta_n^{(k)}$ embeds in a compact, $(k - 1)$-connected $2k$-manifold $M$ with $k$th $\mathbb{Z}_2$-Betti number $b_k(M)$ then

$$\left(\frac{n - k - 1}{k + 1}\right) \leq \left(\frac{2k + 1}{k + 1}\right) b_k(M). \tag{2}$$

The classical Heawood inequality (1) and the van Kampen–Flores Theorem correspond the special cases $k = 1$ and $b_k = 0$, respectively. Kühnel states Conjecture 1 in slightly different form, in terms of Euler characteristic of $M$ rather than $b_k(M)$; our reformulation is equivalent. The $\mathbb{Z}_2$-coefficients are not important in the statement of the conjecture but they are convenient for our new developments.

### 1.1 New results toward Kühnel’s conjecture

In this paper, our main result is an estimate, in the spirit of the generalized Heawood inequality (2), on the largest $n$ such that $|\Delta_n^{(k)}|$ almost embeds into a given $2k$-dimensional manifold. An almost embedding is a relaxation of the notion of embedding that is useful in setting up our proof method.

Let $K$ be a finite simplicial complex and let $|K|$ be its underlying space (geometric realization). We define an almost-embedding of $K$ into a (Hausdorff) topological space $X$ to be a continuous map $f : |K| \rightarrow X$ such that any two disjoint simplices $\sigma, \tau \in K$ have disjoint images, $f(|\sigma|) \cap f(|\tau|) = \emptyset$. In particular, every embedding is an almost-embedding as well. Let us stress, however, that the condition for being an almost-embedding depends on the actual simplicial complex (the triangulation), not just the underlying space. That is, if $K$ and $L$ are two different complexes with $|K| = |L|$ then a map $f : |K| = |L| \rightarrow X$ may be an almost-embedding of $K$ into $X$ but not an almost-embedding of $L$ into $X$. Our main result is the following.

**Theorem 2.** If $\Delta_n^{(k)}$ almost embeds into a $2k$-manifold $M$ then

$$n \leq 2\binom{2k + 2}{k} b_k(M) + 2k + 4,$$

where $b_k(M)$ is the $k$th $\mathbb{Z}_2$-Betti number of $M$.

This bound is weaker than the conjectured generalized Heawood inequality (2) and is clearly not optimal (as we already see in the special cases $k = 1$ or $b_k(M) = 0$).

Apart from applying more generally to almost embeddings, the hypotheses of Theorem 2 are weaker than those of Conjecture 1 in that we do not assume the manifold $M$ to be $(k - 1)$-connected. We conjecture that this connectedness assumption is not necessary for Conjecture 1, i.e., that (2) holds

---

1The inequality (1), which by a direct calculation is equivalent to $n \leq c(M) := \lfloor (7 + \sqrt{1 + 24b_1(M)})/2 \rfloor$, is closely related to the Map Coloring Problem for surfaces (which is the context in which Heawood originally considered the question). Indeed, it turns out that for surfaces $M$ other than the Klein bottle, $c(M)$ is the maximum chromatic number of any graph embeddable into $M$. For $M = S^2$ the 2-sphere (i.e., $b_1(M) = 0$), this is the Four-Color Theorem [AH77, AHK77]; for other surfaces (i.e., $b_1(M) > 0$) this was originally stated (with an incomplete proof) by Heawood and is now known as the Map Color Theorem or Ringel–Youngs Theorem [Rin74]. Interestingly, for surfaces $M \neq S^2$, there is a fairly short proof, based on edge counting and Euler characteristic, that the chromatic number of any graph embeddable into $M$ is at most $c(M)$ (see [Rin74, Thms. 4.2 and 4.8]) whereas the hard part is the tightness of (1).
whenever $\Delta_n^{(k)}$ almost embeds into a $2k$-manifold $M$. The intuition is that $\Delta_n^{(k)}$ is $(k-1)$-connected and therefore the image of an almost-embedding cannot “use” any parts of $M$ on which nontrivial homotopy classes of dimension less than $k$ are supported.

**Previous work.** The following special case of Conjecture 1 was proved by Kühnel [Küh94, Thm. 2] (and served as a motivation for the general conjecture): Suppose that $P$ is an $n$-dimensional simplicial convex polytope, and that there is a subcomplex of the boundary $\partial P$ of $P$ that is $k$-Hamiltonian (i.e., that contains the $k$-skeleton of $P$) and that is a triangulation of $M$, a $2k$-dimensional manifold. Then inequality (2) holds. To see that this is indeed a special case of Conjecture 1, note that $\partial P$ is a piecewise linear (PL) sphere of dimension $n-1$, i.e., $\partial P$ is combinatorially isomorphic to some subdivision of $\partial \Delta_n$ (and, in particular, $(n-2)$-connected). Therefore, the $k$-skeleton of $P$, and hence $M$, contains a subdivision of $\Delta_n^{(k)}$ and is $(k-1)$-connected.

In this special case and for $n \geq 2k+2$, equality in (2) is attained if and only if $P$ is a simplex. More generally, equality is attained whenever $M$ is a triangulated $2k$-manifold on $n+1$ vertices that is $(k+1)$-neighborly (i.e., any subset of at most $k+1$ vertices forms a face, in which case $\Delta_n^{(k)}$ is a subcomplex of $M$). Some examples of $(k+1)$-neighborly $2k$-manifolds are known, e.g., for $k=1$ (the so-called regular cases of equality for the Heawood inequality [Rin74]), for $k=2$ [KL83, KB83] (e.g., a 3-neighborly triangulation of the complex projective plane) and for $k=4$ [BK92], but in general, a characterization of the higher-dimensional cases of equality for (2) (or even of those values of the parameters for which equality is attained) seems rather hard (which is maybe not surprising, given how difficult the construction of examples of equality already is for $k=1$).

### 1.2 Generalization to points covered $q$ times

Kühnel’s conjecture can be recast in a broader setting suggested by a generalization by Sarkaria [Sar91, Thm 1.5] of the van Kampen–Flores Theorem. Sarkaria’s theorem states that if $q$ is a prime, and $d$ and $k$ integers such that $d \leq \frac{q}{q-1} k$, then for every continuous map $f : |\Delta_{qk+2q-2}^{(k)}| \to \mathbb{R}^d$ there are $q$ pairwise disjoint simplices $\sigma_1, \ldots, \sigma_q \in K$ with intersecting images $f(|\sigma_1|) \cap \cdots \cap f(|\sigma_q|) \neq \emptyset$. Sarkaria’s result was generalized by Volovikov [Vol96] for $q$ being a prime power.

Define an almost $q$-embedding of $K$ into a (Hausdorff) topological space $X$ as a continuous map $f : |K| \to X$ such that any $q$ pairwise disjoint faces $\sigma_1, \ldots, \sigma_q \in K$ have disjoint images $f(|\sigma_1|) \cap \cdots \cap f(|\sigma_q|) = \emptyset$. (So almost $2$-embeddings are almost embeddings.) Again, being an almost $q$-embedding depends on the actual simplicial complex (the triangulation), not just the underlying space. Our proof technique also gives an estimate for almost $q$-embeddings when $q$ is a prime power.

**Theorem 3.** Let $q = p^m$ be a prime power. If $\Delta_n^{(k)}$ $q$-almost-embeds into a $d$-manifold $M$ with $d \leq \frac{q}{q-1} k$ then

$$n \leq (q-2)k + 2q-2 \left(\frac{q^k + 2q-2}{k}\right)b_k(M) + (2q-2)k + 4q - 4,$$

where $b_k(M)$ is the $k$th $\mathbb{Z}_q$-Betti number of $M$.

Theorem 3 specializes for $q=2$ to Theorem 2.

### 1.3 Proof technique

Our proof of Theorem 3 strongly relies on a different generalization of the van Kampen–Flores Theorem, due to Volovikov [Vol96] regarding maps into general manifolds but under an additional homological triviality condition.

**Theorem 4 (Volovikov).** Let $q = p^m$ be a prime power. Let $f : |\Delta_{qk+2q-2}^{(k)}| \to M$ be a continuous map where $M$ is a compact $d$-manifold with $d \leq \frac{q}{q-1} k$. If the induced homomorphism

$$f_* : H_k(\Delta_{qk+2q-2}^{(k)}; \mathbb{Z}_q) \to H_k(M; \mathbb{Z}_q)$$

is trivial then $f$ is not a $q$-almost embedding.
Theorem 4 is obtained by specializing the corollary in Volovikov’s article [Vol96] to \( m = d \) and \( s = k + 1 \). Note that Volovikov [Vol96] formulates the triviality condition in terms of cohomology, i.e., he requires that \( f^* : H^k(M; \mathbb{Z}_p) \to H^k(\Delta_{2k+2}^{|\Delta_n|}; \mathbb{Z}_p) \) is trivial. However, since we are working with field coefficients and the (co)homology groups in question are finitely generated, the homological triviality condition (which is more convenient for us to work with) and the cohomological one are equivalent.\(^2\) Note that the homological triviality condition is automatically satisfied if \( H_k(M; \mathbb{Z}_p) = 0 \), e.g., if \( M = \mathbb{R}^{2k} \) or \( M = S^{2k} \). On the other hand, without the homological triviality condition, the assertion is in general not true for other manifolds (e.g., \( K_5 \) embeds into every closed surface different from the sphere, or \( \Delta_5^{(2)} \) embeds into the complex projective plane).

The key idea of our approach is to show that if \( n \) is large enough and \( f \) is a mapping from \( \Delta_n^{(k)} \) to \( M \), then there is a \( q \)-almost-embedding \( g \) from \( \Delta_n^{(k)} \) to \( |\Delta_n^{(k)}| \) for some prescribed value of \( s \) such that the composed map \( f \circ g : \Delta_n \to M \) satisfies Volovikov’s condition. More specifically, the following is our main technical lemma:

**Lemma 5.** Let \( k, s \geq 1 \) and \( b \geq 0 \) be integers. Let \( p \) be a prime number. There exists a value \( n_0 := n_0(k, b, s, p) \) with the following property. Let \( n \geq n_0 \) and let \( f \) be a mapping of \( |\Delta_n^{(k)}| \) into a manifold \( M \) with \( k \)-th \( \mathbb{Z}_p \)-Betti number at most \( b \). Then there exists a subdivision \( D \) of \( \Delta_s^{(k)} \) and a simplicial map \( g_{\text{simp}} : D \to \Delta_n^{(k)} \) with the following properties.

1. The induced map on the geometric realizations \( g : |D| = |\Delta_s^{(k)}| \to |\Delta_n^{(k)}| \) is an almost-embedding from \( \Delta_s^{(k)} \) to \( |\Delta_n^{(k)}| \).
2. The homomorphism \( (f \circ g)_* : H_k(\Delta_s^{(k)}; \mathbb{Z}_p) \to H_k(M; \mathbb{Z}_p) \) is trivial (see Section 2 below for the precise interpretation of \( (f \circ g)_* \)).

The value \( n_0 \) can be taken as \( (\binom{k}{s})b(s - 2k) + 2s - 2k + 1 \).

Therefore, if \( s \geq qk + 2q - 2 \), then \( f \circ g \) cannot be a \( q \)-almost embedding by Volovikov’s theorem. We deduce that \( f \) is not a \( q \)-almost-embedding either, and Theorem 3 immediately follows. This deduction requires the following lemma (proven in Section 2) as in general, a composition of a \( q \)-almost-embedding and an almost-embedding is not always a \( q \)-almost-embedding.

**Lemma 6.** Let \( K \) and \( L \) be simplicial complexes and \( X \) a topological space. Suppose \( g \) is an almost embedding of \( K \) into \( |L| \) and \( f \) is a \( q \)-almost embedding of \( L \) into \( X \) for some integer \( q \geq 2 \). Then \( f \circ g \) is a \( q \)-almost embedding of \( K \) into \( X \), provided that \( g \) is the realization of a simplicial map \( g_{\text{simp}} \) from some subdivision \( K' \) of \( K \) to \( L \).

**Remark 7.** The third author proved in his thesis [Pat15] a slightly better bound on \( n_0 \) in Lemma 5, namely \( n_0 = \left(\binom{k}{s}\right)b(s - 2k) + s + 1 \). The proof, however, uses colorful version of Lemma 12. Since the proof of the colorful version is long and technical and in the end it only improves the bound in Theorem 2 by 2, we have decided to present the more accessible version of the argument.

**Paper organization.** Before we establish Lemma 5 (in Section 4), thus completing the proof of Theorem 3, we first prove a weaker version that introduces the main ideas in a simpler setting, and yields a weaker bound for \( n_0 \), stated in Equation (4). The reader interested only in the case \( q = 2 \) may want to consult a preliminary version of this paper [GMP+15] tailored to that case (where homology computations are without signs and the construction of the subdivision \( D \) is simpler).

## 2 Preliminaries

We begin by fixing some terminology and notation. We will use \( \text{card}(U) \) to denote the cardinality of a set \( U \).

\(^2\)More specifically, by the Universal Coefficient Theorem [Mun84, 53.5], \( H_k(\cdot; \mathbb{Z}_p) \) and \( H^k(\cdot; \mathbb{Z}_p) \) are dual vector spaces, and \( f^* \) is the adjoint of \( f_* \), hence triviality of \( f_* \) implies that of \( f^* \). Moreover, if the homology group \( H_k(X; \mathbb{Z}_p) \) of a space \( X \) is finitely generated (as is the case for both \( \Delta_n^{(k)} \) and \( M \), by assumption) then it is (non-canonically) isomorphic to its dual vector space \( H^k(X; \mathbb{Z}_p) \). Therefore, \( f_* \) is trivial if and only if \( f^* \) is.
We recall that the **stellar subdivision** of a maximal face \( \vartheta \) in a simplicial complex \( K \) is obtained by removing \( \vartheta \) from \( K \) and adding a cone \( a_\vartheta \ast (\partial \vartheta) \), where \( a_\vartheta \) is a newly added vertex, the apex of the cone (see Figure 1).

Throughout this paper we only work with homology groups and Betti numbers over \( \mathbb{Z}_p \), and for simplicity, we mostly drop the coefficient group \( \mathbb{Z}_p \) from the notation. Moreover, we will need to switch back and forth between singular and simplicial homology. More precisely, if \( K \) is a simplicial complex then \( H_*(K) \) will mean the simplicial homology of \( K \), whereas \( H_*(X) \) will mean the singular homology of a topological space \( X \). In particular, \( H_*(|K|) \) denotes the singular homology of the underlying space \( |K| \) of a complex \( K \). We use analogous conventions for \( C_*(K), C_*(X) \) and \( C_*(|K|) \) on the level of chains, and likewise for the subgroups of cycles and boundaries, respectively.\(^3\) Given a cycle \( c \), we denote by \([c]\) the homology class it represents.

![Figure 1: A stellar subdivision of a simplex.](image)

A mapping \( h: |K| \to X \) induces a chain map \( h_*: C_*(|K|) \to C_*(X) \) on the level of singular chains; see [Hat02, Chapter 2.1]. There is also a canonical chain map \( \iota_K: C_*(K) \to C_*(|K|) \) inducing the isomorphism of \( H_*(K) \) and \( H_*(|K|) \), see again [Hat02, Chapter 2.1]. We define \( h_\text{sing}: C_*(K) \to C_*(X) \) as \( h_\text{sing} := h_* \circ \iota_K \). The three chain maps mentioned above also induce maps \( h_* \), \( (\iota_K)_* \), and \( h_* \) on the level of homology satisfying \( h_* = h_* \circ (\iota_K)_* \). We need a technical lemma saying that our maps compose, in a right way, on the level of homology.

**Lemma 8.** Let \( K \) and \( L \) be simplicial complexes and \( X \) a topological space. Let \( j_\text{simp} \) be a simplicial map from \( K \) to \( L \), \( j: |K| \to |L| \) the continuous map induced by \( j_\text{simp} \) and \( h: |L| \to X \) be another continuous map. Then \( h_* \circ (j_\text{simp})_* = (h \circ j)_* \) where \((j_\text{simp})_*: H_*(K) \to H_*(L)\) is the map induced by \( j_\text{simp} \) on the level of simplicial homology and the maps \( h_* \) and \((h \circ j)_*\) are as defined above.

**Proof.** The proof follows from the commutativity of the diagram below.

\[
\begin{array}{ccc}
H_*(|K|) & \xrightarrow{h_\text{sing}} & H_*(X) \\
\downarrow^{(\iota_K)_*} & & \downarrow^{h_*} \\
H_*(K) & \xrightarrow{(j_\text{simp})_*} & H_*(L)
\end{array}
\]

The commutativity of the lower right triangle follows from the definition of \( h_* \). Similarly \((h \circ j)_* = (h \circ j)_\text{sing} \circ (\iota_K)_*\). The fact that \((h \circ j)_\text{sing} = h_\text{sing} \circ j_\text{sing}\) follows from functoriality of the singular homology. The commutativity of the square follows from the naturality of the equivalence of the singular and simplicial homology; see [Mun84, Thm 34.4]. \(\square\)

We now prove the final technical step of our approach, stated in the introduction.

---

\(^3\)We remark that throughout this paper, we will only work with spaces that are either (underlying spaces of) simplicial complexes or topological manifolds. Such spaces are homotopy equivalent to CW complexes [Mil59, Corollary 1], and so on the matter of homology, it does not really matter which (ordinary, i.e., satisfying the dimension axiom) homology theory we use as they are all naturally equivalent for CW complexes [Hat02, Thm. 4.59]. However the distinction between the simplicial and the singular setting will be relevant on the level of chains.
Proof of Lemma 6. Let $σ_1, \ldots, σ_q$ be $q$ pairwise disjoint faces of $K$. Our task is to show $f \circ g(|σ_1|) \cap \cdots \cap f \circ g(|σ_q|) = \emptyset$. Let $φ_i$ be a face of $K'$ that subdivides $σ_i$, for $i \in [q]$. We are done, if we prove
\[
f \circ g(|φ_1|) \cap \cdots \cap f \circ g(|φ_q|) = \emptyset
\] (3)
for every such possible choice of $φ_1, \ldots, φ_q$.

The faces $φ_1, \ldots, φ_q$ are pairwise disjoint since $σ_1, \ldots, σ_q$ are pairwise disjoint. Since $g_{\text{simp}}$ is a simplicial map inducing an almost embedding, the faces $g_{\text{simp}}(φ_1), \ldots, g_{\text{simp}}(φ_q)$ are pairwise disjoint faces of $L$. Consequently, (3) follows from the fact that $f$ is a $q$-almost embedding. □

3 Proof of Lemma 5 with a weaker bound on $n_0$

Let $k, b, s$ be fixed integers. We consider a $2k$-manifold $M$ with $k$th Betti number $b$, a map $f : |Δ_n| \to M$. Recall that although we want to build an almost-embedding, homology is computed over $\mathbb{Z}_p$. The strategy of our proof of Lemma 5 is to start by designing an auxiliary chain map

\[
φ : C_* \left( Δ^{(k)}_n \right) \to C_* \left( Δ^{(k)}_n \right).
\]

that behaves as an almost-embedding, in the sense that whenever $σ$ and $σ'$ are disjoint $k$-faces of $Δ_n$, $φ(σ)$ and $φ(σ')$ have disjoint supports, and such that for every $(k + 1)$-face $τ$ of $Δ_n$, the homology class $[(f_2 \circ φ)(∂τ)]$ is trivial. We then use $φ$ to design a subdivision $D$ of $Δ^{(k)}_n$ and a simplicial map $g_{\text{simp}} : D \to Δ^{(k)}_n$ that induces a map $g : |D| \to |Δ^{(k)}_n|$ with the desired properties: $g$ is an almost-embedding and $(f \circ g), ([∂τ])$ is trivial for all $(k + 1)$-faces $τ$ of $Δ_n$. Since the cycles $∂τ$, for $(k + 1)$-faces $τ$ of $Δ_n$, generate all $k$-cycles of $Δ^{(k)}_n$, this implies that $(f \circ g)_*$ is trivial.

The purpose of this section is to give a first implementation of the above strategy that proves Lemma 5 with a bound of

\[
n_0 \geq \binom{s + 1}{k + 1} p^{b(k + 1)} + s + 1.
\]

(4)

In Section 4 we then improve this bound to $\binom{s}{k} b(s - 2k) + 2s - 2k + 1$ at the cost of some technical complications (note that the improved bound is independent of $p$).

Throughout the rest of this paper we use the following notations. We let $\{v_1, v_2, \ldots, v_{n+1}\}$ denote the set of vertices of $Δ_n$ and we assume that $Δ_n$ is the induced subcomplex of $Δ_n$ on $\{v_1, v_2, \ldots, v_{n+1}\}$. We let $U = \{v_{n+2}, v_{n+3}, \ldots, v_{n+1}\}$ denote the set of vertices of $Δ_n$ unused by $Δ_n$. We let $m = \binom{s+1}{k+1}$ and denote by $σ_1, σ_2, \ldots, σ_m$ the $k$-faces of $Δ_n$, ordered lexicographically.

Later on, when working with homology, we compute the simplicial homology with respect to this fixed order on the vertices of $Δ_n$. In particular, the boundary of a $j$-simplex $ν = \{v_{i_1}, v_{i_2}, \ldots, v_{i_j}\}$ where $i_1 \leq i_2 \leq \cdots \leq i_j$ is

\[
∂ν = \sum_{ℓ=1}^{j+1} (-1)^{j+1} ν \setminus \{v_{i_ℓ}\}.
\]

3.1 Construction of $φ$

For every face $φ$ of $Δ_n$ of dimension at most $k - 1$ we set $φ(φ) = φ$. We then “route” each $σ_i$ by mapping it to its stellar subdivision with an apex $u \in U$, i.e. by setting $φ(σ_i)$ to $σ_i + (-1)^k z(σ_i, u)$ where $z(σ_i, u)$ denotes the cycle $∂(σ_i \cup \{u\})$; see Figure 2 for the case $k = 1$.

We ensure that $φ$ behaves as an almost-embedding by using a different apex $u \in U$ for each $σ_i$. The difficulty is to choose these $m$ apices in a way that $[f_2(φ(∂τ))]$ is trivial for every $(k + 1)$-face $τ$ of $Δ_n$. To that end we associate to each $u \in U$ the sequence

\[
ν(u) := ([f_2(z(σ_1, u))], [f_2(z(σ_2, u))], \ldots, [f_2(z(σ_m, u))]) \in H_k(M)^m,
\]

and we denote by $ν_i(u)$ the $i$th element of $ν(u)$. We work with $\mathbb{Z}_p$-homology, so $H_k(M)^m$ is finite; more precisely, its cardinality equals $p^{bm}$. From $n \geq n_0 = (m - 1)p^{bm} + s + 1$ we get that $\text{card}(U) \geq (m - 1)\text{card}(H_k(M)^m) + 1$. 

6
The pigeonhole principle then guarantees that there exist \( m \) distinct vertices \( u_1, u_2, \ldots, u_m \) of \( U \) such that \( v(u_1) = v(u_2) = \cdots = v(u_m) \). We use \( u_i \) to “route” \( \sigma_i \) and put

\[
\varphi(\sigma_i) := \sigma_i + (-1)^k z(\sigma_i, u_i). \tag{5}
\]

We finally extend \( \varphi \) linearly to \( C_*(\Delta_s^{(k)}) \).

**Lemma 9.** The map \( \varphi \) is a chain map and \( [f_\sharp(\varphi(\partial\tau))] = 0 \) for every \( (k+1) \)-face \( \tau \in \Delta_s \).

Before proving the lemma, we establish a simple claim that will also be useful later.

**Claim 10.** Let \( \tau \) be a \( (k + 1) \)-face of \( \Delta_s \) and let \( u \in U \). Let \( \sigma_{i_1}, \ldots, \sigma_{i_{k+2}} \) be all the \( k \)-faces of \( \tau \) sorted lexicographically, that is, \( i_1 \leq \cdots \leq i_{k+2} \). Then

\[
\partial\tau = z(\sigma_{i_1}, u) - z(\sigma_{i_2}, u) + \cdots + (-1)^{k+1} z(\sigma_{i_{k+2}}, u). \tag{6}
\]

**Proof.** This follows from expanding the equation \( 0 = \partial^2(\tau \cup \{u\}) \). Indeed,

\[
0 = \partial^2(\tau \cup \{u\}) = \partial(\sigma_{i_{k+2}} \cup \{u\}) - \sigma_{i_{k+1}} \cup \{u\} + \cdots + (-1)^{k+1} \sigma_{i_1} \cup \{u\} + (-1)^{k+2} \tau \\
= (-1)^{k+1} (-\partial\tau + z(\sigma_{i_1}, u) - z(\sigma_{i_2}, u) + \cdots + (-1)^{k+1} z(\sigma_{i_{k+2}}, u)).
\]

\( \square \)

**Proof of Lemma 9.** The map \( \varphi \) is the identity on \( \ell \)-chains with \( \ell \leq k-1 \) and Equation (5) immediately implies that \( \partial \varphi(\sigma) = \partial \sigma \) for every \( k \)-simplex \( \sigma \). It follows that \( \varphi \) is a chain map.

Now let \( \tau \) be a \( (k+1) \)-simplex of \( \Delta_s \) and let \( \sigma_{i_1}, \ldots, \sigma_{i_{k+2}} \) be its \( k \)-faces. We have

\[
f_\sharp \circ \varphi(\partial\tau) = f_\sharp \circ \varphi \left( \sum_{j=1}^{k+2} (-1)^{j+k} \sigma_{i_j} \right) = f_\sharp \left( \sum_{j=1}^{k+2} (-1)^{j+k} \left( \sigma_{i_j} + (-1)^k z(\sigma_{i_j}, u_{i_j}) \right) \right) \\
= f_\sharp(\partial\tau) + \sum_{j=1}^{k+2} (-1)^j f_\sharp(z(\sigma_{i_j}, u_{i_j})).
\]

\( [f_\sharp(z(\sigma_{i_\ell}, u_{i_\ell}))] = v_{i_\ell}(u_{i_\ell}) \) is independent of the value \( \ell \). When passing to the homology classes in the above identity, we can therefore replace each \( u_{i_\ell} \) with \( u_1 \), and obtain,

\[
[f_\sharp \circ \varphi(\partial\tau)] = [f_\sharp(\partial\tau)] + \sum_{j=1}^{k+2} (-1)^j [f_\sharp(z(\sigma_{i_j}, u_{i_1}))] = [f_\sharp(\partial\tau + \sum_{j=1}^{k+2} (-1)^j z(\sigma_{i_j}, u_{i_1}))].
\]

This class is trivial by Claim 10. Figure 3 illustrates the geometric intuition behind this proof. \( \square \)
3.2 Subdivisions and orientations

Our next task is the construction of \( D \) and \( g \); however, we first mention a few properties of subdivisions.

Let us consider a simplicial complex \( K \) and a subdivision \( S \) of \( K \). (So \( K \) and \( S \) are regarded as geometric simplicial complexes, and for every simplex \( \eta \) of \( S \) there is a simplex \( \vartheta \) of \( K \) such that \( \eta \subseteq \vartheta \). In this case, we say that \( \eta \) subdivides \( \vartheta \).) There is a canonical chain map \( \rho: C_*(K) \to C_*(S) \) that induces an isomorphism in homology. Intuitively, \( \rho \) maps a simplex \( \vartheta \) of \( K \) to a sum of simplices of \( S \) of the same dimension that subdivide \( \vartheta \). However, we have to be careful about the \( \pm 1 \) coefficients in the sum.

We work with the ordered simplicial homology, that is, we order the vertices of \( K \) as well as the vertices of \( S \). We want to define the mutual orientation \( \mathrm{or}(\eta, \vartheta) \in \{-1, 1\} \) of a \( j \)-simplex \( \vartheta \) of \( K \) and a \( j \)-simplex \( \eta \) of \( S \) that subdivides \( \vartheta \). We set \( \mathrm{or}(\eta, \vartheta) \) to be \( 1 \) if the orientations of \( \vartheta \) and \( \eta \) agree, and \( -1 \) if they disagree; the orientation of each geometric simplex is computed relative to the order of its vertices in \( K \) or \( S \) (with respect to a fixed base of their common affine hull, say). Then we set

\[
\rho(\vartheta) = \sum_{\eta} \mathrm{or}(\eta, \vartheta) \eta
\]

where the sum is over all simplices \( \eta \) in \( S \) of the same dimension as \( \vartheta \) which subdivide \( \vartheta \). Finally, we extend \( \rho \) to a chain map. It is routine to check that \( \rho \) commutes with the boundary operator and that it induces an isomorphism on homology. It is also useful to describe \( \rho \) in the specific case where \( S \) is a stellar subdivision of a complex \( K \) consisting of a single \( k \)-simplex. Here, we assume that \( w_1, \ldots, w_{j+1} \) are the vertices of \( K \) in this order (in \( K \) as well as in \( S \)) and \( a \) is the apex of \( S \), which comes last in the order on \( S \). We also consider \( S \) as a subcomplex of the \((k + 1)\)-simplex on \( w_1, \ldots, w_{k+1}, a \). And we use the notation \( z(\vartheta, a) = \vartheta \cup \{a\} \), analogously as previously in the case of \( k \)-faces of \( \Delta_s \).

**Lemma 11.** In the setting above, let \( \vartheta \) be the \( k \)-face of \( K \). Then \( \rho(\vartheta) = \vartheta + (-1)^{k+1} z(\vartheta, a) \).

**Proof.** Let \( \eta_i := \vartheta \cup \{a\} \setminus \{w_i\} \) for \( i \in [k+1] \). Then \( \eta_i \) are all faces of \( S \) subdividing \( \vartheta \). We have \( \mathrm{or}(\eta_i, \vartheta) = (-1)^{i+k+1} \) as \( \eta_i \) has the same orientation as \( \vartheta \) with respect to a modified order of vertices of \( \vartheta \) obtained by replacing \( w_i \) with \( a \). Therefore \( \rho(\vartheta) = \sum_{i=1}^{k+1} (-1)^{i+k+1} \eta_i \). On the other hand,

\[
z(\vartheta, a) = \partial(\vartheta \cup \{a\}) = \sum_{i=1}^{k+1} (-1)^{i+k+1} \eta_i + (-1)^k \vartheta = (-1)^k (\rho(\vartheta) - \vartheta).
\]

\[\square\]

3.3 Construction of \( D \) and \( g \)

The definition of \( \varphi \), an in particular Equation (5), suggests to construct our subdivision \( D \) of \( \Delta_s^{(k)} \) by simply replacing every \( k \)-face of \( \Delta_s^{(k)} \) by its stellar subdivision. Let \( a_i \) denote the new vertex introduced
when subdividing $\sigma_i$. We fix a linear order on vertices of $D$ in such a way that we reuse the order of vertices that also belong to $\Delta_n^{(k)}$ and then the vertices $u_i$ follow in arbitrary order.

We define a simplicial map $g_{\text{simp}}: D \to \Delta_n^{(k)}$ by putting $g_{\text{simp}}(v) = v$ for every original vertex $v$ of $\Delta_n^{(k)}$, and $g_{\text{simp}}(u_i) = u_i$ for $i \in [m]$. This $g_{\text{simp}}$ induces a map $g: |\Delta_n^{(k)}| \to |\Delta_n^{(k)}|$ on the geometric realizations. Since the $u_i$’s are pairwise distinct, $g$ is an embedding\footnote{We use the full strength of almost-embeddings when proving Lemma 5 with the better bound on $n_0$.}, so Condition 1 of Lemma 5 holds.

In principle, we would like to derive Condition 2 of Lemma 5 by observing that $g$ ‘induces’ a chain map from $C_*(\Delta_n^{(k)})$ to $C_*(\Delta_n^{(k)})$ that coincides with \(\varphi\). Making this a formal statement is thorny because $g$, as a continuous map, naturally induces a chain map $g_\ast$ on singular rather than simplicial chains. We can’t use directly $g_{\text{simp}}$ either, since we are interested in a map from $C_*(\Delta_n^{(k)})$ and not from $C_*(D)$.

We handle this technicality as follows. We consider the chain map $\rho: C_*(\Delta_n^{(k)}) \to C_*(D)$ from (7). This map induces an isomorphism $\rho_*$ in homology. In addition $\varphi = (g_{\text{simp}})_\ast \circ \rho$ where $(g_{\text{simp}})_\ast: C_*(D) \to C_*(\Delta_n^{(k)})$ denotes the (simplicial) chain map induced by $g_{\text{simp}}$. Indeed, all three maps are the identity on simplices of dimension at most $k-1$. For a $k$-simplex $\sigma$, the map $g_{\text{simp}}$ is an order preserving isomorphism when restricted to the subdivision of $\sigma$ (in $D$). Therefore, the required equality $\varphi(\sigma) = (g_{\text{simp}})_\ast \circ \rho(\sigma)$ follows from (5) and Lemma 11.

We thus have in homology

$$f_\ast \circ \varphi_\ast = f_\ast \circ (g_{\text{simp}})_\ast \circ \rho_\ast$$

and since $\rho_\ast$ is an isomorphism and $f_\ast \circ \varphi_\ast$ is trivial by Lemma 9, it follows that $f_\ast \circ (g_{\text{simp}})_\ast$ is also trivial. Since $f_\ast \circ (g_{\text{simp}})_\ast = (f \circ g)_\ast$, by Lemma 8, $(f \circ g)_\ast$ is trivial as well. This concludes the proof of Lemma 5 with the weaker bound.

4 Proof of Lemma 5

We now prove Lemma 5 with the bound claimed in the statement, namely

$$n_0 = \binom{s}{k}b(s - 2k) + 2s - 2k + 1.$$ 

Let $k, b, s$ be fixed integers. We consider a $2k$-manifold $M$ with $k$th $\mathbb{Z}_p$-Betti number $b$, a map $f: |\Delta_n^{(k)}| \to M$, and we assume that $n \geq n_0$.

The proof follows the same strategy as in Section 3: we construct a chain map $\varphi: C_*(\Delta_n^{(k)}) \to C_*(\Delta_n^{(k)})$ such that the homology class $[(f_\ast \circ \varphi_\ast)(\partial \tau)]$ is trivial for all $(k + 1)$-faces $\tau$ of $\Delta_n$, then upgrade $\varphi$ to a continuous map $g: |\Delta_n^{(k)}| \to |\Delta_n^{(k)}|$ with the desired properties.

When constructing $\varphi$, we refine the arguments of Section 3 to “route” each $k$-face using not only one, but several vertices from $U$; this makes finding “collisions” easier, as we can use linear algebra arguments instead of the pigeonhole principle. This comes at the cost that when upgrading $g$, we must content ourselves with proving that it is an almost-embedding. This is sufficient for our purpose and has an additional benefit: the same group of vertices from $U$ may serve to route several $k$-faces provided they pairwise intersect in $\Delta_n^{(s)}$.

4.1 Construction of $\varphi$

We use the same notation regarding $v_1, \ldots, v_{n+1}$, $\Delta_n$, $\Delta_n$, $U$, $m = \binom{s+1}{k+1}$ and $\sigma_1, \sigma_2, \ldots, \sigma_m$ as in Section 3.

Multipoints and the map $v$. As we said we plan to route $k$-faces of $\Delta_n$ through certain collections of vertices from $U$ (weighted); we will call these collections multipoints. It is more convenient to work with them on the level of formal linear combinations. Let $C_0(U)$ denote the $\mathbb{Z}_p$-vector space of formal linear combinations of vertices from $U$. A multpoint is an element of $C_0(U)$ whose coefficients sum to $1$ (in $\mathbb{Z}_p$, of course). The multipoints form an affine subspace of $C_0(U)$ which we denote by $\mathcal{M}$. The support, $\text{sup}(\mu)$, of a multpoint $\mu \in \mathcal{M}$ is the set of vertices $v \in U$ with non-zero coefficient in $\mu$. We say that two multipoints are disjoint if their supports are disjoint.
For any $k$-face $\sigma_i$ and any multipoint $\mu = \sum_{u \in U} \lambda_u u$ we define:

$$z(\sigma_i, \mu) := \sum_{u \in \text{sup}(\mu)} \lambda_u z(\sigma_i, u) := \sum_{u \in \text{sup}(\mu)} \lambda_u \partial(\sigma_i \cup \{ u \}).$$

Now, we proceed as in Section 3 but replace unused points by multipoints of $\mathcal{M}$ and the cycles $z(\sigma_i, \mu)$ with the cycles $z(\sigma_i, \mu)$. Since $\mathbb{Z}_p$ is a field, $H_k(\mathcal{M})^m$ is a vector space and we can replace the sequences $\mathbf{v}(u)$ of Section 3 by the linear map

$$\mathbf{v} : \{ \mathcal{M} \to H_k(\mathcal{M})^m \}$$

**Finding collisions.** The following lemma takes advantage of the vector space structure of $H_k(\mathcal{M})^m$ and the affine structure of $\mathcal{M}$ to find disjoint multipoints $\mu_1, \mu_2, \ldots$ to route the $\sigma_i$‘s more effectively than by simple pigeonhole.

**Lemma 12.** For any $r \geq 1$, any $\mathbb{Z}_p$-vector space $V$, and any affine map $\psi : \mathcal{M} \to V$, if $\text{card}(U) \geq (\dim(\psi(\mathcal{M}))) + 1)(r - 1) + 1$ then $\mathcal{M}$ contains $r$ disjoint multipoints $\mu_1, \mu_2, \ldots, \mu_r$, such that $\psi(\mu_1) = \psi(\mu_2) = \cdots = \psi(\mu_r)$.

**Proof.** Let us write $U = \{ v_{s+2}, v_{s+3}, \ldots, v_{n+1} \}$ and $d = \dim(\psi(\mathcal{M}))$. We first prove by induction on $r$ the following statement:

If $\text{card}(U) \geq (d + 1)(r - 1) + 1$ there exist $r$ pairwise disjoint subsets $I_1, I_2, \ldots, I_r \subseteq U$ whose image under $\psi$ have affine hulls with non-empty intersection.

(This is, in a sense, a simple affine version of Tverberg’s theorem.) The statement is obvious for $r = 1$, so assume that $r \geq 2$ and that the statement holds for $r - 1$. Let $A$ denote the affine hull of $\{ \psi(v_{s+2}), \psi(v_{s+3}), \ldots, \psi(v_{n+1}) \}$ and let $I_r$ denote a minimal cardinality subset of $U$ such that the affine hull of $\{ \psi(v) : v \in I_r \}$ equals $A$. Since $\dim A \leq d$ the set $I_r$ has cardinality at most $d + 1$. The cardinality of $U \setminus I_r$ is at least $(d + 1)(r - 2) + 1$ so we can apply the induction hypothesis for $r - 1$ to $U \setminus I_r$. We thus obtain $r - 1$ disjoint subsets $I_1, I_2, \ldots, I_{r-1}$ whose images under $\psi$ have affine hulls with non-empty intersection. Since the affine hull of $\psi(U \setminus I_r)$ is contained in the affine hull of $\psi(I_r)$, the claim follows.

Now, let $a \in V$ be a point common to the affine hulls of $\psi(I_1), \psi(I_2), \ldots, \psi(I_r)$. Writing $a$ as an affine combination in each of these spaces, we get

$$a = \sum_{u \in J_1} \lambda^{(1)}_u \psi(u) = \sum_{u \in J_2} \lambda^{(2)}_u \psi(u) = \cdots = \sum_{u \in J_r} \lambda^{(r)}_u \psi(u)$$

where $J_j \subseteq I_j$ and $\sum_{u \in J_j} \lambda^{(j)}_u = 1$ for any $j \in [r]$. Setting $\mu_j = \sum_{u \in J_j} \lambda^{(j)}_u u$ finishes the proof.

**Computing the dimension of $\mathbf{v}(\mathcal{M})$.** Having in mind to apply Lemma 12 with $V = H_k(\mathcal{M})^m$ and $\psi = \mathbf{v}$, we now need to bound from above the dimension of $\mathbf{v}(\mathcal{M})$. An obvious upper bound is $\dim H_k(\mathcal{M})^m$, which equals $bm = b\binom{k+1}{k}$. A better bound can be obtained by an argument analogous to the proof of Lemma 9. We first extend Claim 10 to multipoints.

**Claim 13.** Let $\tau$ be a $(k + 1)$-face of $\Delta_s$ and let $\mu \in \mathcal{M}$. Let $\sigma_{i_1}, \ldots, \sigma_{i_{k+2}}$ be all the $k$-faces of $\tau$ sorted lexicographically. Then

$$\partial \tau = z(\sigma_{i_1}, \mu) - z(\sigma_{i_2}, \mu) + \cdots + (-1)^{k+1}z(\sigma_{i_{k+2}}, \mu).$$

(8)

**Proof.** By Claim 10 we know that (8) is true for points. For a multipoint $\mu = \sum_{u \in U} \lambda_u u$, we get (8) as a linear combination of equations (6) for the points $u$ with the ‘weight’ $\lambda_u$ (note that $\sum_{u \in U} \lambda_u = 1$; therefore the corresponding combination of the left-hand sides of (6) equals $\partial \tau$).

**Lemma 14.** $\dim(\mathbf{v}(\mathcal{M})) \leq b\binom{s}{k}$. 

10
Proof. Let \( \tau \) be a \((k+1)\)-face of \( \Delta_s \) and let \( \sigma_1, \ldots, \sigma_{k+2} \) denote its \( k \)-faces. For any multipoint \( \mu \), Claim 13 implies

\[
[f_\mu(\partial \tau)] = \sum_{j=1}^{k+2} (-1)^{j+1} [f_\mu(z(\sigma_j, \mu))] = \sum_{j=1}^{k+2} (-1)^{j+1} v_j(\mu);
\]

therefore

\[
v_{k+2}(\mu) = (-1)^{k+1} [f_\mu(\partial \tau)] + \sum_{j=1}^{k+1} (-1)^{j+k+1} v_j(\mu).
\]

Each vector \( v(\mu) \) is thus determined by the values of the \( v_j(\mu) \)'s where \( \sigma_j \) contains the vertex \( v_1 \). Indeed, the vectors \([f_\mu(\partial \tau)]\) are independent of \( \mu \), and for any \( \sigma_j \) not containing \( v_1 \) we can eliminate \( v_j(\mu) \) by considering \( \tau := \sigma_i \cup \{v_1\} \) (and setting \( \sigma_{k+2} = \sigma_i \)). For each of the \( \binom{k}{2} \) faces \( \sigma_j \) that contain \( v_1 \), the vector \( v_j(\mu) \) takes values in \( H_k(M) \) which has dimension at most \( b \). It follows that \( \dim v(M) \leq b \binom{k}{2} \). \( \square \)

**Coloring graphs to reduce the number of multipoints used.** We could now apply Lemma 12 with \( r = m \) to obtain one multipoint per \( k \)-face, all pairwise disjoint, to proceed with our “routing”. As mentioned above, however, we only need that \( \varphi \) is an almost-embedding, so we can use the same multipoint for several \( k \)-faces provided they pairwise intersect. Optimizing the number of multipoints used reformulates as the following graph coloring problem:

Assign to each \( k \)-face \( \sigma_i \) of \( \Delta_s \) some color \( c(i) \in \mathbb{N} \) such that \( \text{card}\{c(i) : 1 \leq i \leq m\} \) is minimal and disjoint faces use distinct colors.

This question is classically known as Kneser’s graph coloring problem and an optimal solution uses \( s - 2k + 1 \) colors \([\text{Lov78, Mat03}]\). Let us spell out one such coloring (proving its optimality is considerably more difficult, but we do not need to know that it is optimal). For every \( k \)-face \( \sigma_i \) we let \( \min \sigma_i \) denote the smallest index of a vertex in \( \sigma_i \). When \( \min \sigma_i \leq s - 2k \) we set \( c(i) = \min \sigma_i \), otherwise we set \( c(i) = s - 2k + 1 \). Observe that any \( k \)-face with color \( c \leq s - 2k \) contains vertex \( v_c \). Moreover, the \( k \)-faces with color \( s - 2k + 1 \) consist of \( k + 1 \) vertices each, all from a set of \( 2k + 1 \) vertices. It follows that any two \( k \)-faces with the same color have some vertex in common.

**Defining \( \varphi \).** We are finally ready to define the chain map \( \varphi : C_\ast(\Delta_s^{(k)}) \to C_\ast(\Delta_0^{(k)}) \). Recall that we assume that \( n \geq n_0 = (\binom{k}{2}b + 1)(r - 1) + s + 1 \). Using the bound of Lemma 14 we can apply Lemma 12 with \( r = s - 2k + 1 \), obtaining \( s - 2k + 1 \) multipoints \( \mu_1, \mu_2, \ldots, \mu_{s - 2k + 1} \in M \). We set \( \varphi(\vartheta) = \vartheta \) for any face \( \vartheta \) of \( \Delta_s \) of dimension less than \( k \). We then “route” each \( k \)-face \( \sigma_i \) through the multipoint \( \mu_{c(i)} \) by putting

\[
\varphi(\sigma_i) := \sigma_i + (-1)^{k}z(\sigma_i, \mu_{c(i)}),
\]

where \( c(i) \) is the color of \( \sigma_i \) in the coloring of the Kneser graph proposed above. We finally extend \( \varphi \) linearly to \( C_\ast(\Delta_s) \).

We need the following analogue of Lemma 9.

**Lemma 15.** The map \( \varphi \) is a chain map and \([f_\mu(\varphi(\partial \tau))] = 0\) for every \((k+1)\)-face \( \tau \in \Delta_s \).

The proof of Lemma 15 is very similar to the proof of Lemma 9; it just replaces points with multipoints and Claim 10 with Claim 13. We therefore omit the proof. We next argue that \( \varphi \) behaves like an almost embedding.

**Lemma 16.** For any two disjoint faces \( \vartheta, \eta \) of \( \Delta_s^{(k)} \), the supports of \( \varphi(\vartheta) \) and \( \varphi(\eta) \) use disjoint sets of vertices.

**Proof.** Since \( \varphi \) is the identity on chains of dimension at most \((k - 1)\), the statement follows if neither face has dimension \( k \). For any \( k \)-chain \( \sigma_i \), the support of \( \varphi(\sigma_i) \) uses only vertices from \( \sigma_i \) and from the support of \( \mu_{c(i)} \). Since each \( \mu_{c(i)} \) has support in \( U \), which contains no vertex of \( \Delta_s \), the statement also holds when exactly one of \( \vartheta \) or \( \eta \) has dimension \( k \). When both \( \vartheta \) and \( \eta \) are \( k \)-faces, their disjointness implies that they use distinct \( \mu_j \)'s, and the statement follows from the fact that distinct \( \mu_j \)'s have disjoint supports. \( \square \)
Figure 4: Examples of subdivisions for $k = 1$ and $\ell = 3$ (left) and for $k = 2$ and $\ell = 5$ (right). The bottom pictures show the orientations of $|X_i|$ in the given ordering.

4.2 Construction of $D$ and $g$

We define $D$ and $g$ similarly as in Section 3, but the switch from points to multipoints requires to replace stellar subdivisions by a slightly more complicated decomposition.

The subdivision $D$. We define $D$ so that it coincides with $\Delta_s$ on the faces of dimension at most $(k - 1)$ and decomposes each face of dimension $k$ independently. The precise subdivision of a $k$-face $\sigma_i$ depends on the cardinality of the support of the multipoint $\mu_{c(i)}$ used to “route” $\sigma_i$ under $\varphi$, but the method is generic and spelled out in the next lemma; refer to Figure 4.

Lemma 17. Let $k \geq 1$ and $\sigma = \{w_1, w_2, \ldots, w_{k+1}\}$ be a $k$-simplex. For any positive odd integer $\ell \geq 1$ there exist a subdivision $S$ of $\sigma$ in which no face of dimension $k - 1$ or less is subdivided, and a labelling of the vertices of $S$ by $\{w_1, w_2, \ldots, w_{k+1}, x_1, x_2, \ldots, x_\ell\}$ (some labels may appear several times) satisfying the following properties.

1. Every vertex in $S$ corresponding to an original vertex $w_i$ of $\sigma$ is labelled by $w_i$.
2. No $k$-face of $S$ has its vertices labelled by $w_1, w_2, \ldots, w_{k+1}$.
3. for every $j \in [\ell]$, the subdivision $S$ contains exactly one vertex labelled by $x_j$; this vertex appears in a copy $X_j$ of a stellar subdivision of a simplex labelled by $w_1, \ldots, w_{k+1}$ with the apex labelled $x_j$.
4. Let us equip vertices of $S$ with a linear order which respects the order $w_1 \leq w_2 \leq \cdots \leq w_{k+1} \leq x_1 \leq \cdots \leq x_\ell$ of the labels. For each $j \in [\ell]$ considering $|X_j|$ as a simplex in $|S| = |\sigma|$, such $|X_j|$ is oriented coherently with $|\sigma|$ (in the given ordering) if and only if $j$ is odd.

Proof. This proof is done in the language of geometric simplicial complexes (rather than abstract ones).

The case $\ell = 1$ can be done by a stellar subdivision and labelling the added apex $x_1$. The case $k = 1$ is easy, as illustrated in Figure 4 (left). We therefore assume that $k \geq 2$ and build our subdivision and labelling in four steps:
• We start with the boundary of our simplex $\sigma$ where each vertex $w_i$ is labelled by itself. Let $\vartheta$ be the $(k-1)$-face of $\partial \sigma$ opposite vertex $w_2$, i.e. labelled by $w_1, w_3, w_4, \ldots, w_{k+1}$. We create a vertex in the interior of $\sigma$, label it $w_2$, and construct a new simplex $\sigma'$ as the join of $\vartheta$ and this new vertex; this is the dark simplex in Figure 4 (right).

• We then subdivide $\sigma'$ by considering $\ell - 1$ distinct hyperplanes passing through the vertices of $\sigma'$ labelled $w_3, w_4, \ldots, w_{k+1}$ and through an interior point of the edge of $\sigma'$ labelled $w_1, w_2$. These hyperplanes subdivide $\sigma'$ into $\ell$ smaller simplices. We label the new interior vertices on the edge of $\sigma'$ labelled $w_1, w_2$ by alternatively, $w_1$ and $w_2$; since $\ell$ is odd we can do so in a way that every sub-edge is bounded by two vertices labelled $w_1, w_2$.

• We operate a stellar subdivision of each of the $\ell$ smaller simplices subdividing $\sigma'$, and label the added apices $x_1, x_2, \ldots, x_\ell$. This way we obtain a subdivision $\sigma'$ of $\sigma'$.

• We finally consider each face $\eta$ of $\sigma$, subdividing $\partial \sigma$ and other than $\vartheta$ and add the simplex formed by $\eta$ and the (original) vertex $w_2$ of $\sigma$. These simplices, together with $\sigma'$, form the desired subdivision $\sigma$ of $\sigma$.

It follows from the construction that no face of $\partial \sigma$ was subdivided.

Property 1 is enforced in the first step and preserved throughout. We can ensure that Property 2 holds in the following way. First, we have that any $k$-simplex of $S'$ contains a vertex $x_j$ for some $j \in [\ell]$. Next, if we consider a $k$-simplex of $S$ which is not in $S'$ it is a join of a certain $(k-1)$-simplex $\eta$ of $S'$, with $\eta \subset \partial \sigma$, and the vertex $w_2$ of $\sigma$. However, the only such $(k-1)$-simplex labelled by $w_1, w_3, w_4, \ldots, w_{k+1}$ is $\vartheta$, but the join of $\vartheta$ and $w_2$ does not belong to $S$.

Properties 3 and 4 are enforced by the stellar subdivisions of the third step and by alternating the labels $w_1$ and $w_2$ in the second step. No other step creates, destroys or modifies any simplex involving a vertex labelled $x_j$.

Let $S$ be the subdivision of a simplex $\sigma$ from Lemma 17. Similarly as in the case of Lemma 11, we need to describe the chain map $\rho$: $C_*(\sigma) \to C_*(S)$ defined by formula (7). Actually, only a partial information will be sufficient for us, focusing on $k$-simplices of $X_j$.

Since for every $j \in [\ell]$, the apex of $X_j$ is the only vertex labelled by $w_j$, we can use $x_j$ as the name for the apex. Let $\vartheta_j$ be the $k$-simplex on the vertices of $X_j$ except of $x_j$. Note that this simplex does not belong to $S$. Following the usual pattern, we also denote $z(\vartheta_j, x_j) := \partial(\vartheta_j \cup \{x_j\})$.

**Lemma 18.** In the setting above,

$$\rho(\sigma) = \sum_{j=1}^{\ell} (-1)^{j+1} (\vartheta_j + (-1)^j z(\vartheta_j, x_j)) + \sum_\eta \text{or}(\eta, \sigma) \eta$$

(10)

where the second sum is over all $k$-simplices of $S$ which do not belong to any $X_j$.

**Proof.** We expand $\rho(\sigma)$ via (7); however, we further shift the $k$-simplices in some of the $X_j$ to the first sum in (10). This is done via Lemma 11 on each of the $X_j$; the correction term $(-1)^{j+1}$ comes from Property 4 of Lemma 17.

The subdivision $D$ of $\Delta_f(k)$ is now defined as follows. First, we leave the $(k - 1)$-skeleton untouched.

Next for each $k$-simplex $\sigma_i$ we consider the multiplet $\mu = \mu_{c(i)} = \sum_{u \in U} \lambda_u u$ (leaving the dependence on the index $i$ implicit in the affine combination). We recall that $\lambda_u$ are elements of $Z_p$; however, we temporarily consider them as elements of $Z$, in the interval $\{0, 1, \ldots, p - 1\}$. We consider some $u' \in U$, which belongs to the support of $\mu$, and we set $\kappa_u := \lambda_u$ for any $u \in U \setminus \{u'\}$ (as elements of $Z$) whereas we set $\kappa_{u'} := 1 - \sum_{u \in U \setminus \{u'\}} \lambda_u$. It follows that $\kappa_u \equiv \lambda_u \pmod{p}$ for any $u \in U$ as $\sum_{u \in U} \lambda_u \equiv 1 \pmod{p}$ (they sum to 1 as elements of $Z_p$). Next, we set $\ell_i := \sum_{u \in U} |\kappa_u|$. It follows that $\ell_i$ is odd, and we set $S(i)$ to be the subdivision of $\sigma_i$ obtained from Lemma 17 with $\ell := \ell_i$. The final subdivision $D$ is obtained by subdividing each $\sigma_i$ this way. For working with the chains, we need to specify a global linear order on the vertices $D$. We pick an arbitrary such order that respects the prescribed order on each $S(i)$.

According to this subdivision, we have a chain map $\rho$: $C_*(\Delta_f(k)) \to C_*(D)$ defined in Subsection 3.2. On faces of dimension at most $(k - 1)$ it is an identity; on $k$-faces, it is determined by the formula from Lemma 18.
The simplicial map $g_{\text{simp}}$. We now define a simplicial map $g_{\text{simp}}: D \to \Delta^{(k)}_n$. We first set $g_{\text{simp}}(v) = v$ for every vertex $v$ of $\Delta$. Next, we consider some $k$-face $\sigma_i = \{w_1, \ldots, w_{k+1}\}$. We denote by $v_1, v_2, \ldots, v_{k+1}$ the vertices on the boundary of $S(i)$, being understood that each $v_j$ is labelled by $w_j$. We map each interior vertex of $S(i)$ labelled with $w_j$ to $v_j$. It remains to map interior vertices of $S(i)$ labelled $x_j$ for $j \in [\ell]$. Using the notation from the definition of $D$, we consider the integers $\kappa_u$ for $u \in U$ (with respect to our $\sigma_i$). If $\kappa_u > 0$, then we pick $\kappa_u$ vertices $x_j$ with $j$ odd and we map them to $u$. If $\kappa_u < 0$, which may happen only for $u = u'$ (coming again from the definition of $D$), then we pick $-\kappa_u$ vertices $x_j$ with $j$ even and we map them to $u$. Of course, for two distinct elements $u_1$ and $u_2$ from $U$ we pick distinct points $x_j$. The parameter $\ell = \ell_i$ is set up exactly in such a way that we cover all $x_j$. Now we need to know that $\rho$ and $g_{\text{simp}}$ compose to $\varphi$ on the level of chains.

Lemma 19. $(g_{\text{simp}})_\sharp \circ \rho = \varphi$.

Proof. All three maps are the identity on $\Delta^{(k)}_2$, so let us focus on the $k$-faces. Consider a $k$-face $\sigma_i$, the value $\rho(\sigma_i)$ is given by the formula in Lemma 18 with $S = S(i)$. However, for expressing $(g_{\text{simp}})_\sharp \circ \rho(\sigma_i)$ we may ignore the second sum in formula (10) since $a$ does not belong to any $\kappa_j$ contains two vertices with the same label by Lemma 17, which implies that $(g_{\text{simp}})_\sharp (\eta) = 0$.

Therefore

$$(g_{\text{simp}})_\sharp \circ \rho(\sigma_i) = (g_{\text{simp}})_\sharp \left( \sum_{j=1}^{\ell} (\vartheta_j + (-1)^{k-1} z(\vartheta_j, x_j)) \right) = \sum_{u \in U} \kappa_u (\sigma_i + (-1)^k z(\sigma_i, u)).$$

The last equality follows from the definition of $g_{\text{simp}}$ considering that $g_{\text{simp}}$ preserves the prescribed linear orders on $D$ and $\Delta^{(k)}_n$. The sign $(-1)^{j+1}$ disappears as the vertices $x_j$ with $j$ even contribute to $\kappa_u$ with the opposite sign. We know that $\kappa_u \mod p = \lambda_u$ and that $\sum_{u \in U} \kappa_u = 1$. Therefore the expression on the right-hand side of (11) equals $\sigma_i + (-1)^k z(\sigma_i, \mu)$, that is, $\varphi(\sigma_i)$ as required. \hfill \square

The continuous map $g$. Since $D$ is a subdivision of $\Delta^{(k)}_n$, we have $|\Delta^{(k)}_n| = |D|$ and the simplicial map $g_{\text{simp}}: D \to \Delta^{(k)}_n$ induces a continuous map $g: |\Delta^{(k)}_n| \to |\Delta^{(k)}_n|$. All that remains to do is check that $g$ satisfies the two conditions of Lemma 5. Condition 1 follows from a direct translation of Lemma 16; note that in the definition of $g_{\text{simp}}$ we map $x_j$ to $u \in U$ only if $\kappa_u \neq 0$. Condition 2 can be verified by a computation in the same way as in Section 3. Specifically, in homology we have

$$f_* \circ \varphi_* = f_* \circ (g_{\text{simp}})_* \circ \rho_*$$

and we know that $f_* \circ \varphi_*$ is trivial on $\Delta^{(k)}_n$ by Lemma 15. As $\rho_*$ is an isomorphism, this implies that $f_* \circ (g_{\text{simp}})_*$ is trivial. Lemma 8 then implies that $(f \circ g)_*$ is trivial. This concludes the proof of Lemma 5.

Acknowledgement. U.W. learned about Conjecture 1 from Wolfgang Kühnel when attending the Mini Symposia on Discrete Geometry and Discrete Topology at the Jahrestagung der Deutschen Mathematiker-Verenigung in München in 2010. He would like to thank the organizers Frank Lutz and Achill Schürmann for the invitation, and Prof. Kühnel for stimulating discussions.

References


Bounding Helly numbers via Betti numbers∗†‡

Xavier Goaoc1, Pavel Paták2, Zuzana Patáková3, Martin Tancer3, and Uli Wagner4

1Université Paris-Est Marne-la-Vallée, France.  
2Department of Algebra, Charles University, Prague, Czech Republic. 
3Department of Applied Mathematics, Charles University, Prague, Czech Republic. 
4IST Austria, Klosterneuburg, Austria. 

November 11, 2016

Dedicated to the memory of Jiří Matoušek,  
wonderful teacher, mentor, collaborator, and friend.

Abstract

We show that very weak topological assumptions are enough to ensure the existence of a Helly-type theorem. More precisely, we show that for any non-negative integers b and d there exists an integer h(b, d) such that the following holds. If F is a finite family of subsets of Rd such that \( \tilde{\beta}_i(\bigcap G) \leq b \) for any \( G \subseteq F \) and every \( 0 \leq i \leq \lceil d/2 \rceil - 1 \) then F has Helly number at most h(b, d). Here \( \tilde{\beta}_i \) denotes the reduced \( \mathbb{Z}_2 \)-Betti numbers (with singular homology). These topological conditions are sharp: not controlling any of these \( \lceil d/2 \rceil \) first Betti numbers allow for families with unbounded Helly number.

Our proofs combine homological non-embeddability results with a Ramsey-based approach to build, given an arbitrary simplicial complex K, some well-behaved chain map \( C_*(K) \to C_*(\mathbb{R}^d) \).

1 Introduction

Helly’s classical theorem [Hel23] states that a finite family of convex subsets of \( \mathbb{R}^d \) must have a point in common if any \( d + 1 \) of the sets have a point in common. Together with Radon’s and Caratheodory’s theorems, two other “very finite properties” of convexity, Helly’s theorem is a pillar of combinatorial geometry. Along with its variants (eg. colorful or fractional), it underlies many fundamental results in discrete geometry, from the centerpoint theorem [Rad46] to the existence of weak \( \varepsilon \)-nets [ABFK92] or the \((p,q)\)-theorem [AK95].

In the contrapositive, Helly’s theorem asserts that any finite family of convex subsets of \( \mathbb{R}^d \) with empty intersection contains a sub-family of size at most \( d + 1 \) that already has empty intersection. This inspired the definition of the Helly number of a family \( F \) of arbitrary sets. If \( F \) has empty intersection then its Helly number is defined as the size of the largest sub-family \( G \subseteq F \) with the following properties: \( G \) has empty intersection and any proper sub-family of \( G \) has nonempty intersection; if \( F \) has nonempty intersection then its Helly number is, by convention, 1. With this terminology, Helly’s theorem simply states that any finite family of convex sets in \( \mathbb{R}^d \) has Helly number at most \( d + 1 \).

∗2010 Mathematics Subject Classification: 52A35 (Helly-type theorems and geometric transversal theory) and 55U10 (Simplicial sets and complexes). Keywords: Helly-type theorems, Ramsey’s theorem, Embeddings of simplicial complexes, Betti numbers.
†An extended abstract of this work was presented at the 31st International Symposium on Computational Geometry [GPP15].
‡PP, ZP and MT were partially supported by the Charles University Grant GAUK 421511. ZP was partially supported by the Charles University Grant SVV-2014-260103. ZP and MT were partially supported by the ERC Advanced Grant No. 267165 and by the project CE-ITI (GACR P202/12/G061) of the Czech Science Foundation. UW was partially supported by the Swiss National Science Foundation (grants SNSF-200020-138239 and SNSF-PP00P2-138948). Part of this work was done when XG was affiliated with INRIA Nancy Grand-Est and when MT was affiliated with Institutionen for matematik, Kungliga Tekniska Högskolan, then IST Austria.
Helly already realized that bounds on Helly numbers independent of the cardinality of the family are not a privilege of convexity: his topological theorem [Hel30] asserts that a finite family of open subsets of \( \mathbb{R}^d \) has Helly number at most \( d+1 \) if the intersection of any sub-family of at most \( d \) members of the family is either empty or a homology cell.\(^1\) Such uniform bounds are often referred to as Helly-type theorems. In discrete geometry, Helly-type theorems were found in a variety of contexts, from simple geometric assumptions (eg. homothets of a planar convex curve [Swa03]) to more complicated implicit conditions (sets of line intersecting prescribed geometric shapes [Tve89, GHP06, CGHP08], sets of norms making a given subset of \( \mathbb{R}^d \) equilateral [Pet71, Theorem 5], etc.) and several surveys [Eck93, Wen04, Tan13] were devoted to this abundant literature. These Helly numbers give rise to similar finiteness properties in other areas, for instance in variants of Whitney's extension problem [Shv08] or the combinatorics of generators of certain groups [Far09].

Many Helly numbers are established via ad hoc arguments, and decades sometimes go by before a conjectured bound is effectively proven, as illustrated by Tverberg's proof [Tve89] of a conjecture of Grünbaum [Gru58]. This is true not only for the quantitative question (what is the best bound?) but also for the existential question (is the Helly number uniformly bounded?); in this example, establishing a first bound [Kat86] was already a matter of decades. Substantial effort was devoted to identify general conditions ensuring bounded Helly numbers, and topological conditions, as opposed to more geometric ones like convexity, received particular attention. The general picture that emerges is that requiring that intersections have trivial low-dimensional homotopy [Mat97] or have trivial high-dimensional homology [CGG14] is sufficient (see below for a more comprehensive account).

1.1 Problem statement and results

In this paper, we focus on the existential question and give the following new homological sufficient condition for bounding Helly numbers. Throughout the paper, we consider homology with coefficients\(^2\) in \( \mathbb{Z}_2 \), and denote by \( \bar{\beta}_i(X) \) the \( i \)th reduced Betti number (over \( \mathbb{Z}_2 \)) of a space \( X \). Furthermore, we use the notation \( \cap F := \bigcap_{U \in F} U \) as a shorthand for the intersection of a family of sets.

**Theorem 1.** For any non-negative integers \( b \) and \( d \) there exists an integer \( h(b, d) \) such that the following holds. If \( F \) is a finite family of subsets of \( \mathbb{R}^d \) such that \( \bar{\beta}_i(\cap G) \leq b \) for any \( G \subseteq F \) and every \( 0 \leq i \leq [d/2] - 1 \) then \( F \) has Helly number at most \( h(b, d) \).

Our proof, which we sketch in Subsection 1.4, hinges on a general principle, which we learned from Matoušek [Mat97] but which already underlies the classical proof of Helly’s theorem from Radon’s lemma, to derive Helly-type theorems from results of non-embeddability of certain simplicial complexes. The novelty of our approach is to examine these non-embeddability arguments from a homological point of view. This turns out to be a surprisingly effective idea, as homological analogues of embeddings appear to be much richer and easier to build than their homotopic counterparts. More precisely, our proof of Theorem 1 builds on two contributions of independent interest:

- We reformulate some non-embeddability results in homological terms. We obtain a homological analogue of the Van Kampen-Flores Theorem (Corollary 13) and, as a side-product, a homological version of Radon’s lemma (Lemma 15). This is part of a systematic effort to translate various homotopy technique to a more tractable homology setting. It builds on, and extends, previous work on homological minors [Wag11].

- By working with homology rather than homotopy, we can generalize a technique of Matoušek [Mat97] that uses Ramsey’s theorem to find embedded structures. In this step, roughly speaking, we construct some auxiliary (chain) map, with certain homological constraints, inductively by increasing the dimension of the preimage complex while decreasing the size of it. This approach turned out to be also useful in a rather different setting, regarding the (non-)embeddability of skeleta of complexes into manifolds [GMP∗15].

\(^1\)By definition, a homology cell is a topological space \( X \) all of whose (reduced, singular, integer coefficient) homology groups are trivial, as is the case if \( X = \mathbb{R}^d \) or \( X \) is a single point. Here and in what follows, we refer the reader to standard textbooks like [Hat02, Mun84] for further topological background and various topological notions that we leave undefined.

\(^2\)The choice of \( \mathbb{Z}_2 \) as the ring of coefficient ring has two reasons. On the one hand, we work with the van Kampen obstruction to prove certain non-embeddability results, and the obstruction is naturally defined either for integer coefficients or over \( \mathbb{Z}_2 \) (it is a torsion element of order two). On the other hand, the Ramsey arguments used in our proof require working over a fixed finite ring of coefficients to ensure a finite number of color classes (cf. Claim 1).
Our method also proves:

- A bound of $d + 1$ on the Helly number of any family $\mathcal{F}$ of subsets of $\mathbb{R}^d$ such that $\tilde{\beta}_i(\bigcap \mathcal{G}) = 0$ for all $\mathcal{G} \subseteq \mathcal{F}$ and all $i \leq d$ (see Corollary 24), which generalizes Helly’s topological theorem as the sets of $\mathcal{F}$ are, for instance, not assumed to be open. (In the original proof, this assumption is crucial and used to ensure that the union of the sets must have trivial homology in dimensions larger than $d$; this may fail if the sets are not open.)

- A bound of $d + 2$ on the Helly number of any family $\mathcal{F}$ of subsets of $\mathbb{R}^d$ such that $\tilde{\beta}_i(\bigcap \mathcal{G}) = 0$ for all $\mathcal{G} \subseteq \mathcal{F}$ but only for $i \leq \lceil d/2 \rceil - 1$ (see Corollary 23).

In both cases the bounds are tight.

Quantitatively, the bound on $h(b, d)$ that we obtain in the general case is very large as it follows from successive applications of Ramsey’s theorems. The conditions of Theorem 1 relax the conditions of a Helly-type theorem of Amenta [Ame96] (see the discussion below) for which a lower bound of $b(d + 1)$ is known [Lar68]; a stronger lower bound is possible for $h(b, d)$ (see Example 2) but we consider narrowing this gap further to be outside the scope of the present paper. Qualitatively, Theorem 1 is sharp in the sense that all (reduced) Betti numbers $\tilde{\beta}_i$ with $0 \leq i \leq \lceil d/2 \rceil - 1$ need to be bounded to obtain a bounded Helly number (see Example 3).

**Example 2.** First, we observe that for every $d \geq 2$ there is a geometric simplicial complex $\Gamma_d$ with $d + 2$ vertices, embedded in $\mathbb{R}^d$, such that every nonempty induced subcomplex $L$ of $\Gamma_d$ is connected and satisfies $\tilde{\beta}_i(L) = 0$ for $i \neq d - 1$ and $\tilde{\beta}_{d-1}(L) \leq 1$.

Indeed, we can take $\Gamma_d$ to be the stellar subdivision of the $d$-simplex (i.e., the cone over the boundary of the $d$-simplex): Among the vertices of $\Gamma_d$, $d + 1$ of them, say $v_1, \ldots, v_{d+1}$, form a $d$-simplex, and the last one, say $w$, is situated in the barycenter of that simplex. The maximal simplices of $\Gamma_d$ contain $w$ and $d$ of the vertices $v_i$. Given an induced subcomplex $L$, either $L$ misses one of the $v$-vertices, and then $L$ is a $k$-simplex for some $k \leq d$; or $L$ contains all the $v_i$, in which case either $L = \Gamma_d$ or $L$ is the boundary of the simplex spanned by the vertices $v_i$.

Now, let $\Gamma_{b,d}$ be a complex that consists of $b$ disjoint copies of $\Gamma_d$, embedded in $\mathbb{R}^d$. For a vertex $v$ of $\Gamma_{b,d}$, let $U_v$ be the union of all simplices of $\Gamma_{b,d}$ not containing $v$ (i.e., $U_v$ is the geometric realization of the induced subcomplex of $\Gamma_{b,d}$ on all vertices but $v$). We define $\mathcal{F}$ to be the collection of all subcomplexes $F_v$, where $v$ ranges over all vertices of $\Gamma_{b,d}$. Thus, by construction, $\mathcal{F}$ contains $b(d + 2)$ sets, $\bigcap \mathcal{F} = \emptyset$, and for any nonempty proper subsystem $\mathcal{G} \subset \mathcal{F}$, the intersection $\bigcap \mathcal{G}$ is nonempty, and by the properties of $\Gamma_d$, the reduced Betti numbers of $\bigcap \mathcal{G}$ are bounded by $b$.

**Example 3.** Let us fix some $k$ with $0 \leq k \leq \lceil d/2 \rceil - 1$. For $n$ arbitrarily large, consider a geometric realization in $\mathbb{R}^d$ of the $k$-skeleton of the $(n-1)$-dimensional simplex (see [Mat03, Section 1.6]); more

---

3We remark that this construction can be further improved (at the cost of simplicity). For example, for $d = 3$, it is possible to find a geometric simplicial complex $\Gamma'_3$ with six vertices (instead of five) with properties analogous to $\Gamma_3$:

Consider a simplex $\Delta \subseteq \mathbb{R}^3$ with vertices $v_1, v_2, v_3$ and $v_4$. Let $b$ be the barycenter of this simplex and we set $v_5$ to be the barycenter of the triangle $v_1 v_2 v_3$ and $v_6$ to be the barycenter of $v_3 v_4 b$. Finally, we set $\Gamma'_3$ to be the subdivision of $\Delta$ with vertices $v_1, \ldots, v_6$ and with maximal simplices $1245, 1253, 3416, 3426, 5613, 5614, 5623$, and $5624$ where the label $ABCD$ stands for cone $(v_A, v_B, v_C, v_D)$. One can check that this indeed yields a simplicial complex with the required properties.

See the 1-skeleton of $\Gamma'_3$ in Figure 1. We believe that an analogous example can be also constructed for $d \geq 4$. 

3
specifically, let \( V = \{v_1, \ldots, v_n\}\) be a set of points in general position in \( \mathbb{R}^d \) and consider all geometric simplices \( \sigma_A := \text{conv}(A) \) spanned by subsets \( A \subseteq V \) of cardinality \( |A| \leq k + 1 \).

Similarly as in the previous example, let \( U_j \) be the union of all the simplices not containing the vertex \( v_j \), for \( 1 \leq j \leq n \). We set \( F = \{U_1, \ldots, U_n\} \). Then, \( \bigcap F = \emptyset \), and for any proper sub-family \( \mathcal{G} \subseteq F \), the intersection \( \bigcap \mathcal{G} \) is either \( \mathbb{R}^d \) (if \( \mathcal{G} = \emptyset \)) or (homeomorphic to) the \( k \)-dimensional skeleton of a \( (n - 1 - |\mathcal{G}|) \)-dimensional simplex. Thus, the Helly number of \( F \) equals \( n \). Moreover, the \( k \)-skeleton \( \Delta_{m-1} \) of an \( (m - 1) \)-dimensional simplex has reduced Betti numbers \( \tilde{\beta}_i = 0 \) for \( i \neq k \) and \( \tilde{\beta}_k = \binom{m-1}{k+1} \).

Thus, we can indeed obtain arbitrarily large Helly number as soon as at least one \( \tilde{\beta}_k \) is unbounded.

1.2 Relation to previous work

The search for topological conditions that ensure bounded Helly numbers started with Helly’s topological theorem [Hel30] (see also [Deb70] for a modern version of the proof) and organized along several directions related to classical questions in topology. Theorem 1 unifies topological conditions originating from two different approaches:

- Helly-type theorem can be derived from non-embeddability results, in the spirit of the classical proof of Helly’s theorem from Radon’s lemma. Using this approach, Matoušek [Mat97] showed that it is sufficient to control the low-dimensional homotopy of intersections of sub-families to ensure bounded Helly numbers: for any non-negative integers \( b \) and \( d \) there exists a constant \( c(b,d) \) such that any finite family of subsets of \( \mathbb{R}^d \) in which every sub-family intersects in at most \( b \) connected components, each \( ([d/2] - 1) \)-connected, has Helly number at most \( c(b,d) \). (We recall that a topological space \( X \) is \( k \)-connected, for some integer \( k \geq 0 \), if every continuous map \( S^i \to X \) from the \( i \)-dimensional sphere to \( X \), \( 0 \leq i \leq k \), can be extended to a map \( D^{i+1} \to X \) from the \( (i + 1) \)-dimensional disk to \( X \).) By Hurewicz’ Theorem and the Universal Coefficient Theorem [Hat02, Theorem 4.37 and Corollary 3A.6], a \( k \)-connected space \( X \) satisfies \( \tilde{\beta}_i(X) = 0 \) for all \( i \leq k \). Thus, our condition indeed relaxes Matoušek’s, in two ways: by using \( \mathbb{Z}_2 \)-homology instead of the homotopy-theoretic assumptions of \( k \)-connectedness\(^4\), and by allowing an arbitrary fixed bound \( b \) instead of \( b = 0 \).

- Helly’s topological theorem can be easily derived from classical results in algebraic topology relating the homotopy/homotopy of the nerve of a family to that of its union: Leray’s cyclic cover theorem [Bre97, Sections III.4.13, VI.4 and VI.13] for homology, and Borsuk’s Nerve theorem [Bor48, Bjö03] for homotopy (in that case one considers finite open good covers\(^5\)). More general Helly numbers were obtained via this approach by Dugundji [Dug66], Amenta [Ame96]\(^6\), Kalai and Meshulam [KM08], and Colin de Verdière et al. [CGG14]. The outcome is that if a family of subsets of \( \mathbb{R}^d \) is such that any sub-family intersects in at most \( b \) connected components, each a homology cell (over \( \mathbb{Q} \)), then it has Helly number at most \( b(d + 1) \). This therefore relaxes Helly’s original assumption by allowing intersections of sub-families to have \( \tilde{\beta}_0 \)’s bounded by an arbitrary fixed bound \( b \) instead of \( b = 0 \). Theorem 1 makes the same relaxation for the \( \tilde{\beta}_1 \)’s, \( \tilde{\beta}_2 \)’s, \( \ldots \), \( \tilde{\beta}_{[d/2]−1} \)’s and drops all assumptions on higher-dimensional homology, including the requirement that sets be open (which is used to control the \( (> d) \)-dimensional homology of intersections).

Let us highlight two Helly-type results that stand out in this line of research as not subsumed (qualitatively) by Theorem 1. On the one hand, Eckhoff and Nischke [EN09] gave a purely combinatorial argument that derives the theorems of Amenta [Ame96] and Kalai and Meshulam [KM08] from Helly’s convex and topological theorems. On the other hand, Montejano [Mon14] relaxed Helly’s original assumption on the intersection of sub-families of size \( k \leq d + 1 \) from being a homology cell into having

\(^4\)We also remark that our condition can be verified algorithmically since Betti numbers are easily computable, at least for sufficiently nice spaces that can be represented by finite simplicial complexes, say. By contrast, it is algorithmically undecidable whether a given \( 2 \)-dimensional simplicial complex is \( 1 \)-connected, see, e.g., the survey [Soa04].

\(^5\)An open good cover is a finite family of open subsets of \( \mathbb{R}^d \) such that the intersection of any sub-family is either empty or is contractible (and hence, in particular, a homology cell).

\(^6\)The role of nerves is implicit in Amenta’s proof but becomes apparent when compared to an earlier work of Wegner [Weg75] that uses similar ideas.

\(^7\)The result of Colin de Verdière et al. [CGG14] holds in any paracompact topological space; Theorem 1 only subsumes the \( \mathbb{R}^d \) case.
trivial \(d-k\) homology (so only one Betti number needs to be controlled for each intersection, but it must be zero). These results neither contain nor are contained in Theorem 1.

We remark that another non-topological structural condition, known to ensure bounded Helly numbers, also falls under the umbrella of Theorem 1. As observed by Motzkin [Mot55, Theorem 7] (see also Deza and Frankl [DF87]), any family of real algebraic subvarieties of \(\mathbb{R}^d\) defined by polynomials of degree at most \(k\) has Helly number bounded by a function of \(d\) and \(k\) (more precisely, by the dimension of the vector subspace of \(\mathbb{R}[x_1, x_2, \ldots, x_d]\) spanned by these polynomials); since the Betti numbers of an algebraic variety in \(\mathbb{R}^n\) can be bounded in terms of the degree of the polynomials that define it [Mil63, Tho65], this also follows from Theorem 1. We give some other examples in Section 1.3, where we easily derive from Theorem 1 generalizations of various existing Helly-type theorems.

Note that Theorem 1 is similar, in spirit, to some of the general relations between the growth of Betti numbers and fractional Helly theorems conjectured by Kalai and Meshulam [Kal04, Conjectures 6 and 7]. Kalai and Meshulam, in their conjectures, allow a polynomial growth of the Betti numbers in \(\bigcap \mathcal{G}\). As the following example shows, Theorem 1 is also sharp in the sense that even a linear growth of Betti numbers, already in \(\mathbb{R}^1\), may yield unbounded Helly numbers. In particular, the conjectures of Kalai and Meshulam cannot be strengthened to include Theorem 1.

Example 4. Consider a positive integer \(n\) and open intervals \(I_i := (i - 1; i + 0.1)\) for \(i \in [n]\). Let \(X_i := [0, n] \setminus I_i\). The intersection of all \(X_i\) is empty but the intersection of any proper subfamily is nonempty. In addition, the intersection of \(k\) such \(X_i\) can be obtained from \([0, n]\) by removing at most \(k\) open intervals, thus the reduced Betti numbers of such an intersection are bounded by \(k\).

1.3 Further consequences

We conclude this introduction with a few implications of our main result.

New geometric Helly-type theorems. The main strength of our result is that very weak topological assumptions on families of sets are enough to guarantee a bounded Helly number. This can be used to identify new Helly-type theorems, for instance by easily detecting generalizations of known results, as we now illustrate on two Helly-type theorems of Swanepoel.

A first example is given by a Helly-type theorem for hollow boxes [Swa99], which generalizes (qualitatively) as follows:

**Corollary 5.** For all integers \(s, d \geq 1\), there exists an integer \(h'(s, d)\) such that the following holds. Let \(S\) be a set of \(s\) nonzero vectors in \(\mathbb{R}^d\), and let \(\mathcal{F} = \{U_1, U_2, \ldots, U_n\}\) where each \(U_i\) is a polyhedral subcomplex of some polytope \(P_i\) in \(\mathbb{R}^d\) which can be obtained as an intersection of half-spaces with normal vectors in \(S\). Then \(\mathcal{F}\) has Helly number at most \(h'(s, d)\).

Swanepoel’s result corresponds to the case \(S = \{e_1, e_2, \ldots, e_d\}\) where \(e_1, \ldots, e_d\) form a basis of \(\mathbb{R}^d\).

**Proof of Corollary 5.** We verify the assumptions of Theorem 1, i.e., we consider a subfamily \(\mathcal{G} = \{U_i : i \in I\} \subseteq \mathcal{F}\) and we check that \(\beta_i(\bigcap \mathcal{G})\) is bounded by a function of \(s\) and \(d\) for any \(i \geq 0\) (to apply Theorem 1, it would be sufficient to consider \(i \leq \lfloor d/2 \rfloor - 1\), but in the present setting, there is no difference in reasoning for other values of \(i\)).

Let \(\mathcal{P} = \mathcal{P}(S)\) be the set of all polytopes which can be obtained as an intersection of half-spaces with normal vectors in \(S\). Let \(P_i \in \mathcal{P}\) be a polytope such that \(U_i\) is a polyhedral subcomplex of \(P_i\).

Let us consider the polytope \(P = \bigcap_{i \in I} P_i\). From the definition of \(\mathcal{P}\) we immediately deduce that \(P \in \mathcal{P}\). Moreover, the intersection \(U := \bigcap \mathcal{G}\) is a polyhedral subcomplex of \(P\). (The faces \(U\) are of form \(\bigcap_{i \in I} \sigma_i\) where \(\sigma_i\) is a face of \(U_i\); see [RS72, Exercise 2.8(5) + hint].)

Since \(P \in \mathcal{P}\) we deduce that it has at most \(2s\) facets. By the dual version of the upper bound theorem [Zie95, Theorem 8.23], the number of faces of \(P\) is bounded by a function of \(s\) and \(d\). Consequently, \(\beta_i(U)\) is bounded by a function of \(s\) and \(d\), since \(U\) is a subcomplex of \(P\). □

A second example concerns a Helly-type theorem for families of translates and homothets of a convex curve [Swa03], which are special cases of families of pseudo-circles. More generally, a family of pseudo-spheres is defined as a set \(\mathcal{F} = \{U_1, U_2, \ldots, U_n\}\) of subsets of \(\mathbb{R}^d\) such that or any \(\mathcal{G} \subseteq \mathcal{F}\), the intersection \(\bigcap \mathcal{G}\) is homeomorphic to a \(k\)-dimensional sphere for some \(k \in \{0, 1, \ldots, d-1\}\) or to a single point. The case \(b = 1\) of Theorem 1 immediately implies the following:
Corollary 6. For any integer \( d \) there exists an integer \( h(d) \) such that the Helly number of any finite family of pseudo-spheres in \( \mathbb{R}^d \) is at most \( h(d) \).

We note that the special case of Euclidean spheres falls under the umbrella of intersections of real algebraic varieties of bounded degree, for which the Helly number is bounded as observed by Motzkin and others, as discussed above [Mae89, DF87]. For the more general setting pseudo-spheres, however, the above result is new, to the best of our knowledge. An optimal bound \( h(d) = d + 1 \) as soon as the family contains at least \( d + 3 \) pseudo-spheres was obtained by Soares [Sos15], after discussing the contents of Corollary 6 with us.

Generalized linear programming. Theorem 1 also has consequences in the direction of optimization problems. Various optimization problems can be formulated as the minimization of some function \( f : \mathbb{R}^d \to \mathbb{R} \) over some intersection \( \bigcap_{i=1}^n C_i \) of subsets \( C_1, C_2, \ldots, C_n \) of \( \mathbb{R}^d \). If, for \( t \in \mathbb{R} \), we let \( L_t = f^{-1}((-\infty,t]) \) and \( \mathcal{F}_t = \{ C_1, C_2, \ldots, C_n, L_t \} \) then

\[
\min_{x \in \bigcap_{i=1}^n C_i} f(x) = \min \left\{ t \in \mathbb{R} : \bigcap_{i=1}^n \mathcal{F}_i \neq \emptyset \right\}.
\]

If the Helly number of the families \( \mathcal{F}_t \) can be bounded uniformly in \( t \) by some constant \( h \) then there exists a subset of \( h - 1 \) constraints \( C_{i_1}, C_{i_2}, \ldots, C_{i_{h-1}} \) that suffice to define the minimum of \( f \):

\[
\min_{x \in \bigcap_{i=1}^n C_i} f(x) = \min_{x \in \bigcap_{i=1}^{h-1} C_i} f(x).
\]

A consequence of this observation, noted by Amenta [Ame94], is that the minimum of \( f \) over \( C_1 \cap C_2 \cap \ldots \cap C_n \) can\(^8\) be computed in randomized \( O(n) \) time by generalized linear programming [SW92] (see de Loera et al. [dLPS] for other uses of this idea). Together with Theorem 1, this implies that an optimization problem of the above form can be solved in randomized linear time if it has the property that every intersection of some subset of the constraints with a level set of the function has bounded “topological complexity” (measured in terms of the sum of the first \([d/2]\) Betti numbers). Let us emphasize that this linear-time bound holds in a real-RAM model of computation, where any constant-size subproblems can be solved in \( O(1) \)-time; it therefore concerns the combinatorial difficulty of the problem and says nothing about its numerical difficulty.

1.4 Proof outline

Let us briefly sketch the proof of Theorem 1.

Consider the simplified setting where we have subsets \( A_1, A_2, \ldots, A_5 \) of \( \mathbb{R}^2 \) such that any four have non-empty intersection and any three have path-connected intersection. Draw \( K_5 \), the complete graph on 5 vertices, inside the union of the five sets by picking points \( p_i \in \cap_{j \neq i} A_j \) and connecting any two points \( p_u, p_v \) inside the intersection \( \cap_{i \neq u,v} A_i \). The (stronger form of the) non-planarity of \( K_5 \) ensures that two edges that share no vertex must cross, and the intersection point witnesses that \( \cap_{i=1}^5 A_i \) is non-empty (cf. Figure 3). This idea, more systematically, ensures that any family of planar sets with path-connected intersections has Helly number at most 4.

Now consider subsets \( A_1, A_2, \ldots, A_n \) of \( \mathbb{R}^2 \) such that the intersection of any proper subfamily is nonempty and has at most \( b \) path-connected components. We can again pick \( p_i \in \cap_{j \neq i} A_j \). Two points \( p_u, p_v \) may end up in different connected components of \( \cap_{j \neq u,v} A_j \), but among any \( b + 1 \) points \( p_{i_1}, p_{i_2}, \ldots, p_{i_{b+1}} \), two can be connected inside \( \cap_{j \neq i_1,i_2,\ldots,i_{b+1}} A_j \). We can thus still draw a large graph inside the union, but each edge misses an extra set of \( A_i \’s \). A Ramsey-type argument ensures that for \( n \) large enough, we can find a copy of \( K_5 \) where each edge misses distinct extra sets, and therefore that \( \cap_{i=1}^n A_i \) is non-empty.

These arguments generalize to higher dimension: once we can draw \( p_u p_v, p_u p_w \) and \( p_u p_w \) inside the intersection of some family of subsets, we can fill the triangle in that intersection if it is 1-connected (in homotopy). More systematically, given a family of subsets of \( \mathbb{R}^{2k} \) whose proper intersections are

\[^8\]This requires \( f \) and \( C_1, C_2, \ldots, C_n \) to be generic in the sense that the number of minima of \( f \) over \( \cap_{i \in I} C_i \) is bounded uniformly for \( I \subseteq \{1, 2, \ldots, n\} \).
$k$-connected (in homotopy), we can draw $\Delta^{(k)}_{2k+2}$ inside their union and find, via the Van Kampen-Flores theorem, that the complete intersection is non-empty (and similarly in odd dimensions). This is, in short, Matoušek’s theorem [Mat97].

We extend Matoušek’s approach to allow intersections to have bounded but non-trivial homotopy in dimension 1 or more. The main difficulty is that we may not be able to fill any elementary cycle: as illustrated on the right-hand figure, for $n$ arbitrarily large, $K_n$ can be drawn in an annulus so that no triangle can be filled. There still exist cycles that can be filled, for instance 2435; they are simply not boundaries of triangles. Such cycles are more easily found by working with the additive structure of $\mathbb{Z}_2$-homology: the sum of any two homologous cycles is a boundary (and therefore “fillable”), and many pairs of homologous cycles exist because a bounded Betti number ensures a constant number of homology classes.

The key idea is, then, to look for sufficiently large sets of vertices where, as in the example above, every triangle has the same $\mathbb{Z}_2$-homology, and to map the barycentric subdivision of a triangle to these vertices (as described in Figure 7); the resulting sum of evenly many homologous simplices must be a boundary. These large sets of vertices with homologous triangles exist as soon as the Betti number is bounded: indeed, one can simply apply Ramsey’s theorem to the 3-uniform hypergraph on the vertices where every triangle is “colored” by its homology class. This idea generalizes to arbitrary dimension.

Because of the switch to homology, we do not build a map of $\Delta^{(k)}_{2k+2}$ into the target space $\mathbb{R}^d$ ($d = 2k$ or $d = 2k - 1$) but only a chain map from the simplicial chain complex of $\Delta^{(k)}_{2k+2}$ into the singular chain complex of $\mathbb{R}^d$. Hence, we can no longer rely on the classical non-embeddability results and have to develop homological analogs.

We set up our homological machinery in Section 2 (homological almost-embeddings, homological Van Kampen-Flores Theorem, and homological Radon lemma). We then spell out, in Section 3, variations of the technique that derives Helly-type theorems from non-embeddability. We finally introduce our refinement of this technique and the proof of Theorem 1 in Section 4.

1.5 Notation

We assume that the reader is familiar with basic topological notions and facts concerning simplicial complexes and singular and simplicial homology, as described in textbooks like [Hat02, Mun84]. As remarked above, throughout this paper we will work with homology with $\mathbb{Z}_2$-coefficients unless explicitly stated otherwise. Moreover, while we will consider singular homology groups for topological spaces in general, for simplicial complexes we will work with simplicial homology groups. In particular, if $X$ is a topological space then $C_\ast(X)$ will denote the singular chain complex of $X$, while if $K$ is a simplicial complex, then $C_\ast(K)$ will denote the simplicial chain complex of $K$ (both with $\mathbb{Z}_2$-coefficients).

We use the following notation. Let $K$ be a (finite, abstract) simplicial complex. The underlying topological space of $K$ is denoted by $[K]$. Moreover, we denote by $K^{(i)}$ the $i$-dimensional skeleton of $K$, i.e., the set of simplices of $K$ of dimension at most $i$; in particular $K^{(0)}$ is the set of vertices of $K$. For an integer $n \geq 0$, let $\Delta_n$ denote the $n$-dimensional simplex.

Acknowledgments

We would like to express our immense gratitude to Jiří Matoušek, not only for raising the problem addressed in the present paper and valuable discussions about it, but, much more generally, for the privilege of having known him, as our teacher, mentor, collaborator, and friend. Through his tremendous depth and insight, and the generosity with which he shared them, he greatly influenced all of us.

We further thank Jürgen Eckhoff for helpful comments on a preliminary version of the paper, and Andreas Holmsen and Gil Kalai for providing us with useful references.
2 Homological Almost-Embeddings

In this section, we define homological almost-embedding, an analogue of topological embeddings on the level of chain maps, and show that certain simplicial complexes do not admit homological almost-embeddings in $\mathbb{R}^d$, in analogy to classical non-embeddability results due to Van Kampen and Flores. In fact, when this comes at no additional cost we phrase the auxiliary results in a slightly more general setting, replacing $\mathbb{R}^d$ by a general topological space $R$. Readers that focus on the proof of Theorem 1 can safely replace every occurrence of $R$ with $\mathbb{R}^d$.

2.1 Non-Embeddable Complexes

We recall that an embedding of a finite simplicial complex $K$ into $\mathbb{R}^d$ is simply an injective continuous map $[K] \to \mathbb{R}^d$. The fact that the complete graph on five vertices cannot be embedded in the plane has the following generalization.

**Proposition 7** (Van Kampen [vK32], Flores [Flo33]). For $k \geq 0$, the complex $\Delta^{(k)}_{2k+2}$, the $k$-dimensional skeleton of the $(2k+2)$-dimensional simplex, cannot be embedded in $\mathbb{R}^{2k}$.

A basic tool for proving the non-embeddability of a simplicial complex is the so-called Van Kampen obstruction. To be more precise, we emphasize that in keeping with our general convention regarding coefficients, we work with the $\mathbb{Z}_2$-coefficient version of the Van Kampen obstruction, which will be reviewed in some detail in Section 2.3 below. Here, for the benefit of readers who are willing to accept certain topological facts as given, we simply collect those statements necessary to motivate the definition of homological almost-embeddings and to follow the logic of the proof of Theorem 1.

Given a simplicial complex $K$, one can define, for each $d \geq 0$, a certain cohomology class $\sigma^d(K)$ that resides in the cohomology group $H^d(\overline{K})$ of a certain auxiliary complex $\overline{K}$ (the quotient of the combinatorial deleted product by the natural $\mathbb{Z}_2$-action, see below); see the paragraph on obstructions following Lemma 19 for a more proper definition of $\sigma^d(K)$. This cohomology class $\sigma^d(K)$ is called the Van Kampen obstruction to embeddability into $\mathbb{R}^d$ because of the following fact:

**Proposition 8.** Suppose that $K$ is a finite simplicial complex with $\sigma^d(K) \neq 0$. Then $K$ is not embeddable into $\mathbb{R}^d$. In fact, a slightly stronger conclusion holds: there is no almost-embedding $f: [K] \to \mathbb{R}^d$, i.e., no continuous map such that the images of disjoint simplices of $K$ are disjoint.

Another basic fact is the following result (for a short proof see, for instance, [Mel09, Example 3.5]).

**Proposition 9** ([vK32, Flo33]). For every $k \geq 0$, $\sigma^{2k}(\Delta^{(k)}_{2k+2}) \neq 0$.

As a consequence, one obtains Proposition 7, and in fact the slightly stronger statement that $\Delta^{(k)}_{2k+2}$ does not admit an almost-embedding into $\mathbb{R}^{2k}$.

2.2 Homological Almost-Embeddings and a Van Kampen–Flores Result

For the proof of Theorem 1, we wish to replace homotopy-theoretic notions (like $k$-connectedness) by homological assumptions (bounded Betti numbers). The simple but useful observation that allows us to do this is that in the standard proof of Proposition 8, which is based on (co)homological arguments, maps can be replaced by suitable chain maps at every step.\(^{10}\) The appropriate analogue of an almost-embedding is the following.

**Definition 10.** Let $R$ be a (nonempty) topological space, $K$ be a simplicial complex, and consider a chain map \( \gamma : C_*(K) \to C_*(R) \) from the simplicial chains in $K$ to singular chains in $R$.

---

\(^9\)There is also a version of the Van Kampen obstruction with integer coefficients, which in general yields more precise information regarding embeddability than the $\mathbb{Z}_2$-version, but we will not need this here. We refer to [Mel09] for further background.

\(^{10}\)This observation was already used in [Wag11] to study the (non-)embeddability of certain simplicial complexes. What we call a homological almost-embedding in the present paper corresponds to the notion of a homological minor used in [Wag11].

\(^{11}\)We recall that a chain map $\gamma : C_* \to D_*$ between chain complexes is simply a sequence of homomorphisms $\gamma_n : C_n \to D_n$ that commute with the respective boundary operators, $\gamma_{n-1} \circ \partial_D = \partial_C \circ \gamma_n$. 

8
(i) The chain map $\gamma$ is called nontrivial\textsuperscript{12} if the image of every vertex of $K$ is a finite set of points in $\mathbb{R}$ (a 0-chain) of odd cardinality.

(ii) The chain map $\gamma$ is called a homological almost-embedding of a simplicial complex $K$ in $\mathbb{R}$ if it is nontrivial and if, additionally, the following holds: whenever $\sigma$ and $\tau$ are disjoint simplices of $K$, their image chains $\gamma(\sigma)$ and $\gamma(\tau)$ have disjoint supports, where the support of a chain is the union of (the images of) the singular simplices with nonzero coefficient in that chain.

Remark 11. Suppose that $f : |K| \to \mathbb{R}^d$ is a continuous map.

(i) The induced chain map\textsuperscript{13} $f_\#: C_*(K) \to C_*(\mathbb{R}^d)$ is nontrivial.

(ii) If $f$ is an almost-embedding then the induced chain map is a homological almost-embedding.

Moreover, note that without the requirement of being nontrivial, we could simply take the constant zero chain map, for which the second requirement is trivially satisfied.

We have the following analogue of Proposition 8 for homological almost-embeddings.

**Proposition 12.** Suppose that $K$ is a finite simplicial complex with $\sigma^d(K) \neq 0$. Then $K$ does not admit a homological almost-embedding in $\mathbb{R}^d$.

As a corollary, we get the following result, which underlies our proof of Theorem 1.

**Corollary 13.** For any $k \geq 0$, the $k$-skeleton $\Delta_{2k+2}^{(k)}$ of the $(2k + 2)$-dimensional simplex has no homological almost-embedding in $\mathbb{R}^{2k}$.

We conclude this subsection by two facts that are not needed for the proof of the main result but are useful for the presentation of our method in Section 3.

If the ambient dimension $d = 2k + 1$ is odd, we can immediately see that $\Delta_{2k+1}^{(k)}$ has no homological almost-embedding in $\mathbb{R}^{2k+1}$ since it has no homological almost-embedding in $\mathbb{R}^{2k+2}$; this result can be slightly improved:

**Corollary 14.** For any $d \geq 0$, the $[d/2]$-skeleton $\Delta_{d+2}^{([d/2])}$ of the $(d + 2)$-dimensional simplex has no homological almost-embedding in $\mathbb{R}^d$.

**Proof.** The statement for even $d$ is already covered by the case $k = d/2$ of Corollary 13, so assume that $d$ is odd and write $d = 2k + 1$. If $K$ is a finite simplicial complex with $\sigma^d(K) \neq 0$ and if $CK$ is the cone over $K$ then $\sigma^{d+1}(CK) \neq 0$ (for a proof, see, for instance, [BKK02, Lemma 8]). Since we know that $\sigma^{2k}(\Delta_{2k+2}^{(k)}) \neq 0$ it follows that $\sigma^{2k+1}(\Delta_{2k+3}^{(k)}) \neq 0$. Consequently, $\sigma^{2k+1}(\Delta_{2k+3}^{(k+1)}) \neq 0$ since $C\Delta_{2k+2}^{(k)}$ is a subcomplex of $\Delta_{2k+3}^{(k+1)}$ and there exists an equivariant map from the deleted product of the subcomplex to the deleted product of the complex. Proposition 12 then implies that $\Delta_{2k+3}^{(k+1)}$ admits no homological almost-embedding in $\mathbb{R}^{2k+1}$.

The next fact is the following analogue of Radon’s lemma, proved in the next subsection along the proof of Proposition 12.

**Lemma 15** (Homological Radon’s lemma). For any $d \geq 0$, $\sigma^d(\partial\Delta_{d+1}) \neq 0$. Consequently, the boundary of $(d + 1)$-simplex $\partial\Delta_{d+1}$ admits no homological almost-embedding in $\mathbb{R}^d$.

### 2.3 Deleted Products and Obstructions

Here, we review the standard proof of Proposition 8 and explain how to adapt it to prove Proposition 12, which will follow from Lemma 19 and Lemma 20 (b) below. The reader unfamiliar with cohomology and willing to accept Proposition 12 can safely proceed to Section 3.

\textsuperscript{12}If we consider augmented chain complexes with chain groups also in dimension $-1$, then being nontrivial is equivalent to requiring that the generator of $C_{-1}(K) \cong \mathbb{Z}_2$ (this generator corresponds to the empty simplex in $K$) is mapped to the generator of $C_{-1}(\mathbb{R}) \cong \mathbb{Z}_2$.

\textsuperscript{13}The induced chain map is defined as follows: We assume that we have fixed a total ordering of the vertices of $K$. For a $p$-simplex $\sigma$ of $K$, the ordering of the vertices induces a homeomorphism $h_\sigma : [\sigma] \to [K]$. The image $f_\#(\sigma)$ is defined as the singular $p$-simplex $f \circ h_\sigma$.

9
We begin by recalling some basic notions of equivariant topology:

An action of the group \( \mathbb{Z}_2 \) on a space \( X \) is given by an automorphism \( \nu: X \to X \) such that \( \nu \circ \nu = 1_X \); the action is free if \( \nu \) does not have any fixed points. If \( X \) is a simplicial complex (or a cell complex), then the action is called simplicial (or cellular) if it is given by a simplicial (or cellular) map. A space with a given (free) \( \mathbb{Z}_2 \)-action is also called a (free) \( \mathbb{Z}_2 \)-space.

A map \( f: X \to Y \) between \( \mathbb{Z}_2 \)-spaces \((X, \nu)\) and \((Y, \mu)\) is called equivariant if it commutes with the respective \( \mathbb{Z}_2 \)-actions, i.e., \( f \circ \nu = \mu \circ f \). Two equivariant maps \( f_0, f_1: X \to Y \) are equivariantly homotopic if there exists a homotopy \( F: X \times [0,1] \to Y \) such that all intermediate maps \( f_t := F(\cdot, t), 0 \leq t \leq 1 \), are equivariant.

A \( \mathbb{Z}_2 \)-action \( \nu \) on a space \( X \) also yields a \( \mathbb{Z}_2 \)-action on the chain complex \( C_*(X) \), given by the induced chain map \( \nu_2: C_*(X) \to C_*(X) \) (if \( \nu \) is simplicial or cellular, respectively, then this remains true if we consider the simplicial or cellular chain complex of \( X \) instead of the singular chain complex), and if \( f: X \to Y \) is an equivariant map between \( \mathbb{Z}_2 \)-spaces then the induced chain map is also equivariant (i.e., it commutes with the \( \mathbb{Z}_2 \)-actions on the chain complexes).

**Spheres.** Important examples of free \( \mathbb{Z}_2 \)-spaces are the standard spheres \( \mathbb{S}^d, d \geq 0 \), with the action given by antipodality, \( x \mapsto -x \). There are natural inclusion maps \( \mathbb{S}^{d-1} \into \mathbb{S}^d \), which are equivariant. Antipodality also gives a free \( \mathbb{Z}_2 \)-action on the union \( \mathbb{S}^\infty = \bigcup_{d \geq 0} \mathbb{S}^d \), the infinite-dimensional sphere. Moreover, one can show that \( \mathbb{S}^\infty \) is contractible, and from this it is not hard to deduce that \( \mathbb{S}^\infty \) is a universal \( \mathbb{Z}_2 \)-space, in the following sense (see [Mil56] or also [Koz08, Prop. 8.16 and Thm. 8.17] for a more detailed textbook treatment).

**Proposition 16.** If \( X \) is any cell complex with a free cellular \( \mathbb{Z}_2 \)-action, then there exists an equivariant map \( f: X \to \mathbb{S}^\infty \). Moreover, any two equivariant maps \( f_0, f_1: X \to \mathbb{S}^\infty \) are equivariantly homotopic.

Any equivariant map \( f: X \to \mathbb{S}^\infty \) induces a nontrivial equivariant chain map \( f_*: C_* (X) \to C_* (\mathbb{S}^\infty) \). A simple fact that will be crucial in what follows is that Proposition 16 has an analogue on the level of chain maps.

We first recall the relevant notion of homotopy between chain maps: Let \( C_*(X) \) and \( C_*(Y) \) be (singular or simplicial, say) chain complexes, and let \( \varphi, \psi: C_*(X) \to C_*(Y) \) be chain maps. A chain homotopy \( \eta \) between \( \varphi \) and \( \psi \) is a family of homomorphisms \( \eta_t: C_j(X) \to C_{j+1}(Y) \) such that

\[
\varphi_j - \psi_j = \partial_{j+1} \eta_j + \eta_{j-1} \partial^j X
\]

for all \( j \). If \( X \) and \( Y \) are \( \mathbb{Z}_2 \)-spaces then a chain homotopy is called equivariant if it commutes with the (chain maps induced by) the \( \mathbb{Z}_2 \)-actions.

**Lemma 17.** If \( X \) is a cell complex with a free cellular \( \mathbb{Z}_2 \)-action then any two nontrivial equivariant chain maps \( \varphi, \psi: C_*(X) \to C_*(\mathbb{S}^\infty) \) are equivariantly chain homotopic.

**Proof of Lemma 17.** Let the \( \mathbb{Z}_2 \)-action on \( X \) be given by the automorphism \( \nu: X \to X \). For each dimension \( i \geq 0 \), the action partitions the \( i \)-dimensional cells of \( X \) (the basis elements of \( C_i(X) \)) into pairs \( \sigma, \nu(\sigma) \). For each such pair, we arbitrarily pick one of the cells and call it the representative of the pair.

We define the desired equivariant chain homotopy \( \eta \) between \( \varphi \) and \( \psi \) by induction on the dimension, using the fact that all reduced homology groups of \( \mathbb{S}^\infty \) are zero. (This just mimics the argument for the existence of an equivariant homotopy, which uses the contractibility of \( \mathbb{S}^\infty \).)

We start the induction in dimension at \( j = -1 \) (and for convenience, we also use the convention that all chain groups, chain maps, and \( \eta_j \) are understood to be zero in dimensions \( i < -1 \)). Since we assume that both \( \varphi \) and \( \psi \) are nontrivial, we have that \( \varphi_{-1}, \psi_{-1}: C_{-1}(X) \to C_{-1}(\mathbb{S}^\infty) \) are identical, and we set \( \eta_{-1}: C_{-1}(X) \to C_0(\mathbb{S}^\infty) \) to be zero.

Next, assume inductively that equivariant homomorphisms \( \eta_i: C_i(X) \to C_i(\mathbb{S}^\infty) \) have already been defined for \( i < j \) and satisfy

\[
\varphi_i - \psi_i = \eta_{i-1} \partial + \partial \eta_i
\]

Here, we use subscripts and superscripts on the boundary operators to emphasize which dimension and which chain complex they belong to; often, these indices are dropped and one simply writes \( \varphi - \psi = \partial \eta + \eta \partial \).

We also recall that if \( f, g: X \to Y \) are (equivariantly) homotopic then the induced chain maps are (equivariantly) chain homotopic. Moreover, chain homotopic maps induce identical maps in homology and cohomology.

We stress that we work with the cellular chain complex for \( X \).
for all $i < j$ (note that initially, this holds true for $j = 0$).

Suppose that $\sigma$ is a $j$-dimensional cell of $X$ representing a pair $\sigma, \nu(\sigma)$. Then $\partial \sigma \in C_{j-1}(X)$, and so $\eta_{j-1}(\partial \sigma) \in C_{j}(S^\infty)$ is already defined. We are looking for a suitable chain $c \in C_{j+1}(S^\infty)$ which we can take to be $\eta_j(\sigma)$ in order to satisfy the chain homotopy relation (1) also for $i = j$, such a chain $c$ has to satisfy $\partial c = b$, where

$$b := \varphi_j(\sigma) - \psi_j(\sigma) - \eta_{j-1}(\partial(\sigma)).$$

To see that we can find such a $c$, we compute

$$\partial b = \partial \varphi_j(\sigma) = \partial \psi_j(\sigma) + \partial \eta_{j-1}(\partial(\sigma))$$

$$= \varphi_{j-1}(\partial(\sigma)) - \psi_{j-1}(\partial(\sigma)) - \left( \varphi_{j-1}(\partial(\sigma)) - \psi_{j-1}(\partial(\sigma)) - \eta_{j-2}(\partial(\partial(\sigma))) \right) = 0$$

Thus, $b$ is a cycle, and since $H_j(S^\infty) = 0$, $b$ is also a boundary. Pick an arbitrary chain $c \in C_{j+1}(S^\infty)$ with $\partial c = b$ and set $\eta_j(\sigma) := c$ and $\eta_j(\nu(\sigma)) := \nu_j(c)$. We do this for all representative $j$-cells $\sigma$ and then extend $\eta_j$ by linearity. By definition, $\eta_j$ is equivariant and (1) is now satisfied also for $i = j$. This completes the induction step and hence the proof.

**Deleted products and Gauss maps.** Let $K$ be a finite simplicial complex. Then the Cartesian product $K \times K$ is a cell complex whose cells are the Cartesian products of pairs of simplices of $K$. The (combinatorial) deleted product $\tilde{K}$ of $K$ is defined as the polyhedral subcomplex of $K \times K$ whose cells are the products of vertex-disjoint pairs of simplices of $K$, i.e., $\tilde{K} := \{ \sigma \times \tau : \sigma, \tau \in K, \sigma \neq \tau \}$. The deleted product is equipped with a natural free $\mathbb{Z}_2$-action that simply exchanges coordinates, $(x, y) \mapsto (y, x)$. Note that this action is cellular since each cell $\sigma \times \tau$ is mapped to $\tau \times \sigma$.

**Lemma 18.** If $f: |K| \hookrightarrow \mathbb{R}^d$ is an embedding (or, more generally, an almost-embedding) then there exists an equivariant map $\tilde{f}: \tilde{K} \to S^{d-1}$.

**Proof.** Define $\tilde{f}(x, y) := \frac{f(x) - f(y)}{\|f(x) - f(y)\|}$. This map, called the Gauss map, is clearly equivariant.

For the proof of Proposition 12, we use the following analogue of Lemma 18.

**Lemma 19.** Let $K$ be a finite simplicial complex. If $\gamma: C_*(K) \to C_*(\mathbb{R}^d)$ is a homological almost-embedding then there is a nontrivial equivariant chain map (called the Gauss chain map) $\tilde{\gamma}: C_*(\tilde{K}) \to C_*(S^{d-1})$.

The proof of this lemma is not difficult but a bit technical, so we postpone it until the end of this section.

**Obstructions.** Here, we recall a standard method for proving the non-existence of equivariant maps between $\mathbb{Z}_2$-spaces. The arguments are formulated in the language of cohomology, and, as we will see, what they actually establish is the non-existence of nontrivial equivariant chain maps.

Let $\tilde{K}$ be a finite simplicial complex and let $\tilde{K}$ be its (combinatorial) deleted product. By Proposition 16, there exists an equivariant map $\tilde{G}_K: \tilde{K} \to S^\infty$, which is unique up to equivariant homotopy. By factoring out the action of $\mathbb{Z}_2$, this induces a map $\overline{G}_K: \overline{K} \to \mathbb{R}P^\infty$ between the quotient spaces $\overline{K} = \tilde{K}/\mathbb{Z}_2$ and $\mathbb{R}P^\infty = S^\infty/\mathbb{Z}_2$ (the infinite-dimensional real projective space), and the homotopy class of the map $\overline{G}_K$ depends only on $K$. Passing to cohomology, there is a uniquely defined induced homomorphism

$$\overline{G}_K^*: H^*(\mathbb{R}P^\infty) \to H^*(\overline{K}).$$

It is known that $H^d(\mathbb{R}P^\infty) \cong \mathbb{Z}_2$ for every $d \geq 0$. Letting $\zeta^d$ denote the unique generator of $H^d(\mathbb{R}P^\infty)$, there is a uniquely defined cohomology class

$$\sigma^d(K) := \overline{G}_K^*(\zeta^d),$$

---

17 We remark that a classical result due to Haefliger and Weber [Hae63, Web67] asserts that if $\dim K \leq (2d - 3)/3$ (the so-called metastable range) then the existence of an equivariant map from $\tilde{K}$ to $S^{d-1}$ is also sufficient for the existence of an embedding $K \to \mathbb{R}^d$ (outside the metastable range, this fails); see [Sk08] for further background.

18 We stress that this does not mean that there is only one homotopy class of continuous maps $\overline{K} \to \mathbb{R}P^\infty$; indeed, there exist such maps that do not come from equivariant maps $\tilde{K} \to S^\infty$; for instance the constant map that maps all of $\overline{K}$ to a single point.
called the van Kampen obstruction (with $\mathbb{Z}_2$-coefficients) to embedding $K$ into $\mathbb{R}^d$. For more details and background regarding the van Kampen obstruction, we refer the reader to [Mel09]. The basic fact about the van Kampen obstruction (and the reason for its name) is that $K$ does not embed (not even almost-embed) into $\mathbb{R}^d$ if $\sigma^d(K) \neq 0$ (Proposition 8). This follows from Lemma 18 and Part (a) of the following lemma:

**Lemma 20.** Let $K$ be a simplicial complex and suppose that $\sigma^d(K) \neq 0$.

(a) Then there is no equivariant map $\tilde{K} \to S^{d-1}$.

(b) In fact, there is no nontrivial equivariant chain map $C_\ast(\tilde{K}) \to C_\ast(S^{d-1})$.

Together with Lemma 19, Part (b) of the lemma also implies Proposition 12, as desired. The simple observation underlying the proof of Lemma 20 is the following

**Observation 21.** Suppose $\varphi: C_\ast(\tilde{K}) \to C_\ast(S^\infty)$ is a nontrivial equivariant chain map (not necessarily induced by a continuous map). By factoring out the action of $\mathbb{Z}_2$, $\varphi$ induces a chain map $\overline{\varphi}: C_\ast(\tilde{K}) \to C_\ast(\mathbb{R}P^\infty)$. The induced homomorphism in cohomology

$$\overline{\varphi}^*: H^\ast(\mathbb{R}P^\infty) \to H^\ast(\tilde{K})$$

is equal to the homomorphism $\overline{T}_K$ used in the definition of the Van Kampen obstruction, hence in particular

$$\sigma^d(K) = \overline{\varphi}^*(\xi^d).$$

**Proof.** By Lemma 17, $\varphi$ is equivariantly chain homotopic to the nontrivial equivariant chain map $(G_K)_\sharp$ induced by the map $G_K$. Thus, after factoring out the $\mathbb{Z}_2$-action, the chain maps $\overline{\varphi}$ and $(\overline{G}_K)_\sharp$ from $C_\ast(\tilde{K})$ to $C_\ast(\mathbb{R}P^\infty)$ are chain homotopic, and so induce identical homomorphisms in cohomology.

**Proof of Lemma 20.** If there exists an equivariant map $f: \tilde{K} \to S^{d-1}$, then the induced chain map $f_\sharp: C_\ast(\tilde{K}) \to C_\ast(S^{d-1})$ is equivariant and nontrivial, so (b) implies (a), and it suffices to prove the former.

Next, suppose for a contradiction that $\psi: C_\ast(\tilde{K}) \to C_\ast(S^{d-1})$ is a nontrivial equivariant chain map. Let $i: S^{d-1} \to S^\infty$ denote the inclusion map, and let $i_\sharp: C_\ast(S^{d-1}) \to C_\ast(S^\infty)$ denote the induced equivariant, nontrivial chain map. Then the composition $\tilde{\varphi} = (i_\sharp \circ \psi): C_\ast(\tilde{K}) \to C_\ast(S^\infty)$ is also nontrivial and equivariant, and so, by the preceding observation, for the induced homomorphism in cohomology, we get

$$\sigma^d(K) = (i_\sharp \circ \psi)^*(\xi^d) = \overline{\varphi}^*(\overline{T}(\xi^d)).$$

However, $\overline{T}(\xi^d) \in H^d(\mathbb{R}P^{d-1}) = 0$ (for reasons of dimension), hence $\sigma^d(K) = 0$, contradicting our assumption.

**Remark 22.** The same kind of reasoning also yields the well-known Borsuk–Ulam Theorem, which asserts that there is no equivariant map $S^d \to S^{d-1}$, using the fact that the inclusion $i: S^d \to S^\infty$ has the property that $\overline{T}(\xi^d)$, the pullback of the generator $\xi^d \in H^d(\mathbb{R}P^\infty)$, is nonzero. In fact, once again one gets a homological version of the Borsuk–Ulam theorem for free: there is no nontrivial equivariant chain map $C_\ast(S^d) \to C_\ast(S^{d-1})$.

**Proof of Lemma 15.** It is not hard to see that the deleted product $\partial \Delta_{d+1} = \Delta_{d+1}$ of the boundary of $(d + 1)$-simplex is combinatorially isomorphic to the boundary of a certain convex polytope and hence homeomorphic to $S^d$ (respecting the antipodal action), see [Mat03, Exercise 5.4.3]. Thus, the assertion $\sigma^d(\partial \Delta_{d+1}) \neq 0$ follows immediately from the preceding remark (the homological proof of the Borsuk–Ulam theorem). Together with Proposition 12, this implies that there is no homological almost-embedding of $\partial \Delta_{d+1}$ in $\mathbb{R}^d$.

The proof of Proposition 12 is complete, except for the following:

---

\[19\text{In fact, it is known that } H^\ast(\mathbb{R}P^\infty) \text{ is isomorphic to the polynomial ring } \mathbb{Z}_2[\xi], \text{ that } H^\ast(\mathbb{R}P^d) \cong \mathbb{Z}_2[\xi]/(\xi^{d+1}), \text{ and that } \overline{T} \text{ is just the quotient map.} \]
Proof of Lemma 19. Once again, we essentially mimic the definition of the Gauss map on the level of
chains. There is one minor technical difficulty due to the fact that the cells of \( \tilde{K} \) are products of simplices, whereas the singular homology of spaces is based on maps whose domains are simplices, not products of
simplices (this is the same issue that arises in the proof of K"unneth-type formulas in homology).

Assume that \( \gamma: C_*(\tilde{K}) \to C_*(\mathbb{R}^d) \) is a homological almost-embedding. The desired nontrivial equivariant chain map \( \tilde{\gamma}: C_*(\tilde{K}) \to C_*(S^{d-1}) \) will be defined as the composition of three intermediate non-trivial equivariant chain maps

\[
C_*(\tilde{K}) \xrightarrow{\alpha} D_* \xrightarrow{\beta} C_*(\mathbb{R}^d) \xrightarrow{p_*} C_*(S^{d-1}).
\]

These maps and intermediate chain complexes will be defined presently.

We define \( D_* \) as a chain subcomplex of the tensor product \( C_*(\mathbb{R}^d) \otimes C_*(\mathbb{R}^d) \). The tensor product chain complex has a basis consisting of all elements of the form \( s \otimes t \), where \( s \) and \( t \) range over the singular simplices of \( \mathbb{R}^d \), and we take \( D_* \) as the subcomplex spanned by all \( s \otimes t \) for which \( s \) and \( t \) have disjoint supports (note that \( D_* \) is indeed a chain subcomplex, i.e., closed under the boundary operator, since if \( s \) and \( t \) have disjoint supports, then so do any pair of simplices that appear in the boundary of \( s \) and of \( t \), respectively). The chain complex \( C_*(\tilde{K}) \) has a canonical basis consisting of cells \( \sigma \times \tau \), and the chain map \( \alpha \) is defined on these basis elements by “tensoring” \( \gamma \) with itself, i.e.,

\[
\alpha(\sigma \times \tau) := \gamma(\sigma) \otimes \gamma(\tau).
\]

Since \( \gamma \) is nontrivial, so is \( \alpha \), the disjointness properties of \( \gamma \) ensure that the image of \( \alpha \) does indeed lie in \( D_* \), and \( \alpha \) is clearly \( Z_2 \)-equivariant.

Next, consider the Cartesian product \( \mathbb{R}^d \times \mathbb{R}^d \) with the natural \( \mathbb{Z}_2 \)-action given by flipping coordinates. This action is not free since it has a nonempty set of fixed points, namely the “diagonal” \( \Delta = \{(x,x) : x \in \mathbb{R}^d \} \). However, the action on \( \mathbb{R}^d \times \mathbb{R}^d \) restricts to a free action on the subspace \( \mathbb{R}^d := (\mathbb{R}^d \times \mathbb{R}^d) \setminus \Delta \) obtained by removing the diagonal (this subspace is sometimes called the topological deleted product of \( \mathbb{R}^d \)). Moreover, there exists an equivariant map \( p: \mathbb{R}^d \to S^{d-1} \) defined as follows: we identify \( S^{d-1} \) with the unit sphere in the orthogonal complement \( \Delta^\perp = \{(w,w) : w \in \mathbb{R}^d \} \) and take \( p: \mathbb{R}^d \to S^{d-1} \) to be the orthogonal projection onto \( \Delta^\perp \) (which sends \((x,y)\) to \( \frac{1}{2}(x-y,x-y) \)), followed by renormalizing,

\[
p(x,y) := \frac{\frac{1}{2}(x-y,y-x)}{\frac{1}{2}(x-y,y-x)} \in S^{d-1} \subset \Delta^\perp.
\]

The map \( p \) is equivariant and so the induced chain map \( p_* \) is equivariant and nontrivial.

It remains to define \( \beta: D_* \to C_*(\mathbb{R}^d) \). For this, we use a standard chain map

\[
\text{EML}: C_*(\mathbb{R}^d) \otimes C_*(\mathbb{R}^d) \to C_*(\mathbb{R}^d \times \mathbb{R}^d),
\]

sometimes called the Eilenberg–Mac Lane chain map, and then take \( \beta \) to be the restriction to \( D_* \).

Given a basis element \( s \otimes t \) of \( C_*(\mathbb{R}^d) \otimes C_*(\mathbb{R}^d) \), where \( s: \Delta_p \to \mathbb{R}^d \) and \( t: \Delta_q \to \mathbb{R}^d \) are singular simplices, we can view \( s \otimes t \) as the map \( s \otimes t: \Delta_p \times \Delta_q \to \mathbb{R}^d \times \mathbb{R}^d \) with \((x,y)\) \( \mapsto (s(x),t(y)) \). This is almost like a singular simplex in \( \mathbb{R}^d \times \mathbb{R}^d \), except that the domain is not a simplex but a prism (product of simplices). The Eilenberg–Mac Lane chain map is defined by prescribing a systematic and coherent way of triangulating products of simplices \( \Delta_p \times \Delta_q \) that is consistent with taking boundaries; then \( \text{EML}(s \otimes t) \subset C_{p+q}(\mathbb{R}^d \times \mathbb{R}^d) \) is defined as the singular chain whose summands are the restrictions of the map \( \sigma \otimes \tau: \Delta_p \times \Delta_q \to (p+q) \)-simplices that appear in the triangulation of \( \Delta_p \times \Delta_q \). We refer to [GDR05] for explicit formulas for the chain map EML. What is important for us is that the chain map EML is equivariant and nontrivial. Both properties follow more or less directly from the construction of the triangulation of the prisms \( \Delta_p \times \Delta_q \), which can be explained as follows: Implicitly, we assume that the vertex sets \( \{0,1,\ldots,p\} \) and \( \{0,1,\ldots,q\} \) are totally ordered in the standard way. The vertex set of \( \Delta_p \) is the grid \( \{0,1,\ldots,p\} \times \{0,1,\ldots,q\} \), on which we consider the coordinatewise partial order defined by \( (x,y) \leq (x',y') \) if \( x \leq x' \) and \( y \leq y' \). Then the simplices of the triangulation are all totally ordered subsets of this partial order. Thus, if \( \sigma = \{(x_0,y_0),(x_1,y_1),\ldots,(x_r,y_r)\} \) is a simplex that
appears in the triangulation of $\Delta_p \times \Delta_q$ then the simplex $\sigma = \{(y_0, x_0), (y_1, x_1), \ldots, (y_r, x_r)\}$ obtained by flipping all coordinates appears in the triangulation of $\Delta_q \times \Delta_p$; see Figure 2. This implies equivariance of EML (and it is nontrivial since it maps a single vertex to a single vertex).

\[\bigcirc\]

3 Helly-type theorems from non-embeddability

We now detail the technique outlined in Section 1.4 and illustrate it on a few examples before formalizing its ingredients.

**Notation.** Given a set $X$ we let $2^X$ and \( \binom{X}{k} \) denote, respectively, the set of all subsets of $X$ (including the empty set) and the set of all $k$-element subsets of $X$. If $f : X \to Y$ is an arbitrary map between sets then we abuse the notation by writing $f(S)$ for $\{f(s) \mid s \in S\}$ for any $S \subseteq X$; that is, we implicitly extend $f$ to a map from $2^X$ to $2^Y$ whenever convenient.

3.1 Homotopic assumptions

Let $\mathcal{F} = \{U_1, U_2, \ldots, U_n\}$ denote a family of subsets of $\mathbb{R}^d$. We assume that $\mathcal{F}$ has empty intersection and that any proper subfamily of $\mathcal{F}$ has nonempty intersection. Our goal is to show how various conditions on the topology of the intersections of the subfamilies of $\mathcal{F}$ imply bounds on the cardinality of $\mathcal{F}$. For any (possibly empty) proper subset $I$ of $[n] = \{1, 2, \ldots, n\}$ we write $U_I$ for $\cap_{i \in [n] \setminus I} U_i$. We also put $U_{[n]} = \mathbb{R}^d$.

**Path-connected intersections in the plane.** Consider the case where $d = 2$ and the intersections $\bigcap \mathcal{G}$ are path-connected for all subfamilies $\mathcal{G} \subseteq \mathcal{F}$. Since every intersection of $n - 1$ members of $\mathcal{F}$ is nonempty, we can pick, for every $i \in [n]$, a point $p_i$ in $U_{[n] \setminus \{i\}}$. Moreover, as every intersection of $n - 2$ members of $\mathcal{F}$ is connected, we can connect any pair of points $p_i$ and $p_j$ by an arc $s_{i,j}$ inside $U_{[n] \setminus \{i,j\}}$. We thus obtain a drawing of the complete graph on $[n]$ in the plane in a way that the edge between $i$ and $j$ is contained in $U_{[n] \setminus \{i,j\}}$ (see Figure 3). If $n \geq 5$ then the stronger form of non-planarity of $K_5$ implies that there exist two edges $\{i,j\}$ and $\{k,\ell\}$ with no vertex in common and whose images intersect (see Proposition 8 and Lemma 9). Since $U_{[n] \setminus \{i,j\}} \cap U_{[n] \setminus \{k,\ell\}} = \bigcap \mathcal{F} = \emptyset$, this cannot happen and $\mathcal{F}$ has cardinality at most 4.

**[d/2]-connected intersections in $\mathbb{R}^d$.** The previous argument generalizes to higher dimension as follows. Assume that the intersections $\bigcap \mathcal{G}$ are $[d/2]$-connected\(^{20}\) for all subfamilies $\mathcal{G} \subseteq \mathcal{F}$. Then we can build by induction a function $f$ from the $[d/2]$-skeleton of $\Delta_{n-1}$ to $\mathbb{R}^d$ in a way that for any simplex $\sigma$, the image $f(\sigma)$ is contained in $U_{\sigma}$. The previous case shows how to build such a function from the 1-skeleton of $\Delta_{n-1}$. Assume that a function $f$ from the $\ell$-skeleton of $\Delta_{n-1}$ is built. For every

---

\(^{20}\)Recall that a set is $k$-connected if it is connected and has vanishing homotopy in dimension 1 to $k$. 

---
Figure 3: Two edges (arcs) with no common vertices intersect (in this case $s_{1,4}$ and $s_{2,5}$). The point in the intersection then belongs to all sets in $\mathcal{F}$.

For every facet $\tau$ of $\sigma$, we have $f(\tau) \subseteq U_{\tau}$. Thus, the set

$$\bigcup_{\tau \text{ facet of } \sigma} f(\tau)$$

is the image of an $\ell$-dimensional sphere contained in $U_{\tau}$, which has vanishing homotopy of dimension $\ell$. We can extend $f$ from this sphere to an $(\ell + 1)$-dimensional ball so that the image is still contained in $U_{\tau}$. This way we extend $f$ to the $(\ell + 1)$-skeleton of $\Delta_{n-1}$.

The Van Kampen–Flores theorem asserts that for any continuous function from $\Delta^{(k)}_{2k+2}$ to $\mathbb{R}^{2k}$ there exist two disjoint faces of $\Delta^{(k)}_{2k+2}$ whose images intersect (see Proposition 8 and Lemma 9). So, if $n \geq 2[d/2] + 3$, then there exist two disjoint simplices $\sigma$ and $\tau$ of $\Delta^{(d/2)}_{d/2}$ such that $f(\sigma) \cap f(\tau)$ is nonempty. Since $f(\sigma) \cap f(\tau)$ is contained in $U_{\sigma} \cap U_{\tau} = \bigcap \mathcal{F} = \emptyset$, this is a contradiction and $\mathcal{F}$ has cardinality at most $2[d/2] + 2$.

By a more careful inspection of odd dimensions, the bound $2[d/2] + 2$ can be improved to $d + 2$. We skip this in the homotopic setting, but we will do so in the homological setting (which is stronger anyway); see Corollary 23 below.

**Contractible intersections.** Of course, the previous argument works with other non-embeddability results. For instance, if the intersections $\bigcap \mathcal{G}$ are contractible for all subfamilies then the induction yields a map $f$ from the $d$-skeleton of $\Delta_{n-1}$ to $\mathbb{R}^d$ with the property that for any simplex $\sigma$, the image $f(\sigma)$ is contained in $U_{\sigma}$. The topological Radon theorem [BB79] (see also [Mat03, Theorem 5.1.2]) states that for any continuous function from $\Delta^d_{d+1}$ to $\mathbb{R}^d$ there exist two disjoint faces of $\Delta^d_{d+1}$ whose images intersect. So, if $n \geq d + 2$ we again obtain a contradiction (the existence of two disjoint simplices $\sigma$ and $\tau$ such that $f(\sigma) \cap f(\tau) \neq \emptyset$ whereas $U_{\sigma} \cap U_{\tau} = \bigcap \mathcal{F} = \emptyset$), and the cardinality of $\mathcal{F}$ must be at most $d + 1$.

### 3.2 From homotopy to homology

The previous reasoning can be transposed to homology as follows. Assume that for $i = 0, 1, \ldots, k-1$ all subfamilies $\mathcal{G} \subseteq \mathcal{F}$ we have $\tilde{\beta}_i(\bigcap \mathcal{G}) = 0$. We construct a nontrivial chain map $f$ from the simplicial chains of $\Delta^{(k)}_n$ to the singular chains of $\mathbb{R}^d$ by increasing dimension:

- For every $\{i\} \subseteq [n]$ we let $p_i \in U_{\{i\}}$. This is possible since every intersection of $n-1$ members of $\mathcal{F}$ is nonempty. We then put $f(\{i\}) = p_i$ and extend it by linearity into a chain map from $\Delta^{(0)}_n$ to $\mathbb{R}^d$. Notice that $f$ is nontrivial and that for any 0-simplex $\sigma \subseteq [n]$, the support of $f(\sigma)$ is contained in $U_{\sigma}$.

- Now, assume, as an induction hypothesis, that there exists a nontrivial chain map $f$ from the simplicial chains of $\Delta^{(k)}_n$ to the singular chains of $\mathbb{R}^d$ with the property that for any $(\leq \ell)$-simplex
σ ⊆ [n], ℓ < k, the support of f(σ) is contained in \( U_\sigma \). Let σ be a \((\ell + 1)\)-simplex in \( \Delta^{(\ell+1)} \). For every \( \ell \)-dimensional face τ of σ, the support of f(τ) is contained in \( U_\tau \subseteq U_\sigma \). It follows that the support of f(qτ) is contained in \( U_\sigma \), which has trivial homology in dimension \( \ell + 1 \). As a consequence, f(∂σ) is a boundary in \( U_\sigma \). We can therefore extend f to every simplex of dimension \( \ell + 1 \) and then, by linearity, to a chain map from the simplicial chains of \( \Delta^{(\ell+1)} \) to the singular chains of \( \mathbb{R}^d \). This chain map remains nontrivial and, by construction, for any \((\leq \ell + 1)\)-simplex \( \sigma \subseteq [n] \), the support of f(σ) is contained in \( U_\sigma \).

If σ and τ are disjoint simplices of \( \Delta^{(k)}_{n-1} \) then the intersection of the supports of f(σ) and f(τ) is contained in \( U_\sigma \cap U_\tau = \emptyset \) and these supports are disjoint. It follows that f is not only a nontrivial chain map, but also a homological almost-embedding in \( \mathbb{R}^d \). We can then use obstructions to the existence of homological almost-embeddings to bound the cardinality of \( F \). Specifically, since we assumed that \( F \) has empty intersection and any proper subfamily of \( F \) has nonempty intersection, Corollary 14 implies:

**Corollary 23.** Let \( F \) be a family of subsets of \( \mathbb{R}^d \) such that \( \bar{\beta}_i(\bigcap \mathcal{G}) = 0 \) for every \( \mathcal{G} \subseteq F \) and \( i = 0, 1, \ldots, [d/2] - 1 \). Then the Helly number of \( F \) is at most \( d + 2 \).

The homological Radon's lemma (Lemma 15) yields (noting \( \partial \Delta_{d+1} = \Delta^{(d+1)}_{d+1} \)):

**Corollary 24.** Let \( F \) be a family of subsets of \( \mathbb{R}^d \) such that \( \bar{\beta}_i(\bigcap \mathcal{G}) = 0 \) for every \( \mathcal{G} \subseteq F \) and \( i = 0, 1, \ldots, d - 1 \). Then the Helly number of \( F \) is at most \( d + 1 \).

**Remark 25.** The following modification of Example 3 shows that the two previous statements are sharp in various ways. First assume that for some values \( k, n \) there exists some embedding \( f \) of \( \Delta^{(k)}_{n-1} \) into \( \mathbb{R}^d \). Let \( K_i \) be the simplicial complex obtained by deleting the \( i \)-th vertex of \( \Delta^{(k)}_{n-1} \) (as well as all simplices using that vertex) and put \( U_i := f(K_i) \). The family \( F = \{U_1, \ldots, U_n\} \) has Helly number exactly \( n \), since it has empty intersection and all its proper subfamilies have nonempty intersection. Moreover, for every \( \mathcal{G} \subseteq F \), \( \bigcap \mathcal{G} \) is the image through \( f \) of the \( k \)-skeleton of a simplex on \( |F \setminus \mathcal{G}| \) vertices, and therefore \( \bar{\beta}_i(\bigcap \mathcal{G}) = 0 \) for every \( \mathcal{G} \subseteq F \) and \( i = 0, \ldots, k - 1 \). Now, such an embedding exists for:

- \( k = d \) and \( n = d + 1 \), as the \( d \)-dimensional simplex easily embeds into \( \mathbb{R}^d \). Consequently, the bound of \( d + 1 \) is best possible under the assumptions of Corollary 24.

- \( k = d - 1 \) and \( n = d + 2 \), as we can first embed the \((d - 1)\)-skeleton of the \( d \)-simplex linearly, then add an extra vertex at the barycentre of the vertices of that simplex and embed the remaining faces linearly. This implies that if we relax the condition of Corollary 24 by only controlling the first \( d - 2 \) Betti numbers then the bound of \( d + 1 \) becomes false. It also implies that the bound of \( d + 2 \) is best possible under (a strengthening of) the assumptions of Corollary 23.

(Recall that, as explained in Example 3, the \( [d/2] - 1 \) in the assumptions of Corollary 23 cannot be reduced without allowing unbounded Helly numbers.)

**Constrained chain map.** Let us formalize the technique illustrated by the previous example. We focus on the homological setting, as this is what we use to prove Theorem 1, but this can be easily transposed to homotopy.

Considering a slightly more general situation, we let \( F = \{U_1, U_2, \ldots, U_n\} \) denote a family of subsets of some topological space \( R \). As before for any (possibly empty) proper subset \( I \) of \( [n] = \{1, 2, \ldots, n\} \) we write \( U_I \) for \( \bigcap_{i \in [n] \setminus I} U_i \) and we put \( U_\emptyset = R \).

Let \( K \) be a simplicial complex and let \( \gamma : C_*(K) \to C_*(R) \) be a chain map from the simplicial chains of \( K \) to the singular chains of \( R \). We say that \( \gamma \) is constrained by \( (F, \Phi) \) if:

(i) \( \Phi \) is a map from \( K \) to \( 2^{[n]} \) such that \( \Phi(\sigma \cap \tau) = \Phi(\sigma) \cap \Phi(\tau) \) for all \( \sigma, \tau \in K \) and \( \Phi(\emptyset) = \emptyset \).

(ii) For any simplex \( \sigma \in K \), the support of \( \gamma(\sigma) \) is contained in \( U_{\Phi(\sigma)} \).

See Figure 4. We also say that a chain map \( \gamma \) from \( K \) is constrained by \( F \) if there exists a map \( \Phi \) such that \( \gamma \) is constrained by \( (F, \Phi) \). In the above constructions, we simply set \( \Phi \) to be the identity. As we already saw, constrained chain maps relate Helly numbers to homological almost-embeddings (see Definition 10) via the following observation:
Figure 4: An example of a constrained map \( \gamma : K \rightarrow \mathbb{R}^2 \). A label at a face \( \sigma \) of \( K \) denotes \( \Phi(\sigma) \). Note, for example, that the support of \( \gamma(\{a, b, c\}) \) needn’t be a triangle since we work with chain maps. Constrains by \( \Phi \) mean that a set \( U_i \) must contain cover images of all faces without label \( i \). It is demonstrated by \( U_3 \) and \( U_8 \) for example.

**Lemma 26.** Let \( \gamma : C_* (K) \rightarrow C_* (\mathbb{R}) \) be a nontrivial chain map constrained by \( \mathcal{F} \). If \( \bigcap \mathcal{F} = \emptyset \) then \( \gamma \) is a homological almost-embedding of \( K \).

**Proof.** Let \( \Phi : K \rightarrow 2^{[n]} \) be such that \( \gamma \) is constrained by \( (\mathcal{F}, \Phi) \). Since \( \gamma \) is nontrivial, it remains to check that disjoint simplices are mapped to chains with disjoint support. Let \( \sigma \) and \( \tau \) be two disjoint simplices of \( K \). The supports of \( \gamma(\sigma) \) and \( \gamma(\tau) \) are contained, respectively, in \( U_{\Phi(\sigma)} \) and \( U_{\Phi(\tau)} \), and

\[
U_{\Phi(\sigma)} \cap U_{\Phi(\tau)} = U_{\Phi(\sigma) \cap \Phi(\tau)} = U_{\Phi(\sigma \cap \tau)} = U_{\Phi(\emptyset)} = U_{\emptyset} = \bigcap \mathcal{F}.
\]

Therefore, if \( \bigcap \mathcal{F} = \emptyset \) then \( \gamma \) is a homological almost-embedding of \( K \). \( \square \)

### 3.3 Relaxing the connectivity assumption

In all the examples listed so far, the intersections \( \bigcap \mathcal{G} \) must be connected. Matoušek [Mat97] relaxed this condition into “having a bounded number of connected components”, the assumptions then being on the topology of the components, by using Ramsey’s theorem. The gist of our proof is to extend his idea to allow a bounded number of homology classes not only in the first dimension but in any dimension. Let us illustrate how Matoušek’s idea works in two dimension:

**Theorem 27 ([Mat97, Theorem 2 with \( d = 2 \)]).** For every positive integer \( b \) there is an integer \( h(b) \) with the following property. If \( \mathcal{F} \) is a finite family of subsets of \( \mathbb{R}^2 \) such that the intersection of any subfamily has at most \( b \) path-connected components, then the Helly number of \( \mathcal{F} \) is at most \( h(b) \).

Let us fix \( b \) from above and assume that for any subfamily \( \mathcal{G} \subseteq \mathcal{F} \) the intersection \( \bigcap \mathcal{G} \) consists of at most \( b \) path-connected components and that \( \bigcap \mathcal{F} = \emptyset \). We start, as before, by picking for every \( i \in [n] \), a point \( p_i \) in \( U_{\{i\}} \). This is possible as every intersection of \( n - 1 \) members of \( \mathcal{F} \) is nonempty. Now, if we consider some pair of indices \( i, j \in [n] \), the points \( p_i \) and \( p_j \) are still in \( U_{\{i,j\}} \) but may lie in different connected components. It may thus not be possible to connect \( p_i \) to \( p_j \) inside \( U_{\{i,j\}} \). If we, however, consider \( b + 1 \) indices \( i_1, i_2, \ldots, i_{b+1} \) then all the points \( p_{i_1}, p_{i_2}, \ldots, p_{i_{b+1}} \) are in \( U_{\{i_1, i_2, \ldots, i_{b+1}\}} \) which has at most \( b \) connected components, so at least one pair among of these points can be connected by a path inside \( U_{\{i_1, i_2, \ldots, i_{b+1}\}} \). Thus, while we may not get a drawing of the complete graph on \( n \) vertices we can still draw many edges.

To find many vertices among which every pair can be connected we will use the hypergraph version of the classical theorem of Ramsey:
Theorem 28 (Ramsey [Ram29]). For any x, y and z there is an integer \( R_{x,y,z} \) such that any \( x \)-uniform hypergraph on at least \( R_{x,y,z} \) vertices colored with at most \( y \) colors contains a subset of \( z \) vertices inducing a monochromatic sub-hypergraph.

From the discussion above, for any \( b+1 \) indices \( i_1 < i_2 < \ldots < i_{b+1} \) there exists a pair \( \{k, \ell\} \in \binom{[b+1]}{2} \) such that \( p_{i_k} \) and \( p_{i_\ell} \) can be connected inside \( U_{\{i_1, i_2, \ldots, i_{b+1}\}} \). Let us consider the \( (b+1) \)-uniform hypergraph on \( |n| \) and color every set of indices \( i_1 < i_2 < \ldots < i_{b+1} \) by one of the pairs in \( \binom{[b+1]}{2} \) that can be connected inside \( U_{\{i_1, i_2, \ldots, i_{b+1}\}} \) (if more than one pair can be connected, we pick one arbitrarily). Let \( t \) be some integer to be fixed later. By Ramsey’s theorem, if \( n \geq R_{b+1} \binom{[b+1]}{2} \) then there exist a pair \( \{k, \ell\} \in \binom{[b+1]}{2} \) and a subset \( T \subseteq [n] \) of size \( t \) with the following property: for any \( (b+1) \)-element subset \( S \subseteq T \), the points whose indices are the \( k \)th and \( \ell \)th indices of \( S \) can be connected inside \( U_T \).

Now, let us set \( t = 5 + \left( \frac{3}{2} \right)(b-1) = 10b - 5 \). We claim that we can find five indices in \( T \), denoted \( i_1, i_2, \ldots, i_5 \), and, for each pair \( \{i_u, i_v\} \) among these five indices, some \( (b+1) \)-element subset \( Q_{u,v} \subseteq T \) with the following properties:

(i) \( i_u \) and \( i_v \) are precisely in the \( k \)th and \( \ell \)th position in \( Q_{u,v} \), and

(ii) for any \( 1 \leq u, v, u', v' \leq 5 \), \( Q_{u,v} \cap Q_{u',v'} = \{i_u, i_v\} \cap \{i_{u'}, i_{v'}\} \).

We first conclude the argument, assuming that we can obtain such indices and sets. Observe that from the construction of \( T \), the \( i_u \)'s and the \( Q_{u,v} \)'s have the following property: for any \( u, v \in [5] \), we can connect \( p_{i_u} \) and \( p_{i_v} \) inside \( U_{Q_{u,v}} \). This gives a drawing of \( K_5 \) in the plane. Since \( K_5 \) is not planar, there exist two edges with no vertex in common, say \( \{u, v\} \) and \( \{u', v'\} \), that cross. This intersection point must lie in

\[
U_{Q_{u,v}} \cap U_{Q_{u',v'}} = U_{Q_{u,v} \cap Q_{u',v'}} = U_{\{i_u, i_v\} \cap \{i_{u'}, i_{v'}\}} = U_T = \bigcap F = \emptyset,
\]

a contradiction. Hence the assumption that \( n \geq R_{b+1} \binom{[b+1]}{2}, t \) is false and \( F \) has cardinality at most \( R_{b+1} \binom{[b+1]}{2}, 10b - 5 \) - 1, which is our \( h(b) \).

The selection trick. It remains to derive the existence of the \( i_u \)'s and the \( Q_{u,v} \)'s. It is perhaps better to demonstrate the method by a simple example to develop some intuition before we formalize it.

Example. Let us fix \( b = 4 \) and \( \{k, \ell\} = \{2, 3\} \in \binom{[4]}{2} \). We first make a ‘blueprint’ for the construction inside the rational numbers. For any two indices \( u, v \in [5] \) we form a totally ordered set \( Q'_{u,v} \subseteq \mathbb{Q} \) of size \( b+1 = 5 \) by adding three rational numbers (different from \( 1, \ldots, 5 \)) to the set \( \{u, v\} \) in such a way that \( u \) appears at the second and \( v \) at the third position of \( Q'_{u,v} \). For example, we can set \( Q'_{1,4} \) to be \( \{0.5; 1; 1.4; 4.7; 5.13\} \). Apart from this we require that we add a different set of rational numbers for each \( \{u, v\} \). Thus \( Q'_{u,v} \cap Q'_{u',v'} = \{u, v\} \cap \{u', v'\} \). Our blueprint now appears inside the set \( T' := \bigcup_{1 \leq u \leq w \leq 5} Q'_{u,v} \); note that both this set \( T' \) and the set \( T \) in which we search for the sets \( Q_{u,v} \) have 35 elements. To obtain the required indices \( i_u \) and sets \( Q_{u,v} \) it remains to consider the unique strictly increasing bijection \( \pi_0 : T' \to T \) and set \( i_u := \pi_0(u) \) and \( Q_{u,v} := \pi_0(Q'_{u,v}) \).

The general case. Let us now formalize the generalization of this trick that we will use to prove Theorem 1. Let \( Q \) be a subset of \( [w] \). If \( e_1 < e_2 < \ldots < e_w \) are the elements of a totally ordered set \( W \) then we call \( \{e_i : i \in Q\} \) the subset selected by \( Q \) in \( W \).

Lemma 29. Let \( 1 \leq q \leq w \) be integers and let \( Q \) be a subset of \( [w] \) of size \( q \). Let \( Y \) and \( Z \) be two finite totally ordered sets and let \( A_1, A_2, \ldots, A_r \) be \( q \)-element subsets of \( Y \). If \( |Z| \geq |Y| + r(w-q) \), then there exist an injection \( \pi : Y \to Z \) and \( r \) subsets \( W_1, W_2, \ldots, W_r \in \binom{[w]}{r} \) such that for every \( i \in [r] \), \( Q \) selects \( \pi(A_i) \) in \( W_i \). We can further require that \( W_i \cap W_j = \pi(A_i \cap A_j) \) for any two \( i, j \in [r] \), \( i \neq j \).

Proof. Let \( \pi_0 \) denote the monotone bijection between \( Y \) and \( [Y] \). For \( i \in [r] \) we let \( D_i \) denote a set of \( w-q \) rationals, disjoint from \( [Y] \), such that \( Q \) selects \( \pi(A_i) \) in \( D_i \cup \pi_0(A_i) \). We further require that the \( D_i \) are pairwise disjoint, and put \( Z' := |Y| \cup \bigcup_{i \in [r]} D_i \). Since \( |Z| \geq |Y| + r(w-q) = |Z'| \) there exists a strictly increasing map \( \nu : Z' \to Z \). We set \( \pi := \nu \circ \pi_0 \) and \( W_i := \nu(D_i \cup \pi_0(A_i)) \in \binom{[w]}{r} \). The desired condition is satisfied by this choice. See Figure 5. \( \Box \)
Figure 5: Illustration for the proof of Lemma 29. We assume that \( w = 4 \) and \( Q = \{1, 3, 4\} \).

4 Constrained chain maps and Helly number

We now generalize the technique presented in Section 3 to obtain Helly-type theorems from non-embedded results. We will construct constrained chain maps for arbitrary complexes. As above, \( \mathcal{F} = \{U_1, U_2, \ldots, U_n\} \) denotes a family of subsets of some topological space \( \mathbb{R} \) and for \( I \subseteq [n] \) we keep the notation \( U_I \) as used in the previous section (see the beginning of Subsection 3.1). Note that although so far we only used the reduced Betti numbers \( \tilde{\beta} \), in this section it will be convenient to work with standard (non-reduced) Betti numbers \( \beta \), starting with the following proposition.

**Proposition 30.** For any finite simplicial complex \( K \) and non-negative integer \( b \) there exists a constant \( h_K(b) \) such that the following holds. For any finite family \( \mathcal{F} \) of at least \( h_K(b) \) subsets of a topological space \( \mathbb{R} \) such that \( \bigcap \mathcal{G} \neq \emptyset \) and \( \beta_i(\bigcap \mathcal{G}) \leq b \) for any \( \mathcal{G} \subseteq \mathcal{F} \) and any \( 0 \leq i \leq \dim K \), there exists a nontrivial chain map \( \gamma : C_*(K) \to C_*(R) \) that is constrained by \( \mathcal{F} \).

The case \( K = \Delta_{2k+2}^d \), with \( k = \lceil d/2 \rceil \) and \( R = \mathbb{R}^d \), of Proposition 30 implies Theorem 1.

**Proof of Theorem 1.** Let \( b \) and \( d \) be fixed integers, let \( k = \lceil d/2 \rceil \) and let \( K = \Delta_{2k+2}^d \). Let \( h_K(b+1) \) denote the constant from Proposition 30 (we plug in \( b + 1 \) because we need to switch between reduced and non-reduced Betti numbers). Let \( \mathcal{F} \) be a finite family of subsets of \( \mathbb{R}^d \) such that \( \beta_i(\bigcap \mathcal{G}) \leq b \) for any \( \mathcal{G} \subseteq \mathcal{F} \) and every \( 0 \leq i \leq \dim K = \lceil d/2 \rceil - 1 \), in particular \( \beta_i(\bigcap \mathcal{G}) \leq b + 1 \) for such \( \mathcal{G} \). Let \( \mathcal{F}^* \) denote an inclusion-minimal sub-family of \( \mathcal{F} \) with empty intersection: \( \bigcap \mathcal{F}^* = \emptyset \) and \( \bigcap (\mathcal{F}^* \setminus \{U\}) \neq \emptyset \) for any \( U \in \mathcal{F}^* \). If \( \mathcal{F}^* \) has size at least \( h_K(b+1) \), it satisfies the assumptions of Proposition 30 and there exists a nontrivial chain map from \( K \) that is constrained by \( \mathcal{F}^* \). Since \( \mathcal{F}^* \) has empty intersection, this chain map is a homological almost-embedding by Lemma 26. However, no such homological almost-embedding exists by Corollary 13, so \( \mathcal{F}^* \) must have size at most \( h_K(b+1) - 1 \). As a consequence, the Helly number of \( \mathcal{F} \) is bounded and the statement of Theorem 1 holds with \( h(b, d) = h_K(b+1) - 1 \).

The rest of this section is devoted to proving Proposition 30. We proceed by induction on the dimension of \( K \), Section 4.1 settling the case of 0-dimensional complexes and Section 4.3 showing that if Proposition 30 holds for all simplicial complexes of dimension \( i \) then it also holds for all simplicial complexes of dimension \( i + 1 \). As the proof of the induction step is quite technical, as a warm-up, we provide the reader with a simplified argument for the induction step from \( i = 0 \) to \( i = 1 \) in Section 4.2. We let \( V(K) \) and \( v(K) \) denote, respectively, the set of vertices and the number of vertices of \( K \).
4.1 Initialization (dim $K = 0$)

If $K$ is a 0-dimensional simplicial complex then Proposition 30 holds with $h_K(b) = v(K)$. Indeed, consider a family $F$ of at least $v(K)$ subsets of $R$ such that all proper subfamilies have nonempty intersection. We enumerate the vertices of $K$ as $\{v_1, v_2, \ldots, v_{v(K)}\}$ and define $\Phi(\{v_i\}) = \{i\}$; in plain English, $\Phi$ is a bijection between the set of vertices of $K$ and $\{1, 2, \ldots, v(K)\}$. We first define $\gamma$ on $K$ by mapping every vertex $v \in K$ to a point $p(v) \in U_{\gamma(v)}$ then extend it linearly into a chain map $\gamma : C_0(K) \to C_0(R)$. It is clear that $\gamma$ is nontrivial and constrained by $(F, \Phi)$, so Proposition 30 holds when dim $K = 0$.

4.2 Principle of the induction mechanism (dim $K = 1$)

As a warm-up, we now prove Proposition 30 for 1-dimensional simplicial complexes. While this merely amounts to reformulating Matousek’s proof for embeddings [Mat97] in the language of chain maps, it still introduces several key ingredients of the induction while avoiding some of its complications. To avoid further technicalities, we use the non-reduced version of Betti numbers here.

Let $K$ be a 1-dimensional simplicial complex with vertices $\{v_1, v_2, \ldots, v_{v(K)}\}$ and assume that $F$ is a finite family of subsets of a topological space $R$ such that for any $G \subseteq F$, $\bigcap G \neq \emptyset$ and $\beta_0((\bigcap G)) \leq b$. Let $s \in N$ denote some parameter, to be fixed later. We assume that the cardinality of $F$ is large enough (as a function of $s$) so that, as argued in Subsection 4.1, there exist a bijection $\Psi : \Delta_s(0) \to [s + 1]$ and a nontrivial chain map $\gamma' : C_*(\Delta_s(0)) \to C_*(R)$ constrained by $(F, \Psi)$. We extend $\Psi$ to $\Delta_s$ by putting $\Psi(\sigma) = \bigcup_{v \in \sigma} \Psi(v)$ for any $\sigma \in \Delta_s$ and $\Psi(\emptyset) = \emptyset$. Remark that for any $\sigma, \tau \in \Delta_s$ we have $\Psi(\sigma \cap \tau) = \Psi(\sigma) \cap \Psi(\tau)$.

We now look for an injection $f$ of $V(K)$ into $V(\Delta_s)$ such that the chain map $\gamma' \circ f_2 : C_*(\Delta_s(0)) \to C_*(R)$ can be extended into a chain map $\gamma : C_*(K) \to C_*(R)$ constrained by $F$. Let $e = \{u, v\}$ be an edge in $K$. If we could arrange that $\gamma'(f(u) + f(v))$ is a boundary in $U_{\Psi([f(u), f(v)])}$, then we could simply define $\gamma(e)$ to be a chain in $U_{\Psi([f(u), f(v)])}$ bounded by $\gamma'(f(u) + f(v))$ (see Figure 6). Unfortunately this is too much to ask for but we can still follow the Ramsey-based approach of Subsection 3.3: we add “dummy” vertices to $\Psi([f(u), f(v)])$ to obtain a set $W_e$ such that $\gamma'(f(u) + f(v))$ is a boundary in $U_{\Psi([f(u), f(v)])}$. If we use different dummy vertices for distinct edges then setting $\gamma(e)$ to be a chain in $U_{\Psi([f(u), f(v)])}$ bounded by $\gamma'(f(u) + f(v))$ still yields a chain map constrained by $F$. We spell out the details in four steps.
Step 1. Any set $S$ of $2b + 1$ vertices of $\Delta_s$ contains two vertices $u_S, v_S \in S$ such that $\gamma'(u_S + v_S)$ is a boundary in $U_{H_0(S)}$. Indeed, notice first that for any $u \in S$, the support of $\gamma'(u)$ is contained in $U_{\Psi(S)}$. The assumption on $F$ about bounded Betti numbers of intersections of subfamilies of $F$ then ensures that there are at most $2^b$ distinct elements in $H_0(U_{\Psi(S)})$, as $H_0(U_{\Psi(S)}) \simeq \mathbb{Z}_{2^m}$ for some $m \leq b$. Thus, there are two vertices $u_S, v_S \in S$ such that $\gamma'(u_S)$ and $\gamma'(v_S)$ are in the same homology class in $H_0(U_{\Psi(S)})$. Since we consider homology with coefficients over $\mathbb{Z}_2$, the sum of two chains that are in the same homology class is always a boundary. In particular, $\gamma'(u_S + v_S) = \gamma'(u_S) + \gamma'(v_S)$ is a boundary in $U_{\Psi(S)}$.

Step 2. We use Ramsey’s theorem (Theorem 28) to ensure a uniform “$2$-in-$(2^b + 1)$” selection. Let $t$ be some parameter to be fixed in Step 3 and let $H$ denote the $(2^b + 1)$-uniform hypergraph with vertex set $V(\Delta_s)$. For every hyperedge $S \in H$ there exists (by Step 1) a pair $Q_S \in \binom{\binom{2^b + 1}{2}}{2}$ that selects a pair whose sum is mapped by $\gamma'$ to a boundary in $U_{\Psi(S)}$. We color $H$ by assigning to every hyperedge $S$ the “color” $Q_S$. Ramsey’s theorem thus ensures that if $s \geq R_{2^b + 1}\left(\binom{2^b + 1}{2}, t\right)$ then there exist a set $T$ of $t$ vertices of $\Delta_s$ and a pair $Q^* \in \binom{\binom{2^b + 1}{2}}{2}$ so that $Q^*$ selects in any $S \in \binom{T}{2^b + 1}$ a pair $\{u_S, v_S\}$ such that $\gamma'(u_S + v_S)$ is a boundary in $U_{\Psi(S)}$.

Step 3. Now, let $r$ be the number of edges of $K$ and let $\sigma_1, \sigma_2, \ldots, \sigma_r$ denote the edges of $K$. We define

$$h_K(b) = R_{2^b + 1}\left(\binom{\binom{2^b + 1}{2}}{2}, r(2^b - 1) + v(K)\right) + 1$$

and assume that $s \geq h_K(b) - 1$. We set the parameter $t$ introduced in Step 2 to $t = r(2^b - 1) + v(K)$.

We can now apply Lemma 29 with $Y = V(K)$, $Z = T$, $q = 2$, $w = 2^b + 1$, and $A_i = \sigma_i$ for $i \in [r]$. As a consequence, there exist an injection $f : V(K) \to T$ and $W_1, W_2, \ldots, W_r$ in $\binom{T}{2^b + 1}$ such that (i) for each $i$, $Q^*$ selects $f(\sigma_i)$ in $W_i$, and (ii) $W_i \cap W_j = f(\sigma_i \cap \sigma_j)$ for $i, j \in [r], i \neq j$.

Step 4. We define $\Phi$ by

$$\Phi(0) = \emptyset,$$

$$\Phi(\{v_i\}) = \Psi(f(v_i)) \quad \text{for } i = 1, 2, \ldots, v(K),$$

$$\Phi(\sigma_i) = \Psi(W_i) \quad \text{for } i = 1, 2, \ldots, r.$$

We define $\gamma$ on the vertices of $K$ by putting $\gamma(v) = \gamma'(f(v))$ for any $v \in V(K)$. Now remark that for any edge $\sigma_i = \{u, v\}$ of $K$, $\gamma'(f(u) + f(v))$ is a boundary in $U_{\Psi(W_i)}$; this follows from the definition of $T$ and the fact that $Q^*$ selects $\{f(u), f(v)\}$ in $W_i$. We can therefore define $\gamma(\{u, v\})$ to be some (arbitrary) chain in $U_{\Psi(W_i)}$ with boundary $\gamma'(f(u) + f(v))$. We then extend this map linearly into a chain map $\gamma : C_*(K) \to C_*(R)$.

To conclude the proof of Proposition 30 for 1-dimensional complexes it remains to check that the chain map $\gamma$ and the function $\Phi$ defined in Step 4 have the desired properties.

**Observation 31.** $\gamma$ is a nontrivial chain map constrained by $(F, \Phi)$.

**Proof.** First, it is clear from the definition that $\gamma$ is a chain map. Moreover, the definition of $\gamma'$ ensures that for every vertex $v \in K$ the support of $\gamma(v)$ is a finite set of points with odd cardinality. So $\gamma$ is indeed a nontrivial chain map.

The map $\Phi$ is from $K$ to $2^{[r+1]}$ and $\Phi(0)$ is by definition the empty set. The next property to check is that the identity $\Phi(\sigma \cap \tau) = \Phi(\sigma) \cap \Phi(\tau)$ holds for all $\sigma, \tau \in K$. When $\sigma$ and $\tau$ are vertices this follows from the injectivity of $\Psi$ and $f$. When $\sigma$ and $\tau$ are edges this follows from the same identity for $\Psi$ and the fact that Step 4 guaranteed that $W_i \cap W_j = f(\sigma_i \cap \sigma_j)$ for $i, j \in [r], i \neq j$. The remaining case is when $\sigma = \sigma_i$ is an edge and $\tau$ a vertex. Then, by construction, $\tau \in \sigma_i$ if and only if $f(\tau) \in W_i$, and

$$\Phi(\sigma_i) \cap \Phi(\tau) = \Psi(W_i) \cap \Psi(f(\tau)) = \Psi(W_i \cap f(\tau)) = \left\{ \begin{array}{ll} \Psi(0) & \text{if } f(\tau) \notin W_i \\ \Psi(f(\tau)) & \text{if } f(\tau) \in W_i \end{array} \right\} = \Phi(\sigma_i \cap \tau).$$

---

We could require that $\gamma'$ sends every vertex to a point in $U_{\Psi(S)}$, i.e. is a chain map induced by a map, and simply argue that since $U_{H_0(S)}$ has at most $b$ connected components, any $b + 1$ vertices of $\Delta_s$ contains some pair that can be connected inside $U_{\Psi(S)}$. This argument does not, however, work in higher dimension. Since Section 4.2 is meant as an illustration of the general case, we choose to follow the general argument.

21
It remains to check that for any simplex $\sigma \in K$, the support of $\gamma(\sigma)$ is contained in $U_{F(\sigma)}$. When $\sigma = \{v\}$ is a vertex then $\gamma(\sigma) = \gamma'(f(v))$. Since $\gamma'$ is constrained by $(F, \Psi)$, the support of $\gamma'(f(v))$ is contained in $U_{\Psi(f(v))} = U_{\Psi(v)}$, so the property holds. When $\sigma = \sigma_i$ is an edge, $\gamma(\sigma_i)$ is, by construction, a chain in $U_{\Psi(W_3)} = U_{\Psi(\sigma_i)}$ and the property also holds.

4.3 The induction

Let $k \geq 2$, let $K$ be a simplicial complex of dimension $k$ and assume that Proposition 30 holds for all simplicial complexes of dimension $k-1$ or less. Let $F$ be a finite family of subsets of a topological space $R$ such that for any $G \subseteq F$ and any $0 \leq i \leq k-1$, $\|G\| \neq \emptyset$ and $\beta(I \cap G) \leq b$. Assuming that $F$ contains sufficiently many sets, we want to construct a nontrivial chain map $\gamma : C_*(K) \to C_*(R)$ constrained by $F$.

**Preliminary example.** When going from $k = 0$ to $k = 1$, the first step (as described in Section 4.2) is to start with a constrained chain map $\gamma' : C_*(K^{(0)}) \to C_*(R)$ and observe that for some 1-simplices $\{u, v\} \in K$ the chain $\gamma'(\partial\{u, v\})$ must already be a boundary. To see that this is not the case in general, consider the drawing of $\Delta_2^{(2)}$ in an annulus depicted in the figure on the left. Observe that for every triangle $\{i, j, k\} \in \Delta_2^{(2)}$ the image, in this drawing, of $\partial\{i, j, k\}$ is a cycle going around the hole of the annulus and is therefore not a boundary. So, if we start with a chain map $\gamma'$ corresponding to that drawing, we will not be able to extend it by “filling” any triangle directly. This is not a peculiar example, and a similar construction can easily be done with arbitrarily many vertices. Observe, though, that the cycle going from 1 to 2, then 4, then 3 and then back to 1 is a boundary; in other words, if we replace, in the triangle $\partial\{1, 2, 3\}$, the edge from 2 to 3 by the concatenation of the edges from 2 to 4 and from 4 to 3, we build, using a chain map of $\Delta_2^{(2)}$ where no 2-face can be filled, a chain map of $\Delta_2^{(3)}$ where the 2-face can be filled. We systematize this observation using the barycentric subdivision of $K$.

**Barycentric subdivision.** The idea behind the notion of barycentric subdivision is that the geometric realization of a simplicial complex $K'$ can be subdivided by inserting a vertex at the barycentre of every face, resulting in a new, finer, simplicial complex, denoted $sd K'$, that is still homeomorphic to $K'$. Formally, the vertices of $sd K'$ consist of the faces of $K'$, except for the empty face, and the faces of $sd K'$ are the collections $\{\sigma_1, \ldots, \sigma_l\}$ of faces of $K'$ such that

$$\emptyset \neq \sigma_1 \subsetneq \sigma_2 \subsetneq \cdots \subsetneq \sigma_l.$$

In other words, the set of vertices of $sd K'$ is $K' \setminus \{\emptyset\}$ and the faces of $sd K'$ are the chains of $K' \setminus \{\emptyset\}$. For $\sigma \in K'$ we abuse the notation and let $sd \sigma$ denote the subdivision of $\sigma$ regarded as a subcomplex of $sd K'$, that is,

$$sd \sigma = \{\{\sigma_1, \ldots, \sigma_l\} \subseteq K' : \emptyset \neq \sigma_1 \subsetneq \sigma_2 \subsetneq \cdots \subsetneq \sigma_l \subseteq \sigma\}.$$

We will mostly manipulate barycentric subdivisions through the $sd \sigma$. For further reading on barycentric subdivisions we refer the reader, for example, to [Mat03, Section 1.7].

**Overview of the construction of $\gamma$.** Let $s \in \mathbb{N}$ be some parameter depending on $K$ and to be determined later. To construct $\gamma$ we will define three auxiliary chain maps

$$C_*(K^{(k-1)}) \xrightarrow{\alpha} C_*(sd K^{(k-1)}) \xrightarrow{\beta} C_*(\Delta_s^{(k-1)}) \xrightarrow{\gamma'} C_*(R)$$

As before, $\gamma'$ is a chain map from $C_*(\Delta_s^{(k-1)})$ constrained by $F$ and is obtained by applying the induction hypothesis. Unlike in Section 4.2, we do not inject the vertices of $K$ into those of $\Delta_s$ directly but proceed through $sd K$, the barycentric subdivision of $K$. We “inject” $K^{(k-1)}$ into $sd K^{(k-1)}$ by means of a chain map $\alpha$ (which will be the standard chain map corresponding to a subdivision). We then construct an
injection $\beta$ of the vertices of $\text{sd } K$ into the vertices of $\Delta_s$ which we extend linearly into a chain map $\beta^\#$. The key idea is the following:

The boundary of any $k$-simplex $\sigma$ of $K$ is mapped, under $\alpha$, to a sum of $k!$ boundaries of $k$-simplices of $\text{sd } K$, all of which are mapped through $\beta^\#$ to chains with the same homology in some appropriate $U_{\overline{\sigma}}$. Since $k!$ is even and we consider homology with coefficients in $\mathbb{Z}_2$, it follows that $\gamma' \circ \beta^\# \circ \alpha(\sigma)$ is a boundary in $U_{\overline{\sigma}}$. We therefore construct $\gamma'$ as an extension of $\gamma' \circ \beta^\# \circ \alpha$.

**Definition of $\gamma'$.** Since $\Delta_s^{(k-1)}$ has dimension $k - 1$, the induction hypothesis ensures that if the cardinality of $F$ is large enough then there exists a nontrivial chain map $\gamma' : C_*(\Delta_s^{(k-1)}) \to C_*(R)$ constrained by $F$. We denote by $\Psi$ a map such that $\gamma'$ is constrained by $(F, \Psi)$. Remark that $\Psi$ must be monotone over $\Delta_s^{(k-1)}$ as for any $\sigma \subseteq \tau \in \Delta_s^{(k-1)}$ we have $\Psi(\sigma) = \Psi(\sigma \cap \tau) = \Psi(\tau) \subseteq \Psi(\tau)$. It follows that for any $\sigma \in \Delta_s^{(k-1)}$ we have

$$\Psi(\sigma) = \bigcup_{\tau \in \Delta_s^{(k-1)}, \tau \subseteq \sigma} \Psi(\tau)$$

We use this identity to extend $\Psi$ to $\Delta_s$, that is we define:

$$\forall A \subseteq V(\Delta_s), \quad \Psi(A) = \bigcup_{\tau \in \Delta_s^{(k-1)}, \tau \subseteq A} \Psi(\tau).$$

Remark that the extended map still commutes with the intersection:

**Lemma 32.** For any $A, B \subseteq V(\Delta_s)$ we have $\Psi(A) \cap \Psi(B) = \Psi(A \cap B)$.

**Proof.** For any $A, B \subseteq V(\Delta_s)$ we have

$$\Psi(A) \cap \Psi(B) = \left( \bigcup_{\sigma \in \Delta_s^{(k-1)}, \sigma \subseteq A} \Psi(\sigma) \right) \cap \left( \bigcup_{\tau \in \Delta_s^{(k-1)}, \tau \subseteq B} \Psi(\tau) \right)$$

Distributing the union over the intersections we get

$$\Psi(A) \cap \Psi(B) = \bigcup_{\sigma, \tau \in \Delta_s^{(k-1)}, \sigma \subseteq A, \tau \subseteq B} \Psi(\sigma) \cap \Psi(\tau)$$

and as $\Psi(\sigma \cap \tau) = \Psi(\sigma) \cap \Psi(\tau)$ if $\sigma, \tau$ are simplices of $\Delta_s^{(k-1)}$, this rewrites as

$$\Psi(A) \cap \Psi(B) = \bigcup_{\sigma, \tau \in \Delta_s^{(k-1)}, \sigma \subseteq A, \tau \subseteq B} \Psi(\sigma \cap \tau).$$

Finally, observing that

$$\{\sigma \cap \tau : \sigma, \tau \in \Delta_s^{(k-1)}, \sigma \subseteq A, \tau \subseteq B\} = \{\vartheta : \vartheta \in \Delta_s^{(k-1)}, \vartheta \subseteq A \cap B\}$$

we get

$$\Psi(A) \cap \Psi(B) = \bigcup_{\vartheta \in \Delta_s^{(k-1)}, \vartheta \subseteq A \cap B} \Psi(\vartheta) = \Psi(A \cap B)$$

which proves the desired identity. □
Figure 7: The map $\alpha$ applied to a simplex $\sigma$ (left) and to $\partial \sigma$ (right). Significant parts of the boundaries $\partial \tau$ cancel out.

**Definition of $\alpha$.** Now we define a chain map $\alpha : C_*(K^{(k-1)}) \to C_*(\text{sd } K^{(k-1)})$ by first putting

$$\alpha : \sigma \in K^{(k-1)} \mapsto \sum_{\tau \in \text{sd } \sigma, \dim \tau = \dim \sigma} \tau,$$

and then extending that map linearly to $C_*(K^{(k-1)})$. See Figure 7. Remark that $\alpha$ behaves nicely with respect to the differential:

$$\alpha(\partial \sigma) = \sum_{\tau \in \text{sd } \sigma, \dim \tau = \dim \sigma} \partial \tau.$$

Note that the formula above makes sense and is valid even if $\sigma$ is a $k$-simplex although we define $\alpha$ only up to dimension $k - 1$.

**Definition of $\beta$.** We now construct the injection $\beta : V(\text{sd } K) \to V(\Delta_s)$ and, for constraining purposes, an auxiliary function $\kappa$ associating with every $k$-dimensional simplex of $K$ some simplex of $\Delta_s$. We want these functions to satisfy:

(P1) For any simplex $\sigma \in K$, $\kappa(\sigma) \cap \text{Im } \beta = \beta(V(\text{sd } \sigma))$.

(P2) For any $k$-simplices $\sigma, \tau \in K$, $\kappa(\sigma) \cap \kappa(\tau) = \beta(V(\text{sd } \sigma)) \cap \beta(V(\text{sd } \tau))$.

(P3) For any $k$-simplex $\sigma \in K$, when $\tau$ ranges over all $k$-simplices of $\text{sd } \sigma$, all chains $\gamma' \circ \beta_1(\partial \tau)$ have support in $U_{\Psi(\kappa(\sigma))}$ and are in the same homology class in $H_{k-1}(U_{\Psi(\kappa(\sigma))})$.

The intuition behind these properties is that $\kappa(\sigma)$ should augment $\beta(V(\text{sd } \sigma))$ by “dummy” vertices (P1) in a way that distinct simplices use disjoint sets of “dummy” vertices (P2). Property (P3), will allow building $\gamma$ over $k$-simplices as explained in the preceding overview.

We start the construction of $\beta$ and $\kappa$ with a combinatorial lemma. Let $\ell = 2^{k+1} - 1$ stand for the number of vertices of the barycentric subdivision of a $k$-dimensional simplex, and set $m = R_{k+1}(2^k, \ell)$.

**Claim 1.** For any integer $t$, if $s \geq R_m \left(\binom{m}{\ell}, t\right)$ then there exist a set $T$ of $t$ vertices of $\Delta_s$ and a set $Q^* \in \binom{[m]}{\ell}$ such that $Q^*$ selects in any $M \in \binom{T}{m}$ a subset $L_M$ with the following property: when $\sigma$ ranges over all $k$-simplices of $\Delta_s$ with $\sigma \subseteq L_M$, all chains $\gamma'(\partial \sigma)$ are in the same homology class in $H_{k-1}(U_{\Psi(M)})$.

**Proof.** Let $M$ be a subset of $m$ vertices of $\Delta_s$. Since $\gamma'$ is constrained by $(F, \Psi)$, for every $k$-simplex $\sigma \subseteq M$ the support of $\gamma'(\partial \sigma)$ is contained in $U_{\Psi(\partial \sigma)} \subseteq U_{\Psi(\sigma)} \subseteq U_{\Psi(M)}$. We can therefore color the $(k+1)$-uniform hypergraph on $M$ by assigning to every hyperedge $\sigma$ the homology class of $\gamma'(\partial \sigma)$ in $U_{\Psi(M)}$. Since $\beta_{k-1}(U_{\Psi(M)}) \leq b$, there are at most $2^b$ colors in this coloring. As $m = R_{k+1}(2^k, \ell)$, Ramsey’s Theorem implies that there exists a subset $L \subseteq M$ of $\ell$ vertices inducing a monochromatic hypergraph. We let $Q_M$ denote an element of $\binom{[m]}{\ell}$ that selects such a subset $L$. 

24
It remains to find a subset $T$ of vertices of $\Delta_s$ so that all $m$-element subsets $M \subseteq T$ give rise to the same $Q_M$. This is done by another application of Ramsey’s theorem to the $m$-uniform hypergraph on the vertices of $\Delta_s$, where each hyperedge $M$ is colored by the $\ell$-element subset $Q_M$. The subset $T$ can have size $t$ as soon as $s \geq R_m\left(\binom{m}{\ell}, \ell\right)$, which proves the statement.

Now, back to the construction of $\beta$ and $\kappa$. We first want a subset of $V(\Delta_s)$ with a “uniform $\ell$-in-$m$ selection” property of Claim 1 large enough so that we can inject $V(sd K)$ using Lemma 29. We set:

$$t = v(sd K) + r(m - \ell) \quad \text{and} \quad s^* = R_m\left(\binom{m}{\ell}, \ell\right),$$

and assume that $s \geq s^*$; since $s^*$ only depends on $b$ and $K$, this merely requires that $\mathcal{F}$ is large enough, again as a function of $b$ and $K$, so that $\gamma'$ still exists. We let $T$ and $Q^*$ denote the subset of $V(\Delta_s)$ and the element of $\binom{\binom{s}{k}}{l}$ whose existence follows from applying Claim 1. Let $\sigma_1, \sigma_2, \ldots, \sigma_r$ denote the $k$-dimensional simplices of $K$. We apply Lemma 29 with

$$Y = V(sd K), \quad Z = T, \quad A_i = V(sd \sigma_i), \quad q = \ell, \quad \text{and} \quad w = m,$

and obtain an injection $\pi : Y \to Z$ and $W_1, W_2, \ldots, W_r \in \binom{Z}{w}$ such that (i) for every $i \leq r$, $Q^*$ selects $\pi(A_i)$ in $W_i$, and (ii) for any $i \neq j \leq r$, $W_i \cap W_j = \pi(A_i \cap A_j)$. This injection $\pi$ is our map $\beta$ and we put $\kappa(\sigma_i) = W_i$. It is clear that Property (P1) holds, and since

$$\kappa(\sigma_i) \cap \kappa(\sigma_j) = W_i \cap W_j = \pi(A_i \cap A_j) = \beta(V(sd \sigma_i) \cap V(sd \sigma_j)) = \beta(V(sd \sigma_i)) \cap \beta(V(sd \sigma_j)),$$

Property (P2) also holds. The set $Q^*$ selects $\pi(A_i)$ in $W_i$ (Lemma 29) so Claim 1 ensures that when $\tau$ ranges over all $k$-simplices of $\Delta_s$ with $\pi \subseteq \pi(A_i)$, all chains $\gamma'(\partial \tau)$ have support in $U_{\mathcal{F}(\mathcal{W}, r)}$ and are in the same homology class in $H_{k-1}\left(U_{\mathcal{F}(\mathcal{W}, r)}\right)$. Substituting $\pi(A_i) = \beta(V(sd \sigma_i))$ and $W_i = \kappa(\sigma_i)$, we see that (P3) holds.

**Construction of $\gamma$.** Recall that we have the chain maps\textsuperscript{23}:

$$C_s\left(K^{(k-1)}\right) \xrightarrow{\alpha} C_s\left(sd K^{(k-1)}\right) \xrightarrow{\beta_2} C_s\left(\Delta_s^{(k-1)}\right) \xrightarrow{\gamma'} C_s(\mathcal{R}).$$

We define $\gamma = \gamma' \circ \beta_2 \circ \alpha$ as a chain map from $C_s\left(K^{(k-1)}\right)$ to $C_s(\mathcal{R})$. Let $\sigma$ be a $k$-dimensional simplex of $K$. From the definition of $\alpha$ we have

$$\gamma(\partial \sigma) = \sum_{\tau \subseteq sd \sigma, \dim \tau = \dim \sigma} \gamma'(\beta_2(\partial \tau)).$$

By property (P3), all summands in the above chain have support in $U_{\mathcal{F}(\kappa(\sigma))}$ and belong to the same homology class in $H_{k-1}\left(U_{\mathcal{F}(\kappa(\sigma))}\right)$. There is an even number of summands, namely $k!$, and we are using homology over $\mathbb{Z}_2$, so $\gamma' \circ \beta_2 \circ \alpha(\partial \sigma)$ has support in $U_{\mathcal{F}(\kappa(\sigma))}$ and is a boundary in $U_{\mathcal{F}(\kappa(\sigma))}$. We can therefore extend $\gamma$ into a chain map from $C_s(K)$ to $C_s(\mathcal{R})$ in a way that for any $k$-simplex $\sigma$ of $K$, the support of $\gamma(\sigma)$ is contained in $U_{\mathcal{F}(\kappa(\sigma))}$.

**Properties of $\gamma$.** First we verify that $\gamma$ is nontrivial. If $v$ is a vertex of $K$ then $sd v$ consists of a single simplex, also a vertex. The chain $\alpha(v)$ is thus a single vertex of $sd K$, and $\beta_2 \circ \alpha(v)$ is still a single vertex $\beta(sd v)$. Since $\gamma'$ is nontrivial, the support of $\gamma(v)$ is an odd number of points and therefore $\gamma$ is also nontrivial. It remains to argue that $\gamma$ is constrained by $(\mathcal{F}, \Phi)$ where:

$$\Phi : \begin{cases} K & \to 2^F \\ \sigma & \mapsto \begin{cases} \Psi(\beta(V(sd \sigma))) & \text{if } \dim \sigma \leq k - 1 \\ \Psi(\kappa(\sigma)) & \text{if } \dim \sigma = k \end{cases} \end{cases}$$

It is clear that $\Phi(\emptyset) = \Psi(\emptyset) = \emptyset$ by definition of $\Psi$. Also, the construction of $\gamma$ immediately ensures that for any $\sigma \in K$ the support of $\gamma(\sigma)$ is contained in $U_{\mathcal{F}(\sigma)}$. To conclude the proof that $\gamma$ is constrained by $(\mathcal{F}, \Phi)$ and therefore the induction it only remains to check that $\Phi$ commutes with the intersection:

\textsuperscript{23}$\beta_2$ is the chain map induced by $\beta$ restricted to chains of dimension at most $(k - 1)$.
Claim 2. For any $\sigma, \tau \in K$, $\Phi(\sigma \cap \tau) = \Phi(\sigma) \cap \Phi(\tau)$.

Proof. The claim is obvious for $\sigma = \tau$, so from now on assume that this is not the case. First assume that $\sigma$ and $\tau$ have dimension at most $k - 1$. Then,

$$\Phi(\sigma) \cap \Phi(\tau) = \Psi(\beta(V(\text{sd} \sigma))) \cap \Psi(\beta(V(\text{sd} \tau))) = \Psi(\beta(V(\text{sd} \sigma)) \cap \beta(V(\text{sd} \tau))),$$

the last equality following from Lemma 32. Since the map $\beta$ on subsets of $V(\Delta_s)$ is induced by a map $\beta$ on vertices of $\Delta_s$, we have $\beta(V(\text{sd} \sigma) \cap \beta(V(\text{sd} \tau)) = \beta(V(\text{sd} \sigma) \cap V(\text{sd} \tau))$. Moreover, by the definition of the barycentric subdivision we have $V(\text{sd} \sigma) \cap V(\text{sd} \tau) = V(\text{sd}(\sigma \cap \tau))$. Thus,

$$\Psi(\beta(V(\text{sd} \sigma)) \cap \beta(V(\text{sd} \tau))) = \Psi(\beta(V(\text{sd}(\sigma \cap \tau)))) = \Phi(\sigma \cap \tau),$$

and the statement holds for simplices of dimension at most $k - 1$.

Now assume that $\sigma$ and $\tau$ are both $k$-dimensional so that

$$\Phi(\sigma) \cap \Phi(\tau) = \Psi(\kappa(\sigma)) \cap \Psi(\kappa(\tau)) = \Psi(\kappa(\sigma) \cap \kappa(\tau)) = \Psi(\beta(V(\text{sd} \sigma)) \cap \beta(V(\text{sd} \tau))),$$

the last identity following from Property (P2) of the map $\kappa$. Again, from the definition of $\beta$ and the barycentric subdivision we have

$$\beta(V(\text{sd} \sigma)) \cap \beta(V(\text{sd} \tau)) = \beta(V(\text{sd}(\sigma \cap \tau))).$$

We thus obtain

$$\Phi(\sigma) \cap \Phi(\tau) = \Psi \circ \beta \circ V(\text{sd}(\sigma \cap \tau)) = \Phi(\sigma \cap \tau),$$

the last identity following from the definition of $\Phi$ on simplices of dimension at most $k - 1$. The statement also holds for simplices of dimension $k$.

Finally assume that $\sigma$ and $\tau$ are of dimension $k$ and at most $k - 1$ respectively. Then, applying Lemma 32 we have:

$$\Phi(\sigma) \cap \Phi(\tau) = \Psi(\kappa(\sigma)) \cap \Psi(\beta(V(\text{sd} \tau))) = \Psi(\kappa(\sigma) \cap \beta(V(\text{sd} \tau))).$$

Note that $\beta(V(\text{sd} \tau)) \subseteq \text{Im} \beta$ and that, by property (P1), $\kappa(\sigma) \cap \text{Im} \beta = \beta(V(\text{sd} \sigma))$. We thus have

$$\kappa(\sigma) \cap \beta(V(\text{sd} \tau)) = \beta(V(\text{sd} \sigma)) \cap \beta(V(\text{sd} \tau)) = \beta(V(\text{sd}(\sigma \cap \tau))),$$

the last equality following, again, from the definition of barycentric subdivision. As $\sigma \cap \tau$ has dimension at most $k - 1$ we have

$$\Phi(\sigma) \cap \Phi(\tau) = \Psi(\beta(V(\text{sd}(\sigma \cap \tau)))) = \Phi(\sigma \cap \tau)$$

and the statement holds for the last case. 

\[
\square
\]

References


Shortest path embeddings of graphs on surfaces

Alfredo Hubard† Vojtěch Kaluža‡ Arnaud de Mesmay§ Martin Tancer¶

August 12, 2016

Abstract

The classical theorem of Fáry states that every planar graph can be represented by an embedding in which every edge is represented by a straight line segment. We consider generalizations of Fáry’s theorem to surfaces equipped with Riemannian metrics. In this setting, we require that every edge is drawn as a shortest path between its two endpoints and we call an embedding with this property a shortest path embedding. The main question addressed in this paper is whether given a closed surface $S$, there exists a Riemannian metric for which every topologically embeddable graph admits a shortest path embedding. This question is also motivated by various problems regarding crossing numbers on surfaces.

We observe that the round metrics on the sphere and the projective plane have this property. We provide flat metrics on the torus and the Klein bottle which also have this property.

Then we show that for the unit square flat metric on the Klein bottle there exists a graph without shortest path embeddings. We show, moreover, that for large $g$, there exist graphs $G$ embeddable into the orientable surface of genus $g$, such that with large probability a random hyperbolic metric does not admit a shortest path embedding of $G$, where the probability measure is proportional to the Weil-Petersson volume on moduli space.

Finally, we construct a hyperbolic metric on every orientable surface $S$ of genus $g$, such that every graph embeddable into $S$ can be embedded so that every edge is a concatenation of at most $O(g)$ shortest paths.

1 Introduction

Fáry’s theorem and joint crossing numbers. A famous theorem of Fáry [11] states that any simple planar graph can be embedded so that edges are represented by straight line segments. In this article we investigate analogues of this theorem in the context of graphs embedded into surfaces. We focus on the following problem: Given a surface $S$, is there a metric on $S$ such that every graph embeddable into $S$ can be embedded so that edges are represented by shortest paths?

---

*The project was partially supported by the Czech-French collaboration project EMBEDS (CZ: 7AMB15FR003, FR: 33936TF). The research of A. H. was funded by GUDHI, geometric understanding in higher dimensions. The research of A. dM. leading to these results has received funding from the People Programme (Marie Curie Actions) of the European Union’s Seventh Framework Programme (FP7/2007-2013) under REA grant agreement n [291734]. V. K. and M. T. were partially supported by the project CE-ITI (GA ČR P202/12/G061).

†INRIA Sophia-Antipolis and Université Paris-Est Marne-la-Vallée, France, alfredo.hubard@inria.fr
‡Department of Applied Mathematics, Charles University, Malostranské náměstí 25, Prague 1, Czech Republic, 
kaluza@kam.mff.cuni.cz
§CNRS, Gipsa-Lab, 11 rue des Mathématiques, 38 400 Saint Martin d’Hres, France, arnaud.de-mesmay@gipsa-lab.fr
¶Department of Applied Mathematics and Institute of Theoretical Computer Science, Charles University, Malostranské náměstí 25, Prague 1, Czech Republic, tancer@kam.mff.cuni.cz
We call such an embedding a shortest path embedding, and such a metric a universal shortest path metric.¹

Before being enticed by this question, we were motivated to consider it by a number of problems involving joint embeddings of curves or graphs on surfaces arising from seemingly disparate settings. The literature on the subject goes back at least 15 years with Negami’s work related to diagonal flips in triangulations [25]. He conjectured that there exists a universal constant $c$ such that for any pair of graphs $G_1$ and $G_2$ embedded in a surface $S$, there exists a homeomorphism $h : S \to S$ such that $h(G_1)$ and $G_2$ intersect transversely at their edges and the number of edge crossings satisfies $cr(h(G_1), G_2) \leq c|E(G_1)| \cdot |E(G_2)|$.

Recently, on one hand, Matoušek, Sedgwick, Tancer, and U. Wagner [20, 21], working on decidability of embeddability of 2-complexes into $\mathbb{R}^3$ and on the other hand, Geelen, Huynh, and Richter [13], in a quest for explicit bounds for graph minors, were faced with a similar question and provided bounds for related problems. Joint crossing number type problems are dually equivalent to problems of finding a graph with a specific pattern within an embedded graph while bounding the multiplicity of the edges used. This is a fundamental concern of computational topology of surfaces where one is interested in finding objects with a fixed topology and minimal combinatorial complexity, e.g., short canonical systems of loops [19], short pants decompositions [6] or short octagonal decompositions [4]; see also [5].

Negami provided the upper bound $cr(h(G_1), G_2) \leq cg|E(G_1)| \cdot |E(G_2)|$, and despite subsequent discoveries [1, 26], his conjecture is still open. In a paper that refines Negami’s work [26], Richter and Salazar wrote “this conjecture seems eminently reasonable: why should two edges be forced to cross more than once?” The connection with our work is that if two graphs are embedded transversally by shortest path embeddings, then indeed no two edges cross more than once, since otherwise one of them could be shortcut. We note that prior to our work, Schaefer [27, paragraph on Geodesic crossing numbers] had considered similar questions, mainly for drawing edges of a graph by geodesics. We provide the details below including answers to some of Schaefer’s questions. We also note that our methods easily yield a new proof of Negami’s theorem for orientable surfaces; see Corollary 20.

Beyond crossing numbers, the existence or non-existence of shortest path universal metrics might be relevant in curvature free and extremal Riemannian geometry.

Related work. Various results in graph drawing [32] revolve around generalizing Fáry’s theorem to find drawings of graphs with additional constraints, for instance drawing the edges with polylines with few bends. On the other hand, only few extensions to graphs embedded in surfaces are known. Two classical avatars of Fáry’s theorem in the plane are of relevance to our work: Tutte’s barycentric embedding theorem [33] and the Koebe-Andreev-Thurston circle packing theorem (see, for example, the book of Stephenson [28]). Both have been generalized to surfaces, providing positive answers to the following questions:

1. Given a surface $S$, a metric $m$, and a graph $G$ embeddable into $S$, can we embed the graph $G$ so that every edge is represented by a geodesic with respect to $m$?

2. Given a graph $G$ embeddable into $S$, does there exist a metric $m$ on $S$ so that $G$ embeds into $S$ with shortest paths?

The first question was considered by Schaefer in [27], a positive answer for many metrics had been previously given by Y. Colin de Verdière [7] who generalized Tutte’s barycentric embedding approach using a variational principle. The idea behind this approach is to start with a topological embedding of the graph, replace the edges by springs, and let the system reach an equilibrium. Y. Colin de Verdière proved that for any metric of non-positive curvature,

¹We do not require that these shortest paths are unique but as we will see later on, in the case of our positive results, i.e., Theorem 1 and 4, the uniqueness of the shortest paths can be obtained as well.)
the edges become geodesics with disjoint interiors when the system reaches stability; moreover, this embedding is essentially unique within its homotopy class. However, geodesics need not be shortest paths, and two geodesics can intersect an arbitrarily large number of times, see Figure 1. Yet, these examples do not provide a negative answer to the second question, or to our main question, since we could change the embedding by a homeomorphism of the torus (thus even preserving the combinatorial map) to obtain a shortest path embedding.

The second question also has a positive answer, which can be proved via a generalization of the circle packing theorem to closed surfaces [28]. Namely, for every triangulation $T$ of a surface, there exists a metric of constant curvature so that $T$ can be represented as the contact graph of a family of circles. The representation of the triangulation that places a vertex at the center of its corresponding circle is an embedding with shortest paths. Such a representation can be computed efficiently and can be used as a tool for representing graphs on surfaces [23]. However, the metric is determined by the triangulation, which makes this approach ill-suited for our purpose.

**Our results.** Our objective here is a mix of these last two results. On the one hand, we require shortest paths and not geodesics, on the other hand, we want a single metric for each surface and not one which depends on the triangulation. We will also consider the relaxation of our problem where we are allowed to use concatenations of shortest paths: we say that a metric is a $k$-universal shortest path metric if every topologically embeddable graph can be represented by an embedding in which edges are drawn as concatenations of $k$ shortest paths. This is akin to various problems in graph drawing where graphs are embedded with polylines with a bounded number of bends instead of straight lines [9, 31].

Our results focus on Riemannian metrics of constant curvature, and our techniques are organized by the sign of the curvature. We first observe that for the sphere and the projective plane, since there is a unique Riemannian metric of curvature 1, the circle packing approach applies to all graphs. Then, with the aid of irreducible triangulations, we provide flat metrics (i.e., of zero curvature) on the torus and the Klein bottle for which every graph admits a shortest path embedding.

**Theorem 1.** The sphere $S^2$, the projective plane $\mathbb{R}P^2$, the torus $T^2$, and the Klein bottle $K$ can be endowed with a universal shortest path metric.

This result could lead to the idea that shortest path embeddings can be achieved for any metric, i.e., that every metric is a universal shortest path metric. We prove that this is not the case already for the unit square flat metric on the Klein bottle (arguably the first example to consider).

**Theorem 2.** Let $K$ denote the Klein bottle endowed with the unit square flat metric on the polygonal scheme $aba^{-1}b$. Then there exists a graph embeddable into $K$ which cannot be embedded into $K$ so that the edges are shortest paths.
In higher genus, the number of irreducible triangulations is too large to check all cases by hand. Hyperbolic surfaces of large genus are hard to comprehend, but the probabilistic method allows us to show that if there exists universal shortest path metrics of constant curvature $-1$ at all, their fraction tends to 0 as the genus tends to infinity.

**Theorem 3.** For any $\varepsilon > 0$, with probability tending to 1 as $g$ goes to infinity, a random hyperbolic metric is not a $O(g^{1/3-\varepsilon})$-universal shortest path metric. In particular, with probability tending to 1 as $g$ goes to infinity, a random hyperbolic metric is not a universal shortest path metric.

Here the probability measure on the space hyperbolic surfaces is proportional to the Weil-Petersson volume, see Section 5. Our proof is an application of deep results on this volume by Mirzakhani [22] and Guth, Parlier, and Young [15].

For a given graph $G$ and a metric $m$ on $S$, Schaefer [27] defines the geodesic crossing number of $G$ as the minimal number of crossings of any drawing of $G$ in $S$ in which edges are represented by geodesics. Schaefer asks if this definition is equivalent to the analogous definition with shortest paths instead of geodesics. Notice that the examples in Theorems 2 and 3 have nonpositive curvature, hence, combined with the aforementioned result of Y. Colin de Verdière imply that some graphs have geodesic crossing number zero but shortest path crossing number nonzero, answering Schaefer’s question.

For genus $g > 1$ we do not know if there exist shortest path universal metrics. But relaxing the question to concatenations of shortest paths and combining ideas from hyperbolic geometry and computational topology, we provide for every orientable surface of genus $g$ an $O(g)$-universal shortest paths metric. The proof relies on the octagonal decompositions of É. Colin de Verdière and Erickson [4] and a variant of the aforementioned theorem of Y. Colin de Verdière [7].

**Theorem 4.** For every $g > 1$, there exists an $O(g)$-universal shortest path hyperbolic metric $m$ on the orientable surface $S$ of genus $g$.

In this article we focused on Riemannian metrics of constant curvature, but we remark that both of our last results also hold in some setting of piecewise-Euclidean metrics as well. For the upper bound, it suffices to replace hyperbolic hexagons with Euclidean ones, and the rest of the proof works similarly. The lower bound can be derived following the heuristic strong parallels between the Weil-Petersson volume form on moduli space and the counting measure on the space of $N = 4g$ Euclidean triangles randomly glued together. In particular the results that we use have analogs in this latter space: see Brooks and Makover [2] and the second half of the article of Guth, Parlier, and Young [15].

We have stated our results for graphs in this introduction. We note that one could consider the problem of shortest path embeddings for a graph with a fixed embedding up to a homeomorphism of the surface (i.e., for a combinatorial map), which is more in the spirit of Negami’s and Mohar’s conjectures. Our positive results can be stated in this stronger version; i.e., in our proofs the map is preserved. Our negative results would be weaker if the map had to be preserved, and in fact the proofs deal firstly with the statements for maps and then we derive the analog for graphs with some extra work.

**Open questions.** The main open question is the existence of universal shortest path metrics, or $O(1)$-universal shortest path metrics. Natural candidates for these are given by certain celebrated extremal metrics like the ones occurring as lower bounds for Gromov’s systolic inequality [3, 14].

Lists of irreducible triangulations exist for the double torus and the non-orientable surface of genus up to four [29]. While the numbers are too big to be investigated by hand as we did for the torus and the Klein bottle, it may be possible to investigate some computerized approach to test their shortest path embeddability for some well chosen hyperbolic metric.
Our Theorem 4 only deals with orientable surfaces. A similar approach might work for non-orientable surfaces as well, the key issue being to generalize the octagonal decompositions of É. Colin de Verdière and Erickson [4] to the non-orientable setting. We leave this as an open problem.

Outline. After introducing the main definitions in Section 2, we will prove Theorems 1, 2, 3, and 4 in Sections 3, 4, 5, and 6, respectively.

2 Preliminaries

In this article we only deal with compact surfaces without boundaries. By the classification theorem, these are characterized by their orientability and their genus, generally denoted by $g$. Orientable surfaces of genus 0 and 1 are respectively the sphere $S^2$ and the torus $T^2$, while non-orientable surfaces of genus 1 and 2 are the projective plane $\mathbb{R}P^2$ and the Klein bottle $K$. The orientable surface of genus $g$ is denoted by $S_g$. The Euler genus is equal to the genus for non-orientable surfaces and equals twice the genus for orientable surfaces.

By a path on a surface $S$ we mean a continuous map $p : [0, 1] \to S$, and a closed curve denotes a continuous map $\gamma : S^1 \to S$. These are simple if they are injective. We will be using occasionally the notions of homotopy, homology, and universal cover, we refer to Hatcher [16] for an introduction to these concepts. All the graphs that we consider in this paper are simple graphs unless specified otherwise, i.e., loops and multiple edges are disallowed. An embedding of a graph $G$ into a surface $S$ is, informally, a crossing-free drawing of $G$ on $S$. We refer to Mohar and Thomassen [24] for a thorough reference on graphs on surfaces, and only recall the main definitions. A graph embedding is cellular if its faces are homeomorphic to open disks. Euler’s formula states that $v - e + f = 2 - g$ for any graph with $v$ vertices, $e$ edges, and $f$ faces cellularly embedded in a surface $S$ of Euler genus $g$. When the graph is not cellularly embedded, this becomes an inequality: $v - e + f \geq 2 - g$. A triangulation of a surface is a cellular graph embedding such that all the faces are adjacent to three edges. By a slight abuse of language, we will sometimes refer to an embedding of a triangulation, by which we mean an embedding of its underlying graph which is homeomorphic to the given triangulation. A pants decomposition of a surface $S$ is a family of disjoint curves $\Gamma$ such that cutting $S$ along all of the curves of $\Gamma$ gives a disjoint union of pairs of pants, i.e., spheres with three boundaries. Every surface except the sphere, the projective plane, the torus, and the Klein bottle admits a pants decomposition with $3g - 3$ closed curves and $2g - 2$ pairs of pants. Note that all the pants decompositions are not topologically the same, i.e., are not related by a self-homeomorphism of the surface. A class of pants decompositions equivalent under such homeomorphisms will be called the (topological) type of the pants decomposition. We say that an embedding $f : G \to S$ contains a pants decomposition if there exists a subgraph $H \subseteq G$ such that $f : H \to S$ is a pants decomposition of $S$.

In this article, we will also be dealing with notions coming from Riemannian geometry, we refer to the book of do Carmo for more background [8]. By a metric we always mean a Riemannian metric, which associates to every point of a surface the curvature at this point. The Gauss-Bonnet theorem ties geometry and topology; it implies that the sign of a metric of constant curvature that a topological surface accepts is determined solely by its Euler genus.

A Riemannian metric induces a length functional on paths and closed curves. A path or a closed curve is a geodesic if the functional is locally minimal. Shortest paths between two points are global minima of the length functional. Unlike in the plane, geodesics are not, in general, shortest paths; in addition, neither geodesics nor shortest paths are unique in general. If we have a shortest path embedding of a graph where every edge is drawn as the unique shortest path between its endpoints, we speak of shortest paths embedding with uniqueness.
3 Shortest path embeddings through minimal triangulations

Theorem 1. The sphere $S^2$, the projective plane $\mathbb{R}P^2$, the torus $T^2$, and the Klein bottle $K$ can be endowed with a universal shortest path metric.

In the theorem above, for $S^2$ and $\mathbb{R}P^2$ we use the round metric of positive constant curvature scaled to 1. In the case of torus we use the flat metric obtained by the identification of the opposite edges of the square. In the case of the Klein bottle we can show that an analogous result fails with the flat square metric on the polygonal scheme $aba^{-1}b$, as we will see in Section 4. But we can get the result for the metric obtained by the identification of the edges of a rectangle of dimensions $1 \times b$ where $b = \sqrt{4/3 + \varepsilon}$ for some small $\varepsilon > 0$. (The edges of length 1 are identified coherently, whereas the edges of length $b$ are identified in opposite directions.)

In all cases we can get shortest path embeddings with uniqueness. Actually, for the torus and the Klein bottle, uniqueness will be a convenient assumption for inductive proofs.

The sphere and the projective plane. The circle packing theorem states that any triangulation of the sphere can be represented as the contact graph of circles on the sphere [28, Theorem 4.3], endowed with the usual metric. Since any graph can be extended into a triangulation (adding new vertices if needed), and paths between centers of circles are shortest paths, this proves Theorem 1 for $S^2$.

For the projective plane, a similar circle packing theorem follows from the spherical case, yet since we could not find a reference in the literature we include a proof here.

Proposition 5. The 1-skeleton of every triangulation $T$ of the projective plane is the contact graph of a packing of circles on the projective plane, endowed with the spherical metric.

Proof. The main ingredient is that the circle packing theorem for the sphere not only states that for every graph $G$ triangulating the sphere, there exists a packing of circles $P$ such that $G$ is isomorphic to the intersection graph of $P$, but also that $P$ is unique up to a M"obius transformation (see Stephenson [28]).

The proposition can be derived from this and the following fact about M"obius maps acting on the sphere: for every M"obius map $\phi$ such that $\phi \circ \phi = Id$ and $\phi$ is fixed-point free there exists a M"obius map $\tau$ such that $\tau^{-1} \phi \tau$ is the antipodal transformation $x \mapsto -x$, see [34].

Indeed, let $T$ be a simplicial complex homeomorphic to the projective plane, i.e., $f: T \to \mathbb{R}P^2$ is a homeomorphism. Let $\pi: S^2 \to \mathbb{R}P^2$ be the universal covering map. Let $\hat{T}$ be the simplicial complex homeomorphic to the sphere such that $\hat{f}: \hat{T} \to S^2$ is a homeomorphism and $f^{-1} \circ \pi \circ \hat{f}: T \to T$ is the simplicial double cover of $T$. The antipodal map on $S^2$ (together with $f$ and with the fact that $\hat{T}$ comes from $T$) induces a fixed point free map $a: T \to T$ such that $a \circ a = id$. There exists a packing $P$ that has the 1-skeleton of $\hat{T}$ as its intersection graph. The same packing has also the 1-skeleton of $a(\hat{T})$ as its intersection graph (where the labels of the vertices are changed). The uniqueness of the packing implies that there exists a M"obius transformation $\phi$ that preserves $P$ as an unlabelled collection circles, but switches the circle corresponding to vertex $v$ in $T$ with the circle corresponding to $a(v)$. Since $a$ is fixed point free and $a \circ a = id$, the same holds for $\phi$. Hence there exists $\tau$ such that $\tau^{-1} \phi \tau$ is the antipodal map, and so $\tau^{-1}(P)$ is an antipodally invariant circle packing whose intersection graph is the 1-skeleton of $\hat{T}$. Passing to the quotient $\pi: S^2 \to \mathbb{R}P^2$, we obtain the desired packing on $\mathbb{R}P^2$.

Minimal triangulations. Let $S$ be a surface and $T$ be a triangulation of it. The triangulation $T$ is called reducible, if it contains an edge $e$ such that the contraction of $e$ yields again a triangulation, which we denote by $T/e$. We refer to $e$ as a contractible edge (we do not mean contractibility in a topological sense). On the other hand, a triangulation is minimal
(or irreducible), if no edge can be contracted this way. For every surface there is a finite list of minimal triangulations. In particular, for the torus $T^2$ this list consists of 21 triangulations found by Lawrencenko [17] and for the Klein bottle $K$ there are 29 minimal triangulations found by Sulanke [30].

The strategy of the proof of Theorem 1 for $T^2$ and $K$ is to show that it is sufficient to check Theorem 1 for minimal triangulations with appropriate fixed metric; see Lemma 6. Then, since every embedded graph can be extended to a triangulation (possibly with adding new vertices), we finish the proof by providing the list of shortest path embeddings of the minimal triangulations.

**Lemma 6.** Let $S$ be a surface equipped with a flat metric. Let $T$ be a reducible triangulation with contractible edge $e$. Let us assume that $T/e$ admits a shortest path embedding with uniqueness into $S$. Then $T$ admits a shortest path embedding with uniqueness into $S$ as well.

The restriction on flat metrics in the lemma above does not seem essential, but this is all we need and this way the proof is quite simple.

**Proof.** Let $v$ be the vertex of $T/e$ obtained by the contraction of $e$. The idea is to consider the shortest path drawing of $T/e$. Then we perform the appropriate vertex splitting of $v$ (the inverse operation to the contraction) in a close neighborhood of $v$ so that we get a shortest path embedding of $T$. In order to see that this is indeed possible, let us consider an edge $f = uv$ of $T/e$. Since the shortest paths are unique, a simple compactness argument shows that there is an $\varepsilon$-neighborhood $N_f$ of $f$ in $S$ which is isometric to an $\varepsilon$-neighborhood of a segment of the same length in $\mathbb{R}^2$ and such that for every $v'$ in the $\varepsilon$-neighborhood of $v$, the straight line segment connecting $v$ and $v'$ inside $N_f$ is the unique shortest path between $v$ and $v'$ in $S$. Therefore, considering such a neighborhood for every edge, it is sufficient to perform the vertex splitting of $v$ in sufficiently small neighborhood of $v$ so that we do not introduce new intersections as on the picture.

The minimal triangulations of $T^2$ and $K$. In Figure 2 we provide a list of shortest path embeddings with uniqueness of minimal triangulations of the torus with a flat metric obtained by identifying the opposite edges of the unit square. They are in the same order as in the book of Mohar and Thomassen [24, Figure 5.3]. The black (thin) edges are the edges of the triangulation whereas the green (thick) edges are the identified boundaries of the unit square which are not parts of the edges of the triangulations. We just skip drawings of the triangulations 7 to 17, because they are all analogous to the triangulation 6, they only have different patterns of diagonals. It is clear that every edge is a geodesic. In order to check that each of them is drawn as a shortest path, it is sufficient to verify that each edge projects vertically and horizontally to a segment of length less than $\frac{1}{2}$.

For the Klein bottle $K$, we also provide a metric such that all the minimal triangulations admit shortest path embeddings with uniqueness. We obtain this metric as the identification of the edges of the rectangle $R = [0, a] \times [0, b]$, where $a = 1$ and $b = \sqrt{4/3} + \varepsilon$ for sufficiently small $\varepsilon$. The edges of length 1 are identified in coherent directions. The edges of length $b$ are identified in the opposite directions. The value $b = \sqrt{4/3} + \varepsilon$ is set up in such a way that if we consider the points $p = (0, \frac{3}{4}b) = (1, \frac{b}{3})$ and $q = (\frac{1}{3}, \frac{b}{2})$ of $K$, then the shortest path between $p$ and $q$ is the horizontal path of height $\frac{1}{4}b$. However, when we shift $p$ along the boundary of $R$ a little bit closer to the center, say by $\frac{1}{1000}$, then the shortest path becomes the diagonal edge connecting the left copy of $p$ and $q$, see Figure 4.
Figure 2: Minimal triangulations of the torus.

Figure 3: Minimal triangulations of the Klein bottle.
There are 29 minimal triangulations of the Klein bottle. A list of 25 of them was first found by Lawrencenko and Negami [18]. Later on, Sulanke [30] found a gap in the claimed completeness of this list and provided a complete list containing 4 additional triangulations. These triangulations split into two classes. The triangulations of the first class are named Kh1–Kh25 and there are four triangulations Kc1–Kc4 in the second class. The triangulations from the second class differ by the property that they contain a cycle of length 3 which splits the Klein bottle into two Möbius bands.

We start with the examination of the triangulations in the first class. We present shortest path embeddings with uniqueness for 15 of them; see the top three lines of Figure 3. We omit the triangulations Kh15-Kh24 that are very analogous to Kh14, only the diagonal edges form a different pattern. The vertices of the triangulations are positioned in lattice points of the lattice generated by vectors \((\frac{1}{12}, 0)\) and \((0, \frac{b}{12})\). In some cases an additional shift is necessary by a small value \(\frac{1}{1000}\) (but this value is large compared to \(\varepsilon\)): this is indicated by arrows next to the vertices. (The pair of arrows in Kh25 indicates a shift by \(\frac{2}{1000}\).) Most of the drawings are very similar to the drawings by Negami, Lawrencenko, and Sulanke. Only for the drawings of Kh3, Kh12, Kh13, and Kh25 we did more significant movements. It is routine (but tedious) to check that all the edges are indeed drawn as shortest paths. For many edges this can be checked easily. For few not so obvious cases the general recipe is to use the universal cover approach and Lemma 8.

Now let us focus on the triangulations in the second class. All of them are obtained by gluing two triangulations of the Möbius bands along their boundaries. In our case, we split \(K\) into two bands by a cycle depicted on the bottom left picture of Figure 3. It is perhaps the most natural way how to split the Klein bottle into two Möbius bands; a particular advantage for us is that there is an isometric homeomorphism which maps one band to another and which preserves the common boundary pointwise. Therefore, it is sufficient to present the shortest paths embeddings with uniqueness into the bands, as on the middle three pictures. Then we get drawings of Kc1-Kc4 using this homeomorphism. For example, Kc1 is obtained by gluing two copies of Mb1 together. The vertices on the pictures are the lattice points of the same lattice as above with exception of two points of Mb3. The points on the ‘central’ cycle of Mb3 have coordinates \((\frac{1}{6}, \frac{b}{2}), (\frac{5}{6}, \frac{b}{2}), (\frac{1}{2}, \frac{b}{2})\). Note that we have redrawn the original drawings of Lawrencenko and Negami [18] quite significantly, but it is easy to check that we get the same triangulations, because these ‘halves’ of triangulations of \(K\) are quite small.

4 Square flat metric on the Klein bottle

The task of this section is to prove the following theorem.

**Theorem 2.** Let \(K\) denote the Klein bottle endowed with the unit square flat metric on the polygonal scheme \(aba^{-1}b\). Then there exists a graph embeddable into \(K\) which cannot be embedded into \(K\) so that the edges are shortest paths.

We consider the minimal triangulation Kc1 (see Figure 3) and we denote by \(G\) the underlying graph for this triangulation. We will prove that \(G\) does not admit a shortest path embedding into
with the square metric. First, we observe that the triangulation $Kc1$ is the only embedding of $G$ into $K$.

**Proposition 7.** $G$ has a unique embedding into the Klein bottle.

*Proof of Proposition 7.* $G$ has 9 vertices and their degrees are $(8, 8, 8, 5, 5, 5, 5, 5, 5)$. Since no other irreducible triangulation of the Klein bottle has this degree sequence, any other hypothetical embedding of $G$ into the Klein bottle is either non-cellular or has a reducible edge. In the first case, it means that $G$ is cellurally embeddable into the sphere or the projective plane, which is not the case. Indeed, it is obtained as the gluing of two copies of $K_6$ along a triangle, and therefore contains $K_5 \oplus K_5$ (two copies of $K_5$ identified along an edge minus that edge) as a minor, which does not embed into the projective plane [24, Figure 6.4]. In the latter case, we observe that an edge contraction cannot decrease the degree of all three degree 8 vertices, and thus we reach a contradiction since a triangulation on 8 vertices cannot have a degree 8 vertex.

For contradiction, let us assume that $G$ admits a shortest paths embedding into $K$. We know that $Kc1$ is obtained by gluing two triangulations of a Möbius band along a cycle of length 3 (the triangle corresponding to this cycle is not part of the triangulation). Let $abc$ be this cycle. With a slight abuse of notation we identify this cycle with its image in the (hypothetical) shortest path embedding into $K$. Our strategy is to show that already $abc$ cannot be embedded into $K$ with shortest path edges, which will give the required contradiction. By Proposition 7, we know that $abc$ splits $K$ into two Möbius bands.

Let $X = \mathbb{R}^2$ be the universal cover of $K$ (with standard Euclidean metric). Let $\pi: X \to K$ be the isometric projection corresponding to the cover. We will represent the Klein bottle with the flat-square metric as the unit square $[0, 1]^2$ with suitable identification of the edges ($aba^{-1}b$, as in the previous section). We will use the convention that $\pi(((0, 1)^2) = (0, 1)^2$; that is, the projection is the identity on the interior of this square. See Figure 5.

Given a point $p \in K$ we set $X_p := \pi^{-1}(p)$. Finally, let $\mathcal{V}_p$ be the Voronoi diagram in $X$ corresponding to the set $X_p$.

**Lemma 8.** Let $p$ and $q$ be two points in $K$ and $\gamma$ be an arc (edge) connecting them, considered as a subset of $K$. Then $\gamma$ is the unique shortest path between $p$ and $q$ if and only if there are $p' \in X_p$, $q' \in X_q$ such that $\gamma = \pi(p'q')$ where $p'q'$ denotes the straight edge connecting $p'$ and $q'$ in $X$ and $q'$ belongs to the open Voronoi cell for $p'$ in $\mathcal{V}_p$.

*Proof of Lemma 8.* Any path $\kappa$ with endpoints $p$ and $q$ lifts to some path $\kappa'$ with endpoints $p' \in X_p$ and $q' \in X_q$ ($\kappa'$, $p'$, and $q'$ are not determined uniquely). This lift preserves the length of the path. Vice versa, any path connecting a point in $X_p$ with a point in $X_q$ projects to a path connecting $p$ and $q$ (not necessarily simple), again preserving the length.

Therefore, $\gamma$ is the shortest path in $K$ connecting $p$ and $q$ if and only if it lifts to a straight edge realizing the distance between $X_p$ and $X_q$ in $X$. Such an edge connects $p' \in X_p$ and $q' \in X_q$. By symmetry, we can fix $q'$ arbitrarily and we look for the closest $p'$. Then, a point $p'$ is the unique point of $X_p$ closest to $q'$ if and only if $q'$ belongs to the open Voronoi cell for $p'$ in $\mathcal{V}_p$. This is what we need.

Now let us lift the cycle $abc$ to a path $a'b'c'a''$ in $X$; see Figure 5. Given a curve in $X$, we call the length of its projection to the $x$-axis, the “horizontal length” of the curve; similarly we speak about the horizontal distance and the vertical distance of two points in $X$.

**Lemma 9.** The horizontal distance between $a'$ and $a''$ is at least 2.

*Proof of Lemma 9.* If we consider the point $a'$ fixed, then the position of $a''$ in $X_a$ determines the homotopy class of the cycle $abc$ in $\pi_1(K)$. Therefore, it also determines the homology classes.
of this cycle in $H_1(K;\mathbb{Z}_2)$ and in $H_1(K;\mathbb{Z})$. We note that the cycle $abc$ must be homologically trivial in $H_1(K;\mathbb{Z}_2)$ because it bounds a Möbius band; however, it is homologically nontrivial in $H_1(K;\mathbb{Z})$ because it bounds a Möbius band (which is non-orientable) on both sides. In addition the cycle $abe$ is two sided, that is, its (regular) neighborhood is an annulus and not a Möbius band.

The horizontal distance between $a'$ and $a''$ must be a non-negative integer. We will rule out the cases when this distance is 0 or 1.

If this distance is 1, then the cycle $abc$ is not two-sided (this can be read on the lift), a contradiction.

If the horizontal distance is 0 and the vertical distance is odd, then $abc$ is homologically nontrivial in $H_1(K;\mathbb{Z}_2)$. (It is sufficient to consider the segment connecting $a'$ and $a''$ and project it to a cycle $z$ in $K$. Then $z$ is homotopy equivalent to $abc$.) A contradiction.

Similarly, if the horizontal distance is 0 and the vertical distance is even, then $abc$ is homologically trivial in $H_1(K;\mathbb{Z})$. (Again we project the segment connecting $a'$ and $a''$.) A contradiction.

**Lemma 10.** Let $\gamma$ be a unique shortest path in $K$ connecting points $p$ and $q$. Let $\gamma'$ be a lift of $\gamma$ with endpoints $p'$ and $q'$. Then the horizontal distance in $X$ between $p'$ and $q'$ is less than $\frac{5}{8}$.

**Proof.** Let $C$ be the open Voronoi cell for $p'$ in $V_p$. By Lemma 8, $q'$ belongs to $C$. Therefore, it is sufficient to check that every point $c'$ of $C$ has horizontal distance less than $\frac{5}{8}$ from $p'$.

Without loss of generality, we may assume that the $x$-coordinate of $p'$ equals 0 since shifting $p'$ in horizontal direction only shifts $X_p$ and $V_p$ (note that this is not true for the vertical direction). For contradiction, there is a $c'$ in $C$ at distance at least $\frac{5}{8}$ and without loss of generality the $x$-coordinate of $c'$ is positive. Let $p''$ be the point of $X_p$ with $x$-coordinate equal 1 which is vertically closest to $c'$ (pick any suitable point in case of draw); see the picture on the left. The vertical distance between $c'$ and $p''$ is at most $\frac{1}{2}$. A simple calculation, using the Pythagoras theorem, gives that $p''$ is at most as far from $c'$ as $p'$. A contradiction.

Finally, we summarize how the previous lemmas yield a contradiction. By Lemma 9, the horizontal distance between $a'$ and $a''$ is at least 2. On the other hand, Lemma 10 gives that
the horizontal length of each of the edges $a'b'$, $b'c'$, and $c'a''$ is at most $\frac{5}{8}$, altogether at most $\frac{15}{8}$. 
This gives the required contradiction, which finishes the proof of Theorem 2.

5 Asymptotically almost all hyperbolic metrics are not universal

Before stating the main theorem of this section, we will give some very quick background on the geometry of surfaces, we refer to Farb and Margalit [10] for a proper introduction. The Teichmüller space $\mathcal{T}_g$ of a surface $S$ of genus $g$ denotes the set of hyperbolic metrics on $S$, such that two metrics are equivalent if they are related by an isometry isotopic to the identity. In some contexts, like ours, one might also want to identify metrics related by an isometry (not necessarily isotopic to the identity). The corresponding space is called the moduli space $\mathcal{M}_g$ of the surface, and is obtained by quotienting $\mathcal{T}_g$ by the mapping class group of $S$, i.e., its group of homeomorphisms. This moduli space can be endowed with multiple structures, here we will be interested in a particular one, called the Weil-Petersson metric. This metric provides $\mathcal{M}_g$ with a Riemannian structure of finite volume, and therefore by renormalizing, we obtain a probability space, allowing to choose a random metric. We can now state the main theorem of this section.

**Theorem 3.** For any $\varepsilon > 0$, with probability tending to 1 as $g$ goes to infinity, a random hyperbolic metric is not a $O(g^{1/3-\varepsilon})$-universal shortest path metric. In particular, with probability tending to 1 as $g$ goes to infinity, a random hyperbolic metric is not a universal shortest path metric.

The proof is a consequence of two important results on random hyperbolic metrics. The first is a small variant of a theorem of Guth, Parlier, and Young [15, Theorem 1] that relies on the work of Wolpert [35]. Before stating it, we need some definitions.

Given a hyperbolic metric $m$ on a surface $S$, we say that $m$ has total pants length at least $\ell$ if in any pants decomposition $\Gamma$ of $S$, the lengths of the closed curves of $\Gamma$ sum up to at least $\ell$. We say that $m$ has total pants length of type $\xi$ at least $\ell$ if in any pants decomposition $\Gamma$ of $S$ of type $\xi$, the lengths of the closed curves of $\Gamma$ sum up to at least $\ell$.

**Theorem 11.** For any $\varepsilon > 0$ and any family of types of pants decomposition $\xi_g$, a random metric on $\mathcal{M}_g$ has total pants length of fixed type $\xi_g$ at least $g^{4/3-\varepsilon}$ with probability tending to 1 as $g \to \infty$.

**Proof of Theorem 11.** This bound is obtained with a similar technique as the proof of Theorem 1 of Guth, Parlier, and Young [15]. We refer to their article for more details, and as in their proof, we will discard non super-exponential terms, e.g., $n! \approx n^n$. For every $a, b, c \in \mathbb{R}^+$ there exists a unique hyperbolic metric on a pair of pants with boundary lengths $a, b, c$. For a pants decomposition of fixed type $\xi_g$, the Weil-Petersson volume form on moduli space is the push forward of the form $d\ell_1 \wedge \ldots \wedge d\ell_{3g-3} \wedge d\tau_1 \wedge \ldots \wedge d\tau_{3g-3}$ on Teichmüller space which is identified with $\mathbb{R}^{6g-6}$ and the $\ell_i$ denote the lengths of the (geodesic) boundaries of the pants decomposition, while the $\tau_i$ quantify how much the metric twists around each geodesic. Since every full twist gives a homeomorphic metric, the subset of Teichmüller space $\{(\ell_i, \tau_i) \mid \sum \ell_i \leq L, 0 \leq \tau_i \leq \ell_i\}$ projects surjectively onto the region of moduli space corresponding to surfaces with total pants length of type $\xi_g$ at most $L$. The volume of this set is bounded by $\leq (\frac{L}{g})^{6g}$, which is to be compared with the total volume of moduli space $\approx g^{2g}$. For $L$ smaller than $g^{4/3-\varepsilon}$, the ratio tends to zero, which proves the theorem.

The following is an immediate corollary of this theorem.

**Corollary 12.** Let $T_g$ be a family of triangulations of $S_g$, such that every member of $T_g$ contains a pants decomposition of fixed type $\xi_g$. For any $\varepsilon > 0$, with probability tending to 1 as $g \to \infty$, a shortest embedding of $T_g$ into a random hyperbolic surface of genus $g$ has length at least $\Omega(g^{4/3-\varepsilon})$.
The next theorem was proved by Mirzakhani [22, Theorem 4.10].

**Theorem 13.** With probability tending to 1, the diameter of a random hyperbolic surface of genus $g$ is $O(\log g)$.

Theorem 3 is proved by providing an explicit family of graphs $G_g$ which will embed badly. It is defined in the following way for $g \geq 2$. Let $\xi_g$ be a type of pants decompositions for every value of $g$.

- We start with a pants decomposition of type $\xi_g$ of a surface $S_g$.
- We place four vertices on every boundary curve.
- We triangulate each pair of pants with a bounded size triangulation so that each cycle of length 3 bounds a triangle in the triangulation, and any path connecting two boundary components of the pair of pants has length at least 4 (in particular $G_g$ is a simple graph and each cycle of length 3 in the graph $G_g$ bounds a triangle in the triangulation).

The following proposition controls the issues related to the flexibility of embeddings of graphs into surfaces.

**Proposition 14.** There is a unique embedding of $G_g$ into $S_g$, up to a homeomorphism; in particular every embedding contains a pants decomposition of type $\xi_g$.

**Proof of Proposition 14.** Let $v$ be the number of vertices, $e$ be the number of edges and $t$ be the number of triangles of the triangulation in the definition of $G_g$ (triangles in the graph-theoretical sense). By Euler’s formula and by the construction we get $v - e + t = \chi$ where $\chi$ is the Euler characteristic of $S_g$. Let us consider an embedding $\Psi$ of $G_g$ into $S_g$. Let $f$ be the number of faces of this embedding and $F$ be the set of faces. Euler’s formula for this embedding gives $v - e + f \geq \chi$ (we get an inequality because some of the faces need not be embedded cellulary). In particular, we get $f \geq t$. On the other hand, we get $2e = 3t$ and $2e = \sum_{\sigma \in F} \deg \sigma \geq 3f$ since each edge is in exactly two faces. This gives $3t \geq 3f$. Therefore, both of the aforementioned inequalities have to be equalities. In particular, each $\sigma \in F$ is a triangle bounded by a cycle of length 3 in $G_g$. Since the number of cycles of length 3 in $G_g$ equals $t = f$, we deduce that $\Psi$ coincides with the embedding from the definition of $G_g$ up to a homeomorphism. □

**Remark:** We preferred to use a hands-on construction of the graphs $G_g$, but another approach could be to rely on the theory of LEW-embeddings and use one of its results on uniqueness of embeddings, see for example Mohar and Thomassen [24, Corollary 5.2.3].

With these three results at hand we are ready to provide a proof of the theorem.

**Proof of Theorem 3.** We use the family of graphs $G_g$ previously defined. Since there are $O(g)$ curves in a pants decomposition, it contains $O(g)$ edges, and every embedding of $G_g$ into $S_g$ contains a pants decomposition of type $\xi_g$ by Proposition 14.

Now, by Corollary 12, for every $\varepsilon > 0$, and for $g$ large enough, the probability that the shortest possible embedding of $G_g$ into a random metric has length at least $O(g^{4/3-\varepsilon})$ is at least $1 - \varepsilon/2$. In particular, since there are $O(g)$ edges in $G_g$, some edge $e_g$ in this embedding must have length at least $\Omega(g^{1/3-\varepsilon})$. By Theorem 13, we can choose $g$ large enough so that with probability at least $1 - \varepsilon/2$, the random hyperbolic metric has diameter $O(\log g)$. Hence, by the union bound, with probability $1 - \varepsilon$ both properties hold. Therefore, for every $\varepsilon > 0$, there exists some value $g_0$ such that for any $g \geq g_0$, in any embedding of $G_g$, there exists an edge $e_g = (x, y)$ such that $\ell_m(e_g) = \Omega(g^{1/3-\varepsilon})$, but $d_m(x, y) \leq \text{diam}(m) \leq O(\log g)$. This implies that $e$ is not drawn by a shortest path. Similarly, subdividing each edge $O(g^{1/3-\varepsilon})$ times will run into the same issue. This concludes the proof. □
6 Higher genus: positive results

Theorem 4. For every $g > 1$, there exists an $O(g)$-universal shortest path hyperbolic metric $m$ on the orientable surface $S$ of genus $g$.

Our approach to prove Theorem 4 is to cut the surface $S_g$ with a hexagonal decomposition $\Delta$, so that every edge of $G$ is cut $O(g)$ times by this decomposition $\Delta$. The construction to do this is a slight modification of the octagonal decompositions provided by É. Colin de Verdière and Erickson [4, Theorem 3.1]. Each of the hexagons is then endowed with a specific hyperbolic metric $m_H$, and pasting these together yields the hyperbolic metric $m$ on $S_g$. The hyperbolic metric $m_H$ is chosen so that the hexagons are convex, i.e., the shortest paths between points of a hexagon stay within this hexagon. Therefore, there only remains to embed the graph $G$ cut along $\Delta$, separately in every hexagon with shortest paths. To do this, we use a variant of a theorem of Y. Colin de Verdière [7] which generalizes Tutte’s barycentric method to metrics of nonpositive curvature.

Hexagonal decompositions. A hexagonal decomposition, respectively an octagonal decomposition of $S_g$ is an arrangement of closed curves on $S_g$ that is homeomorphic to the one pictured in Figure 6.b., respectively Figure 6.a. In particular, every vertex has degree four and every face has six sides, respectively eight sides.

Octagonal decompositions were introduced by É. Colin de Verdière and Erickson [4] where they showed how to compute one that does not cross the edges of an embedded graph too many times. We restate their theorem in our language.

Theorem 15 ([4, Theorem 3.1]). Let $G$ be a graph embedded in a surface $S_g$ for $g \geq 2$. There exists an octagonal decomposition $\Gamma$ of $S_g$ such that each edge of $G$ crosses each closed curve of $\Gamma$ a constant number of times.

We observe that this octagonal decomposition can be upgraded to a hexagonal decomposition that still does not cross $G$ too much:

Corollary 16. Let $G$ be a graph embedded in a surface $S_g$. There exists a hexagonal decomposition $\Delta$ of $S_g$ such that each edge of $G$ crosses each closed curve of $\Delta$ a constant number of times, except for maybe one closed curve which is allowed to cross each edge of $G$ at most $O(g)$ times. In particular, the number of crossing between every edge of $G$ and $\Delta$ is $O(g)$.
Figure 7: The intersection graph $I$ and the two involutions: $\sigma_1$ is the symmetry about the dashed horizontal line, and $\sigma_2$ swaps every disk and its adjacent star.

**Proof.** The decomposition $\Delta$ is simply obtained by taking the decomposition $\Gamma$ and adding a single curve that follows closely a concatenation of $O(g)$ subpaths of curves of $\Gamma$, see Figure 6c. The resulting arrangement of curves has the topology of a hexagonal decomposition, and the bounds on the number of crossings results directly from the construction. \(\square\)

**The hyperbolic metric.** We first endow each hexagon of the hexagonal decomposition with the hyperbolic metric $m_H$ of an equilateral right-angled hyperbolic hexagon. Since the hexagons have right angles and the vertices of a hexagonal decomposition have degree 4, this metric can be safely pasted between hexagons to endow $S_g$ with a hyperbolic metric $m$. The main property of this metric that we will use is the following one:

**Proposition 17.** Every hexagon $H$, viewed as a subset of $S_g$ endowed with $m$, is convex, i.e., every path between $x, y \in H$ that is a shortest path in $H$ is also a shortest path in $S_g$.

**Proof.** The proof relies on an exchange argument based on the symmetries of the hexagonal decomposition.

The intersection graph $I$ of the hexagonal decomposition is defined by taking one vertex for each hexagon and edges between adjacent hexagons (we allow multiple edges). We are interested in two graph automorphisms which are also involutions. These are pictured in Figure 7:

- The symmetry $\sigma_1$ about the horizontal axis, corresponding to the so-called hyperelliptic involution of the surface $S$.
- The automorphism $\sigma_2$ swapping every hexagon with its neighbor in the octagonal decomposition.

Since all the hexagons are isometric, these involutions correspond naturally to isometric involutions of $S$.

Now, let $\gamma$ be a shortest path between two vertices $x$ and $y$ in a hexagon $H_1$, let us assume without loss of generality that $H_1$ is in the upper part of $I$. This path $\gamma$ naturally induces a walk in $I$ obtained by taking each hexagon of which interior is met by $\gamma$. This walk does not backtrack at some hexagon $H$: otherwise one could shortcut $\gamma$ by staying on the boundary of $H$.

From $\gamma$, one can build a path $\gamma'$ between $x$ and $y$ which stays in the upper half of the graph: for every maximal subpath of the graph in the lower half, one applies the isometry $\sigma_1$, effectively mirroring these paths in the upper half. Similarly, by applying $\sigma_2$, one obtains a path $\gamma''$, which only uses half of the hexagons. The walk in $I$ corresponding to $\gamma''$ lies now in a path. Since it does not backtrack, it is necessarily trivial and never leaves hexagon $H_1$. We have thus found a shortest path in $H_1$ connecting $x$ and $y$. \(\square\)

**Finishing the proof.** We prove in this paragraph how to reembed a graph embedded in a hexagon so that its edges are shortest paths. This allows us to finish the proof.

**Theorem 18.** Let $G$ be a graph embedded as a triangulation in a hyperbolic hexagon $H$ endowed with the metric $m_H$. If there are no dividing edges in $G$, i.e., edges between two non-adjacent
vertices on the boundary of $H$, then $G$ can be embedded with geodesics, with the vertices on the boundary of $H$ in the same positions as in the initial embedding.

Let us postpone the proof of this theorem for now, and show how to conclude the proof of Theorem 4. We first show how to upgrade a graph embedded in a disk to a triangulation.

**Lemma 19.** For any graph $G$ embedded in a disk without dividing edges, there exists a triangulation $G'$ of the disk containing $G$ as a subgraph and that does not contain any dividing edges.

**Proof.** For every face $F$ of $G$, we start by adding a vertex in $F$ and edges connecting it to the vertices adjacent to the face. This does not add loops or dividing edges, but may add multiple edges if one vertex occurs multiple times on the boundary of a face. These are taken care of by subdividing them once again and triangulating.

All the pieces are now in place for the proof of Theorem 4.

**Proof of Theorem 4.** By Corollary 16, one can embed $G$ into $S_g$ such that every edge of $G$ is cut $O(g)$ times by the hexagonal decomposition $\Delta$. This defines a graph $G' = \bigcup_i G'_i$ such that each of the graphs $G'_i$ is embedded in a single hexagon and $G'$ is obtained from $G$ by subdividing every edge $O(g)$ times. If there are dividing edges in $G'_i$, they can be removed by subdividing the edge once. By Lemma 19, one can upgrade all the $G'_i$ to triangulations. We can then apply Theorem 18 in each of the hexagons separately, yielding embeddings with shortest paths. Since the vertices on the boundary did not move during the reembedding, this defines an embedding of $G$ into $S_g$. Since $H$ is simply connected and $m_H$ is hyperbolic, there is a unique geodesic connecting any two points, and this geodesic is a shortest path. Therefore the edges of $G'$ are shortest paths in $H$. By Proposition 17, each edge of $G'$ is also a shortest path in $S_g$. Therefore each edge of $G$ is embedded as a concatenation of $O(g)$ shortest paths.

We note that by subdividing each edge once more, the shortest paths we obtain are unique.

The proof of Theorem 18 is obtained in a spirit similar to the proof of the one of the celebrated spring theorem of Tutte [33]. However, there are two main differences which prevent us from directly appealing to the literature: on the one hand the metric is not Euclidean but hyperbolic, and on the other hand the boundary of the input polygon is not strictly convex, since there may be multiple vertices of $G$ on a geodesic boundary of $H$. The hypothesis on dividing edges is tailored to circumvent the second issue, and in a Euclidean setting it was proved by Floater [12] that the correspond embedding theorem holds. Regarding the first issue, Y. Colin de Verdière stated a Tutte embedding theorem [7, Theorem 3] for the hyperbolic setting with strictly convex boundary, yet he actually did not provide a proof for it. In Appendix A we show how to prove Theorem 18 in the generality that we need following the ideas laid out by Y. Colin de Verdière in the rest of his article [7]. This concludes the proof of Theorem 4.

Finally, we remark that this proof technique provides an alternative proof of Negami’s Theorem [25] for orientable surfaces. If $G_1$ and $G_2$ are two graphs embedded on the orientable surface of genus $g$, a crude application of Theorem 4 shows that one can reembed both graphs with a homeomorphism such that each edge is realized as a concatenation of $O(g)$ shortest paths for our hyperbolic metric. Since hyperbolic shortest paths in general position cross at most once, this gives embeddings of $G_1$ and $G_2$ such that there are $O(g^2)$ crossings between each edge of $G_1$ and each edge of $G_2$. Negami proved that $O(g)$ crossings are actually enough, and a deeper look at our construction also achieves this better bound: it is easy to see that in our reembeddings, each edge is actually cut into $O(1)$ subedges realized as shortest paths in each hexagon. Since there are $O(g)$ hexagons, there are in total $O(g)$ crossings between each edge of $G_1$ and $G_2$, which yields the following:
Corollary 20. There exists an absolute constant $c > 0$ such that if $S_g$ is an orientable surface of genus $g$, for any two embedded graphs $G_1, G_2 \to S_g$ there exists a homeomorphism $h : S_g \to S_g$ such that $cr(h(G_1), G_2) \leq cg|E(G_1)| \cdot |E(G_2)|$.

Acknowledgements We are grateful to Éric Colin de Verdière for his involvement in the early stages of this research. We also thank Sergio Cabello, Francis Lazarus, Bojan Mohar, Eric Sedgwick and Uli Wagner for helpful discussions and Xavier Goaoc for organizing the workshop that led to this work.

References


18
A Tutte’s embedding theorem in a hyperbolic setting

In this section, we explain the proof of the subsequent theorem, following the arguments of Y. Colin de Verdière [7].

Theorem 18. Let $G$ be a graph embedded as a triangulation in a hyperbolic hexagon $H$ endowed with the metric $m_H$. If there are no dividing edges in $G$, i.e., edges between two non-adjacent vertices on the boundary of $H$, then $G$ can be embedded with geodesics, with the vertices on the boundary of $H$ in the same positions as in the initial embedding.

As announced, the proof follows from a spring-like construction, i.e. we think of the edges of the graph $G$ as springs with some arbitrary stiffness, the vertices which are not on the boundary are allowed to move and we prove that the equilibrium state for this physical system is an embedding of the graph.

For an embedding $\varphi : G \to H$, denote by $e_{ij}$ the map $[0, 1] \to H$ representing the edge $(i, j)$. Starting with an embedding $\varphi_0 : G \to H$ and given assignments $c_{i,j} : E(G) \to \mathbb{R}^+$, we are interested in the map $\varphi : G \to H$ minimizing the energy functional

$$E_\varphi = \sum_{(i,j) \in E} \int_0^1 c_{ij}||e'_{ij}(t)||^2 dm_H$$

with fixed vertices on the boundary of $H$. This is the equilibrium state of the spring system with the $c_{i,j}$ coefficients specifying the stiffness of the springs. We claim that $\varphi$ is an embedding such that the edges are geodesics.

**Step 1: Existence.** The existence of $\varphi$ follows from classical compactness considerations, since an Arzelà-Ascoli argument proves the compactness of sets with bounded energy. Then an extremum of $E_\varphi$ corresponds to a $\varphi$ where all the arcs $e_{i,j}$ are geodesics. Furthermore, every vertex $\varphi(x)$ which is not on the boundary lies in the strict hyperbolic convex hull of its neighbors which are not mapped to the same point.

**Step 2: Curvature considerations.** Since $\varphi_0$ provides an embedding of $G$ into $H$, $G$ can be seen as a topological subspace of $H$. The corresponding simplicial complex will be denoted by $X$ (it is of course homeomorphic to $H$) and its set of vertices, edges and triangles by $V$, $E$, and $T$. By extending $\varphi$ separately with a local homeomorphism in the interior of each non-degenerate triangle, we can extend it into a map $\Phi : X \to H$ agreeing with $\varphi$ on $G$.

Now, the map $\Phi : X \to H$ provides values for the angles of the non-degenerate triangles in $X$. For degenerate triangles, values of the angles are taken arbitrarily so that they sum to $\pi$ (therefore morally their hyperbolic area is zero). For an interior vertex $v$, let us define the curvature $K(v) = 2\pi - \sum_i \alpha^i_v$, where $\alpha^i_v$ are the angles adjacent to $v$. For a vertex $v$ on the interior of a geodesic boundary, we define it by $K(v) = \pi - \sum_i \alpha^i_v$, and on the six vertices of $H$, we take it to be $K(v) = \pi/2 - \sum_i \alpha^i_v$.

The area of a geodesic hyperbolic triangle is $\pi$ minus the sum of its angles. Summing over all the triangles of $\Phi(X)$, we obtain $|T|\pi - \sum_v \sum_i \alpha^i_v = \sum_{t \in T} \text{Area}(t)$. With Euler’s formula and double counting, this gives $\sum_{t \in T} \text{Area}(t) = \pi + \sum_v K(v)$. Since the boundary is fixed, $\Phi$ has degree one and is thus surjective, therefore the sum of the areas of the triangles is at least the area of the hexagon, which is $\pi$ since it is right-angled. Therefore $\sum_v K(v) \geq 0$.
Step 3: Punctual degeneracies. In this step we investigate which subcomplexes of $X$ can be mapped to a single point. We show that no triangle can be mapped to a single point, and that a set of edges mapped to a single point forms a path subgraph in $G$.

Let $X_1$ be a maximal connected subcomplex of $X$ which is mapped to a point $x$ by $\Phi$. This subcomplex has to be simply connected, otherwise the region inside could be mapped to $x$ as well which would reduce the value of $E_\varphi$. Since the boundary edges are fixed by $\varphi$, $X_1$ does not contain any edge on the boundary or triangle adjacent to the boundary.

For every vertex $v$ in $\Phi^{-1}(x)$, $\Phi(v) = \varphi(v)$ lies in the strict convex hull of its neighbors which are not mapped to $x$, as was observed in Step 1. Therefore the angles of the non-degenerate triangles adjacent to $v$ sum up to at least $2\pi$. Indeed the angular opening at $\varphi(v)$ has to be at least $\pi$ by the convexity hypothesis, but any map $S^1 \to S^1$ is either surjective or at least two-to-one everywhere, in which case this angular opening of at least $\pi$ amounts to at least $2\pi$ in the sum of angles around $v$. This shows that $K(x) := \sum_{v \in \Phi^{-1}(x)} K(v)$ is nonpositive. Since the boundary edges are fixed, we also have $K(v) \leq 0$ for the vertices on the boundary.

Summing over all the values of $x$, we obtain that $\sum_v K(v) \leq 0$, and thus this sum is zero by the previous paragraph, and each of the $K(x)$ is also zero.

From that we infer that $X_1$ contains no triangle: if it did, there would be at least 3 preimages of $x$ for which the angles of the adjacent non-degenerate triangles would sum up to at least $2\pi$. Summing them into $K(x)$ we would obtain a nonzero value. Similarly, $X_1$ can only be a linear subgraph of $G$, and every triangle adjacent to a $X_1$ not reduced to a point is degenerate.

Step 4: Linear degeneracies. Now that we showed that triangles can not be mapped to points, we show that triangles are not mapped to lines either, or equivalently that edges are not mapped to points.

Let $X_2$ be a maximal connected subcomplex of $X$ such that the image of the triangles of $X_2$ by $\varphi$ are degenerate. Let us assume that $X_2$ is non-empty. Then the image $\Phi(X_2)$ is an arc of a geodesic of $H$: indeed if there was a broken line in $\Phi(X_2)$, around the breaking points there would be non-degenerate triangles adjacent to a $X_1$ not reduced to a point, which is absurd by the previous paragraph.

If this geodesic is not a boundary geodesic of $H$, two of the points on the boundary of $X_2$ are mapped to the endpoints of the arc of geodesic, and all the other vertices have their adjacent edges within $X_2$ because of the convexity condition. Therefore, there must be two arcs connecting the two boundary points, as in the top of Figure 8, which is impossible in the simplicial complex $X$.

If this geodesic is on the boundary of $H$, then by the same convexity argument, two vertices of $\partial X$ must map to the endpoints of this arc of geodesic, and the other vertices have all their
edges within $X_2$. Therefore there is a dividing edge connecting these two vertices, as in the bottom of Figure 8, which is a contradiction.

**Step 5: Conclusion.** Since $X_2$ is empty, no triangle in the image of $\Phi$ is degenerate. Furthermore, all the $X_1$ are reduced to a single point and thus $K(v)$ is zero for all the vertices $v$. The only remaining possible pathology is if all the triangles adjacent to a non-boundary vertex $v$ are mapped to a half-plane around $\Phi(v)$. By the convexity constraint, this can only happen if the edges adjacent to $v$ are aligned, but this would yield degenerate triangles. Therefore $\Phi$ is a local homeomorphism of degree 1, hence it is a global homeomorphism and $\varphi$ is an embedding.
Embeddability in the 3-sphere is decidable

Jiří Matoušek\textsuperscript{1,2,a,b}, Eric Sedgwick\textsuperscript{3,a}, Martin Tancer\textsuperscript{1,4,5,a}, and Uli Wagner\textsuperscript{4,a,c}

\textsuperscript{1}Department of Applied Mathematics, Charles University, Malostranské nám. 25, 118 00 Praha 1, Czech Republic.
\textsuperscript{2}Department of Computer Science, ETH Zürich, 8092 Zürich, Switzerland
\textsuperscript{3}School of Computing, DePaul University, 243 S. Wabash Ave, Chicago, IL 60604, USA
\textsuperscript{4}IST Austria, Am Campus 1, 3400 Klosterneuburg, Austria
\textsuperscript{5}Part of this work was done when the third author was affiliated with Institutionen för matematik, Kungliga Tekniska Högskolan, Linstedsvägen 25, 100 44 Stockholm
\textsuperscript{a}Supported by the ERC Advanced Grant No. 267165.
\textsuperscript{b}Partially supported by Grant GRADR Eurogiga GIG/11/E023.
\textsuperscript{c}Supported by the Swiss National Science Foundation (Grants SNSF-PP00P2-138948 and SNSF-200020-138230).

Abstract

We show that the following algorithmic problem is decidable: given a 2-dimensional simplicial complex, can it be embedded (topologically, or equivalently, piecewise linearly) in $\mathbb{R}^3$? By a known reduction, it suffices to decide the embeddability of a given triangulated 3-manifold $X$ into the 3-sphere $S^3$. The main step, which allows us to simplify $X$ and recurse, is in proving that if $X$ can be embedded in $S^3$, then there is also an embedding in which $X$ has a short meridian, i.e., an essential curve in the boundary of $X$ bounding a disk in $S^3 \setminus X$ with length bounded by a computable function of the number of tetrahedra of $X$.

1 Introduction

The embeddability problem. Let $\text{EMBED}_{k \to d}$ be the following algorithmic problem: given a finite simplicial complex $K$ of dimension at most $k$, does there exist a (piecewise linear) embedding of $K$ into $\mathbb{R}^d$?

A systematic investigation of the computational complexity of this problem was initiated in [MTW11]; earlier it was known that $\text{EMBED}_{1 \to 2}$ (graph planarity) is solvable in linear time, so is $\text{EMBED}_{2 \to 2}$ [GR79], and for every $k \geq 3$ fixed, $\text{EMBED}_{k \to 2k}$ can be decided in polynomial time (this is based on the work of Van Kampen, Wu, and Shapiro; see [MTW11]).

For dimension $d \geq 4$, there is now a reasonably good understanding of the computational complexity of $\text{EMBED}_{k \to d}$: for all $k$ with $(2d - 2)/3 \leq k \leq d$ it is NP-hard (and even undecidable if $k \geq d - 1 \geq 4$) [MTW11], while for $k < (2d - 2)/3$ it is polynomial-time solvable, assuming $d$ fixed, as was shown in a series of papers on computational homotopy theory [ČKM+13, ČKM+12, KMS13, ČKV13]. (However, the cases with $(2d - 2)/3 \leq k$ known to be NP-hard but not proved undecidable are still intriguing.)
Thus, the most significant gap up until now has been the cases $d = 3$ and $k = 2, 3$, and in particular, after graph planarity (EMBED$1\rightarrow2$), the problem EMBED$2\rightarrow3$ can be regarded as the most intuitive and probably practically most relevant case.

**Embeddability in $\mathbb{R}^3$.** Here we close this gap, at least as far as decidability is concerned.

**Theorem 1.1.** The problem EMBED$2\rightarrow3$ is algorithmically decidable. That is, there is an algorithm that, given a 2-dimensional simplicial complex $K$, decides whether $K$ can be embedded (piecewise linearly, or equivalently, topologically) in $\mathbb{R}^3$.

Let us remark that one can naturally consider (at least) three different kinds of embeddings of a simplicial complex $K$ in $\mathbb{R}^d$, illustrated in the next picture for a 1-dimensional complex (graph):

```
linear  piecewise linear (PL)  topological
```

For *linear embeddings*, also referred to as *geometric realizations*, each simplex of $K$ should be mapped affinely to a (straight) geometric simplex in $\mathbb{R}^d$. This kind of embeddability is decidable in PSPACE regardless of the dimensions, and it is not what we consider here.

For *piecewise linear*, or *PL*, embeddings, one seeks a linear embedding of some (arbitrarily fine) subdivision of $K$. Finally, for a *topological embedding*, $K$ is embedded by an arbitrary injective continuous map.

While topological and PL embeddability need not coincide for some ranges of dimensions, for ambient dimension $d = 3$, they do,\(^1\) and this is the notion of embeddability considered here.

An algorithm for EMBED$3\rightarrow3$ can be obtained from Theorem 1.1 by a simple reduction, given in Section 12.

**Corollary 1.2.** The problem EMBED$3\rightarrow3$ is decidable as well.

**Thickening to 3-manifolds.** For a 2-complex $K$, (PL) embeddability in $\mathbb{R}^3$ is easily seen to be equivalent to embeddability in $S^3$, and from now on, we work with $S^3$ as the target.

The first step in our proof of Theorem 1.1 is testing whether a given simplicial 2-complex $K$ embeds in any 3-dimensional manifold at all.

Let us suppose that there is an embedding $f: K \to M$ for some 3-manifold $M$ (without boundary), and take a sufficiently small closed neighborhood $X$ of the image $f(K)$ in $M$—the technical term here is a *regular neighborhood*. Then $X$ is a 3-manifold with boundary, called a *3-thickening* of $K$.

There is an algorithm, due to Neuwirth [Neu68] (see also [Sko95] for an exposition) that, given $K$, tests whether it has any 3-thickening, and if yes, produces a finite list of all possible 3-thickenings, up to homeomorphism, as triangulated 3-manifolds with boundary *(without the knowledge of M)*. Then $K$ embeds in $S^3$ iff one of its 3-thickenings does. Hence it suffices to prove the following.

\(^1\) For complexes of dimension $k = 2$, this follows from [Bin59, Pap43], see also [MTW11] for more details and references; for complexes of dimension $k = 3$, this follows from the reduction in Section 12.
Theorem 1.3. There is an algorithm that, given a triangulated 3-manifold $X$ with boundary, decides whether $X$ can be embedded in $S^3$.

Concerning the running time. Our proof does provide an explicit running time bound for the algorithm, but currently a rather high one, certainly primitive recursive but even larger than an iterated exponential tower. Thus, we prefer to keep the bounds unspecified, in the interest of simplicity of the presentation.

By refining our techniques, it might be possible to show the problem to lie in the class NP. Going beyond that may be quite challenging: indeed, as observed in [MTW11], EMBED$_{2 \rightarrow 3}$ is at least as hard as the problem of recognizing $S^3$ (that is, given a simplicial complex, decide whether it is homeomorphic to $S^3$). The latter problem is in NP [Iva08, Sch04], but nothing more seems to be known about its computational complexity (e.g., polynomiality or NP-completeness).

Related work. There is a vast amount of literature on computational problems for 3-manifolds and knots. Here we give just a sample; further background and references can be found in the sources cited below and in [AHT06]. A classical result is Haken’s algorithm deciding whether a given polygonal knot in $\mathbb{R}^3$ is trivial [Hak61]. More recently, this problem was shown to lie in NP [HLP99], and, assuming the Generalized Riemann Hypothesis, in coNP as well [Kup11]. The knot equivalence problem is also decidable [Hak61, Hem79, Mat97], but nothing seems to be known about its complexity status.

Closer in spirit to the problem investigated here are algorithms for deciding whether a given 3-manifold is homeomorphic to $S^3$, already mentioned above [Rub95, Tho94, Iva08, Sch04].

An important special case of Theorem 1.3 is testing embeddability into $S^3$ for an $X$ whose boundary is a single torus; this amounts to recognizing knot complements and was solved in [JS03]. Some of the ideas in that work are used in our proof, but most of the argument is fairly different.

In a different direction, Tonkonog [Ton11] provided an algorithm for deciding whether a given 3-manifold $X$ with boundary embeds into some homology 3-sphere\(^2\) (which may depend on $X$). His methods are completely different from ours (except for using a 3-thickening to pass from 2-dimensional complexes to 3-manifolds), and it seems to be only loosely related to the problems investigated here.

Future directions. Besides the obvious questions of finding a more efficient algorithm, say one in NP, and/or proving hardness results, one may consider embeddability into other 3-manifolds $M$ besides $S^3$. We believe that this may be within reach of the methods used here, but definitely a number of issues would have to be settled.

The main technical contribution. Our algorithm relies on a large body of work in 3-dimensional topology.

When we talk about a surface in $X$, unless explicitly stated otherwise, we always mean a 2-dimensional manifold $F$ with boundary properly embedded in $X$, that is, with $\partial F \subset \partial X$. Similarly, curves are considered properly embedded in a surface, so a connected curve can be a loop in the interior of the surface or an arc connecting two points of the boundary. Two properly embedded surfaces $F$ and $F'$ are isotopic if they are embeddings of the same surface $F_0$ and there is a continuous family of proper embeddings $F_0 \to X$ starting with $F$ and ending

\(^2\)A 3-manifold whose homology groups are the same as those of $S^3$. 
with $F'$. An similar definition of isotopy applies to curves embedded in surfaces.

As in almost all algorithms working with 3-manifolds, we use Haken’s method of normal curves and surfaces, actually in a slightly extended form. Here we recall them very briefly; we refer to [Hem76, JT95] for background, and in Section 5 below we will discuss a variant.

A normal curve in a triangulated 2-dimensional surface $F$ intersects every triangle in finitely many disjoint pieces, which we can think of as straight segments, as in the left picture:

The main point is that such a curve is described, up to isotopy, purely combinatorially: namely, for every triangle $T$, there are just three types of segments of the curve inside, and it is enough to specify the number of segments for each type, for each $T$. In the picture, the numbers are 5, 2, 1.

Similarly, a normal surface in a triangulated 3-manifold intersects each tetrahedron in finitely many of disjoint pieces, each of them a triangle or a quadrilateral, as in the right picture above. This time there are seven types of pieces, four triangular and three quadrilateral, per tetrahedron (although no two types of quadrilateral pieces may coexist in a single tetrahedron, since they would have to intersect, which is not allowed). So a normal surface $F$ in a 3-manifold with $t$ tetrahedra can be described by a vector of $7t$ nonnegative integers. This vector is called the normal vector of $F$.

A normal isotopy is an isotopy during which the intermediate curve or surface stays normal; in particular, it may not cross any vertex of the triangulation.

Going back to embeddings, we first simplify the situation using a result of Fox [Fox48], which allows us to assume that the complement of the supposed embedding of $X$ in $S^3$ is a union of handlebodies.\(^3\) (These handlebodies may be knotted or linked in $S^3$, though, as in the picture at the beginning of the next section.) This assumption is quite important and nontrivial; for example, we note that if $X$ is a solid torus, it can also be embedded in $S^3$ in a knotted way, so that the complement is not homeomorphic to a solid torus.

Thus, now we ask if there is a way of “filling” each component of $\partial X$ with a handlebody so that the resulting closed manifold is homeomorphic to $S^3$. Spherical boundary components are easy, since there is only one way, up to homeomorphism, of filling a spherical boundary component with a ball. However, already for a toroidal component there are infinitely many nonequivalent ways of filling it with a solid torus. Indeed, the filling can be done in such a way that a circle $\alpha$ on the toroidal component of $\partial X$, as in the left picture,

is identified with a curve $\beta$ on the boundary of the solid torus, shown in the right picture, where $\beta$ may wind around the solid torus as many times as desired. For boundary components

\[^3\]A handlebody is a ball with (solid) handles, or equivalently, a 3-thickening of a finite connected 1-dimensional complex (graph).
of higher genus, there are also infinitely many ways of filling, and their description is still more complicated. For every specific way of filling the boundary components of $X$ with handlebodies we could test whether the resulting closed manifold is an $S^3$, but we cannot test all of the infinitely many possibilities. This is the main difficulty we have to overcome to get an algorithm.

Next, by more or less standard considerations, we can make sure that there is no “way of simplifying $X$ by cutting along a sphere or disk”—in technical terms, we may assume that $X$ is irreducible, that is, every 2-sphere embedded in $X$ bounds a ball in $X$, and that $X$ has an incompressible boundary, i.e., any curve in $\partial X$ bounding a disk in $X$ also bounds a disk in $\partial X$.

For dealing with such an $X$, the following result is the key:

**Theorem 1.4.** Let $X$ be an irreducible 3-manifold, neither a ball nor an $S^3$, with incompressible boundary and with a 0-efficient triangulation $\mathcal{T}$. If $X$ embeds in $S^3$, then there is also an embedding for which $X$ has a short meridian $\gamma$, i.e., an essential normal curve $\gamma \subset \partial X$ bounding a disk in $S^3 \setminus X$ such that the length of $\gamma$, measured as the number of intersections of $\gamma$ with the edges of $\mathcal{T}$, is bounded by a computable function of the number of tetrahedra in $\mathcal{T}$.

In this theorem, 0-efficient triangulation is a technical term introduced in [JR03], whose definition will be recalled later in Section 7. We are using 0-efficient triangulations in order to exclude non-trivial normal disks and 2-spheres in $X$.

We should also mention that the triangulations commonly used in 3-dimensional topology, and also here, are not simplicial complexes in the usual sense—they are still made by gluing (finitely many) tetrahedra by their faces, but any set of gluings that produces a manifold is allowed, even those that identify faces of the same tetrahedron. As a result, a particular tetrahedron may not have four distinct faces, six distinct edges and four distinct vertices. In particular, 0-efficient triangulations of the manifolds we consider have a single vertex in each boundary component and none in the interior, all edges in the boundary form loops. This is the necessary result of modifying a triangulation by collapsing simplices, a triangular face to an edge or to a vertex, etc.; see [JR03, Sec. 2.1] for a thorough discussion. There is even a mind-boggling one-tetrahedron one-vertex triangulation of the solid torus obtained by gluing a pair of faces of a single tetrahedron, see [JS03].

Let us remark that $X$ as in the theorem need not have a short meridian for every possible embedding, even if we assume that the complement consists of handlebodies. For example, if $X$ is a thickened torus (a torus times an interval), we can embed it so that the curves bounding disks in $S^3 \setminus X$ are arbitrarily long w.r.t. a given triangulation of $X$. We must sometimes change the embedding to get short meridians.

It is also worth mentioning that this problem does not occur if $\partial X$ is a single torus, i.e., the knot complement case. Here a celebrated theorem of Gordon and Luecke [GL89] makes sure that there is only one embedding, up to a self-homeomorphism of $S^3$, and the meridian is unique up to isotopy. This is why the single-torus boundary case solved in [JS03] is significantly easier than the general case.

---

4Meaning that $\gamma$ does not bound a disk in $\partial X$.
2 An outline of the arguments

Our algorithm for Theorem 1.3, deciding the embeddability of a given 3-manifold \( X \) in \( S^3 \), for the case of \( X \) irreducible and with incompressible boundary, consists in testing every possible normal curve \( \gamma \subset \partial X \) of length bounded as in Theorem 1.4. For each such candidate \( \gamma \), we construct a new manifold \( X' = X'(\gamma) \) by adding a 2-handle to \( X \) along \( \gamma \), which means that we glue a disk bounded by \( \gamma \) to the outside of \( X \) and thicken it slightly, as illustrated in the following picture:

Here \( X \) is the complement of the union of two (linked) handlebodies, a knotted solid 3-torus and a solid torus, and for \( X' \), the solid 3-torus in the complement has been changed to a solid double torus.

Then we test the embeddability of each \( X'(\gamma) \) recursively, and \( X \) is embeddable iff at least one of the \( X'(\gamma) \) is. It is not hard to show that the algorithm terminates, using the vector of genera of the boundary components of \( X \); see Section 3.

The proof of Theorem 1.4 occupies most of the paper and has many technical steps. In this section we give an outline.

We assume \( X \) to be embedded in \( S^3 \), the complement being a union of handlebodies, and we apply a result of Li [Li10] stating that there is a planar surface (i.e., a disk with holes) \( P \subset X \) that is “stuck” in its position in a suitable sense (namely, \( P \) is either essential,\(^5\) or strongly irreducible and boundary strongly irreducible) and is meridional or almost meridional.

Here an essential curve \( \gamma \subset \partial X \) is a meridian in a given embedding of \( X \) in \( S^3 \) if it bounds a disk in \( S^3 \setminus X \). The surface \( P \) is meridional if each component of \( \partial P \) is a meridian, and it is almost meridional if all components of \( \partial P \) but one are meridians. (Actually, Li has yet another case in his statement, but as we will check, that case can be reduced to the ones given above; see Lemma 4.4.) The next picture illustrates a meridional \( P \) in the case where \( X \) is embedded in \( S^3 \) as the complement of a solid torus neighborhood of the figure ‘8’ knot:

Next, by choosing \( P \) as above with suitable minimality properties, one can make sure that \( P \) is normal or almost normal\(^6\) for the given triangulation. For the case of \( P \) essential, this

---

\(^5\)The precise definitions of essential, strongly irreducible, and boundary strongly irreducible are somewhat complicated and we postpone them to Section 4.

\(^6\)An almost normal surface is like a normal surface except that in at most one tetrahedron we also allow, in addition to the triangular and quadrangular pieces, one of two types of exceptional pieces, namely, a tube or an octagon; see Section 5.
is an old result going back to Haken and Schubert (and for our notion of complexity of $P$, a proof is given in Section 7), while for $P$ strongly irreducible and boundary strongly irreducible this follows from [BDTS12]; also see [Sto00] for the case of a strongly irreducible surface in a closed manifold. It remains to show that, in this setting, at least one of the meridians in $\partial P$ must be short.

Here we apply an average length estimate, which is an idea of Jaco and Rubinstein appearing in [JS03, JRS09].

Let $\gamma_1, \ldots, \gamma_b$ be the components of $\partial P$, and let $\ell(\partial P) = \sum_{i=1}^b \ell(\gamma_i)$ be the boundary length of $P$. We know that all the $\gamma_i$ but at most one are meridians. The length of the shortest meridian is bounded by the average $\ell(P)/(b-1)$, and we want to bound this average by a (computable) function of $t$, the number of tetrahedra in the triangulation $T$ of $X$.

Now by the theory of normal surfaces, the (almost) normal surface $P$ can be written as a normal sum\(^7\) of fundamental surfaces in $X$,

$$P = \sum_i k_i F_i,$$

where the $k_i$ are positive integers and the $F_i$ are surfaces from a finite collection; their number, as well as $\ell_{\text{max}} := \max_i \ell(\partial F_i)$ can be bounded by a (computable) function of $t$ alone, and does not depend on $P$.

Since the boundary length is additive w.r.t. normal sum, we have $\ell(\partial P) = \sum_i k_i \ell(\partial F_i) \leq \ell_{\text{max}} K$, where $K := \sum_i k_i$ is the number of fundamental summands in the expression for $P$, and so it suffices to show that $K \leq Ct$, with some computable function $C = C(t)$.

The basic version of the average-length estimate uses the Euler characteristic $\chi$ as an accounting device. Since $\chi$ is additive as well, $\chi(P) = \sum_i k_i \chi(F_i)$. Since $P$ is a planar surface with $b$ boundary components, we have $\chi(P) = 2 - b$.

Now an ideal situation for the average-length estimate (which we cannot guarantee in our setting) is when $\chi(F_i) \leq -1$ for every $i$; in other words, none of the summands is a disk, 2-sphere, annulus, Möbius band, or torus (or projective plane or Klein bottle, but these cannot occur in $X$ embedded in $S^3$). Then we get $b - 2 - \chi(P) = \sum_i k_i (-\chi(F_i)) \geq K$, and we are done (even with $C = 1$).

In our actual setting, the summands with $\chi > 0$, i.e., spheres and disks, are excluded by the 0-efficient triangulation of $X$. We also need not worry about torus summands, since they have empty boundary and thus do not contribute to $\ell(\partial P)$. The real problem are annuli (and Möbius bands, but since twice a Möbius band, in the sense of normal sum, is an annulus, Möbius bands can be handled easily once we deal with annuli).

There are two kinds of annuli, which need very different treatment: the essential ones, and the boundary parallel ones. Here an annulus $A \subset X$ is boundary parallel if it can be isotoped to an annulus $A' \subset \partial X$ with $\partial A' = \partial A$ while keeping the annulus boundary fixed. Boundary parallel annuli do not occur for $P$ essential, but they might occur for the case of $P$ strongly irreducible and boundary strongly irreducible.

To deal with the annulus summands, we first construct what we call an annulus curve $\alpha \subset \partial X$. This is the boundary of a maximal collection $\mathcal{A}$ of essential annuli, maximal in the sense that each of the two boundary curves of every other essential annulus, after a suitable

\(^7\)For normal surfaces $F, F_1, F_2$ in a triangulation $T$ of $X$, $F$ is called the normal sum of $F_1$ and $F_2$ if $\vec{v}(F) = \vec{v}(F_1) + \vec{v}(F_2)$, where $\vec{v}(F)$ denotes the normal vector of $F$. Similarly for almost normal surfaces, where we have extra coordinates in $\vec{v}(F)$ for the exceptional types of pieces; in this case, at least one of $F_1, F_2$ has to be normal. Also see Section 5.
normalization, either intersects $\alpha$ or is normally isotopic to a component of $\alpha$. We bound the length of $\alpha$ by a computable function of $t$, and $|\alpha \cap P|$, the number intersections of $\alpha$ with $P$, by $C'b$, for some computable $C' = C'(t)$, again assuming $P$ minimal in a suitable sense. For obtaining this bound we may need to change the embedding of $X$, and we also use results about “untangling” a system of curves on a surface by a boundary-fixing self-homeomorphism from [MSTW13].

Similarly, we construct a collection $\Gamma$ of curves that helps to deal with boundary parallel annuli: those that have minimal boundary in a suitable sense either intersect $\alpha$, or their boundaries are normally isotopic to components of $\alpha$ or curves from $\Gamma$.

Having constructed such an $\alpha$ and $\Gamma$, we work with normal curves and surfaces in a “marked” sense, which also takes into account the position of the curves and surfaces w.r.t. $\alpha$ and $\Gamma$. This, in particular, makes the number of intersections with $\alpha$ additive w.r.t. $\alpha$ and $\Gamma$. This, in particular, makes the number of intersections with $\alpha$ additive w.r.t. the marked normal sum, which in turn allows us to bound the number of annulus summands in (1), both boundary parallel and essential, that intersect $\alpha$ by $C'b$.

Then we might have boundary-parallel annulus summands that avoid $\alpha$, but we show that those do not occur at all, since they would contradict the minimality of $P$.

Finally, there remain essential annuli that have a boundary component parallel to a component of $\alpha$. Here we show that if such an annulus had the coefficient $k_i$ in (1) at least $|\alpha \cap \partial P| \leq C'b$, then there is a self-homeomorphism of $X$, namely, a Dehn twist in the annulus, that makes $P$ simpler, contradicting its supposed minimality. (Here we may again modify the assumed embedding of $X$ in $S^3$ in order to get a short meridian—and, as we have remarked, some such modification is necessary in the proof, since some embeddings may not have short meridians.) Hence for these essential annuli, too, the coefficients are bounded by a linear function of $b$. This concludes the proof.

3 The algorithm

If $X$ embeds in $S^3$, then it is orientable, and orientability can easily be tested algorithmically (e.g., by a search in the dual graph of the triangulation, or by computing the relative homology group $H_3(X, \partial X)$). So from now on, we assume $X$ orientable. In this situation, the boundary of $X$ is a compact orientable 2-manifold, and thus each component is a 2-sphere with handles.

We describe a recursive procedure $\text{EMB}(X)$ that accepts a triangulated orientable 3-manifold with boundary and returns TRUE or FALSE depending on the embeddability of $X$ in $S^3$. (With some more effort, for the TRUE case, we could also recover a particular embedding, but we prefer simplicity of presentation.) The procedure works as follows.

1. (Each component separately) Let $X_1, \ldots, X_k$ be the connected components of $X$. If $k > 1$, test if there is an $S^3$ among the $X_i$ (several algorithms are available for that [Rub95, Tho94, Iva08, Sch04]), and if yes, return FALSE. Otherwise, still for $k > 1$, return the conjunction $\text{EMB}(X_1) \land \cdots \land \text{EMB}(X_k)$.

2. (Fill spherical holes) Now we have $X$ connected. If it is an $S^3$, return TRUE. If there are components of $\partial X$ that are $S^2$'s, form $X'$ by attaching a 3-ball to each spherical component of $\partial X$, and return $\text{EMB}(X')$. 

8
3. (Connected sum) Form a decomposition $X = X_1 \# \cdots \# X_k$ of $X$ into a connected sum\(^8\) of prime manifolds\(^9\) that are not 3-spheres.\(^{10}\) If $k > 1$, i.e., $X$ is not prime, return $\text{EMB}(X_1) \wedge \cdots \wedge \text{EMB}(X_k)$. If $X$ is prime but not irreducible, i.e., contains an $S^2$ that does not bound a ball, then return FALSE.

4. (Boundary compression) Test if there is a compressing disc $D$ for $\partial X$ (i.e., $\partial D \subset \partial X$ does not bound a disk in $\partial X$).\(^{11}\) If yes, cut $X$ along $D$, obtaining a new manifold $X'$. Three cases may occur:

   (a) If $X'$ has two components, $X_1'$ and $X_2'$, return $\text{EMB}(X_1') \wedge \text{EMB}(X_2')$. This case may occur, for example, for $X$ a handlebody with two handles (a “thickened $S^2$”) when $D$ separates the two handles.

   (b) If $X'$ is connected and the two “scars” after cutting along $D$ lie in the same component of $\partial X'$, return $\text{EMB}(X')$. This case may occur, e.g., for $X$ a solid torus.

   (c) If neither of the previous two cases occur, then $X'$ is connected but the scars lie in different components of $\partial X'$. Return FALSE. To get an example of $X$ fitting this case, we can start with a thickened torus (i.e., torus times $[-1,1]$) and connect the two boundary components by a 1-handle—which cannot be done in $\mathbb{R}^3$, but it does give a 3-manifold (with double torus boundary).

5. (Short meridian) Now $X$ is irreducible and with incompressible boundary. Using \cite[Thm. 5.20]{JR03}, retriangulate $X$ with a 0-efficient triangulation. Then proceed as described at the beginning of Section 2: let $\gamma_1, \ldots, \gamma_n$ be a list of all closed essential normal curves in $\partial X$ up to the length bound as in Theorem 1.4, for each $i$ form $X'({\gamma_i})$ by attaching a 2-handle along $\gamma_i$, and return the disjunction $\text{EMB}(X'(\gamma_1)) \vee \cdots \vee \text{EMB}(X'(\gamma_n))$.

**Lemma 3.1.** The above procedure always terminates and returns a correct answer, assuming the validity of Theorem 1.4.

**Proof.** First we show that the algorithm always terminates. Let $C_1, \ldots, C_k$ be the components of $\partial X$ numbered so that $g(C_1) \geq \cdots \geq g(C_k)$, where $g(\cdot)$ stands for the genus, and let $\vec{g}(X)$ be the vector $(g(C_1), \ldots, g(C_k))$. We consider these vectors ordered lexicographically (if two vectors have a different length, we pad the shorter one with zeros on the right).

Let us think of the computation of the algorithm as a tree, with nodes corresponding to recursive calls. The branching degree is finite, so it suffices to check that every branch is finite.

It is easy to see that $\vec{g}(X)$ cannot increase by passing to a connected component or to a prime summand, and that it decreases strictly by a boundary compression and also by the short meridian step. Indeed, we observe that in the boundary compression step or the short meridian step, exactly one of the boundary components $C_i$ is affected, and it is either split into two components $C'$ and $C''$ of nonzero genus and with $g(C_i) = g(C') + g(C'')$, or it

---

\(^8\)For two 3-manifolds $X$ and $Y$, the connected sum $X \# Y$ is obtained by removing a small ball from the interior of $X$, another small ball from the interior of $Y$, and identifying the boundaries of these two balls.

\(^9\)A prime 3-manifold is one that has no decomposition as a connected sum $X \# Y$ with neither $X$ nor $Y$ an $S^2$.

\(^{10}\)The algorithm for prime decomposition goes back to Schubert \cite{Sch49}, for closed manifolds it is presented in detail in \cite{JT95}, and a version for manifolds with boundary is implicit in \cite{JR03}.

\(^{11}\)The idea of an algorithm is due to Haken, and the algorithm is implicit in \cite{JR03}.
remains in one piece but the genus decreases by one. Since after steps 1–3 we have a connected irreducible manifold without spherical boundary components, for which the next step either finishes the computation or reduces $\tilde{g}_2(X)$ strictly, every branch is finite as needed.

It remains to show that the returned answer is correct. For Step 2, we need that there is a unique way of filling a spherical hole; this is well known and can be inferred, for example, from the fact that there is only one orientation-preserving self-homeomorphism of $S^2$ up to isotopy [FM11, Sec. 2.2].

For Step 3, it is easily checked that a connected sum embeds iff the summands do. Moreover, every $S^2$ embedded in $S^3$ separates it, and hence if $X$ contains a non-separating $S^2$, then it is not embeddable.

For Step 4, it is clear that if $X$ is embeddable, then so is $X'$.

If, in case (4a), $X_1'$ and $X_2'$ are both embedded, then it is easy to construct an embedding of $X$: Denote $D$’s scars by $D_1$ and $D_2$. Then a regular neighborhood of $D_i$ is a ball $B_i$ with boundary $S_i = \partial B_i$, and that meets both $X_i'$ and $S^3 \setminus X_i'$ in balls. Think of each $X_i'$ as embedded in its own copy of $S^3$, and take a connected sum of these two $S^3$’s so that $S^3 = S^3 # S^3 \supset X_1' # D_1 = D_2 X_2' = X$. Similarly, if $X'$ is embedded in case (4b), then we can connect the scars by a thin handle in $S^3 \setminus X'$ and obtain an embedding of $X$.

In case (4c), let $C_1 \neq C_2$ be the components of $\partial X'$ containing the scars. Since the disk $D$ does not separate $X$, we can choose a loop $\delta \subset X$ meeting $D$ in a single point, and such that $\delta$ also meets $C_1$ in a single point. But then, if $X$ were embedded in $S^3$, $C_1$ would yield a nonseparating surface in $S^3$—a contradiction.

Finally, if one of the $X'(\gamma_i)$ is embeddable in Step 5, then so is $X$ (since in $X'$ we have a handle that was added to $X$, and we can just assign it to the complement of $X$), and if $X$ is embeddable, then at least one of the $X'(\gamma_i)$ is by Theorem 1.4. \hfill \qedsymbol

4 Intersections of curves and surfaces

In this section we collect terminology, definitions and basic results concerning properly embedded curves in surfaces and properly embedded surfaces in 3-manifolds. In particular, for latter sections we need that any pair of properly embedded surfaces, each either essential, or, strongly irreducible and boundary strongly irreducible, can be isotoped to intersect essentially. There are few new results in this section. The reader is referred to Hempel [Hem76] and Jaco [Jac80] for more background.

We assume throughout that all curves and surfaces have been isotoped to have transverse intersection.

4.1 Curves

A curve is a properly embedded 1-dimensional manifold in a surface $F$, each component either a loop, which is closed, or an arc, which has two endpoints in $\partial F$.

A loop is trivial if it bounds a disk in $F$ and an arc is trivial if it co-bounds a disk in $F$ with some arc in $\partial F$. A curve is essential if no component is trivial.

Pairs of curves are assumed to intersect transversally. If $\alpha$ and $\beta$ are a pair of curves, then their geometric intersection number $i(\alpha, \beta) = \min(|\alpha' \cap \beta'|)$ taken over all pairs of curves $(\alpha', \beta')$ where $\alpha'$ and $\beta'$ are isotopic to $\alpha$ and $\beta$ within $F$, respectively. (The isotopies are also allowed to move endpoints of arcs within the boundary.)
We say that $\alpha$ and $\beta$ bound a bigon if there is a disk bounded by a pair of sub-arcs, one from each curve; see Figure 3 in Section 6 below. We say that they bound a half-bigon if there is a disk bounded by a pair of sub-arcs, one from each curve, along with an arc in $\partial F$. If $\alpha$ and $\beta$ bound a bigon or half-bigon, then they can be isotoped to reduce their intersection.

We need this converse, a mild generalization of Farb and Margalit’s bigon criterion:

**Lemma 4.1** (Bigon criterion [FM11]). A pair of curves $\alpha$ and $\beta$ realize their geometric intersection number if and only if they do not bound a bigon or half-bigon.

**Proof.** Farb and Margalit show that any pair of connected loops that intersect non-minimally form a bigon. They also note that this extends to disconnected curves consisting of loops.

If either curve has an arc component, then the doubled curves are properly embedded closed curves in the double of the surface. If they intersect non-minimally in the original, they intersect non-minimally in the double and hence bound a bigon there. Thus, they bound a half-bigon in the original. \qed

### 4.2 Essential surfaces

We will assume that our surfaces are properly embedded in a 3-manifold $X$ that is irreducible, i.e., every sphere embedded in $X$ bounds a ball in $X$, and boundary incompressible, i.e., any curve in $\partial X$ bounding a disk in $X$ also bounds a disk in $\partial X$ (is trivial).

Let $F$ be a surface properly embedded in $X$. A *compressing disk* for a $F$ is an embedded disk $D \subset X$ whose interior is disjoint from $F$ and whose boundary is an essential loop in $F$. A *boundary compressing disk* is an embedded disk $D \subset X$ whose boundary, $\partial D = f \cup x$, is the union of $f = \partial D \cap F = D \cap F$, an essential arc properly embedded in $F$, and $x = \partial D \cap \partial X = D \cap F$, an arc properly embedded in $\partial X$. Here is an illustration:

![Illustration of a compressing disk and a boundary compressing disk](image)

A surface $F$ is *compressible* if it has a compressing disk, *boundary compressible* if it has a boundary compressing disk, and *incompressible* and *boundary incompressible* if not, respectively. A surface is *essential* if it is incompressible, boundary incompressible, and not a sphere bounding a ball, or a disk co-bounding a ball with a disk in $\partial X$.

We establish some basic facts about surfaces in $X$.

**Proposition 4.2.** The following statements hold for properly embedded surfaces in $X$, an irreducible, orientable 3-manifold with non-empty incompressible boundary:

(i) Every disk co-bounds a ball with a disk in $\partial X$.

(ii) Every connected surface with an inessential boundary curve is either compressible or a disk.

---

12 Meaning that we glue two copies of $F$ by identifying their boundaries.
(iii) The boundary curve of every compressible annulus is trivial in $\partial X$ boundary.

(iv) Every boundary compressible annulus is boundary parallel (parallel to an annulus in $\partial X$).

(v) No surface is a projective plane.

(vi) Every Möbius band is essential.

Proof. Because $X$ has incompressible boundary, the boundary of a properly embedded disk bounds a disk in $\partial X$. The union of these disks is a sphere that, because $X$ is irreducible, bounds a ball, yielding (i).

For (ii), suppose that some boundary curve of connected $F$ bounds a disk in $\partial X$. Then, among disks in $\partial X$ bounded by boundary curves of $F$, an innermost such disk can be pushed slightly into the interior of $X$ while keeping its boundary in $F$. The boundary curve is either trivial in $F$, in which case $F$ is a disk, or essential in $F$, in which case $F$ is compressible.

Concerning (iii), let $D$ be a compressing disk for an annulus $A$. Then $\partial D$ separates $A$ into two annuli $A'$ and $A''$. So $D \cup A'$ and $D \cup A''$ are properly embedded disks, each with one boundary curve of $A$. Because $\partial X$ is incompressible, the boundary curves are both trivial in $\partial X$.

As for (iv), let $B$ be a boundary compressing disk for an annulus $A$. Then $\partial N(A \cup B)$, the boundary of a regular neighborhood of their union, has two components, an annulus isotopic to $A$ and a disk. By (i), the disk co-bounds a ball with a disk in $\partial X$. But then the union of $N(A \cup B)$ with the ball is a solid torus, across which $A$ is parallel to an annulus in $\partial X$.

Concerning (v), if $P$ is a projective plane, then $\partial N(P)$, the boundary of its regular neighborhood, is a sphere which separates $P$ from $\partial X$. Then $X$ is reducible, for the sphere cannot bound a ball—no ball has interior boundary or contains an embedded projective plane.

Finally, in (vi), let $M$ be a Möbius band. Suppose first that $M$ is compressible and let $D$ be a compressing disk for an annulus $A$. Then $\partial N(A \cup B)$, the boundary of a regular neighborhood of their union, has two components, an annulus isotopic to $A$ and a disk. By (i), the disk co-bounds a ball with a disk in $\partial X$. But then the union of $N(A \cup B)$ with the ball is a solid torus, across which $A$ is parallel to an annulus in $\partial X$.

Suppose a Möbius band $M$ is boundary compressible. Then $\partial N(M)$ is a boundary compressible annulus $A$. By (iv), $A$ is boundary parallel, and co-bounds an solid torus with an annulus in the boundary. But the parallel region cannot contain the Möbius band $M$, and hence $X$ is the union of two solid tori, $N(M)$ and the solid torus parallel region.

We say that a pair of surfaces, $F$ and $G$, intersect essentially if each component of the curve $F \cap G$ is essential in both $F$ and $G$ (they are allowed to be disjoint). It is well known that essential surfaces can be arranged to intersect essentially:

**Lemma 4.3.** Let $F$ and $G$ be properly embedded essential surfaces in an irreducible manifold with incompressible boundary. Then $G$ can be isotoped so that they intersect essentially.

Proof. Assume that we have isotoped $G$ to minimize the number of curve components in $F \cap G$. We will show by contradiction that $F$ and $G$ intersect essentially.

We first note that if there is an intersection curve that is inessential in $F$, then there is an intersection curve that is inessential in $G$ and vice-versa: If an intersection curve bounds a disk in $F$, choose one whose disk is innermost. Since $G$ is incompressible, this disk is not a
compressing disk for \( G \) and it follows that its boundary, an intersection loop, is inessential in \( G \). The same observation applies to inessential intersection arcs.

Then, assuming that some intersection loop is trivial, we can pass to one that is innermost on \( F \), i.e., choose \( \alpha \) to be an intersection loop that bounds a disk \( D \subset F \) whose interior is disjoint from \( G \). Since \( G \) is not compressible, \( \alpha \) also bounds a disk \( D' \subset G \). The union \( D \cup D' \) is a sphere that, because \( X \) is irreducible, bounds a ball. And there is an isotopy of \( G \) that is restricted to a neighborhood of \( D' \), and that pushes \( D' \) across the ball and past \( D \). This isotopy of \( G \) eliminates \( \alpha \) and any other intersection curves in the interior of \( F \cap D' \), and it does not introduce any new intersection curves since \( \alpha \) was innermost.

Now assume some intersection arc is trivial in one of the surfaces, and as noted, we can let \( \alpha \) denote such an arc that is outermost in \( F \). That is, \( \alpha \) cuts off a disk \( D \subset F \) whose interior is disjoint from \( G \) and whose boundary meets \( \partial X \) in an arc. And \( \alpha = D \cap G \) cuts off a, not necessarily outermost, disk \( D' \subset G \) that also meets \( \partial X \) in an arc.

The union \( D \cup D' \) is a disk with its boundary in \( \partial X \) that, because \( \partial X \) is incompressible, bounds a disk \( D'' \subset \partial X \). Since \( X \) is irreducible, \( D \cup D' \cup D'' \) is a sphere bounding a ball. Moreover, there is an isotopy of \( G \) that pushes a neighborhood of \( D' \) past \( D \) and outside the ball. \( \square \)

4.3 Almost meridional surfaces

Suppose that \( X \) is an irreducible manifold with incompressible boundary that is embedded in \( S^3 \). We recall that an essential curve \( \mu \subset \partial X \) is a meridian if it bounds a disk in \( S^3 \setminus X \). A properly embedded surface is meridional if each of its boundary curves is a meridian, and almost meridional if all but exactly one of its boundary curves is a meridian.

Let \( D \) be a boundary compressing disk for an orientable surface \( P \). Then \( \partial N(P \cup D) \) is a surface with at least two components. One component is isotopic to \( P \); let \( P' \) be the union of the other components. Then \( P' \) is said to be the result of boundary compressing \( P \) along \( D \).

**Lemma 4.4.** Suppose that a manifold \( X \) is embedded in \( S^3 \). If \( P \) is a connected almost meridional planar surface properly embedded in \( X \), then any surface \( P' \) obtained by boundary compressing \( P \) contains an almost meridional component.

**Proof.** Let \( P' \) be obtained from \( P \) by boundary compressing along the disk \( D \). What happens to \( \partial P' \)? The disk \( D \) meets at most two boundary components of \( \partial P \). Any component not met by \( D \) has two parallel copies in \( \partial N(P \cup D) \), one for \( P \) and one for \( P' \), so those are unchanged. Let \( \beta \) be the one or two loops meet by the arc \( x = D \cap \partial X \). Since \( D \) lies on one side of the the 2-sided planar surface \( P \), when \( \beta \) is a single loop, \( x \) approaches it twice from the same side. It follows that \( \partial N(x \cup \beta) \) is a pair of pants, i.e., an \( S^2 \) with three holes bounded by loops. One of these loops belongs to \( P \) and two to \( P' \), or vice-versa.

If any two of these three loops are meridians, then so is the third, since it bounds a disk, namely the union of the pants and the two disks pushed slightly into \( S^3 \setminus \text{interior}(X) \).

We apply this “two meridians implies three meridians” principle to show that \( P' \) has an almost meridional component, regardless of how the boundary compressing disk meets the boundary components of \( P \).

If the boundary compressing disk meets the non-meridional component twice, then the compression eliminates the non-meridional curve, and creates two new curves, each belonging to a separate component of \( P' \). At least one of the new curves is not meridional, and hence its component is almost meridional.

13
If the boundary compressing disk meets a meridian and the non-meridian, then the compression does not separate $P$, and trades these curves for a new non-meridional curve. Thus $P'$ is almost meridional.

If the boundary compressing disk meets two distinct meridians, then they are eliminated and a new one is created. The connected surface $P'$ is almost meridional.

If the boundary compressing disk meets a single meridian twice, then $P'$ has two components, each with one of the two new curves, either both meridional or both non-meridional. If both are meridional, then the component with the original non-meridian on its boundary is almost meridional. If both are non-meridional, then the component without the original non-meridian is almost meridional. One of the two components of $P'$ is almost meridional.

**Lemma 4.5.** Suppose that $X$, an irreducible manifold with incompressible boundary, is embedded in $S^3$. If $X$ contains an incompressible, almost meridional planar surface, then $X$ contains an essential almost meridional planar surface.

**Proof.** An incompressible almost meridional surface can be sequentially boundary compressed until it is incompressible and boundary incompressible. By the prior lemma, each surface in the sequence, hence the final one, has an almost meridional component. This final component is not a disk because $X$ is boundary incompressible. Hence it is an essential almost meridional planar surface. □

### 4.4 Strongly irreducible surfaces

A two-sided surface properly embedded in $X$ is **bi-compressible** if it has a compressing pair $(D_+, D_-)$, a pair of disks, each a compressing or boundary compressing disk, one for each side of the surface. The pair is **simultaneous** if $\partial D_+ \cap \partial D_- = \emptyset$.

A surface is **weakly reducible** if it is simultaneously bi-compressible using compressing disks only. A **strongly irreducible** surface is one that is bi-compressible using compressing disks but not simultaneously so. A surface is **boundary weakly reducible** if it is simultaneously bi-compressible using any combination of compressing disks and boundary compressing disks. A surface is **boundary strongly irreducible** if it is bi-compressible, using any combination of compressing or boundary compressing disks, but not simultaneously so.

Some of our results assume that a surface is both strongly irreducible and boundary strongly irreducible. It may seem that the strongly irreducible hypothesis is vacuous. But this is not the case—it guarantees that the surface has at least one (non-boundary) compressing disk for each side.

**Lemma 4.6** ([BDTS12], Lemma 3.8). In an irreducible manifold with incompressible boundary, the boundary of a strongly irreducible surface is essential in the boundary of the manifold.

We state here a special case of Lemma 4.2 of [Bac13]. This

**Lemma 4.7** (Lemma 4.2 of [Bac13]). Let $F$ be an essential surface and $G$ a surface that is strongly irreducible and boundary strongly irreducible. Then $G$ may be isotoped so that $F$ and $G$ intersect essentially.

Let us remark that Bachman does not give a proof but claims it to be a direct generalization of [Bac09, Corol. 3.8]. He has also provided us with an unpublished manuscript with a proof.
5 Theory of normal curves and surfaces in a marked triangulation

In this section we introduce a mild generalization of the theory of normal curves and surfaces.

**Definition 5.1.** A marked triangulation is a pair \((T, M)\) consisting of a triangulation \(T\) of a 2- or 3-manifold along with a marking \(M \subset T^1\), a finite set of points along the edges of \(T\).

If \(M = \emptyset\), then \((T, M)\) is a triangulation in the usual sense and we will usually omit \(M\) and refer directly to \(T\). Similarly, when \(M = \emptyset\), we will describe objects as being normal rather than \(M\)-normal, and note that our definitions restrict to the standard ones.

An arc in a triangle is \(M\)-normal if its endpoints lie in distinct edges of the face and it misses \(M\). A properly embedded curve \(\alpha\) in the boundary surface is \(M\)-normal if it is the union of \(M\)-normal arcs. The length of \(\alpha\), \(\ell(\alpha) = |\alpha \cap T^1|\), is its number of intersections with the 1-skeleton.

There are several types of elementary surfaces contained in a tetrahedron \(\Delta\). An \(M\)-normal disk is a disk in \(\Delta\) whose boundary is an \(M\)-normal curve of length 3 or 4 in \(\partial \Delta\). We also consider two types of \(M\)-exceptional pieces: An octagon is a disk in \(\Delta\) whose boundary is an \(M\)-normal curve of length 8 in \(\partial \Delta\). A tube is an unknotted annulus in \(\Delta\) whose boundary consists of two \(M\)-normal curves whose total length is at most 8.

An \(M\)-normal surface is a properly embedded surface that is the union of \(M\)-normal disks. An almost \(M\)-normal surface is a properly embedded surface that is the union of a single \(M\)-exceptional piece and a collection of \(M\)-normal disks.

The weight of an (almost) \(M\)-normal surface \(A\) is \(\text{wt}(A) = |A \cap T^1|\), the number of intersections with the 1-skeleton. Its length is the length of its boundary: \(\ell(A) := \ell(\partial A)\).

An \(M\)-normal isotopy is a normal isotopy that does not pass through any point in \(M\). An \(M\)-type is the equivalence class of an \(M\)-normal arc in a face, or, an \(M\)-normal disc or \(M\)-exceptional piece in a tetrahedron. Two types are \(M\)-compatible if they have disjoint representatives. A pair of curves or surfaces are \(M\)-compatible if each pair of types they possess are \(M\)-compatible. That is, a pair of curves is \(M\)-compatible if, for each face of the triangulation, their arcs in that face are pairwise disjoint after an \(M\)-normal isotopy. An analogous statement holds for surfaces.

We note that \(M\)-compatibility is a local condition; in general it may not be possibly to make \(M\)-compatible curves or surfaces globally disjoint by an \(M\)-normal isotopy.

The \(M\)-normal vector or \(M\)-normal coordinates of an \(M\)-normal curve, surface, or almost \(M\)-normal surface \(A\) is a uniquely determined vector \(\vec{v}_M(A)\), indexed over the set of normal types and with each entry recording the number of \(M\)-normal objects of the index type.

If \(A, B, C\) are \(M\)-normal surfaces such that \(\vec{v}_M(C) = \vec{v}_M(A) + \vec{v}_M(B)\), then \(C\) is an \(M\)-normal sum of \(A\) and \(B\), and we write \(C = A + B\). The same definition applies if \(A\) is \(M\)-normal and \(B\) and \(C\) almost \(M\)-normal, or if \(A, B, C\) are \(M\)-normal curves.
We note that not every two $M$-normal surfaces, for example, can be normally added—this is possible exactly if they are $M$-compatible.

If $A$ and $B$ are $M$-compatible, then one can construct an $M$-normal sum as follows. In each face or tetrahedron $\Delta$, the $M$-normal pieces $A \cap \Delta$ and $B \cap \Delta$ can be $M$-normally isotoped to be disjoint, and then attached across each facet of $\Delta$ to the pieces in an adjacent face/tetrahedron. This produces a properly embedded $M$-normal curve, $M$-normal surface, or almost $M$-normal surface, respectively, which is the $M$-normal sum.

However, in our considerations, we will mostly use a different geometric construction of an $M$-normal sum, where we assume that the curves or surfaces in question intersect minimally, in a suitable sense, but then we do not isotope them to be disjoint as above, but rather they stay in place and we deal with their intersections as well; see Section 5.1 below.

It is well known that Euler characteristic, weight and length are all additive with respect to normal sum, and this works without change for the $M$-normal case. If $A$ and $B$ are compatible (almost) $M$-normal curves or surfaces then the following hold:

1. $\chi(A + B) = \chi(A) + \chi(B)$
2. $\text{wt}(A + B) = \text{wt}(A) + \text{wt}(B)$
3. $\ell(A + B) = \ell(A) + \ell(B)$.

An (almost) $M$-normal curve or surface is fundamental if it cannot be expressed as the sum of other (almost) $M$-normal curves or surfaces. Every (almost) $M$-normal curve/surface is a non-negative integer combination of fundamentals.

An (almost) $M$-normal curve or surface is fundamental if it cannot be expressed as the sum of other (almost) $M$-normal curves or surfaces. Every (almost) $M$-normal curve/surface is a non-negative integer combination of fundamentals.

Here $M$-normal curve theory differs from standard normal curve theory. While all normal curves are compatible, $M$-normal curves have distinct compatibility classes, and this increases the number of fundamentals. In Figure 1, we see the boundary of a tetrahedron with two marked points, one on each of a pair of opposite edges. Let $\gamma$ be the length 8 $M$-normal curve that meets each of the sub-edges once. As a normal curve $\gamma$ is not fundamental—it is the sum of the two distinct length 4 curves $\alpha$ and $\beta$. But in the marked triangulation these curves are incompatible and $\gamma$ is fundamental.

If $\alpha$ is an (almost) $M$-normal curve or surface, then $\vec{v}_M(\alpha)$ is a solution to a set of matching equations: For a triangulated surface, this set consists of one equation for each sub-edge in the interior of the surface. It sets equal the sum of those coordinates meeting the sub-edge from one side to the sum of those meeting it on the other side. In a triangulated 3-manifold,
the set of matching equations consists of one equation for each $M$-normal arc type contained in an interior face. This equation sets equal the sum of the coordinates for elementary types using the arc type on one side to the sum of those using it on the other side.

We say that a vector $\vec{v}_M$ of the correct dimension is $M$-admissible if all its coordinates are non-negative, it satisfies the matching equations, and is self-compatible, i.e., it does not possess non-zero coordinates for any pair of non-$M$-compatible types. If $\vec{v}_M$ is an $M$-admissible vector, there is an (almost) $M$-normal curve/surface $\alpha$ for which $\vec{v}_M = \vec{v}_M(\alpha)$.

The following proposition is a straightforward generalization of a well known fact from normal surface theory to $M$-normal surfaces; see [HLP99] for a nice exposition.

**Proposition 5.2.** Given a 3-manifold with a marked triangulation $T_M$, the set $F$ of fundamental $M$-normal surfaces is computable, and both $|F|$ and the maximum weight $\text{wt}(F)$ of an $F \in F$ are bounded by a computable function of $t$ and $m$.

The bound on $|F|$ and $\text{wt}(F)$ has the form $\exp(p(t, m))$, where $p(t, m)$ is a suitable polynomial.

**Proof.** It is well known that without the marking, there are $7t$ normal disk types, $3t$ exceptional octagons and $25t$ exceptional tubed pairs of disks. Moreover, the presence of a tubed pair of disks may split one type of normal disks into two, but certainly we have no more than $42t$ types in total.

The points of $M$ divide each edge into at most $m + 1$ subarcs. In order to specify an $M$-normal type of a triangle, for example, we need to specify the subarc containing each of the three vertices, which leads to the bound $(m + 1)^3$. The worst bound is obtained for tubes and octagons, with $(m + 1)^8$, so a rough bound for the total number of $M$-types is $42t(m + 1)^8$.

A similar way of counting applies to the number of matching equations, which represent compatibility of the coordinates of the $M$-normal vector across the pieces of the edges of $T$ delimited by the points of $M$. Indeed, the matching equations correspond to $M$-arc types. There are at most $4t$ interior faces, each with 3 underlying normal arc types. A given $M$-arc type is thus determined by this normal type and by the sub-arcs it meets, and so there are at most $12t(m + 1)^2$ matching equations.

Then, reasoning as in [HLP99, Sec. 6], using a Hilbert basis of the appropriate integral cone, we obtain the bounds of the claimed form. 

5.1 Snug pairs of curves and surfaces, Haken sums, and normal sums

The normal sum $F + G$ of a pair of (almost) $M$-normal curves or surfaces $F$ and $G$ has been defined, if they are $M$-compatible, to be an $M$-normal surface whose $M$-normal vector is the sum of the $M$-normal vectors of $F$ and $G$, $\vec{v}_M(F + G) = \vec{v}_M(F) + \vec{v}_M(G)$.

It is desirable to show that qualities of the sum, such as essentiality or minimality, also apply to the summands. Here we describe a well known geometric interpretation of the sum that makes this possible; also see, for example, [JO84, JT95]. We also present some related material.

**Snug pairs.** We begin with a definition of a “placement with no unnecessary intersections” for a pair of curves or surfaces.

**Definition 5.3.** A pair $(F, G)$ of properly embedded curves or surfaces is snug if it is transverse and the number of components of the intersection $F \cap G$ is minimized over pairs $(F', G')$, 

17
where $F'$ and $G'$ are isotopic to $F$ and $G$, respectively. The pair $(F, G)$ is locally snug if $F \cap G$ is disjoint from the 1-skeleton $T^1$, and, they are snug in the interior of each simplex of the triangulation (here we only allow isotopies moving each intersection of $F$ or $G$ with a face only within that face).

If $F$ and $G$ are locally snug $M$-normal surfaces then it follows that:

1. each pair of $M$-normal arcs, one from $F$ and one from $G$, meets in 0 or 1 points;
2. each pair of $M$-normal disks, one from $F$ and one from $G$, meets in 2 or fewer arcs, and the union of the arcs has at most one endpoint in any face;
3. no loop of $F \cap G$ lies inside a tetrahedron.

Any pair of compatible $M$-normal curves or surfaces can be made locally snug by $M$-normal isotopies that first make their intersections with edges disjoint and then “straighten” them so that: normal arcs are straight, normal triangles are flat, and normal quads are the union of two flat triangles. We do not define locally snug when $F$ is an almost normal surface and $G$ is a normal surface, for in that case we require only the definition of the normal sum $F + G$ and not its geometric interpretation.

**Haken sum and normal sum of curves.** Now, for a while, we deal only with curves, and we develop a geometric interpretation of their normal sum. Here we consider only unmarked triangulations, i.e., $M = \emptyset$.

Let $D$ be a regular neighborhood of an intersection point $x$ of a pair of transverse curves $\alpha$ and $\beta$. We can remove the intersection by deleting the arcs in the interior of the disk and then attaching $\alpha$ to $\beta$ along a pair of antipodal sub-arcs of $\partial D$. Thus, we replace the “×” in $\alpha \cup \beta$ with either “)”(“ or “≈”. This is called an exchange or a switch at $x$. A curve is said to be a Haken sum $\alpha \wedge \beta$ of $\alpha$ and $\beta$ if it is obtained by an exchange at each of their intersection points. Of course, $\alpha \wedge \beta$ is dependent on the direction of the switches and is therefore not well determined.

If, however, $\alpha$ and $\beta$ are locally snug normal curves, then each intersection point is of the form $x = \alpha' \cap \beta'$ where $\alpha'$ and $\beta'$ are normal arcs in some face. Then $\alpha'$ and $\beta'$ meet at least one common edge $e$ of the face. The regular exchange is the exchange that does not produce an abnormal arc, a non-normal arc with both endpoints attached to $e$; see Fig. 2 top.

As we will see, the normal sum of $\alpha + \beta$ of locally snug curves can be obtained by doing all the regular exchanges.

**Lemma 5.4.** Let $\alpha$ and $\beta$ be locally snug normal curves. Then the Haken sum $\alpha \tilde{\wedge} \beta$ obtained by making all the regular exchanges is the normal sum $\alpha + \beta$; i.e., $\vec{v}(\alpha + \beta) = \vec{v}(\alpha) + \vec{v}(\beta)$.

**Proof.** We show that the result holds in each face of the triangulation. In an abuse of notation, let $\alpha$ and $\beta$ be restriction of the curves to a particular face. For contradiction suppose that they are a counterexample that minimizes $|\alpha \cap \beta|$. Then $\alpha$ and $\beta$ are not disjoint, for in that case, the union is normal and normal vectors add.

Since they intersect in a face, we can identify an outermost half-bigon bounded by a sub-arcs of $\alpha$ and $\beta$ and an edge of the face; see Figure 3. The regular exchange trades these sub-arcs and results in a pair of normal curves, $\alpha'$ normally isotopic to $\alpha$ and $\beta'$ normally isotopic to $\beta$, that are locally snug but with fewer intersections. By assumption, these $\alpha'$ and $\beta'$ satisfy the conclusion, hence so do $\alpha$ and $\beta$. 

18
Lemma 5.5. Let $\tilde{\alpha} + \tilde{\beta}$ be a Haken sum of locally snug properly embedded curves. Then $\tilde{\alpha} + \tilde{\beta}$ is normal if and only if $\alpha$ and $\beta$ are normal and all switches are regular, i.e., $\tilde{\alpha} + \tilde{\beta} = \alpha + \beta$. In addition, if $\alpha$ and $\beta$ are normal and $\tilde{\alpha} + \tilde{\beta}$ contains at least one irregular switch, then $\tilde{\alpha} + \tilde{\beta}$ contains an abnormal arc.

Proof. $(\Leftarrow)$ This is by Lemma 5.4. $(\Rightarrow)$ We show that if either $\alpha$ or $\beta$ is not normal then neither is $\tilde{\alpha} + \tilde{\beta}$ for any Haken sum of the curves.

Suppose then that $\alpha' \subset \alpha$ is an outermost non-normal arc, one that co-bounds a disk with a sub-arc of an edge $e' \subset e$. If $\beta$ meets the disk, it meets it in a collection of $n$ arcs, each with one endpoint in $\alpha'$ and one endpoint in $e'$, because $\alpha'$ is outermost and $\alpha$ and $\beta$ are snug. Let $D$ be a regular neighborhood of the disk. Then, regardless of the switches, $\alpha + \beta$ meets $D$ in a collection of $n + 1$ arcs that have $n + 2$ endpoints along the edge and $n$ endpoints not on the edge. It follows that at least one arc meets the edge in 2 points and is not normal. A symmetric argument applies if the outermost non-normal arc belongs to $\beta$. Nor can either $\alpha$ or $\beta$ possess a loop in a face. Local snugness implies that any loop is disjoint from the other curve and survives any Haken sum.

We now know that $\alpha$, $\beta$ are normal. To conclude the proof, it is sufficient to show that $\tilde{\alpha} + \tilde{\beta}$ contains an abnormal arc if at least one switch is irregular (this contradicts the normality of $\alpha + \beta$, and thus proves the last claim of the lemma). In an abuse of notation, let $\alpha$ and $\beta$ refer to the collection of normal arcs in a particular face.

We perform the specified switches in order according to the following scheme: If $\alpha \cap \beta \neq \emptyset$, then $\alpha$ and $\beta$ form an outermost half bigon $B$ with an edge as in Figure 3. The regular switch produces collections $\alpha'$ and $\beta'$ that are normally isotopic to $\alpha$ and $\beta$, but with one fewer intersections. An irregular switch produces a disjoint abnormal arc that survives any and all additional exchanges. If each exchange is regular, we can continue and the process produces a disjoint union $\alpha \sqcup \beta$. If any exchange is not regular, the resulting curve contains an abnormal arc. \qed

---

Figure 2: Switches for curves and surfaces
Normal sign. When \( \alpha \) and \( \beta \) lie in an oriented surface, for example the boundary of an oriented manifold, we can define the normal sign of each point of \( \alpha \cap \beta \). Viewing \( \alpha \) as horizontal and \( \beta \) as vertical, the regular exchange at the point connects a pair of quadrants. The point has positive sign if the exchange connects the southwest quadrant to the northeast quadrant, and it has negative sign if it connects the northwest to the southeast; see Figure 4. This is equivalent to the definition given in [BDTS12]. The definition depends on the ordering of the pair of curves and on an orientation on the surface: reversing the order or the orientation reverses every sign.

Normal sum of surfaces. Similar to the case of curves above, one can also construct the normal sum \( F + G \) of normal surfaces geometrically, using suitable switches. We assume that \( F \) and \( G \) are locally snug.

We construct \( F + G \) by specifying its intersection with the 1-, 2-, and 3-skeleta of the triangulation, respectively. First, we let the intersection of \( F + G \) with the 1-skeleton to be the union of the intersection points from \( F \) and those from \( G \).

Second, in each face we perform regular switches on all intersecting pairs of arcs \((f,g)\), where \( f \) comes from \( F \) and \( g \) from \( G \).

Finally, we construct the normal sum \( F + G \) in the interior of each tetrahedron \( T \). As discussed earlier, each normal disk is either a flat triangle or a quadrilateral made of 2 flat triangles. It follows that every intersection between normal disks from compatible surfaces is either 1 or 2 arcs, not necessarily straight. Compatibility ensures that the regular switches prescribed at the endpoints of each arc are consistent with each other and can be extended across the entire arc of intersection. The normal sum \( F \) and \( G \) is the result of performing such regular switches along every such arc of intersection.

Note that any intersection arc between normal disks can be extended from a tetrahedron through a face to a neighboring tetrahedron. In its entirety this intersection curve between \( F \) and \( G \) is either a loop, or an arc with both endpoints in \( \partial X \). Compatibility ensures that the regular switches in each face and through the interior of each tetrahedron agree. Thus we can regard the switch as a regular switch along the entire intersection curve.
Exchange arcs and surfaces, trace curves. Here we introduce some additional terminology. First, we consider a regular switch of two curves. Inside the neighborhood where the regular switch was performed, we identify an exchange arc that connects the points of the newly formed arcs corresponding to the former intersection points; see Figure 2.

Next, we consider two locally snug normal surfaces $F$ and $G$. A patch is a component of $F \cup G \setminus F \cap G$. Regular switches reconnect the patches, and trace curves are the seams between patches after performing regular switches along all intersection curves; see Figures 2 and 5.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{exchange.png}
\caption{The exchange between surfaces along an intersection curve.}
\end{figure}

If the intersection curve $\alpha$ is an arc, then, after performing a regular switch, we can identify an exchange rectangle, a rectangle whose top and bottom, say, are bounded by trace arcs and whose left and right sides are exchange arcs lying in $\partial X$.

If $\alpha$ is a loop, then our assumption that $X$ is orientable means that a regular neighborhood of $\alpha$ is a solid torus, not a solid Klein bottle. Again, since $X$ is orientable, $\alpha$ is either orientation preserving in both $F$ and $G$, or, orientation reversing in both $F$ and $G$. In the former case, there is an exchange annulus, a zero-weight annulus bounded by the trace curves and with core $\alpha$. In the latter case, there is a single trace curve which bounds an exchange Möbius band (we will be able to exclude this case in our proofs, though).

As observed in [Hat82] and, in the context of normal surfaces, in [JS03], every intersection arc between surfaces connects intersection points of the boundary curves that have opposite normal sign:

\begin{lemma}
\label{lem:intersections}
Let $A = B + C$ be a normal sum of surfaces in an orientable manifold $X$ with an induced orientation on $\partial X$. Then every arc in $B \cap C$ joins a pair of points in $\partial B \cap \partial C$ with opposite normal sign.
\end{lemma}

\section{Complexity and tight curves}

In this section, we consider properly properly embedded curves in a triangulated surface. We assume that they are transverse to the 1-skeleton but, a priori, they are not assumed to be normal.

Fix, once and for all, an ordering of all normal arc types of the triangulated surface. For this purpose we do not take into account any marking present. As in the previous section, a normal curve $\alpha$ determines a vector $\vec{v}(\alpha)$ which records the number of normal arcs of the indexed type. Order these normal vectors lexicographically.
Recall that the length of a properly embedded curve $\alpha$ is the number of intersections with the 1-skeleton, $\ell(\alpha) = |\alpha \cap T^1|$. We say that a curve is least length if it minimizes length over all curves to which it is isotopic.

**Lemma 6.1.** A least length essential curve is normal.

**Proof.** A loop in face demonstrates that the curve is not essential and any abnormal arc is either inessential or yields an isotopy reducing the length.

If $\alpha$ is a normal curve, then we define its complexity to be the pair consisting of its length and its normal vector, 
$$cpx(\alpha) := (\ell(\alpha), \vec{v}(\alpha)).$$

We reiterate that we do not take into account any marking $M$ in the definition of complexity. If $\alpha$ is not normal, we define its complexity to be $cpx(\alpha) = (\ell(\alpha), \vec{0})$. Complexities will also be ordered lexicographically.

**Definition 6.2.** A curve $\alpha$ is tight if it minimizes the complexity $cpx(\alpha)$ over all curves to which it is isotopic.

The interior of a connected inessential curve can be made disjoint from the 1-skeleton, so a tight inessential loop has $cpx = (0, \vec{0})$ and a tight inessential arc has $cpx = (2, \vec{0})$.

**Lemma 6.3.** A tight essential curve is normal and unique up to normal isotopy.

**Proof.** Indeed, the complexities of two normal curves are equal if and only if their normal vectors are identical.

**Lemma 6.4.** Let $\alpha \bowtie \beta$ be a Haken sum of locally snug properly embedded curves. Then $cpx(\alpha \bowtie \beta) \leq cpx(\alpha) + cpx(\beta)$ with equality holding if and only if $\alpha, \beta$ and $\alpha \bowtie \beta$ are all normal, or, all not normal.

**Proof.** The curve $\alpha \bowtie \beta$ is constructed by performing an exchange at every intersection point of $\alpha \cup \beta$. This is done away from the 1-skeleton, so we have $\ell(\alpha \bowtie \beta) = \ell(\alpha) + \ell(\beta)$. Thus any difference in complexity is determined solely by the normal vectors of the curves. If $\alpha \bowtie \beta$ is normal, then by the previous two lemmas $\vec{v}(\alpha + \beta) = \vec{v}(\alpha) + \vec{v}(\beta)$ and equality holds. If $\alpha \bowtie \beta$ is not normal, then its normal vector is $\vec{0}$. Then complexity is additive when both $\alpha$ or $\beta$ are not normal, and sub-additive otherwise.

If a tight curve is written as a sum, then the exchange arcs are essential in the complement of the curve.

**Lemma 6.5.** Suppose that a tight normal curve is written as a sum $\alpha + \beta$ of two normal curves. Then no exchange arc co-bounds a disk with a sub-arc of the curve.

**Proof.** Perform an irregular exchange only at the intersection corresponding to this exchange arc. The new curve is a Haken sum $\alpha \tilde{+} \beta$ with one component a trivial loop, the rest isotopic to $\alpha + \beta$, and the same total length. It follows that the trivial loop has zero length otherwise $\alpha + \beta$ would not be tight. Therefore $\alpha + \beta$ and the second component of $\alpha \tilde{+} \beta$ are normally isotopic. However, by Lemma 5.5 there is an abnormal arc in the second component of $\alpha \tilde{+} \beta$, a contradiction.

22
Lemma 6.6. Let $B$ be a bigon or half-bigon bounded by a pair of locally snug normal curves $\alpha$ and $\beta$; see Figure 3. Let $\alpha'$ and $\beta'$ be the pair of isotopic curves obtained by corner exchange(s) that trade the sides of $B$. Then one of the following holds:

1. $\cpx(\alpha') = \cpx(\alpha), \cpx(\beta') = \cpx(\beta')$, thus $\alpha'$ and $\beta'$ are normally isotopic to $\alpha$ and $\beta$, respectively, and $|\alpha' \cap \beta'| < |\alpha \cap \beta|$;
2. $\cpx(\alpha') < \cpx(\alpha)$;
3. $\cpx(\beta') < \cpx(\beta)$.

Proof. Let $\alpha, \alpha', \beta$ and $\beta'$ be as indicated in Figure 3. Note that the exchange doesn’t add or remove intersections with the 1-skeleton, and so the total length is unchanged. If the traded arcs differ in length then one curve increases and the other decreases in length, hence in complexity. In this case, either (2) or (3) holds. So we continue assuming $\ell(\alpha) = \ell(\alpha')$ and $\ell(\beta) = \ell(\beta')$.

If any exchange is irregular, then one of the curves, say $\alpha'$, is not normal. Then its complexity $\cpx(\alpha') = \ell(\alpha') + \bar{\ell}(\alpha') < \ell(\alpha) + \bar{\ell}(\alpha) = \cpx(\alpha)$ has decreased, yielding conclusion (2). Conclusion (3) results when $\beta'$ is not normal.

We are left in the case that the exchange trades length fairly and $\alpha'$ and $\beta'$ are both normal. Because length and normal vectors are both additive with respect to normal addition, we have $\cpx(\alpha) + \cpx(\beta) = \cpx(\alpha + \beta) = \cpx(\alpha') + \cpx(\beta')$. If $\cpx(\alpha') = \cpx(\alpha)$, then $\cpx(\beta') = \cpx(\beta)$ and by Lemma 6.3 the trade yields normally isotopic curves, conclusion (1). Otherwise, either (2) or (3) holds. □

Lemma 6.7. Let $\alpha$ be a tight essential curve and $C$ set of pairwise snug, tight essential curves. Then, after a normal isotopy of $\alpha$, \{\alpha\} $\cup$ $C$ is pairwise snug.

Proof. Normally isotope $\alpha$ to minimize the total of all intersections with $C$. By way of contradiction, suppose some pair is not snug, that there is $\beta \in C$ for which $|\alpha \cap \beta| > i(\alpha, \beta)$. Among all such $\beta$ take one that, together with $\alpha$, determines an innermost bigon; then any other curves from $C$ meeting that bigon meet it in arcs that run straight across.

Apply Lemma 6.6. Since all curves are tight, we must have the first conclusion. But, trading across the bigon reduces intersections between $\alpha$ and $\beta$ without raising intersections of any other pair—a contradiction. □

Lemma 6.8. Suppose that a tight essential normal curve is a normal sum $\alpha + \beta$. Then $\alpha$ and $\beta$ are tight, essential, and after a normal isotopy, snug.

Proof. Normally isotope $\alpha$ and/or $\beta$ to minimize $|\alpha \cap \beta|$. This does not change their sum.

First we show that the pair is snug: If not, then some pair of sub-arcs of $\alpha$ and $\beta$ bound a bigon or half-bigon $B$. Apply Lemma 6.6. The first conclusion does not hold, so without loss of generality assume that $\cpx(\alpha') < \cpx(\alpha)$. Isotope $\alpha'$ back slightly so that $\alpha'$ and $\beta$ still overlap and form a very thin bigon. Since $\alpha \cup \beta$ and $\alpha' \cup \beta$ are isotopic as graphs, $\alpha + \beta$ is isotopic to some Haken sum $\alpha' + \beta$. But by Lemma 6.4, $\cpx(\alpha' + \beta) \leq \cpx(\alpha') + \cpx(\beta) < \cpx(\alpha) + \cpx(\beta) = \cpx(\alpha + \beta)$, a contradiction.

It follows that $\alpha$ and $\beta$ are both essential. If either possesses a component that bounds a disk, then the fact that $\alpha$ and $\beta$ are snug implies that this component misses the other curve, survives normal addition, and $\alpha + \beta$ contains an inessential component, a contradiction.
It remains to show that each summand is tight. Without loss of generality, suppose that \( \alpha \) is not tight, that there is a tight curve \( \alpha_t \) with lower complexity, \( \text{cpx}(\alpha_t) < \text{cpx}(\alpha) \), that is isotopic to \( \alpha \) but not normally so. Isotope \( \alpha_t \) to intersect \( \alpha \cup \beta \) minimally.

Then any innermost (half-) bigon in the complement of \( \alpha_t \cup \alpha \cup \beta \) is bounded by \( \alpha \) and \( \alpha_t \), since it cannot be bounded by \( \alpha \) and \( \beta \), which are snug. And because any patch of \( \beta \) is a sub-arc of \( \alpha + \beta \), any innermost (half-) bigon bounded by \( \beta \) and \( \alpha_t \) is also a (half-) bigon bounded by the tight curves \( \alpha + \beta \) and \( \alpha_t \) which, using Lemma 6.6 again, can be eliminated by a normal isotopy of \( \alpha_t \). This contradicts the minimality of the intersection between \( \alpha_t \) and \( \alpha \cup \beta \). Then, sub-curves of \( \alpha \) and \( \alpha_t \) co-bound a product region \( R \) as in Figure 6.

If they are not snug, \( R \) is a bigon or half-bigon. If they are snug, \( R \) is a rectangle when \( \alpha \) is an arc, an annulus when \( \alpha \) is a two-sided loop, and a bigon with corners identified when \( \alpha \) is a one-sided loop.

In all of these cases, as observed above, no arc of \( \beta \) forms a (half-) bigon inside \( R \), and must therefore run across \( R \) and have an endpoint in both \( \alpha \) and \( \alpha_t \).

In the non-snug case, let \( \alpha' \) be the curve of less complexity obtained by routing \( \alpha \) along \( \alpha_t \) when it meets the bigon or half-bigon. In the snug case, let \( \alpha' = \alpha_t \). In either case, \( \text{cpx}(\alpha') < \text{cpx}(\alpha) \). Moreover, the complex \( \alpha' \cup \beta \) is isotopic to \( \alpha \cup \beta \) and because they are isotopic, there are exchanges, not necessarily regular, so that the Haken sum \( \alpha' \mathbin{\tilde{+}} \beta \) is a curve isotopic to \( \alpha + \beta \). But by Lemma 6.4, \( \text{cpx}(\alpha' \mathbin{\tilde{+}} \beta) \leq \text{cpx}(\alpha') + \text{cpx}(\beta) < \text{cpx}(\alpha) + \text{cpx}(\beta) = \text{cpx}(\alpha + \beta) \). This contradicts the fact that \( \alpha + \beta \) is tight.

Rails and fences. Now we again consider a triangulation with a marking \( M \), and auxiliary curves in it that, unlike \( M \)-normal curves, go through the points of \( M \).

A rail is a normal arc with its endpoints in \( M \), and a fence is a normal curve that is the union of rails.

We note that if a face contains an \( M \)-normal arc \( \alpha \) and a rail \( \mu \) that are locally snug, then \(|\alpha \cap \mu|\) is either 0 or 1, depending only on the endpoints of \( \mu \) and the \( M \)-normal type of \( \alpha \).

The following lemma can also be considered obvious:

**Lemma 6.9.** Intersection number with fences is additive with respect to normal addition of \( M \)-normal curves: If \( \mu \) is an fence and \( \alpha \) and \( \beta \) are \( M \)-compatible, \( M \)-normal curves, then \(|(\alpha + \beta) \cap \mu| = |\alpha \cap \mu| + |\beta \cap \mu|\).

**Proposition 6.10.** Let \( \mu \) be a fence that is a tight essential curve (w.r.t. the unmarked triangulation). Suppose that a sum \( \alpha + \beta \) of \( M \)-normal curves is tight, essential and snug with \( \mu \). Then

1. \( \alpha \) and \( \beta \) are both snug with respect to \( \mu \);
2. \( i(\alpha + \beta, \mu) = i(\alpha, \mu) + i(\beta, \mu) \) where \( i(\ldots) \) is the geometric intersection number;
(3) If $\beta$ is two-sided, connected and normally isotopic to $\mu$ then, after a normal isotopy, every point of $\alpha \cap \beta$ has the same normal sign.

Proof. Among counterexamples to conclusion (1) of the proposition, choose one that minimizes $|\alpha \cap \beta|$. Then $\alpha$, $\beta$ and $\mu$ are pairwise locally snug, and we will show that they are in fact pairwise snug. Suppose not and let $B$ be an innermost (half-) bigon bounded by some pair of the curves.

If $B$ is bounded by $\mu$ and either of the other curves, say $\alpha$, then every sub-arc of $\beta$ in $B$ crosses $B$ and meets both $\alpha$ and $\mu$. Let $\alpha'$ be the result of rerouting $\alpha$ around $B$ as in Lemma 6.6. Then $\alpha + \beta$ is isotopic to some Haken sum $\alpha' + \beta$ that has fewer intersections with $\mu$. This contradicts our assumption that $\alpha + \beta$ and $\mu$ are snug.

If $B$ is bounded by $\alpha$ and $\beta$, then every sub-arc of $\mu$ in $B$ crosses $B$ and meets both $\alpha$ and $\beta$. Let $\alpha'$ and $\beta'$ be the curves given by Lemma 6.6. Because $\alpha + \beta$ is tight, $\alpha$ and $\beta$ are locally snug. The isotopy doesn't create intersections, and so $\alpha'$ and $\beta'$ are also locally snug. Then $\alpha + \beta = \alpha' + \beta'$ for some generalized Haken sum of $\alpha'$ and $\beta'$. By Lemma 5.5 that sum is a normal sum, $\alpha + \beta = \alpha' + \beta'$.

Note that $\alpha'$ and $\beta'$ are snug with $\mu$ if and only if $\alpha$ and $\beta$ are, as the move did not change the number of times they meet $\mu$. Since $|\alpha \cap \beta| > |\alpha' \cap \beta'|$ we obtain a contradiction and establish conclusion (1).

Since $\alpha$, $\beta$ and their sum are all snug with respect to $\mu$ and intersections with respect to $\mu$ are additive, we have additivity of geometric intersection number, conclusion (2).

We now prove the final statement of the proposition. Assume that $\beta$ is normally isotopic to $\mu$. Then $i(\beta, \mu) = 0$ since $\beta$ is two-sided. By (2) and the fact that $\alpha + \beta$ and $\alpha$ are both snug with $\mu$ we have: $|(\alpha + \beta) \cap \mu| = i(\alpha + \beta, \mu) = i(\alpha, \mu) = |\alpha \cap \mu|$.

Figure 7: Mixed normal signs imply that $i(\alpha + \beta, \beta) < i(\alpha, \beta)$.

And since $\beta$ is normally isotopic to $\mu$, there is a normal, not necessarily $M$-normal, isotopy taking $\alpha + \beta$ to $\alpha + \mu$. Now $\alpha$ cuts across a thin regular neighborhood of $\mu$ in a collection of arcs that span (cut across) the neighborhood. Together they cut this neighborhood into rectangles; see Figure 7.

Each regular exchange connects a pair of rectangles at a corner of each. In fact, every rectangle that meets $\alpha \cap \mu$ twice must be attached to another rectangle at one of its corners.
Otherwise, an arc of $\alpha + \mu$ bounds the unattached rectangle, showing that the arc it is trivial in the neighborhood of $\mu$ and can be isotoped out of it. This would imply $i(\alpha + \beta, \mu) < |\alpha \cap \mu|$, contradicting the equality shown earlier. The fact that each rectangle is attached at exactly one corner implies that as we follow $\mu$ every intersection with $\alpha$ must have the same normal sign. Since $\beta$ is normally isotopic to $\mu$ we have our desired conclusion (3).

\section{Normal summands of incompressible annuli}

We would like to apply two well known results from normal surface theory: (1) an essential surface is isotopic to a normal surface, and (2) every summand of a least weight essential normal surface is also least weight and essential (Theorem 6.5 of Jaco and Tollefson [JT95]). But, as will be seen shortly, our notion of surface complexity prioritizes the reduction of boundary complexity over the reduction of total surface weight. Thus the results (1) and (2) cannot be applied as stated.

Proposition 7.1 recovers the first result using our notion of complexity. Proposition 7.2 gives a weaker version of the second for incompressible annuli. While we expect the full version to hold with our notion of complexity, we prove a restricted version both to simplify the proof and to incorporate boundary parallel annuli which are non-essential. Our proof follows the strategy of [JO84] and [JT95].

The complexity $\text{cpx}(F)$ of a properly embedded surface $F$ is the triple

$$\text{cpx}(F) = (\text{cpx}(\partial F), |F \cap T^1|, |F \cap T^2|) = ((\ell(\partial F), \bar{v}(\partial F)), |F \cap T^1|, |F \cap T^2|).$$

We compare complexities lexicographically. Thus, the complexity of $F$ is measured first by the complexity of its boundary, then by the weight of $F$, $\text{wt}(F) = |F \cap T^1|$, and then by the number of components of the intersections with the 2-skeleton of $T$.

A normal surface is \textit{least complexity} if it minimizes complexity among normal surfaces to which it is isotopic (but not necessarily normally isotopic).

A surface is \textit{tight} if it minimizes complexity, ranging over \textit{all} those surfaces to which it is isotopic.

A tight normal surface is clearly least complexity, and as a consequence of Proposition 7.1, a normal essential surface of least complexity is tight. But, this does not hold in general for surfaces that are not essential: for example, a normal boundary parallel annulus may be least complexity but after tightening no longer normal.

We first recover normalization of an essential surface. We will apply this with surfaces whose boundaries are tight, hence least length.

\textbf{Proposition 7.1.} Suppose that $X$ is a triangulated, irreducible manifold with incompressible boundary. If $F \subset X$ is a tight, properly embedded, essential surface, then $F$ is normal.

\textbf{Proof.} To prove $F$ is normal we must show that it meets each tetrahedron $\Delta$ in a collection of disks whose boundaries are normal curves of length 3 or 4. We adopt the view taken in [BDTS12], showing $F$ meets each tetrahedron in pieces that are incompressible and edge incompressible.

If any component of $F \cap \Delta$ is compressible in $\Delta$, then, by an innermost disk argument, we obtain a compressing disk avoiding all other components of $F \cap \Delta$, and hence $F \cap \Delta$ is compressible inside $\Delta$.  

26
Because $F$ is essential, the boundary of any compressing disk $D$ for $F \cap \Delta$ is trivial in $F$. Because $X$ is irreducible, compressing along $D$ yields a surface $F'$ that is isotopic to $F$, but for which either $|F \cap T^1|$ or $|F \cap T^2|$ has been reduced, a contradiction. It follows that $F \cap \Delta$ is the union of disks.

An edge compressing disk for a surface in $\Delta$ is an embedded disk $E$ whose boundary $\partial E = e \cup f$, consists of two arcs, $e \subset T^1$ and $f = E \cap F = \partial E \cap F$; see [BDTS12].

If some component of $F \cap \Delta$ has an edge compressing disk then, by an innermost disk argument, there is an edge compressing disk $E$ for $F \cap \Delta$. If $e \subset \partial X$ then, because $F$ is not boundary compressible, $f$ is trivial in $F$. But compressing along $E$ yields an isotopic surface $F_0$ ($X$ is irreducible and has incompressible boundary) whose boundary length is reduced by at least two, contradicting the fact that $\partial F = \partial F_0$ is least length. And if $e$ lies in an interior edge, then $E$ can be used to guide an isotopy reducing $|F \cap T^1|$, also a contradiction.

Then $F$ meets each face in normal arcs. For otherwise, there is an arc whose ends both lie in the same edge, and an outermost such arc bounds an edge compressing disk. Then $F$ meets the boundary of each tetrahedron in normal curves. And it is well known, see Thompson [Tho94], that if any such curve has length greater than 4 we see an edge compressing disk for $F$ in the boundary of the tetrahedron.

0-efficient triangulations. First we recall the definition of 0-efficient triangulations from [JR03]. A triangulation of a manifold $X$ with nonempty boundary is 0-efficient if every normal disk is vertex-linking. (A normal disk is vertex-linking at vertex $v$ if it consists of precisely one normal triangle from each tetrahedral corner meeting $v$.)

Moreover, if no boundary component of $X$ is an $S^2$, then $X$ does not contain any normal 2-spheres [JR03, Prop. 5.15]. In our setting, we use 0-efficient triangulations only in the situations without $S^2$ boundary components (since in the algorithm, we fill each such component with a ball). Note also that in the proposition below we can assume that $X$ does not contain $S^2$ boundary components even if do not explicitly claim that $X$ is obtained in an intermediate stage of the algorithm. Indeed, we assume that $X$ is irreducible. Then an $S^2$ boundary component implies that $X$ is a ball; however, the proposition also assumes that $X$ contains an essential annulus or Möbius band.

We now establish the second result, that some summand of a non-fundamental incompressible annulus is an essential annulus. This applies to boundary compressible as well as essential annuli.

**Proposition 7.2.** Let $X$ be a triangulated, orientable, irreducible manifold with incompressible boundary and a 0-efficient triangulation. Let $A$ be an incompressible annulus or Möbius band that has tight boundary and is least complexity and normal. Suppose that $A$ can be written as a non-trivial sum $A = B + C$ where $B$ is connected and $\partial B \neq \emptyset$. Then $B$ is an essential annulus or Möbius band with tight boundary.

7.1 Proof of Proposition 7.2

**Sketch of the proof.** Our proof is loosely modeled on Jaco and Tollefson’s proof of [JT95, Th. 6.5]. Apart from using slightly different notion of complexity, we also have to add additional ingredients when $A$ is a boundary parallel annulus.

As we will see, the core of the proof is to show that $B$ is essential. For contradiction we assume that $B$ is not essential. The first important step is to find out what are the possible
patches when \( A \) is decomposed by trace curves from the normal sum \( A = B + C \); see Figure 8 left. If \( A \) is essential (annulus or Möbius band), then disk patches as well as half-disk patches can be ruled out following [JT95] (disk patches avoid \( \partial X \) whereas half-disk patches contain a single arc on \( \partial X \)); see Lemmas 7.8 and 7.9. After ruling out such patches we can deduce that every intersection curve is essential in \( B \), that is a spanning arc or a core curve. This already mean that \( C \) intersects \( B \) in a very specific way and both cases can be ruled out along [JT95]; see Lemma 7.11.

If \( A \) is not essential, then \( A \) is a boundary parallel annulus by Proposition 4.2. In this case we do not know how to rule out disk patches but we still can rule out half-disk patches (Lemma 7.9); here we use that simplification of the boundary has higher priority than simplification of the interior in our notion of complexity. Since \( A \) is boundary parallel, there is an annulus \( A_{2X} \) to which \( A \) is parallel and together they bound a solid torus \( T \) in \( X \). Because there are no half-disk patches, we can show that one of the exchange rectangles for the sum \( A = B + C \) is inside this torus and it meets \( A \) and \( A_{2X} \) only in essential arcs. However, with such a rectangle \( A \) cannot be boundary parallel; see Lemma 7.10 for details. This finishes the sketch of the proof and now we provide the details.

Because \( A \) is incompressible, \( \partial A \) is essential by Proposition 4.2. Without loss of generality, we will assume that the sum \( A = B + C \) lexicographically minimizes \((|\partial B \cap \partial C|, |B \cap C|)\), the number of boundary intersections and the total number of intersection curves, over pairs \((B', C')\) where \( B' \) and \( C' \) are locally snug surfaces isotopic to \( B \) and \( C \), respectively. Since \( \partial A (= \partial B + \partial C) \) is assumed tight, we have, by Lemma 6.8, that \( \partial B \) and \( \partial C \) are tight, and because \( |\partial B \cap \partial C| \) is minimized, snug.

**Lemma 7.3.** Either the conclusion of Proposition 7.2 holds, or \( B \) is a boundary parallel annulus and every component of \( C \) is an incompressible annulus, Möbius band, torus or Klein bottle.

**Proof.** No component of \( C \) has Euler characteristic \( \chi > 0 \): Because \( X \) is 0-efficient, no normal surface is a sphere, nor a projective plane, for then its normal double would be a normal sphere. And, also by 0-efficiency, any disk has boundary a trivial vertex linking curve that survives normal addition, and is present in \( \partial A \)—a contradiction.
Then every component has $\chi = 0$ and it is an annulus, Möbius band, torus, or Klein bottle. No component is a compressible annulus since these have a trivial boundary component (Proposition 4.2) and this contradicts the fact that both summands have essential boundary.

Since $B$ is connected and $\partial B \neq \emptyset$, $B$ is either an annulus or Möbius band. Also by Proposition 4.2, a Möbius band is essential and satisfies the conclusion of Proposition 7.2. So does an annulus, unless it is boundary compressible, and hence boundary parallel, by Proposition 4.2.

We proceed with the proof of Proposition 7.2 under the assumption that $B$ is a boundary parallel annulus.

When $A$ is formed as the normal sum $B + C$, it is partitioned into patches coming from $B$ and $C$, as was discussed in Section 5.1, and we have exchange surfaces attached to the curves separating the patches; see Figure 8 left.

It follows that no exchange surface is a Möbius band. As noted in Section 5.1, this occurs only when an intersection loop is one-sided in both summands.

Define a half disk to be a disk that is halfway properly embedded in $X$, that is, an embedded disk whose boundary meets $\partial X$ in a single arc. Note that a boundary compressing disk for a surface is a half disk whose boundary meets the surface in the complementary arc, but the reverse does not hold in general, for the arc may not be essential in the surface.

An exchange rectangle or annulus $E$ meets four patches of $A$. A pair $D,D'$ of these patches are said to be adjacent across $E$ if they meet opposite boundary curves $\sigma$ and $\tau$ of $E$, but from the same side (we again refer to Figure 8 left).

**Lemma 7.4.** $\sigma$ bounds (half) disk in $A$ if and only if $\tau$ both bounds a (half) disk in $A$.

**Proof.** We prove that if $\sigma$ bounds a (half) disk in $A$, then $\tau$ bounds a (half) disk in $A$. The reverse implication is proved by interchanging $B$ and $C$ and remarking that in this proof we do not use the extra assumptions on $B$.

The surface $A$ is either essential or a boundary parallel annulus, and it is incompressible by the assumptions.

Suppose that $\sigma$ bounds a disk $D$ in $A$. Then $E \cup D$ is a disk which, after a slight isotopy, meets $A$ only in $\tau$. Since $A$ is incompressible, $\tau$ bounds a disk in $A$ as claimed.

The same argument works when $A$ is essential and, say, $\sigma$ bounds a half disk $H$. Then $E \cup H$ is a half disk which, after a slight isotopy, meets $A$ only in $\tau$. Since $A$ is not boundary compressible, $\tau$ bounds a half disk in $A$.

We conclude by showing that $\tau$ bounds a half disk when $\sigma$ bounds a half disk and $A$ is a boundary parallel annulus (assumed to have tight boundary). To obtain a contradiction, suppose that $\sigma$ bounds a half disk $H$ but $\tau$ is an essential arc in $A$. Then $E \cup H$ is, after a slight isotopy, a boundary compressing disk meeting $A$ in the arc $\tau$. Since $A$ is parallel to an annulus $A_{\partial X} \subset \partial X$, their union bounds a solid torus in $X$. The rectangle $E$ is a disk properly embedded in this solid torus.

Indeed, if $E$ is outside the solid torus, consider a boundary compressing disk $D_A$ for $A$ meeting $A$ in $\tau$ inside the solid torus. Then the disk $D_A \cup E \cup H$ meets the core curve of $A_{\partial X}$ exactly once, implying that it is a non-separating disk and therefore a compressing disk for $\partial X$—a contradiction.

As soon as we know that $E$ is inside the solid torus, we have that the boundary of $E$ meets $A_{\partial X}$ in a pair of exchange arcs that each span $A_{\partial X}$ by Lemma 6.5, and meets the annulus $A$ in one curve $\sigma$ that is trivial in $A$ and the other $\tau$ that is a spanning arc for $A$. Therefore,
when restricting to $A_{\partial X}$, we get that two corners of $E$ are in one component of $\partial A$ and the other two in the second one. When restricting to $A$, we get that three corners are in the same component and the other corner in the second—a contradiction.

Lemma 7.5. Suppose $A$ is essential. If $\sigma$ and $\tau$ bound disks in $A$, then they are adjacent across $E$.

Proof. Suppose that $\sigma$ and $\tau$ bound non-adjacent disks $D$ and $D'$. They either are disjoint, or one is a sub-disk of the other, say $D' \subset D$.

When disjoint, the union $D \cup E \cup D'$ is a sphere that, after a slight isotopy, separates components of $A$ (since spheres separate in irreducible manifolds)—a contradiction. See Figure 8 right.

Suppose then that $D' \subset D$; see Figure 9. Let $A'$ be the surface obtained from $A$ by removing $D$ and replacing it with $E \cup D'$.

The union of the disk $D$ and a slight offset of the disk $E \cup D'$ bounds a ball, across which the disks are isotopic ($X$ is irreducible). So $A'$, the result of this disk swap, is isotopic to $A$. Also note that performing an irregular exchange (fold) at this intersection loop produces a surface with two components: one is $A'$, and the other, $A''$, is a torus obtained by identifying the ends of the annulus $D \setminus D'$.

Because $E$ has zero weight, we have $\text{wt}(A) = \text{wt}(A') + \text{wt}(A'')$. But, because this was not a regular exchange, $A' \cup A''$ is not normal, and there is an abnormal arc bounding a half disk in some face by Lemma 5.5. If this half disk meets $\partial X$, then $A' \cup A''$ either is boundary compressible or is not least length, both contradictions.

Thus, the half disk lies in the interior and can be used to guide an isotopy of $A' \cup A''$ that removes two intersections with the 1-skeleton. But this implies that $A' \cup A''$ can be isotoped to have strictly less weight than $A$. This is a contradiction since the component $A'$ has lower complexity, but is isotopic to the tight surface $A$.

Unfortunately, the above proof contradicts minimal interior weight and does not apply when $A$ is a (non-essential) boundary parallel annulus, which may not be normal when tight. Fortunately, the half disk version contradicts tight boundary and can be applied when $A$ is essential or a boundary parallel annulus.

Lemma 7.6. If $\sigma$ and $\tau$ both bound half disks in $A$, then they are adjacent.
Proof. Suppose to the contrary that \( \sigma \) and \( \tau \) bound half disks \( H \) and \( H' \) that are not adjacent across \( E \). The half disks \( H \) and \( H' \) are either disjoint, or, say, \( H' \subset H \).

If disjoint, then the union \( H \cup E \cup H' \) is a properly embedded disk, that after a slight isotopy, separates components of \( A \), which is a contradiction.

Now suppose that \( H' \subset H \); see Figure 9. Replacing \( H \) with \( E \cup H' \) is a disk swap across a ball that produces a surface \( A' \) isotopic to \( A \). But notice that performing an irregular rather than regular switch at this intersection curve produces a surface with two components: one is \( A' \), and the other is an annulus \( A'' \) formed by identifying the ends of the rectangle \( H \setminus H' \). The irregular switch on the intersection arc yields irregular switches at the endpoints which are intersections of the boundary curves.

So while \( \ell(\partial A) = \ell(\partial A') + \ell(\partial A'') \), the curve \( \partial A' \cup \partial A'' \) is not normal, contains an abnormal arc by Lemma 5.5, and so there is an isotopy reducing its length. Since \( \partial A' \cup \partial A'' \) is isotopic to a curve of length strictly lower than \( \partial A \), each of its components has length strictly lower than \( \partial A \), contradicting the minimality of the length of \( \partial A \).

Figure 10: Adjacent disks; here \( D' \subset D \). The union \( D' \cup E \cup D \) bounds a ball after slight isotopy.

In the next two lemmas we will utilize the following observation:

**Observation 7.7.** Suppose that \( D \) and \( D' \) are disks that are adjacent across the exchange annulus \( E \). If the disks are disjoint, then \( D \cup E \cup D' \) is, possibly after a slight isotopy, a sphere bounding a ball.

**Proof.** If \( D \) and \( D' \) are disjoint, we get the observation immediately, since \( X \) is irreducible. If not, say if \( D' \subset D \), then fix \( D \) and slightly isotope the interior of the disk \( E \cup D' \) off \( D \) to the side of the exchange annulus; see Figure 10. After the isotopy \( D \cup E \cup D' \) is a sphere bounding a ball.

Similarly, if \( H \) and \( H' \) are half disks adjacent across the exchange rectangle \( E \), then the union \( H \cup E \cup H' \) together with \( \partial X \) bound a ball, possibly after a slight isotopy. Indeed, the union \( H \cup E \cup H' \) is a properly embedded disk, after first perhaps slightly isotoping, say \( E \cup H' \), when \( H \) and \( H' \) are not disjoint. Its boundary \( \partial H \cap \partial E \cap \partial H' \), perhaps slightly isotoped, bounds a disk in \( \partial X \) and together these disks bound a ball in \( X \).

We also describe how a surface \( A \) obtained as a normal sum \( A = B + C \) can be obtained, under certain additional conditions, as a normal sum \( A = B' + C' \) guided by some of the normal exchanges of \( A = B + C \). Let \( \mathcal{E} \) be the set of exchange bands for the normal sum.
Figure 11: Summing $B$ and $C$ (upper left picture) yields $A$ (upper right picture) with the corresponding set of exchange bands/arcs $E = \{E_1, E_2, E_3, E_4\}$. Some subset $E' \subset E$ may not be consistent (two lower left pictures); however, if it is consistent (two lower right pictures), then summing $B'$ and $C'$ yields $A$ again.

$A = B + C$. Note that $|E| = |B \cap C|$. We say that a subset $E' \subset E$ is a consistent subset if the induced patches, components of $A \setminus E'$, can be bicolored so that two patches have different colors if they either lie on opposite sides of the same trace curve, or, are adjacent across an exchange band. (If this happens for two sides of a same patch, then in particular it cannot be bicolored.) In this case, we can see that $E'$ is the set of exchange bands for a normal sum $A = B' + C'$ where $B'$ and $C'$ are each the union of patches of a single color, connected across $E'$. See Figure 11. The same analysis holds for subsets of exchange arcs for a curve sum $\alpha = \beta + \gamma$.

**Lemma 7.8.** If $A$ is essential, then there are no disk patches. Every intersection loop is essential in $A$, $B$, and $C$.

This lemma does not apply to the boundary parallel case, because we cannot assume that disks are adjacent.

**Proof.** This argument appears in [JO84] and [JT95]. If there is a disk patch bounded by a trace curve, then by Lemmas 7.4 and 7.5, it is adjacent across an exchange annulus $E$ to another disk. The adjacent disk is not a single patch, since if it were, $E \setminus E$ would be a consistent subset, and we could express $A$ as a sum $A = B' + C'$, where $B'$ and $C'$ are isotopic to $B$ and $C$, respectively, but $|B' \cap C'| < |B \cap C|$, a contradiction. Thus, the adjacent disk contains trace loops. Pass to an innermost trace loop bounding a disk patch and repeat.

This process can be continued arbitrarily, and therefore must eventually repeat. Thus there is a shortest cycle of adjacent disks, each biting the tail patch of its predecessor across some exchange annulus; see Figure 12. Note that the cycle has length 1 when $D' \subset D$ for...
some pair of adjacent disks $D$ and $D'$. By irreducibility, the union of a pair of adjacent disks and their exchange annulus is a sphere bounding a ball, and the union of these balls is a solid torus.

Let $E'$ be the subset of $E$ along which the cycle of disks is adjacent. Then $E'$ is seen to be a consistent subset of $E$ by coloring all annulus patches on the boundary of the solid torus with one color, and all other patches with the other. This expresses $A$ as the sum of surfaces $A = B' + C'$ where $B'$ is a normal torus bounding a solid torus.

Adding $B'$ to $C'$ corresponds to a fractional Dehn twist in $B'$. Since $B'$ bounds a solid torus, there is an isotopy of the solid torus that undoes the twist and carries $C'$ to $A$. So $C'$ is isotopic to $A$ but with strictly lower complexity $\text{cpx}(C') = \text{cpx}(A) - \text{cpx}(B')$—a contradiction.

We repeat the above argument to work in the context of trace arcs and half disks. This argument does apply to boundary parallel annuli, as their half disks are known to be adjacent. The proof will reach a contradiction to the tightness of $\partial A$.

**Lemma 7.9.** No trace arc bounds a half disk in $A$.

**Proof.** When $A$ is a boundary parallel annulus, Lemma 7.8 does not apply, and trivial trace loops may be present. So we will refer to outermost (in $A$) half disks instead of half disk patches. An outermost half disk in $A$ does not contain trace arcs, but, is not a patch if it contains trace loops. If there is an outermost half disk $H$ bounded by a trace arc, then by Lemmas 7.4 and 7.6, it is adjacent across an exchange rectangle $E$ to another half disk $H'$.

The adjacent half disk $H'$ is not outermost: If $H$ and $H'$ are both outermost, then $H \cup E \cup H'$ is a properly embedded disk that meets $\partial X$ in a disk bounding the curve $h \cup e \cup h' \cup e'$, where $h = H \cap \partial X$, $h' = H' \cap \partial X$ and $e$ and $e'$ are exchange arcs, the two components of $E \cap \partial X$. Since $h$ and $h'$ are sub-arcs of $\partial B$ and $\partial C$, we can see, by undoing the exchanges $e$ and $e'$, that $\partial B$ and $\partial C$ bound a bigon and are thus not snug, a contradiction.

Thus, the adjacent half disk is not outermost and contains at least one trace arc. Pass to an innermost trace loop bounding a outermost half disk and repeat. This process can be continued arbitrarily, and therefore must eventually repeat. Thus, there is a shortest cycle of adjacent half disks, each biting the tail patch of its predecessor across some exchange

![Figure 12: A cycle of adjacent half disks. Double the figure for a cycle of disks.](image)
rectangle; see Figure 12. Note that the cycle has length 1 when \( H' \subset H \) for some pair of adjacent half disks \( H \) and \( H' \). By irreducibility and incompressibility of \( \partial X \), the union of a pair of adjacent half disks and their exchange rectangle is a disk that co-bounds a ball with a disk in \( \partial X \), and the union of all these balls is a solid torus meeting \( \partial X \) in an annulus.

Let \( \mathcal{E}' \) be the subset of \( \mathcal{E} \) along which the cycle of half disks is adjacent. Then \( \mathcal{E}' \) is seen to be a consistent subset of \( \mathcal{E} \) by coloring all rectangle patches on the boundary of the solid torus with one color, and all other patches with the other. This expresses \( A \) as the sum of surfaces \( A = B' + C' \) where \( B' \) is a normal boundary parallel annulus.

Adding \( B' \) to \( C' \) corresponds to a fractional Dehn twist in \( B' \). Since \( B' \) is boundary parallel, there is an isotopy of the solid torus that undoes the twist and carries \( C' \) to \( A \). So \( C' \) is isotopic to \( A \) but with shorter length \( \ell(C') = \ell(A) - \ell(B') \), a contradiction. □

We know that \( A \) is incompressible, and thus by Proposition 4.2 it is either an essential annulus or Möbius band, or a boundary parallel annulus. We deduce that in both cases it means that \( B \) is essential.

**Lemma 7.10.** If \( A \) is a boundary parallel annulus, then \( B \) is essential (not a boundary parallel annulus).

*Proof.* Suppose that \( B \) is a boundary parallel annulus. If \( \partial C = \emptyset \) then \( \partial B = \partial A \), and hence \( A \) and \( B \) are isotopic, contradicting the fact that \( A \) was chosen to have least complexity.

So we have \( A = B + C \), where all three have non-empty tight boundary and \( A \) and \( B \) are both boundary parallel annuli. Then \( \partial A = 2a \) and \( \partial B = 2b \), where \( a \) and \( b \) are tight essential curves since \( A \) and \( B \) have tight boundaries. Since all normal coordinates of \( \partial A \) and \( \partial B \) are even, it follows that \( \partial C = 2c \) for some tight essential curve \( c \) by Lemma 6.8.

Each pair, \( 2b \) and \( 2c \), of parallel curves bounds an annulus in \( \partial X \). Normally isotope \( B \) and \( C \) so that these annuli are very thin and intersect in a collection of squares, each contained in a face of \( \partial X \). Pick a particular square. Each of its corners is an endpoint of an intersection arc between \( B \) and \( C \). All corners have the same normal sign. It follows that the exchange rectangles for corners on the same edge lie on opposite sides of \( A \); see Figure 13.

Thus, there is an exchange rectangle \( R \) properly embedded inside the solid torus \( T \) that is bounded by \( A \) and \( A_{\partial X} \subset \partial X \), the annulus into which \( A \) is isotopic. Then \( \partial R \) is a compressing disk for \( T \) since it meets \( A \) in a pair of arcs that are essential by Lemma 7.9 and meets \( A_{\partial X} \) in a pair of arcs that are essential by Lemma 6.5. But this contradicts the fact that \( A \) is boundary parallel to \( A_{\partial X} \). The unique, up to isotopy, compressing disk for \( T \) meets \( A \) and \( A_{\partial X} \) each in a single essential arc. □

**Lemma 7.11.** If \( A \) is essential then \( B \) is essential (not a boundary parallel annulus).

*Proof.* We first note, as in [JT95], that each disk patch is incompressible and boundary incompressible. Any (boundary) compressing disk for a patch has its boundary in the essential surface \( A \) and therefore meets it in a trivial curve. This, in turn, implies the existence of a (half) disk patch.

In contradiction, suppose \( B \) is a boundary parallel annulus. By Lemma 7.8, \( B \) and \( C \) intersect in curves that are essential in both. These curves are either core loops or spanning arcs for \( B \).

But they cannot be spanning arcs: If so, then there is a boundary compressing disk \( D \) for \( B \) so that \( \partial D \cap B \cap \partial C = \emptyset \). Choose \( D \) to be such a boundary compressing disk that meets \( C \)
in the minimal number of curves. Let $D'$ be a disk bounded by an innermost loop, outermost arc, or when $C \cap D = \emptyset$, let $D = D'$. By minimality $D'$ meets a patch in an essential curve and is thus a (boundary) compressing disk for a patch. This contradicts Observation 7.7.

We proceed assuming that $B \cap C$ is a collection of loops essential in both $B$ and $C$. Then the patches of $B$ and $C$ are annuli because each component of $C$ has zero Euler characteristic. As before, choose a boundary compressing disk $D$ for $B$ that meets $C$ in the minimal number of curves.

By minimality of $D$ and essentiality of patches, no curve of intersection is a closed loop or an arc with endpoints both in $\partial X$. It follows that $\gamma$, an outermost arc of intersection, co-bounds a bigon or half bigon $D'$, with an arc $\beta \subset B \cap \partial D$. (Figure 3, with different labelling). By minimality of $|D \cap C|$, $\beta$ is a spanning (essential) arc for an annulus patch $A_B \subset B$ and $\gamma$ is a spanning arc for a patch $A_C \subset C$. The patches meet in either one (half bigon) or two (bigon) intersection loops. Moreover, because $X$ is irreducible and has incompressible boundary, $A_B$ and $A_C$ are parallel across the solid torus that they bound. Thus, Figure 3 is a cross-section of the total intersection, $(\text{half})\text{bigon} \times S^1$.

What does the normal addition $A = B + C$ do with the patches $A_B$ and $A_C$? It cannot trade them as in the figure, because then, performing only the trade and no other exchanges produces normal surfaces $B'$ and $C'$ isotopic to $B$ and $C$ but with fewer intersections, contradicting our assumption that we had minimized $|B \cap C|$. Nor, in the half bigon case, can it attach $A_B$ and $A_C$. For this means $A$ has a boundary parallel annulus component. This rules out the half bigon case. And, in the bigon case, it cannot attach $A_B$ to $A_C$ along both curves, for if it did, a component of $A$ would be a torus bounding a solid torus. So $A_B$ and $A_C$ are attached along one intersection loop, but not along the other. But then $A_B \cup A_C$ is parallel to the zero weight exchange annulus for the intersection loop where they were not attached.

Form a Haken sum $A'$ by performing all regular exchanges, except perform an irregular exchange (fold) along the curve corresponding to the zero weight annulus. This is similar to the situation in Figure 9, although the context is a bit different. Then one component of $A'$ is isotopic to $A$ and the other, call it $A''$, is a torus bounding a solid torus. But $A' \cup A''$ is not normal because of the irregular switch. As in the end of the proof of Lemma 7.5, we conclude that $A' \cup A''$ is either boundary compressible, not least length, or, $A$ is not tight. All are contradictions.

This completes the proof of Proposition 7.2.
8 Constructing an annulus curve $\alpha$

As usual, we assume that $X$ is irreducible, orientable with incompressible boundary and presented via a 0-efficient triangulation.

**Definition 8.1.** An annulus curve $\alpha$ is a properly embedded (multicomponent) normal curve in $\partial X$ with the following property: There exists a collection $A$ of pairwise disjoint properly embedded essential annuli in $X$ such that $\alpha \subseteq \partial A$ and $\alpha$ represents all normal isotopy classes of boundary components of $A$ exactly once, i.e., for every annulus $A \in A$ and every component $\gamma$ of $\partial A$, there is exactly one component of $\alpha$ that is normally isotopic to $\gamma$.

The following proposition provides an annulus curve that can be used to track essential annuli. Each boundary curve of a tight essential annulus either appears in the curve or meets the curve.

**Proposition 8.2.** Let $X$ be an irreducible, orientable manifold with incompressible boundary presented via a 0-efficient triangulation with $t$ tetrahedra. Then there is a tight normal annulus curve $\alpha$ so that:

1. $\alpha$ is maximal, by which we mean that if $A \subset X$ is an essential annulus or Möbius band whose boundary is tight and disjoint from $\alpha$, then each boundary component of $\partial A$ is normally isotopic to a component of $\alpha$.

2. $|\alpha|$, the number of components of $\alpha$, is smaller than $4t$.

3. $\ell(\alpha)$ is bounded by a computable function of $t$.

The bound for $\ell(\alpha)$ we obtain from our proof is an $O(t)$-times iterated exponential, and this is currently a bottleneck of the whole algorithm.

The proposition follows in a simple way from the next two lemmas.

**Lemma 8.3.** Suppose $\alpha_0$ is a tight, normal annulus curve in the boundary of an irreducible, orientable manifold with incompressible boundary and a 0-efficient triangulation. Let $A$ be an essential annulus with tight boundary that is disjoint from $\alpha_0$, and such that a component of $\partial A$ is not normally isotopic into $\alpha_0$.

Then there is a tight normal annulus curve $\alpha$ such that $\alpha_0 \subsetneq \alpha$ and $\ell(\alpha)$ is bounded by a computable function of $t$ and $\ell(\alpha_0)$.

**Proof.** The curve $\alpha_0$ can be regarded as a fence in the marked triangulation $(T, \alpha_0 \cap T^1)$. The annulus $A$ has tight boundary disjoint from the fence $\alpha_0$. By isotoping the interior of $A$ (if necessary) while keeping its boundary $\partial A$ fixed, we may assume that $A$ is tight and hence, by Proposition 7.1, normal.

Write $A = F_1 + F_2 + \cdots + F_k$, a sum of connected fundamentals for the marked triangulation $(T, \alpha_0 \cap T^1)$. By Lemma 6.8 and by Proposition 6.10, respectively, each boundary $\partial F_i$ is tight and disjoint from the fence $\alpha_0$. Since $\ell(\partial F_i) \leq |F \cap T^1|$, by Proposition 5.2, $\ell(\partial F_i)$ is bounded by a computable function of $t$ and $\ell(\alpha_0)$, the number of marking points. By Proposition 7.2, each $F_i$ with non-empty boundary is an essential annulus or Möbius band.

Moreover, there must be a summand with a boundary component that is not normally isotopic into $\alpha_0$. Otherwise, each component of the boundary sum $\partial A = \partial F_1 + \partial F_2 + \cdots + \partial F_k$, would be normally isotopic to a component of the fence $\alpha_0$. This would imply that the
summands are pairwise disjoint after a normal isotopy and, that each component of \( \partial A \) is itself normally parallel to a component of the fence \( \alpha_0 \), contradicting our assumption.

Fix such a summand with a boundary component \( \alpha' \) not normally isotopic into \( \alpha_0 \). We will construct a new annulus curve \( \alpha \) that contains both \( \alpha_0 \) and \( \alpha' \) (which ensures \( \alpha_0 \not\subset \alpha \)).

Let \( A' \) be the summand if it is an essential annulus or twice the summand if it is a Möbius band.\(^{13}\)

Let \( A_0 \) be a collection of pairwise disjoint essential annuli witnessing that \( \alpha_0 \) is an annulus curve, i.e., \( \alpha_0 \not\subset \partial A_0 \) and \( \alpha_0 \) represents all normal isotopy classes of boundary components of \( A_0 \) exactly once. By construction of the annulus \( A' \), its boundary \( \partial A' \) is disjoint from the fence \( \alpha_0 \) and hence from \( \partial A_0 \), and one boundary component of \( A' \) is the curve \( \alpha' \).

We isotope \( A' \), leaving the boundary fixed, to minimize components of intersection \( |A' \cap A_0| \), and we distinguish two cases.

The first case is that \( A' \) misses \( A_0 \). Then \( A = A_0 \cup A' \) is a collection of pairwise disjoint properly embedded essential annuli. We define \( \alpha \) to be the annulus curve corresponding to \( A \).

More precisely, \( \alpha' \) is one of the boundary components of \( A' \). If the other boundary component is normally isotopic to \( \alpha' \) or to some component of \( \alpha_0 \), we set \( \alpha = \alpha_0 \cup \alpha' \), and otherwise, we set \( \alpha = \alpha_0 \cup \partial A' \). Thus, \( \alpha \) is an annulus curve, as witnessed by \( A \), \( \alpha_0 \not\subset \alpha \), and \( \ell(\alpha) \) is bounded by a computable function of \( t \) and \( \ell(\alpha_0) \) since \( \ell(\partial A') \) is.

The second case is that \( A' \) intersects \( A_0 \). In this case, since \( \partial A \) and \( \partial A_0 \) are disjoint, a standard innermost loop argument shows that all curves of intersection are essential, i.e., core curves in the annuli; see the proof of Lemma 4.3.

If \( A' \) meets \( A_0 \), let \( \alpha'' \subset A' \) be the core intersection curve closest to \( \alpha' \), and let \( \tilde{A} \) be the corresponding annulus in \( A_0 \). Then \( \alpha' \) and \( \alpha'' \) co-bound a sub-annulus \( A'' \subset A' \) whose interior misses \( \partial A_0 \).

Note that \( \alpha'' \) splits the annulus component \( \tilde{A} \) of \( A_0 \) into annuli \( A_1 \) and \( A_2 \). Then \( \alpha' \) is a boundary curve of both of the annuli \( A'' \cup A_1 \) and \( A'' \cup A_2 \), both of these are disjoint from \( A_0 \), and at least one of them is essential.\(^{14}\)

Suppose w.l.o.g. that \( A'' = A'' \cup A_1 \) is essential. Then \( A = A_0 \cup A'' \) is a collection of pairwise disjoint essential annuli witnessing that \( \alpha := \alpha_0 \cup \alpha' \) is an annulus curve (since one boundary component of \( A'' \) is \( \alpha' \) and the other one is a boundary component of \( \tilde{A} \in A_0 \) and hence already represented in \( \alpha_0 \)). Moreover, as in the first case, \( \ell(\alpha) \) is bounded by a computable function of \( t \) and \( \ell(\alpha_0) \) since \( \ell(\partial A') \) and hence \( \ell(\alpha') \) is.

\(\square\)

**Lemma 8.4.** Let \( \alpha \) be an essential curve embedded in a a connected closed orientable triangulated surface \( F \) with \( f \) faces, such that no pair of components of \( \alpha \) are isotopic. Then \( |\alpha| < f \).

\(^{13}\)The double of any normal surface \( F \) can be obtained by offsetting two copies, one to each side, of each normal disk of \( F \). It follows that \( 2F \) is on the boundary of \( F \times [-1, 1] \), an interval bundle over \( F \). When the manifold is orientable, the bundle is a non-twisted product if and only if the surface is orientable. When the summand \( F \) is a Möbius band, \( 2F \) is the annulus \( F \times \{ \pm 1 \} \). Since every curve in the boundary of an orientable manifold is two-sided, the boundary of the annulus is two copies of \( \alpha' \).

\(^{14}\)First note that for each of these annuli, any core curve is parallel to \( \alpha'' \), so a compression disk for either of these annuli would yield a compression disk for \( A' \) (and for \( \tilde{A} \) as well), a contradiction. Thus, if the annuli are not essential, they both must be boundary compressible. Since we assume that \( \partial X \) is incompressible, it follows that each of the two annuli is boundary parallel, and thus each is of them co-bounds a solid-torus with an annulus in the boundary and is a longitudinal annulus in its respective solid torus (longitudinal means that it meets each meridional disk of the solid torus once). Then the union of the two solid tori is a solid torus in which \( \tilde{A} = A_1 \cup A_2 \) is longitudinal, demonstrating that \( \tilde{A} \) is boundary parallel and contradicting the assumption that \( A_0 \) consists of essential annuli.
Proof. The result holds trivially for a torus, which requires at least two faces to triangulate and allows $\alpha$ to have at most one component.

Now suppose that the genus of the surface is $g \geq 2$. We claim that then $\alpha$ has at most $3g - 3$ components. We may assume that $\alpha$ is a maximal collection of non-parallel curves, and hence it decomposes $F$ into $p$ pairs of pants (spheres with 3 holes) [FM11, Sec. 8.3.1]. Each pair of pants has Euler characteristic $-1$, and because the Euler characteristic of the boundary of a pair of pants is zero, the Euler characteristic of the surface is additive over the pants, $\chi(F) = -p = 2 - 2g$. Because each curve is on the boundary of two pairs of pants, we have $|\alpha| = \frac{3}{2}p = 3g - 3$.

We have $\chi(F) = 2 - 2g = f - e + v$, where $e = \frac{3}{2}f$ is the number of edges and $v$ is the number of vertices. Then $g = \frac{f}{4} + 1 - v \leq \frac{f}{4}$ and $|\alpha| \leq 3g - 3 \leq \frac{3}{2}f < f$, as desired. \hfill \Box

We now complete the proof of the main proposition.

Proof of Proposition 8.2. The annulus curve $\alpha$ can be constructed iteratively, starting with $\alpha = \emptyset$. If, at any stage, the maximality property (1) is not satisfied, we apply Lemma 8.3 to add a distinct component to $\alpha$.

We claim that the process terminates after adding at most $4t$ components: By Lemma 8.4, each boundary component of $X$ contains fewer components of $\alpha$ than it has faces (here, we are using that $\alpha$ is tight, so by Lemma 6.3, the fact that no two components of $\alpha$ are normally isotopic also implies that no two of them are isotopic). Thus, in total, $\alpha$ has fewer components than $X$ has boundary faces and the number of boundary faces is bounded by $4t$. This is (2).

By Lemma 8.3, the first component added has length bounded by a computable function of $t$. Every subsequent component added has length bounded by a computable function of $t$ and the total length of the preceding components. Since the number of components is less than $4t$, the total length of the curve is bounded by a computable function of $t$. \hfill \Box

9 Curves bounding boundary parallel annuli

In the previous section we constructed the annulus curve $\alpha$, which will be used to bound the coefficients of essential annulus summands in the planar (almost) meridional surface $P$. In this section we construct $\Gamma$, a collection of curves bounding normal boundary parallel annuli. Later, the curves of $\Gamma$ will act as fences, and will be used to rule out boundary parallel annulus summands altogether.

Proposition 9.1. Suppose $X$ is an irreducible, orientable manifold with incompressible boundary and presented via a 0-efficient triangulation with $t$ tetrahedra. Let $\alpha$ be the tight normal annulus curve given by Proposition 8.2. Then there is a finite set $\Gamma$ of tight essential curves, possibly mutually intersecting, such that:

1. If $B$ is a normal boundary parallel annulus with tight boundary that is disjoint from $\alpha$, then each boundary component of $B$ is normally isotopic either to a component of $\alpha$ or to a curve of $\Gamma$.

2. $|\Gamma|$ and $\max_{\gamma \in \Gamma} \ell(\gamma)$ are bounded by a computable function of $t$. 

38
Proof. Let \( \alpha \) be given by Proposition 8.2. Then \( \alpha \) is a fence in the marked triangulation \((T, \alpha \cap T^1)\). Let \( \Gamma \) be the set of the boundaries of all boundary parallel annuli that are fundamental in the marked triangulation \((T, \alpha \cap T^1)\) and disjoint from the fence \( \alpha \). Then (2) follows from Proposition 5.2.

Now we want to verify (1). Let \( B \) be a normal boundary parallel annulus with tight boundary disjoint from the fence \( \alpha \). By isotoping \( B \) we can assume that \( B \) is least complexity. If \( B \) is a fundamental, then its boundary has already been included in \( \Gamma \) and we are done.

If not, then \( B \) can be written as a sum of fundamentals for \((T, \alpha \cap T^1)\), \[ B = F_1 + F_2 + \cdots + F_k. \]

By Proposition 7.2, each \( F_i \) with boundary is an essential annulus or Möbius band. Since \( B \) is disjoint from the fence \( \alpha \), each \( F_i \) has boundary disjoint from \( \alpha \). Hence, by Proposition 8.2, each \( F_i \) has boundary components normally isotopic to \( \alpha \). But as observed in the proof of Proposition 8.2, this implies that each boundary component of \( B \) is normally isotopic to \( \alpha \), as required.

10 Planar meridional surfaces

In this section we consider a planar (almost) meridional surface \( P \) in \( X \) and a collection \( \mathcal{A} \) of disjoint essential annuli. The collection \( \partial \mathcal{A} \) of the boundaries of the annuli in \( \mathcal{A} \) forms a collection of disjoint curves (loops) in \( \partial X \), and \( \partial P \) is another collection of disjoint loops.

We want to move \( P \) by means of a self-homeomorphism \( h: X \to X \) in such a way that the number of intersections of these two collections, \( \partial \mathcal{A} \) and \( \partial P \), becomes bounded; more precisely, we need a bound of the form \( C(t) |\mathcal{A}| \cdot |\partial P| \). This is formulated in Proposition 10.3 below; the self-homeomorphism \( h \) is going to be one of two ways of changing the original embedding of \( X \) in \( S^3 \) in order to get a short meridian.

First we collect auxiliary results. We begin with a corollary of the main result of [MSTW13], which was developed for the purpose of proving a result in the spirit of Proposition 10.3.

Lemma 10.1 ([MSTW13, Cor. 1.6]). Let \( S \) be connected surface, i.e., a connected compact 2-manifold with boundary, of genus \( g \). Let \( (\alpha_1, \ldots, \alpha_m) \) be a system of disjoint curves (properly embedded arcs and loops) in \( S \), and let \( (\beta_1, \ldots, \beta_n) \) be another such system. Then there is a homeomorphism \( \varphi: S \to S \) fixing \( \partial S \) pointwise such that the total number of intersections of \( \alpha_1, \ldots, \alpha_m \) with \( \varphi(\beta_1), \ldots, \varphi(\beta_n) \) is at most \( K(g) mn \), where \( K(g) \) is a computable function depending only on \( g \) (in fact, \( K(g) = O(g^4) \)).

We remark that for our further approach a bound of the form \( K(g,m)n \) would also be sufficient. Such a bound, even independent of \( g \), was obtained independently by Geelen, Huynh, and Richter [GHR13], but only under the additional assumption that the union of the \( \beta_i \) does not separate \( S \). Thus, we cannot directly use their result here; the extra assumption could probably removed, but it is easier to use the bounds from [MSTW13].

We also need the following, probably standard, lemma.

Lemma 10.2. Let \( G = (V, E) \) be a graph with \( n \geq 2 \) vertices embedded in \( S^2 \), possibly with loops and multiple edges. Let us assume that no two parallel edges (connecting the same two vertices) and no two parallel loops (attached to the same vertex) are isotopic by an isotopy fixing the end-vertices and avoiding the other vertices. We also assume that there is no contractible loop \( \ell \); that is, both the interior and exterior of each loop contain a vertex. Then \( |E| \leq 3n - 6 \).
Figure 14: Removing loops and parallel edges.

Proof. If $G$ contains neither loops nor multiple edges, then this is just the usual bound for the number of edges of a simple planar graph. It remains to resolve loops and multiple edges.

First let $\ell$ be a loop with an endpoint $v$. It splits $S^2$ into two regions $X$ and $Y$. Let $F_X$ resp. $F_Y$ be the face of $G$ inside $X$ resp. $Y$ bounded by $\ell$. Then $F_X$ has to contain a vertex $v_X \neq v$, for otherwise, $\ell$ can be contracted to $v$ or it is isotopic to another loop with endpoint $v$.

Similarly, we have a vertex $v_Y$ in $F_Y$. Note that $v_X$ and $v_Y$ are not connected with an edge. We can remove $\ell$ from the graph and connect $v_X$ and $v_Y$ with an edge, keeping the graph embedded in $S^2$ and satisfying the isotopy assumptions. This way we can remove all loops without increasing the number of edges; see Figure 14.

Similarly, if we have two parallel edges, we can remove one of them and add a new edge as compensation, reducing the number of pairs of parallel edges. In this way, we get a simple planar graph and the desired bound.

Proposition 10.3. Let $X$ be an orientable, irreducible manifold with incompressible boundary. Let $P \subset X$ be a properly embedded planar surface that is either essential, or strongly irreducible and boundary strongly irreducible. Let $A \subset X$ be a collection of pairwise disjoint essential annuli. Then there is a homeomorphism $h: X \to X$ so that $|\partial h(P) \cap \partial A| < C|A| \cdot |\partial P|$, where $C = C(t)$ is a computable function of the number $t$ of tetrahedra in the triangulation of $X$.

Proof. Using either Lemma 4.3 or Lemma 4.7, we may isotope $P$ so that its intersection with $A$ is essential, that is every component of $P \cap A$ is a curve that is essential in both $P$ and $A$. This implies that the result holds when $P$ is a disk, for then $P$ contains no essential curves, and thus $P \cap A$ is empty. We proceed assuming $|\partial P| > 1$.

In $A$ every intersection arc is a spanning arc and every intersection loop is the core curve of an annulus. This is illustrated below in the left picture, while the right picture shows the intersection curves in $P$:
Say that two arcs belong to the same parallel class if they are isotopic in $P$. If $|\partial P| = 2$, then $P$ is an annulus and there is at most one parallel class of intersection arcs. When $|\partial P| > 2$, form a planar graph by treating each boundary component as a vertex and each parallel class of arcs as an edge. The number of edges, hence parallel classes, is bounded by $3(|\partial P| - 2)$ by Lemma 10.2. We can cover the cases when $P$ is an annulus or disk by reducing this last bound slightly. In all cases the number of parallel classes of arcs is bounded by $3(|\partial P| - 1) < 3|\partial P|$.

A band in $X$ is an embedded, but not properly embedded, rectangle meeting $\partial X$ in precisely its top and bottom sides. For each parallel class of intersection arcs in $P$, we may choose a band $B_j$ that is a sub-surface of $P$, contains all intersection arcs in the class, and meets no other curves of intersection. Then $B$, the union of all such bands, has at most 3$|\partial P|\$ components and contains all arcs, but no loops, of the intersection $P \cap \mathcal{A}$.

Next, let us draw the core curve $\alpha_i$ for every annulus in $A_i \in \mathcal{A}$, and a curve $\beta_j$ parallel to the top and bottom sides (those in $\partial X$) in the middle of each band $B_j$. Let us think of these $\alpha_i$ and $\beta_j$ as being (locally) horizontal and lying in the same level; then, again locally, $A_i$ is a vertical “wall” through $\alpha_i$ and $B_j$ is a vertical “wall” through $\beta_j$. We have the $A_i$ and $B_j$ fibered with segments, as in the left picture, and so the union $\mathcal{A} \cup \mathcal{B}$ has the structure of an $I$-bundle $M_0$ over $(\bigcup \alpha_i) \cup (\bigcup \beta_j)$, where $I$ is the interval $[-1, 1]$; see the left picture below:

As the picture illustrates, some of the $A_i$ or $B_j$ may be twisted between the intersections with the others.

Next the $I$-bundle structure on $\mathcal{A} \cup \mathcal{B}$ can be extended to a sufficiently small regular neighborhood $N(\mathcal{A} \cup \mathcal{B})$. Indeed, we can consider the regular neighborhood as the star of $\mathcal{A} \cup \mathcal{B}$ (say in the second barycentric subdivision of some triangulation); see [RS72, Chapter 3]. Therefore, $N(\mathcal{A} \cup \mathcal{B})$ has locally structure as product of $\mathcal{A} \cup \mathcal{B}$ with $I$.

We obtain an $I$-bundle $M$ over a base surface $S$ forming a narrow ribbon along the $\alpha_i$ and the $\beta_j$. This is illustrated locally in the right picture above.

The plan is now to use Lemma 10.1 (untangling curves in a surface) for the systems of curves $\alpha_i$ and $\beta_j$ within $S$, which yields a self-homeomorphism $\varphi: S \to S$ fixing $\partial S$ pointwise, such that the number of intersections of the $\alpha_i$ with the $\varphi(\beta_j)$ is suitably bounded. Then we want to extend $\varphi$ to a bundle self-homeomorphism $h: M \to M$ that is the identity over $\partial S$ (i.e., on the vertical walls bounding $M$ in the picture). After that, $h$ can be extended identically to $X \setminus M$ and we will be done.

There are two issues to be handled. First, in order to use Lemma 10.1, bound the genus of each component of $S$ by a computable function of $t$; we will actually obtain an $O(t)$ bound.
To this end, we observe that $S$ is double-covered by a surface $\tilde{S} := N(A \cup B) \cap \partial X$. Let $K$ be a component of $S$ and $\tilde{K}$ be the corresponding double cover of $K$ in $\tilde{S}$. In particular, $\chi(\tilde{K}) = 2\chi(K)$. For a surface $F$, we let $b(F)$ denote the number of boundary components, and define $g_e(F) := 2 - \chi(F) - b(F)$. If $F$ is connected, this value is known as the Euler genus of $F$. Then we get

$$g_e(K) = 1 + g_e(\tilde{K})/2 - b(K) + b(\tilde{K})/2 \leq 1 + g_e(\tilde{K})/2.$$

Let $\tilde{Q}$ be a component of $\tilde{K}$ ($\tilde{K}$ has two components if $K$ is orientable). Since $\tilde{Q} \subseteq \partial X$, we have $g_e(\tilde{Q}) \leq g_e(X^3_{\tilde{Q}})$ where $X^3_{\tilde{Q}}$ is the component of $\partial X$ containing $\tilde{Q}$. The Euler genus of $X^3_{\tilde{Q}}$ is bounded by $O(t)$, since $\partial X^3_{\tilde{Q}} = \emptyset$ and $\mid \chi(X^3_{\tilde{Q}}) \mid = O(t)$ (note that the number of triangles, edges and vertices in triangulation of $X$ are all bounded by $O(t)$).

Altogether $g_e(\tilde{Q}) = O(t)$, and since $\tilde{K}$ has at most two components, $g_e(\tilde{K}) = O(t)$. Since the genus of a surface is at most twice the Euler genus, we also obtain $\chi(Q) = O(t)$.

By applying Lemma 10.1 as announced above, working in each component $K$ of $S$ separately and then summing up, we obtain a self-homeomorphism $\varphi$ of $S$, fixed pointwise on $\partial S$, such that the total number of intersections of the $\alpha_i$ with the $\varphi(\beta_j)$ is at most $C_0(t)|A| \cdot |B| \leq 3C_0(t)|A| \cdot |\partial P|$, where $C_0(t)$ is a computable function of $t$.

It remains to deal with the second and last issue, namely, showing that $\varphi$ extends to a bundle self-homeomorphism $h: M \to M$ that is identical over $\partial S$. Here we may assume w.l.o.g. that $S$, and hence $M$, are connected.

By the assumption, $M$ is embedded in $X$, and so it is orientable. It follows that if the surface $S$ is orientable, then $M$ is actually the product $S \times [-1, 1]$, and the extension of $\varphi$ to $h$ is obvious.

So let $S$ be non-orientable; then $M$ is non-trivially twisted and there are no global coordinates. For a subset $S'$ of $S$ we will use a notation $S' \times I$ for the subbundle of $M$ consisting of points of $M$ that project to $S'$. (In particular, we also regard $M$ as $S \times I$.)

For any connected non-orientable surface $S$ there is a non-separating arc $\sigma \subset S$ with both endpoints on the same boundary component for which $S_{\sigma}$, which is $S$ cut along $\sigma$, is an orientable surface. We also let $f_{\sigma}: S_{\sigma} \to S$ be the map gluing $S_{\sigma}$ back to $S$.\(^{15}\)

Thus, after cutting $M$ along $\sigma \times I$, we obtain a product bundle $M_{\sigma}$, homeomorphic to $S_{\sigma} \times I$. In the boundary of $M_{\sigma}$ we see two rectangle scars from cutting along $\sigma \times I$. We get the twisted bundle $M$ back by gluing $M_{\sigma}$ to itself along the rectangles so that the top of one is glued to the bottom of the other.

Now the given homeomorphism $\varphi: S \to S$ also takes $\sigma$ to a curve $\varphi(\sigma)$ that has the same separation properties. We define $S_{\varphi(\sigma)}$ and $f_{\varphi(\sigma)}$ in the same way as $S_{\sigma}$ and $f_{\sigma}$ above. The homeomorphism $\varphi$ also induces a homeomorphism $\varphi': S_{\sigma} \to S_{\varphi(\sigma)}$ satisfying $f_{\varphi(\sigma)} \circ \varphi' = \varphi \circ f_{\sigma}$.

The homeomorphism $\varphi'$ can be extended to a homeomorphism of the product bundles $h': S_{\sigma} \times I \to S_{\varphi(\sigma)} \times I$ in two ways (by either keeping $I$ or swapping it). By gluing back along the rectangular scars, $h'$ induces a homeomorphism $h: M \to M$.

Recall that $M$ is orientable since it embeds into $X$. Since we had two choices for $h'$ we select one for which $h$ is an orientation preserving automorphism. It follows that whenever $K$ is a boundary component of $S$, then $h$ is the identity on $K \times I$ (it cannot flip $I$ here since such a flip would reverse the orientation on $\partial M$, hence on $M$).

\(^{15}\)It is not difficult to find such an arc $\sigma$ in the projective plane or the Klein bottle with a single hole. Any other nonorientable surface with nontrivial boundary can be obtained by adding handles (not across the desired arc $\sigma$) and holes to one of the two surfaces above (adding a handle increases the non-orientable genus by 2).
11 Proof of the short-meridian theorem

We already have almost all of the ingredients ready to prove Theorem 1.4, following the outline from Section 2.

We assume that $X$ is irreducible, has incompressible boundary (which we may assume to be nonempty), embeds in $S^3$, and has a $0$-efficient triangulation with $t$ tetrahedra. Note that the second conclusion of the following lemma may require a re-embedding of $X$ into $S^3$.

**Lemma 11.1.** $X$ contains a planar surface $P$ so that:

1. $P$ is essential, or, strongly irreducible and boundary strongly irreducible, and
2. $\partial P$ is meridional or almost meridional in some embedding of $X$ in $S^3$.

**Proof.** Since $X$ embeds in $S^3$, we can apply the result of Fox [Fox48] that shows $X$ may be embedded so that $S^3 \setminus \text{interior}(X)$ is a collection of handlebodies.

Then we may view $X$ as the exterior, $X = S^3 \setminus N(\Gamma)$, where $\Gamma$ is a graph consisting of a spine of each handlebody. In this context, we may apply Theorem 3 of Li [Li10] that states $X$ contains a planar surface that is either: (1) meridional, strongly irreducible, and boundary strongly irreducible, or (2) almost meridional and essential, or (3) non-separating, almost meridional, and incompressible. By Lemma 4.5, case (3) reduces to case (2), and the lemma follows.

**Lemma 11.2.** Let $P$ be a surface satisfying the conditions of Lemma 11.1, and let $h : X \to X$ be a homeomorphism. Then $h(P)$ satisfies the conditions of Lemma 11.1 for some re-embedding of $X$.

**Proof.** Because the homeomorphism $h$ maps any disk in $X$ to a disk in $X$, $P$ is essential if and only if $h(P)$ is essential; and, $P$ is strongly irreducible and boundary strongly irreducible if and only if $h(P)$ is strongly irreducible and boundary strongly irreducible.

Let $e : X \to S^3$ be the embedding for which $P$ is (almost) meridional. Then $r := e \circ h^{-1} : X \to S^3$ is a re-embedding of $X$, and any component $e(\mu) \subset e(\partial P)$ bounds a disk in $S^3 \setminus X$ if and only if $r(h(\mu))$ bounds a disk in $S^3 \setminus (r \circ h(X))$. Then $P$ is (almost) meridional in the original embedding if and only if $h(P)$ is meridional in the re-embedding.

We can thus place additional constraints on $P$. Let $\alpha$ be the tight essential annulus curve given by Proposition 8.2 and let $\Gamma$ be the set of tight essential curves bounding boundary parallel annuli given by Proposition 9.1.

**Assumption 11.3.** Among planar surfaces $P$ satisfying the conclusion of Lemma 11.1 choose $P$ to minimize, in this order:

1. $|\partial P \cap \alpha|$;
2. $\operatorname{cpx}(\partial P)$, and hence $\ell(\partial P)$; and
3. $|\partial P \cap (\alpha \cup \Gamma)|$.

The next lemma shows that $P$’s intersections with $\alpha$ are bounded by a linear function of $\chi(P)$ for $t$ fixed.

**Lemma 11.4.** Under Assumption 11.3, we have
(1) \(|\partial P \cap \alpha| \leq C_0(t)|\partial P|\), where \(C_0(t)\) depends on \(t\), the number of tetrahedra;
(2) \(\partial P\) is tight and essential; and
(3) \(\{\partial P\} \cup \{\alpha\} \cup \Gamma\) is pairwise snug.

Proof. We have \(\alpha \subset \partial A\), where \(A\) is a collection of pairwise disjoint essential annuli. We choose \(A\) to minimize \(|A|\) subject to \(\alpha \subset \partial A\). Then \(|A| \leq |\alpha|\), because each \(A \in A\) must contribute at least one unique component to \(\alpha\).

By Proposition 10.3, there is a homeomorphism of \(X\) so that the image of the planar surface, call it \(P\), satisfies \(|\partial P \cap \partial A| \leq C(t)|A| \cdot |\partial P|\) for a suitable \(C(t)\). Now Proposition 8.2 guarantees that \(|\alpha| \leq 4t\), and hence \(|\partial P \cap \partial A| \leq 4C(t)t|\partial P|\) by Assumption 11.3.

Because \(P\) is either essential or strongly irreducible, \(\partial P\) consists of essential curves (Lemma 4.6). Thus, \(\partial P\) can be tightened; this may possibly increase \(|P \cap \partial A|\).

However, since \(\alpha\) and \(\Gamma\) are tight, using Lemma 6.7 repeatedly, we can make \(\{\partial P\} \cup \{\alpha\} \cup \Gamma\) pairwise snug within their normal isotopy classes. In particular, after this step \(|P \cap \partial A| = i(\partial P, \partial A)\) where \(i(\ldots)\) is the geometric intersection number. This again guarantees that \(|P \cap \partial A|\) is minimized.

Therefore, we can simultaneously achieve \(\partial P\) tight and \(\{\partial P\} \cup \{\alpha\} \cup \Gamma\) pairwise snug. Hence both of these properties hold under Assumption 11.3.

Lemma 11.5. \(P\) can be isotoped, without changing \(\partial P\), to be normal or almost normal.

Proof. If \(P\) is essential, then, since \(\partial P\) is tight, \(P\) itself can be tightened without changing \(\partial P\). Then \(P\) is normal by Proposition 7.1.

If \(P\) is strongly irreducible and boundary strongly irreducible, then the main result of [BDTS12] states that \(P\) is isotopic to an almost normal surface. Moreover, in the proof of Proposition 3.1 of [BDTS12] it is assumed that \(\partial P\) is least length (see Lemma 3.9), which is satisfied when \(\partial P\) is tight. The additional normalization steps taken there isotope the interior of \(P\) without changing its boundary, so \(\partial P\) is also fixed in the almost normal case.

The average length argument. We mark the triangulation \(T\) of \(X\) with marking \(M = (\alpha \cup \Gamma) \cap T^1\). Thus, the (almost) meridional, (almost) normal planar surface \(P\) can be written as a sum of fundamental (almost) \(M\)-normal surfaces, \(P = \sum k_i F_i\), and its boundary is the sum of the boundary curves of the fundamentals:

\[\partial P = \sum k_i \partial F_i.\]

Since \(\partial P\) is essential and tight, the boundary of each summand is essential and tight by Lemma 6.8. Each \(F_i\) falls into at least one of the following categories:

1. \(\partial F_i = \emptyset\);
2. \(F_i\) is almost normal;
3. \(\chi(F_i) > 0\) and \(F_i\) is normal;
4. \(\chi(F_i) = 0\), \(F_i\) is normal, and one of the following hold:
   (a) \(F_i\) is a compressible annulus;
(b) \( F_i \) meets \( \alpha \);
(c) \( F_i \) is a boundary parallel annulus disjoint from \( \alpha \);
(d) \( F_i \) is an essential annulus or Möbius band disjoint from \( \alpha \).

5. \( \chi(F_i) < 0 \) and \( F_i \) is normal.

We will bound the total length \( \ell(\partial P) \) by bounding the coefficients of each of these types in the boundary sum \( \partial P = \sum k_i \partial F_i \). Obviously, we can ignore summands with empty boundary (Case 1). Since there can be at most one exceptional piece in the almost normal case, we have \( k_i \leq 1 \) in Case 2.

**Lemma 11.6.** There are no normal summands with \( \chi(F_i) > 0 \).

**Proof.** Such summands are either spheres, or projective planes, or disks. As we mentioned in Section 7, normal spheres contradict 0-efficiency of the triangulation of \( X \).

Projective planes are excluded because \( X \) is irreducible. As for disks, since \( X \) has incompressible boundary, they also have trivial boundary. But this contradicts the fact that each summand has essential boundary.

This excludes Case 3, and we proceed with Case 4a.

**Lemma 11.7.** No summand is a compressible annulus.

**Proof.** Because \( X \) has incompressible boundary, a compressible annulus has trivial boundary by Proposition 4.2. This would contradict the fact that every summand of \( \partial P \) is essential and tight by Lemma 6.8.

The next lemma supplies a bound for Case 4b, although it does not need the assumptions of Case 4b in full strength.

**Lemma 11.8.** \( \sum k_i < C_0(t)|\partial P| \), where the sum is restricted to those \( F_i \) for which \( \partial F_i \cap \alpha \neq \emptyset \).

**Proof.** Addition of (almost) \( M \)-normal surfaces implies \( M \)-normal addition of their boundary curves which is additive with respect to intersections with the tight essential fence \( \alpha \) by Proposition 6.10. Thus the sum of the coefficients \( k_i \) of fundamentals \( F_i \) meeting \( \alpha \) is bounded by the total number of intersection with \( \alpha \), which in turn is bounded by Lemma 11.4.

Case 4c can be excluded:

**Lemma 11.9.** No \( F_i \) is a normal boundary parallel annulus disjoint from \( \alpha \).

**Proof.** If not, then we can write \( P = B + P' \), an (almost) \( M \)-normal sum, where \( B = F_i \) is a boundary parallel annulus. All summands of \( P \) have tight boundary by Lemma 6.8, so \( \partial B = 2b \) a pair of normally parallel tight essential curves. By the construction of \( \Gamma \) (Proposition 9.1), \( b \) is normally isotopic to either a component of \( \alpha \) or an element of \( \Gamma \), and therefore to a fence. By Proposition 6.10, each point of \( \partial P' \cap b \) has the same normal sign. But since \( \partial B = 2b \), all intersections in \( \partial B \cap \partial P' \) have the same normal sign. This contradicts Lemma 5.6.

Next, we want a bound for Case 4d.
Lemma 11.10. For each $F_i$ that is a normal essential annulus or Möbius band disjoint from $\alpha$, with a suitable $C_1(t)$.

Proof. If $F_i$ is a Möbius band, let $F'_i := 2F_i$ and $k'_i := \lfloor k_i/2 \rfloor$, and otherwise, let $F'_i := F_i, k'_i := k_i$. Then we write $P = P' + k'_iF'_i$, where $F'_i$ is an essential annulus. We wish to show that $k'_i \leq |\partial P \cap \alpha| - 1$; then the result follows from Lemma 11.4.

So we proceed by contradiction, assuming $k'_i \geq |\partial P \cap \alpha|$. Let $f_1$ and $f_2$ be the components of $\partial F'_i$.

Proposition 8.2 guarantees that $f_1$ and $f_2$ are each normally parallel to a component of the fence $\alpha$, and thus, by Proposition 6.10, $\partial P'$ meets each component $f_j$ in points with the same normal sign.

Since intersection arcs join intersection points of opposite sign (Lemma 5.6), each arc component of $P' \cap F'_i$ meets both boundary components of $F'_i$ and is thus a spanning arc of $F'_i$. There are $n = \frac{1}{2}|P' \cap \alpha|$ such spanning arcs of intersection, and $\partial P'$ meets, say, $f_i$ in $n$ positive intersections and $f_2$ in $n$ negative intersections. From the view of boundary curves $\partial P = \partial P' + k'_i F'_i$, where adding copies of $f_1$ and $f_2$ is a fractional Dehn twist (with fraction $k'_i/2$) in each of those curves.

We have assumed that $k'_i \geq |\partial P \cap \alpha| = |\partial P' \cap \alpha| = 2n$, so the fraction is greater than 1. Then $\partial P' + (k'_i - n)f_1 + (k'_i - n)f_2$ is homeomorphic to $\partial P$. Moreover, the homeomorphism can be extended over the annulus $F'_i$ to a homeomorphism of $X$ that is a Dehn twist in $F'_i$. But this homeomorphism takes $P$ to a surface with shorter length and the same number of intersection with $\alpha$, contradicting our choice in Assumption 11.3. (This is another place where we may change the embedding of $X$.)

Finally, it is straightforward to bound those summands in Case 5.

Lemma 11.11. We have $\sum k_i \leq -\chi(P) < |\partial P|$, where the sum is restricted to those $F_i$ with $\chi(F_i) < 0$.

Proof. We have observed that all summands of $P$ have $\chi \leq 0$. Those with $\chi = 0$ do not contribute to $\chi(P)$, and so $\chi(P) = \sum k_i \chi(F_i)$ for the summands with $\chi \leq -1$. It follows that $\sum k_i \leq -\chi(P)$ for these summands.

We are ready to bound the average length of a component of $\partial P$.

Lemma 11.12. We have

$$\ell(\partial P) \leq L(t)|\partial P|,$$

where $L$ is a computable function of the number of tetrahedra $t$.

Proof. Recall that we wrote $\partial P$ as a sum of boundaries of (almost) normal $M$-fundamentals for the marked triangulation $(T, M)$, where $M = (\alpha \cup \Gamma) \cap T^1$. Proposition 5.2 bounds both the weight of any fundamental solution and the total number of fundamental solutions by computable functions of $t$ and the number of marked points $m = |M| = \ell(\alpha) + \ell(\Gamma)$. These lengths are bounded by computable functions of $t$ by Propositions 8.2 and 9.1, respectively. Thus the weight of any $M$-fundamental solution and the total number of $M$-fundamental solutions are bounded by computable functions of $t$ only.

As in the proof outline in Section 2, let $\ell_{\text{max}} := \max\{\ell(\partial F_i)\}$, the maximum taken over all normal or almost normal $M$-fundamental surfaces $F_i$ (in the marked triangulation of $X$).
Because the length of a surface’s boundary is at most its weight, $\ell_{\text{max}}$ is bounded by a computable function of $t$.

Because length is additive, we have

$$\ell(\partial P) = \sum k_i \ell(\partial F_i),$$

where the sum is restricted to surfaces $F_i$ with non-empty boundary.

If $F_i$ is one of the four types of fundamentals that contribute to $\ell(\partial P)$, then $k_i \leq C_2(t)|\partial P|$, where $C_2(t) = \max(C_0(t), C_1(t))$, by Lemmas 11.8, 11.10, and 11.11 (sometimes the bound is much better).

Since the number of distinct fundamentals is bounded by a computable function of $t$, call it $C_3(t)$, we have $\sum k_i \leq C_3(t)C_2(t)|\partial P|$ over all summands that contribute to $\partial P$. The total length is then bounded by $\ell_{\text{max}} \cdot C_3(t)C_2(t)|\partial P|$ as the lemma claims.

Theorem 1.4 now follows. Because $P$ is meridional or almost meridional, at least $|\partial P| - 1$ of its boundary components are meridians (note that $|\partial P| > 1$ by Lemma 11.6). Hence the average length of a meridian is at most

$$\frac{\ell(\partial P)}{|\partial P| - 1} \leq \frac{L(t)|\partial P|}{|\partial P| - 1} = \frac{|\partial P|}{|\partial P| - 1}L(t) \leq 2L(t).$$

Unlike the children of Lake Wobegon, some meridian must be at most average, and hence its length is bounded by $2L(t)$, a computable function of $t$. This completes the proof of Theorem 1.4.

## 12 Embedding 3-dimensional complexes

In this section we prove Corollary 1.2: we provide an algorithm for $\text{EMBED}_{3\to3}$. It uses the algorithm for $\text{EMBED}_{2\to3}$, as well as an $S^3$ recognition algorithm and an algorithm for $\text{EMBED}_{2\to2}$.

Let $K$ be a 3-complex for which we want to test embeddability in $\mathbb{R}^3$. We assume, w.l.o.g., that $K$ is connected. The idea is to replace every 3-simplex of $K$ by a suitable 2-dimensional structure so that an embedding of this 2-structure ensures the embeddability of the 3-simplex.

We call a vertex $v$ of $K$ a cut vertex if removing $v$ from $K$ disconnects $K$. We let $K' := (\text{sd } K)^{(2)}$ to be the 2-skeleton of the barycentric subdivision of $K$ (see the paragraph below the description of the algorithm). We will show that if $K$ is connected and without cut vertices, then $K$ embeds in $\mathbb{R}^3$ if and only if $K'$ does. And we will also show that the assumption that $K$ does not contain cut vertices is achievable.

**Description of the algorithm (assuming $K$ connected).**

1. If $K$ is homeomorphic to $S^3$ (which can be tested, as in the algorithm for $\text{EMBED}_{2\to3}$), return FALSE.

2. If there is a vertex whose link\(^{16}\) is not embeddable in $S^2$, return FALSE. (The embeddability in $S^2$ can be tested using [GR79] and $S^2$ recognition, for example.)

\(^{16}\)We recall that the link of a vertex $v$ in a simplicial complex $K$ consists of all simplices $\sigma$ of $K$ that do not contain $v$ and such that $\sigma$ together with $v$ forms a simplex of $K$. 

47
3. If \( K \) contains a cut vertex \( v \), consider two connected induced subcomplexes \( K_1 \) and \( K_2 \) of \( K \) such that \( K_1 \cup K_2 = K \) and \( K_1 \cap K_2 = \{v\} \), \( K_1, K_2 \neq K \). (Note that such \( K_1, K_2 \) exist: after removing \( v \) from \( K \) we can possibly obtain more than two components, but we can merge them into two groups.) Run the algorithm for \( K_1 \) and \( K_2 \) separately and return TRUE if and only if both \( K_1 \) and \( K_2 \) embed in \( \mathbb{R}^3 \).

4. Run the algorithm for \( \text{EMBED}_{2 \rightarrow 3} \) with \( K' := (\text{sd } K)^{(2)} \) and return its answer.

**Geometric realizations and the barycentric subdivision.** In this section we need to carefully distinguish a simplicial complex \( K \) and its geometric realization \( |K| \). (In this section we use \( |\cdot| \) solely for geometric realizations, although earlier it meant the number of connected components.) Given a complex \( K \), we denote its barycentric subdivision by \( \text{sd } K \). See the next picture for an example of barycentric subdivision, and e.g. [Mat03] or almost any textbook on algebraic topology for a detailed treatment of this notion.

![Barycentric subdivision example](image)

Given a subcomplex (or a face) \( L \) of \( K \) we also denote \( \text{sd } L \) the barycentric subdivision of \( L \) regarded as a subcomplex of \( \text{sd } K \). The geometric realizations of \( K \) and \( \text{sd } K \) can be canonically chosen so that \( |K| = |\text{sd } K| \) (and \( |L| = |\text{sd } L| \) for every subcomplex); we assume this canonical choice.

**Correctness of the algorithm.** Now we argue that the algorithm is correct modulo two lemmas proved below. In first step we exclude the case \( K = S^3 \) and thus, we can freely use that PL embeddability of \( K \) in \( \mathbb{R}^3 \) is equivalent to PL embeddability of \( K \) in \( S^3 \).

Further, if \( K \) PL embeds in \( \mathbb{R}^3 \) then the links of vertices PL embed in \( S^2 \),\(^{17} \) so the answer in Step 2 is correct, and further we may assume that all the links embed in \( S^2 \).

The next lemma shows correctness of Step 3.

**Lemma 12.1.** Let \( K \) be a connected simplicial complex such that link of each vertex embeds in \( S^2 \), and let \( K_1 \) and \( K_2 \) be two connected induced subcomplexes as in Step 3. Then \( K \) PL embeds in \( S^3 \) if and only if \( K_1 \) and \( K_2 \) PL embed in \( S^3 \).

Finally, the correctness of Step 4 relies on the next lemma.

**Lemma 12.2.** Let \( K \) be a connected simplicial complex without cut vertices, and let \( K' = (\text{sd } K)^{(2)} \). Then the following conditions are equivalent.

(i) \( K \) PL embeds in \( S^3 \);

(ii) \( K \) topologically embeds in \( S^3 \); and

(iii) \( K' \) topologically embeds in \( S^3 \).

\(^{17}\)Intersecting the PL embedding of \( K \) with a sufficiently small 2-sphere around the image of a vertex \( v \), we get an embedding of the link of \( v \) in \( K \).
Proofs of the lemmas. To finish the proof of correctness of the algorithm, it is sufficient to prove Lemmas 12.1 and 12.2. The proofs rely on the 3-dimensional PL Schoenflies theorem (see, e.g., [Bin83, Theorem XIV.1]):

**Theorem 12.3** (PL Schoenflies theorem for $\mathbb{R}^3$). Let $f: S^2 \to \mathbb{R}^3$ be a PL embedding.\(^{18}\) Then there is a PL homeomorphism $h: \mathbb{R}^3 \to \mathbb{R}^3$ such that $h \circ i$ is a standard embedding of $S^2$ in $\mathbb{R}^3$ (as the boundary of a 3-simplex $\Delta^3$). Moreover, $h$ can be chosen to be the identity outside any given open set $U$ that contains the bounded component of $\mathbb{R}^3 \setminus f(S^2)$. In particular, the bounded component of $\mathbb{R}^3 \setminus f(S^2)$ is a PL ball.

This easily implies the following version for $S^3$ (which is also standard but we did not find a reference exactly in this setting):

**Corollary 12.4** (PL Schoenflies for $S^3$). If $f: S^2 \to S^3$ is a PL embedding, then there is a PL homeomorphism $g: S^3 \to S^3$ such that $g \circ f$ is the standard inclusion of $S^2$ as the boundary of a hemisphere. In particular, the closures of both components of $S^3 \setminus f(S^2)$ are PL balls with boundary $f(S^2)$.

**Proof of Corollary 12.4.** Choose a sufficiently fine PL triangulation of $S^3$ such that $f(S^2)$ avoids one of the closed $d$-simplices $\sigma$ of $S^3$. By Newman’s theorem [RS72, Cor. 3.13], the closure of the complement of a PL 3-ball in $S^3$ is a PL 3-ball, i.e., PL homeomorphic to a 3-simplex; in particular, $S^3 \setminus \sigma$ is PL homeomorphic to a 3-simplex $\Delta^3$.

Fix such a PL homeomorphism $j: S^3 \setminus \sigma \cong \Delta^3 \subset \mathbb{R}^3$. Then $j \circ f$ is a PL embedding of $S^2$ in $\mathbb{R}^3$. Thus, by Theorem 12.3, there is a PL homeomorphism $h$ of $\mathbb{R}^3$ such that $h \circ j \circ f$ is a standard embedding of $S^2$ as the boundary of some 3-simplex $\Delta^3$. Moreover, the PL homeomorphism $h \circ j$ witnesses that the closure of the component of $S^3 \setminus f(S^2)$ avoiding $\sigma$ is a PL ball $B^3 \subset S^3$ with boundary $f(S^2)$. Furthermore, there is a PL homeomorphism $k$ from $\Delta^3$ to the closed lower hemisphere $H^3_\circ \subset S^3$ (e.g., with $S^3$ triangulated as the octahedral 3-sphere). By [RS72, Cor. 3.15], the PL homeomorphism $k \circ h \circ j: B^3 \cong H^3_\circ$ can be extended to a PL homeomorphism $g: S^3 \to S^3$, which has the desired property.

**Proof of Lemma 12.1.** If $K$ PL embeds in $S^3$, then both $K_1$ and $K_2$ PL embed in $S^3$ since they are subcomplexes of $K$. In sequel we assume that $K_1$ and $K_2$ PL embed in $S^3$ and we want to prove that $K$ PL embeds in $S^3$.

The idea is very simple, we just want to transform an embedding of $K_1$ and $K_2$ so that the common vertex $v$ protrude on the boundary of each and thus they can be joined together; see Figure 15 in one dimension less. It remains to show that such a transformation can be found.

From the assumptions on links of vertices and from $K_1 \neq K$ we deduce that the link of $v$ in $K_1$ is planar (and thus different from $S^2$). Indeed, if it were homeomorphic to $S^2$, then the link of $v$ in $K$ would not embed in $S^2$ since the link of $v$ in $K_2$ must be nonempty ($K_2 \neq \{v\}$ since $K_1 \neq K$). Similarly we can deduce that the link of $v$ in $K_2$ is planar.

Let $f_1: [K_1] \to S^3$ be a PL embedding. By the previous observation we deduce that $f_1(v)$ is on the boundary of $f_1([K_1])$. Therefore, there is a geometric simplex $\sigma$ in a small neighborhood of $f_1(v)$ such that $\sigma \cap f_1([K_1]) = \{v\}$. Consequently, by the PL Schoenflies theorem (Corollary 12.4), there is a PL automorphism $\psi$ of $S^3$ mapping $\partial \sigma$ to $S^2 \subset S^3$. In

\(^{18}\)Formally, the standard PL model of $S^d$ is the boundary $\partial \Delta^{d+1}$, and $f$ is a PL map of the complex $\partial \Delta^3$ in $\mathbb{R}^3$. 49
addition, we can assume that it maps the interior of $\sigma$ to the upper hemisphere of $S^3$ and $f_1(v)$ to a pre-chosen point $x$ on $S^2$. Altogether, $g_1 := \psi \circ f_1$ is a PL embedding mapping $|K_1|$ to the lower hemisphere of $S^3$ such that $g_1(|K_1|) \cap S^2 = \{x\}$.

Similarly, we can find a PL map $g_2 : |K_2| \to S^3$ such that $|K_2|$ is mapped to the upper hemisphere of $S^3$ and $g_2(|K_2|) \cap S^2 = \{x\}$. Finally, we can construct the desired PL embedding $g$ of $|K|$ by setting $g(y) := g_1(y)$ if $y \in |K_1|$ and $g(y) := g_2(y)$ if $y \in |K_2|$. 

**Proof of Lemma 12.2.** Clearly (i)$\Rightarrow$(ii), and (ii)$\Rightarrow$(iii) since $K'$ is a subcomplex of a subdivision of $K$. It remains to show (iii)$\Rightarrow$(i).

Since $K'$ is 2-dimensional and since topological and PL embeddability coincide for embedding 2-complexes in $S^3$, there is an PL embedding $f' : |K'| \to S^3$. Let $f_0$ be the restriction of $f'$ to $|K^{(2)}|$ (which is a subspace of $|K'|$). We want to extend $f_0$ to a PL embedding $f : |K| \to S^3$.

We will describe how to extend $f_0$ to each tetrahedron independently, and then we argue that these extensions can be done simultaneously, which yields the desired $f$. The argument is illustrated in Figure 16.

Let $\tau$ be a tetrahedron of $K$. By the PL Schoenflies theorem (Corollary 12.4), $f_0(\partial \tau)$ splits the sphere $S^3$ in two open components whose closures are PL homeomorphic to $B^3$. Let $b_\tau$ be the barycentre of $\tau$ and let $C_\tau$ be the component that contains $f'(b_\tau)$. We will argue that $f_0(|K^{(2)}|) \cap C_\tau = \emptyset$.

Recall that $f_0(\partial \tau)$ is disjoint from $C_\tau$. For contradiction, let us assume that there is a component $C_0$ of $|K^{(2)}| \setminus \partial \tau$ such that $f_0(C_0) \cap C_\tau \neq \emptyset$, that is $f_0(C_0) \subseteq C_\tau$. Let $X_\tau := |(sd \tau)^{(2)}| \setminus |(sd \partial \tau)|$. We have $f'(X_\tau) \subseteq C_\tau$ since $X_\tau$ is connected. Let $V_0$ be the set of those vertices $v$ of $\tau$ which belong to the closure of $C_0$. We have $|V_0| > 0$ since $K$ is connected. Furthermore, $|V_0| > 1$ since $K$ does not contain cut vertices. Therefore $|V_0| \geq 2$ and there
is a path $P$ inside $|K^{(2)}|$ starting in a vertex $v_1$ of $\tau$, ending in a vertex $v_2$ of $\tau$, and with interior points in $C_0$.

Let $D$ be the subcomplex of $(sd \tau)^{(2)}$, homeomorphic to a disk, consisting of the simplices in the plane of symmetry of $v_1$ and $v_2$ (considering $\tau$ as a regular simplex); it passes through the other two vertices of $\tau$ and the midpoint of $v_1v_2$. By a double application of the PL Schoenflies theorem and using that the interior of $|D|$ belongs to $X_\tau$ (which maps to $C_\tau$ under $f'$), we have that $f'(|D|)$ splits the closure of $C_\tau$ into two parts, both PL homeomorphic to a ball. Since $f_0(v_1)$ and $f_0(v_2)$ are in different parts, $f'(|D|) \cap f_0(C_0) \neq \emptyset$, which is impossible since $f'$ is an embedding.

We have deduced that $f_0(|K^{(2)}|) \cap C_\tau = \emptyset$ for each tetrahedron $\tau$. By the PL Schoenflies theorem $f_0$ can be extended to a PL embedding of $K^{(2)} \cup \{\tau\}$ so that the interior of $\tau$ is mapped to $C_\tau$. Moreover, if we consider two distinct tetrahedra $\tau_1, \tau_2 \in K$, then $C_{\tau_1} \cap C_{\tau_2} = \emptyset$. Indeed, if $C_{\tau_1} \cap C_{\tau_2} \neq \emptyset$, then $C_{\tau_1} \subseteq C_{\tau_2}$ because $\partial C_{\tau_2} = f_0(\partial \tau_2)$ is disjoint from $C_{\tau_1}$. Similarly we deduce $C_{\tau_2} \subseteq C_{\tau_1}$ implying $\partial C_{\tau_1} = \partial C_{\tau_2}$ contradicting the assumption that $\tau_1$ and $\tau_2$ are distinct simplices.

Altogether, we can extend $f_0$ to every tetrahedron of $K$ independently, obtaining the required PL embedding of $K$.

\section*{Acknowledgment}

We thank David Bachman, Ryan Derby-Talbot, and William Jaco for very helpful conversations. We further thank Tony Huynh for kind answers to our questions, Helena Nyklová for help with many pictures, and Isaac Mabillard for remarks on a preliminary write-up of the algorithm for EMBED$3 \to 3$.

\section*{References}


UNTANGLING TWO SYSTEMS OF NONCROSSING CURVES

BY

JIŘÍ MATOUŠEK*,**
Department of Applied Mathematics, Charles University
Malostranské nám. 25, 118 00 Praha 1, Czech Republic
and
Department of Computer Science, ETH Zürich, 8092 Zürich, Switzerland
e-mail: matousek@kam.mff.cuni.cz

AND

ERIC SEDGWICK*
School of Computing, DePaul University
243 S. Wabash Ave, Chicago, IL 60604, USA
e-mail: esedgwick@cdm.depaul.edu

AND

MARTIN TANCER*†
Department of Applied Mathematics, Charles University
Malostranské nám. 25, 118 00 Praha 1, Czech Republic
and
Institutionen för matematik, Kungliga Tekniska Högskolan
100 44 Stockholm, Sweden
e-mail: tancer@kam.mff.cuni.cz

AND

ULI WAGNER*††
IST Austria, Am Campus 1, 3400 Klosterneuburg, Austria
e-mail: uli@ist.ac.at

* Supported by the ERC Advanced Grant No. 267165.
** Partially supported by Grant GRADR Eurogiga GIG/11/E023.
† Supported by a Göran Gustafsson postdoctoral fellowship.
†† Supported by the Swiss National Science Foundation (Grant SNSF-PP00P2-138948).

Received June 12, 2013 and in revised form March 7, 2014
ABSTRACT
We consider two systems \((\alpha_1, \ldots, \alpha_m)\) and \((\beta_1, \ldots, \beta_n)\) of simple curves drawn on a compact two-dimensional surface \(M\) with boundary.

Each \(\alpha_i\) and each \(\beta_j\) is either an arc meeting the boundary of \(M\) at its two endpoints, or a closed curve. The \(\alpha_i\) are pairwise disjoint except for possibly sharing endpoints, and similarly for the \(\beta_j\). We want to “untangle” the \(\beta_j\) from the \(\alpha_i\) by a self-homeomorphism of \(M\); more precisely, we seek a homeomorphism \(\varphi: M \to M\) fixing the boundary of \(M\) pointwise such that the total number of crossings of the \(\alpha_i\) with the \(\varphi(\beta_j)\) is as small as possible. This problem is motivated by an application in the algorithmic theory of embeddings and 3-manifolds.

We prove that if \(M\) is planar, i.e., a sphere with \(h \geq 0\) boundary components (“holes”), then \(O(mn)\) crossings can be achieved (independently of \(h\)), which is asymptotically tight, as an easy lower bound shows.

In general, for an arbitrary (orientable or nonorientable) surface \(M\) with \(h\) holes and of (orientable or nonorientable) genus \(g \geq 0\), we obtain an \(O((m + n)^4)\) upper bound, again independent of \(h\) and \(g\).

The proofs rely, among other things, on a result concerning simultaneous planar drawings of graphs by Erten and Kobourov.

1. Introduction

Let \(M\) be a surface, by which we mean a two-dimensional compact manifold with (possibly empty) boundary \(\partial M\). (Unless stated otherwise, we work with connected surfaces.)

By the classification theorem for surfaces, if \(M\) is orientable, then \(M\) is homeomorphic to a sphere with \(h \geq 0\) holes and \(g \geq 0\) attached handles (see Fig. 2); the number \(g\) is also called the orientable genus of \(M\). If \(M\) is nonorientable, then it is homeomorphic to a sphere with \(h \geq 0\) holes and with \(g \geq 0\) cross-caps;\(^1\) in this case, the integer \(g\) is known as the nonorientable genus of \(M\). In the sequel, the word “genus” will mean orientable genus for orientable surfaces and nonorientable genus for nonorientable surfaces.

We will consider curves in \(M\) that are properly embedded, i.e., every curve is either a simple arc meeting the boundary \(\partial M\) exactly at its two endpoints, or a simple closed curve avoiding \(\partial M\). An almost-disjoint system of curves in

---

\(^1\) A cross-cap is obtained by removing a small disc from \(M\) and gluing in a Möbius band along its boundary to the boundary circle of the resulting hole.
\( \alpha_1 \alpha_2 \alpha_3 \beta_1 \beta_2 \beta_3 \) (a)

\( \alpha_1 \phi(\beta_2) \alpha_2 \beta_1 \beta_3 \phi(\beta_1) \) (b)

Figure 1. Systems \( A \) and \( B \) of curves on a surface \( \mathcal{M} \), with \( g = 0 \) and \( h = 3 \) (a), and a re-drawing of \( B \) via a \( \partial \)-automorphism \( \varphi \) (composed of an isotopy and a Dehn twist of the darkly shaded annular region, see below) so that the number of intersections is reduced (b).

\( \mathcal{M} \) is a collection \( A = (\alpha_1, \ldots, \alpha_m) \) of curves that are pairwise disjoint except for possibly sharing endpoints.\(^2\)

In this paper we consider the following problem: We are given two almost-disjoint systems \( A = (\alpha_1, \ldots, \alpha_m) \) and \( B = (\beta_1, \ldots, \beta_n) \) of curves in \( \mathcal{M} \), where the curves of \( B \) intersect those of \( A \) possibly very many times, as in Fig. 1 (a). We would like to “redraw” the curves of \( B \) in such a way that they intersect those of \( A \) as little as possible.

We consider re-drawings only in a restricted sense, namely, induced by \( \partial \)-automorphisms of \( \mathcal{M} \), where a \( \partial \)-automorphism is a homeomorphism \( \varphi: \mathcal{M} \to \mathcal{M} \) that fixes the boundary \( \partial \mathcal{M} \) pointwise.\(^3\) Thus, given the \( \alpha_i \) and the \( \beta_j \), we are looking for a \( \partial \)-automorphism \( \varphi \) such that the number of intersections (crossings) between \( \alpha_1, \ldots, \alpha_m \) and \( \varphi(\beta_1), \ldots, \varphi(\beta_n) \) is as small as possible (where sharing endpoints does not count). We call this minimum number of crossings achievable through any choice of \( \varphi \) the entanglement number of the two systems \( A \) and \( B \).

In the orientable case, let \( f_{g,h}(m,n) \) denote the maximum entanglement number of any two systems \( A = (\alpha_1, \ldots, \alpha_m) \) and \( B = (\beta_1, \ldots, \beta_n) \) of almost-disjoint curves on an orientable surface of genus \( g \) with \( h \) holes. Analogously, we define \( \hat{f}_{g,h}(m,n) \) as the maximum entanglement number of any two systems

\(^2\) We use ordered collections of curves just because of the convenience of the notation.

\(^3\) In general, by an automorphism we mean a self-homeomorphism.
A and B of m and n curves, respectively, on a nonorientable surface of genus g with h holes. It is easy to see that \( f \) and \( \hat{f} \) are nondecreasing in m and n, which we will often use in the sequel.

To give the reader some intuition about the problem, let us illustrate which re-drawings are possible with a \( \partial \)-automorphism and which are not. In the example of Fig. 1, it is clear that the two crossings of \( \beta_3 \) with \( \alpha_3 \) can be avoided by sliding \( \beta_3 \) aside.\(^4\) It is perhaps less obvious that the crossings of \( \beta_2 \) can also be eliminated: To picture a suitable \( \partial \)-automorphism, one can think of an annular region in the interior of \( \mathcal{M} \), shaded darkly in Fig. 1 (a), that surrounds the left hole and \( \beta_1 \) and contains most of the spiral formed by \( \beta_2 \). Then we cut \( \mathcal{M} \) along the outer boundary of that annular region, twist the region two times (so that the spiral is unwound), and then we glue the outer boundary back. Here is an example of a single twist of an annulus; straight-line curves on the left are transformed to spirals on the right. (this kind of homeomorphism is often called a Dehn twist).\(^5\)

\[ \begin{array}{c}
\text{Fig. 1.} \\
\end{array} \]

On the other hand, it is impossible to eliminate the crossings of \( \beta_1 \) or \( \beta_3 \) with \( \alpha_2 \) by a \( \partial \)-automorphism. For example, we cannot re-route \( \beta_1 \) to go around the right hole and thus avoid \( \alpha_2 \), since this re-drawing is not induced by any \( \partial \)-automorphism \( \varphi \): indeed, \( \beta_1 \) separates the point \( x \) on the boundary of the left hole from the right hole, whereas \( \alpha_2 \) does not separate them; therefore, the curve \( \alpha_2 \) has to intersect \( \varphi(\beta_1) \) at least twice, once when it leaves the component containing \( x \) and once when it returns to this component.

\(^4\) This corresponds to an isotopy of the surface that fixes the boundary pointwise.

\(^5\) Formally, if we consider the circle \( S^1 = \mathbb{R}/2\pi\mathbb{Z} \) parameterized by angle, then a single Dehn twist of the standard annulus \( A = S^1 \times [0, 1] \) is the \( \partial \)-automorphism of \( A \) given by \( (\theta, r) \mapsto (\theta + 2\pi r, r) \). Being a \( \partial \)-automorphism of the annulus, a Dehn twist of an annular region contained in the interior of a surface \( \mathcal{M} \) can be extended to a \( \partial \)-automorphism of \( \mathcal{M} \) by defining it to be the identity map outside the annular region.
A rather special case of our problem, with \( m = n = 1 \) and only closed curves, was already considered by Lickorish [Lic62], who showed that the intersection of a pair of simple closed curves can be simplified via Dehn twists (and thus a \( \partial \)-automorphism) so that they meet at most twice (also see Stillwell [Sti80]). The case with \( m = 1, n \) arbitrary, only closed curves, and \( \mathcal{M} \) possibly nonorientable was proposed in 2010 as a Mathoverflow question [Huy10] by T. Huynh. In an answer A. Putman proposes an approach via the “change of coordinates principle” (see, e.g., [FM11, Sec. 1.3]), which relies on the classification of 2-dimensional surfaces—we will also use it at some points in our argument.

The results. A natural idea for bounding \( f_{g,h}(m,n) \) and \( \hat{f}_{g,h}(m,n) \) is to proceed by induction, employing the change of coordinates principle mentioned above. This does indeed lead to finite bounds, but the various induction schemes we have tried always led to bounds at least exponential in one of \( m, n \). Independently of our work, Geelen, Huynh, and Richter [GHR13] also recently proved bounds of this kind; see the discussion below. Partially influenced by the results on exponentially many intersections in representations of string graphs and similar objects (see [KM91, SSS03]), we first suspected that an exponential behavior might be unavoidable. Then, however, we found, using a very different approach, that polynomial bounds actually do hold.

For planar \( \mathcal{M} \), i.e., \( g = 0 \), we obtain an asymptotically tight bound:

**Theorem 1.1:** For planar \( \mathcal{M} \), we have \( f_{0,h}(m,n) = O(mn) \), independent of \( h \).

Here and in the sequel, the constants implicit in the \( O \)-notation are absolute, independent of \( g \) and \( h \).

A simple example providing a lower bound of \( 2mn \) is obtained, e.g., by replicating \( \alpha_2 \) in Fig. 1 \( m \) times and \( \beta_1 \) \( n \) times. We currently have no example forcing more than \( 2mn \) intersections.

In general, we obtain the following bounds:

**Theorem 1.2:**

(i) For the orientable case,

\[ f_{g,h}(m,n) = O((m + n)^4). \]

(ii) For the nonorientable case,

\[ \hat{f}_{g,h}(m,n) = O((m + n)^4). \]
Both parts of Theorems 1.2 are derived from the planar case, Theorem 1.1. In the orientable case, we use the following results on genus reduction. For a convenient notation, let us set $L = \max(m, n)$.

**Proposition 1.3** (Orientable genus reductions): (i) For $g > L$, we have

$$f_{g,h}(m,n) \leq f_{L,g+h-L}(m,n).$$

(ii) $f_{g,h}(m,n) \leq f_{0,h+1}(cg(m+g),cg(n+g))$ for a suitable constant $c > 0$.

To derive Theorem 1.2 (i), for $g > L$, we use Proposition 1.3(i), then (ii), and then the planar bound:

$$f_{g,h}(m,n) \leq f_{L,g+h-L}(m,n) \leq f_{0,g+h+1-L}(2cL^2,2cL^2) = O(L^4).$$

For $g \leq L$, we omit the first step.

In the nonorientable case, Theorem 1.2 (ii) is derived in two steps. First, analogous to Proposition 1.3 (i), we have the following reduction:

**Proposition 1.4** (Nonorientable genus reduction): For $g > 4L + 2$, we have

$$\hat{f}_{g,h}(m,n) \leq \hat{f}_{g',h'}(m,n),$$

where $g' = 4L + 2 - (g \mod 2)$ and $h' = h + \lceil g/2 \rceil - 2L - 1$.

The second step is a reduction to the orientable case.

**Proposition 1.5** (Orientability reduction): There is a constant $c$ such that

$$\hat{f}_{g,h}(m,n) \leq \hat{f}_{(g-1)/2+1+g \mod 2}(c(g + m),c(g + n)).$$

Now we can derive Theorem 1.2 (ii). We set $L := \max(m, n)$. For $g > 4L + 2$, we use Proposition 1.4, then Proposition 1.5. We also use monotonicity of the entanglement numbers in $m$ and $n$. We obtain

$$\hat{f}_{g,h}(m,n) \leq \hat{f}_{4L+2-(g \mod 2),\vartheta_1(g,h,m,n)}(m,n) \leq f_{2L,\vartheta_2(g,h,m,n)}(6cL,6cL)$$

where $\vartheta_1$ and $\vartheta_2$ are functions that, for simplicity, we do not evaluate explicitly. Then we use Proposition 1.3 and the planar bound, Theorem 1.1, to obtain an $O(L^4)$ bound similarly as in the orientable case. For $g \leq 4L + 2$, we omit the first step. Table 1 summarizes the proof of Theorem 1.2.
Table 1. A summary of the proof.

(1) For a planar surface, temporarily remove the holes not incident to any \( \alpha_i \) or \( \beta_j \), and contract the remaining “active” holes, augment the resulting planar graphs to make them 3-connected. Make a simultaneous plane drawing of the resulting planar graphs \( G_1 \) and \( G_2 \) with every edge of \( G_1 \) intersecting every edge of \( G_2 \) at most \( O(1) \) times. Decontract the active holes and put the remaining holes back into appropriate faces (Theorem 1.1; Section 2).

(2) If the genus is larger than \( c(m+n) \), find handles or cross-caps avoided by the \( \alpha_i \) and \( \beta_j \), temporarily remove them, untangle the \( \alpha_i \) and \( \beta_j \), and put the handles or cross-caps back (Propositions 1.3 (i) and 1.4; Section 3).

(3) If the surface is nonorientable, make it orientable by cutting along a suitable curve that intersects the \( \alpha_i \) and \( \beta_j \) at most \( O(m+n) \) times, untangle the resulting pieces of the \( \alpha_i \) and \( \beta_j \), and glue back (Proposition 1.5; Section 5).

(4) Make the surface planar by cutting along a suitable system of curves (canonical system of loops), untangle the resulting pieces of the \( \alpha_i \) and \( \beta_j \), and glue back (Proposition 1.3 (ii); Section 4).

Motivation. We were led to the question concerning untangling curves on surfaces while working on a project on 3-manifolds and embeddings. Specifically, we are interested in an algorithm for the following problem: given a 3-manifold \( M \) with boundary, does \( M \) embed in the 3-sphere? A special case of this problem, with the boundary of \( M \) a torus, was solved in [JS03]. The general version of the problem is motivated, in turn, by the question of algorithmically testing the embeddability of a 2-dimensional simplicial complex in \( \mathbb{R}^3 \); see [MTW11].

Very recently, we showed that these embeddability problems are algorithmically decidable; see [MSTW14]. For the proof, we use the following upper bound on \( f_{g,h}(m,n) \) and \( \hat{f}_{g,h}(m,n) \), which we state here as a separate corollary in the specific form used in [MSTW14], for convenience of reference.
Corollary 1.6: Both $f_{g,h}(m,n)$ and $\hat{f}_{g,h}(m,n)$ are bounded from above by $K(g)mn$, where $K(g)$ is a computable function of $g$, independent of $h$ (in fact, $K(g) = O(g^4)$).

Proof. By Theorem 1.1, for planar $\mathcal{M}$, we have $f_{0,h} = O(mn)$. By Proposition 1.3 (ii), in case of an orientable surface of arbitrary genus,

$$f_{g,h}(m,n) \leq f_{0,h+1}(cg(m+g), cg(n+g)) = O(g^2(m+g)(n+g)) = O(g^4 mn).$$

For the nonorientable case, Proposition 1.5 gives

$$\hat{f}_{g,h}(m,n) \leq f_{[(g-1)/2],h+1+(g \mod 2)}(c(m+g), c(n+g)) = O(g^4 mn)$$

as well.

Independently of the application to embeddability, we consider the problem investigated in this paper interesting in itself and contributing to a better understanding of combinatorial properties of curves on surfaces.

As mentioned above, the question studied in the present paper has also been investigated independently by Geelen, Huynh and Richter [GHR13], with a rather different and very strong motivation stemming from the theory of graph minors, namely the question of obtaining explicit upper bounds for the graph minor algorithms of Robertson and Seymour. Phrased in the language of the present paper, Geelen et al. [GHR13, Theorem 2.1] show that $f_{g,h}(m,n)$ and $\hat{f}_{g,h}(m,n)$ are both bounded by $n3^m$, but only under the assumption that $\mathcal{M} \setminus (\beta_1 \cup \cdots \cup \beta_n)$ is connected.\(^6\)

Further work. We suspect that the bound in Theorem 1.2 should also be $O(mn)$. The possible weak point of the current proof is the reduction in Proposition 1.3(ii), from genus comparable to $m+n$ to the planar case.

This reduction uses a result of the following kind: given a graph $G$ with $n$ edges embedded on a compact 2-manifold $\mathcal{M}$ of genus $g$ (without boundary), one can construct a system of curves on $\mathcal{M}$ such that cutting $\mathcal{M}$ along these curves yields one or several planar surfaces, and at the same time, the curves have a bounded number of crossings with the edges of $G$ (see Section 4). Concretely,

\(^6\) We remark that without this additional assumption, the bounds proved by Geelen et al. (or even weaker ones of the form $K(g,m)n$) could also be used for the application to the algorithmic embeddability problem, but due to the extra assumption their results cannot be directly applied to [MSTW14] (even though it might be possible to remove the extra assumption).
we use a result of Lazarus et al. [LPVV01], where the system of curves is of a special kind, forming a canonical system of loops. (This result is in fact essentially due to Vegter and Yap [VY90]; however, the formulation in [LPVV01] is more convenient for our purposes.) Their result is asymptotically optimal for a canonical system of loops, but it may be possible to improve it for other systems of curves. This and similar questions have been studied in the literature, mostly in an algorithmic context (see, e.g., [CM07, DFHT05, Col03, Col12] for some of the relevant works), but we haven’t found any existing result superior to that of Lazarus et al. for our purposes.

2. Planar surfaces

In this section we prove Theorem 1.1. In the proof we use the following basic fact (see, e.g., [MT01]).

**Lemma 2.1:** If \( G \) is a maximal planar simple graph (a triangulation), then for every two planar drawings of \( G \) in \( S^2 \) there is an automorphism \( \psi \) of \( S^2 \) converting one of the drawings into the other (and preserving the labeling of the vertices and edges). Moreover, if an edge \( e \) is drawn by the same arc in both of the drawings, w.l.o.g. we may assume that \( \psi \) fixes this arc pointwise.

Let us introduce the following piece of terminology. Let \( G \) be as in the lemma, and let \( D_G, D'_G \) be two planar drawings of \( G \). We say that \( D_G, D'_G \) are **directly equivalent** if there is an orientation-preserving automorphism of \( S^2 \) mapping \( D_G \) to \( D'_G \), and we call \( D_G, D'_G \) **mirror-equivalent** if there is an orientation-reversing automorphism of \( S^2 \) converting \( D_G \) into \( D'_G \).

We will also rely on a result concerning simultaneous planar embeddings; see [BKR13]. Let \( V \) be a vertex set and let \( G_1 = (V, E_1) \) and \( G_2 = (V, E_2) \) be two planar graphs on \( V \). A planar drawing \( D_{G_1} \) of \( G_1 \) and a planar drawing \( D_{G_2} \) of \( G_2 \) are said to form a **simultaneous embedding** of \( G_1 \) and \( G_2 \) if each vertex \( v \in V \) is represented by the same point in the plane in both \( D_{G_1} \) and \( D_{G_2} \) (in particular, any edge drawn in \( D_{G_1} \) may intersect any edge drawn in \( D_{G_2} \)).

We note that \( G_1 \) and \( G_2 \) may have common edges, but they are not required to be drawn in the same way in \( D_{G_1} \) and in \( D_{G_2} \). If this requirement is added, one speaks of a simultaneous embedding with **fixed edges**. There are pairs of planar graphs known that do not admit any simultaneous embedding with fixed edges (and consequently, no simultaneous straight-line embedding). An
important step in our approach is very similar to the proof of the following result.

**Theorem 2.2** (Erten and Kobourov [EK05]): Every two planar graphs $G_1 = (V, E_1)$ and $G_2 = (V, E_2)$ admit a simultaneous embedding in which every edge is drawn as a polygonal line with at most 3 bends.

We will need the following result, which follows easily from the proof given in [EK05]. For the reader’s convenience, instead of just pointing out the necessary modifications, we present a full proof.

**Theorem 2.3:** Every two planar graphs $G_1 = (V, E_1)$ and $G_2 = (V, E_2)$ admit a simultaneous, piecewise linear embedding in which each edge of $G_1$ and each edge of $G_2$ intersect at least once and at most $C$ times, for a suitable constant $C$.

In addition, if both $G_1$ and $G_2$ are maximal planar graphs, let us fix a planar drawing $D'_{G_1}$ of $G_1$ and a planar drawing $D'_{G_2}$ of $G_2$. The planar drawing of $G_1$ in the simultaneous embedding can be required to be either directly equivalent to $D'_{G_1}$, or mirror-equivalent to it, and similarly for the drawing of $G_2$ (each of the four combinations can be prescribed).

**Proof.** For the beginning, we assume that both graphs are Hamiltonian. Later on, we will drop this assumption.

Let $v_1, v_2, \ldots, v_n$ be the order of the vertices as they appear on (some) Hamiltonian cycle $H_1$ of $G_1$. Since the vertex set $V$ is common for $G_1$ and $G_2$, there is a permutation $\pi \in S(n)$ such that $v_{\pi(1)}, \ldots, v_{\pi(n)}$ is the order of the vertices as they appear on some Hamiltonian cycle $H_2$ of $G_2$.

We draw the vertex $v_i$ in the grid point $p_i = (i, \pi(i))$, $i = 1, 2, \ldots, n$. Let $S$ be the square $[1, n] \times [1, n]$. A **bispiked** curve is an $x$-monotone polygonal curve with two bends such that it starts inside $S$; the first bend is above $S$, the second bend is below $S$ and it finishes in $S$ again.

The $n - 1$ edges $v_i v_{i+1}$, of $H_1$, $i = 1, 2, \ldots, n - 1$, are drawn as bispiked curves starting in $p_i$ and finishing in $p_{i+1}$. In order to distinguish edges and their drawings, we denote these bispiked curves by $c(i, i + 1)$.

---

7 An obvious bound from the proof is $C \leq 36$, since every edge in this embedding is drawn using at most 5 bends. By a more careful inspection, one can easily get $C \leq 25$, and a further improvement is probably possible.
Similarly, we draw the edges $v_{\pi(i)}v_{\pi(i+1)}$ of $H_2$, $i = 1, 2, \ldots, n - 1$, as $y$-monotone analogs of bispiked curves, where the first bend is on the left of $S$ and the second is on the right of $S$; here is an example:

We continue only with a description of how to draw $G_1$; $G_2$ is drawn analogously with the grid rotated by 90 degrees.

Let $D_{G_1}'$ be a planar drawing of $G_1$. Every edge from $E_1$ that is not contained in $H_1$ is drawn either inside $D_{H_1}'$ or outside. Thus, we split $E_1 \setminus E(H_1)$ into two sets $E_1'$ and $E_1''$.

Let $P_0$ be the polygonal path obtained by concatenation of the curves $c(1, 2)$, $c(2, 3), \ldots, c(n - 1, n)$. Now our task is to draw the edges of $E_1' \cup \{v_1v_n\}$ as bispiked curves, all above $P_0$, and then the edges of $E_1''$ below $P_0$.

We start with $E_1'$ and we draw edges from it one by one, in a suitably chosen order, while keeping the following properties.

(P1) Every edge $v_iv_j$, where $i < j$, is drawn as a bispiked curve $c(i, j)$ starting in $p_i$ and ending in $p_j$.

(P2) The $x$-coordinate of the second bend of $c(i, j)$ belongs to the interval $[j - 1, j]$.

(P3) The polygonal curve $P_k$ that we see from above after drawing the $k$th edge is obtained as a concatenation of some curves $c(1, i_1), c(i_1, i_2), \ldots, c(i_\ell, n)$.

Here is an illustration; the square $S$ is deformed for the purposes of the drawing:
Initially, before drawing the first edge, the properties are obviously satisfied.

Let us assume that we have already drawn \(k-1\) edges of \(E_1'\), and let us focus on drawing the \(k\)th edge. Let \(e = v_i v_j \in E_1'\) be an edge that is not yet drawn and such that all edges below \(e\) are already drawn, where “below \(e\)” means all edges \(v_{i'} v_{j'} \in E_1'\) with \(i \leq i' < j' \leq j\), \((i,j) \neq (i',j')\). (This choice ensures that we will draw all edges of \(E_1'\).)

Since \(D_{G_1}'\) is a planar drawing, we know that there is no edge \(v_{i'} v_{j'} \in E_1'\) with \(i < i' < j < j'\) or \(i' < i < j' < j\), and so the points \(p_i\) and \(p_j\) have to belong to \(P_{k-1}\). The subpath \(P'\) of \(P_{k-1}\) between \(p_i\) and \(p_j\) is the concatenation of curves \(c(i, \alpha_1), c(\alpha_1, \alpha_2), \ldots, c(\alpha_s, j)\) as in the inductive assumptions. In particular, the \(x\)-coordinate of the second bend \(b^*\) of \(c(\alpha_s, j)\) belongs to the interval \([j-1, j]\).

We draw \(c(i, j)\) as follows: The second bend of \(c(i, j)\) is slightly above \(b^*\) but still below the square \(S\). The first bend of \(S\) is sufficiently high above \(S\) (with the \(x\)-coordinate somewhere between \(i\) and \(j-1\)) so that the resulting bispiked curve \(c(i, j)\) does not intersect \(P_{k-1}\). The properties (P1) and (P2) are obviously satisfied by the construction. For (P3), the path \(P_k\) is obtained from \(P_{k-1}\) by replacing \(P'\) with \(c(i, j)\).

After drawing the edges of \(E_1'\), we draw \(v_1 v_n\) in the same way. Then we draw the edges of \(E_1''\) in a similar manner as those of \(E_1'\), this time as bispiked curves below \(P_0\). This finishes the construction for Hamiltonian graphs.

Now we describe how to adjust this construction for non-Hamiltonian graphs, in the spirit of [EK05].

First we add edges to \(G_1\) and \(G_2\) so that they become planar triangulations. This step does not affect the construction at all, except that we remove these edges in the final drawing.

Next, we subdivide some of the edges of \(G_i\) with dummy vertices. Moreover, we attach two new extra edges to each dummy vertex, as in the following illustration:

![Diagram]

By choosing the subdivided edges suitably, one can obtain a 4-connected, and thus Hamiltonian, graph; see [EK05, Proof of Theorem 2] for details (this idea
previously comes from [KW02]). An important property of this construction is that each edge of $G_i$ is subdivided at most once.

In this way, we obtain new Hamiltonian graphs $G'_1$ and $G'_2$, for which we want to construct a simultaneous drawing as in the first part of the proof. A little catch is that $G'_1$ and $G'_2$ do not have the same vertex sets, but this is easy to fix. Let $d_i$ be the number of dummy vertices of $G'_i$, $i = 1, 2$, and say that $d_1 \geq d_2$. We pair the $d_2$ dummy vertices of $G'_2$ with some of the dummy vertices of $G'_1$. Then we iteratively add $d_1 - d_2$ new triangles to $G'_2$, attaching each of them to an edge of a Hamiltonian cycle. This operation keeps Hamiltonicity and introduces $d_1 - d_2$ new vertices, which can be matched with the remaining $d_1 - d_2$ dummy vertices in $G'_1$.

After drawing the resulting graphs, we remove all extra dummy vertices and extra edges added while introducing dummy vertices. An original edge $e$ that was subdivided by a dummy vertex is now drawn as a concatenation of two bispiked curves. Therefore, each edge is drawn with at most 5 bends.

Two edges with 5 bends each may in general have at most 36 intersections, but in our case, there can be at most 25 intersections, since the union of the two segments before and after a dummy vertex is both $x$-monotone and $y$-monotone.

Because of the bispiked drawing of all edges, it is also clear that every edge of $G_1$ crosses every edge of $G_2$ at least once.

Finally, the requirements on directly equivalent or mirror-equivalent drawings can easily be fulfilled by interchanging the role of top and bottom in the drawing of $G_1$ or left and right in the drawing of $G_2$. Theorem 2.3 is proved.

\begin{proof}

Proof of Theorem 1.1. Let a planar surface $\mathcal{M}$ and the curves $\alpha_1, \ldots, \alpha_m, \beta_1, \ldots, \beta_n$ be given; we assume that $\mathcal{M}$ is a subset of $S^2$. Furthermore, by eventually applying some $\partial$-automorphism moving the curves $\beta_j$, we can assume that for every $i$ and $j$ the curves $\alpha_i$ and $\beta_j$ meet on the boundary in the endpoints or in the interior transversally and in a finite number of points. From this we construct a set $V$ of $O(m + n)$ vertices in $S^2$ and planar drawings $D_{G_1}$ and $D_{G_2}$ of two simple graphs $G_1 = (V, E_1)$ and $G_2 = (V, E_2)$ in $S^2$, as follows.

1. We put all endpoints of the $\alpha_i$ and of the $\beta_j$ into $V$ (note that some of them can be shared).

2. We choose a new vertex in the interior of each $\alpha_i$ and each $\beta_j$, or two distinct vertices if $\alpha_i$ or $\beta_j$ is a loop with a single endpoint, or three vertices if $\alpha_i$ or $\beta_j$ is a closed curve, and we add all of these vertices
to \( V \). These new vertices are all distinct and do not lie on any curves other than where they were placed.

(3) If the boundary of a hole in \( M \) already contains a vertex introduced so far, we add more vertices so that it contains at least 3 vertices of \( V \). This finishes the construction of \( V \).

(4) To define the edge set \( E_1 = E(G_1) \) and the planar drawing \( D_{G_1} \), we take the portions of the curves \( \alpha_1, \ldots, \alpha_m \) between consecutive vertices of \( V \) as edges of \( E_1 \). Similarly, we make the arcs of the boundaries of the holes into edges in \( E_1 \); these will be called the hole edges. By the choice of the vertex set \( V \) above, this yields a simple plane graph.

(5) Then we add new edges to \( E_1 \) so that we obtain a drawing \( D_{G_1} \) in \( S^2 \) of a maximal planar simple graph \( G_1 \) (i.e., a triangulation) on the vertex set \( V \). While choosing these edges, we make sure that all holes containing no vertices of \( G \) lie in faces of \( D_{G_1} \) adjacent to some of the \( \alpha_i \). New edges drawn in the interior of a hole are also called hole edges.

(6) We construct \( G_2 = (V, E_2) \) and \( D_{G_2} \) analogously, using the curves \( \beta_1, \ldots, \beta_n \). We make sure that all hole edges are common to \( G_1 \) and \( G_2 \).

After this construction, each hole of \( M \) contains either no vertex of \( V \) on its boundary or at least three vertices. In the former case, we speak of an inner hole, and in the latter case, of a subdivided hole. A face \( f \) of \( D_{G_1} \) or \( D_{G_2} \) is a non-hole face if it is not contained in a subdivided hole. An inner hole \( H \) has its signature, which is a pair \((f_1, f_2)\), where \( f_1 \) is the unique non-hole face of \( D_{G_1} \) containing \( H \), and \( f_2 \) is the unique non-hole face of \( D_{G_2} \) containing \( H \). By the construction, each \( f_1 \) appearing in a signature is adjacent to some \( \alpha_i \), and each \( f_2 \) is adjacent to some \( \beta_j \).

In the following claim, we will consider different drawings \( D_{G_1}' \) and \( D_{G_2}' \) for \( G_1 \) and \( G_2 \). By Lemma 2.1, the faces of \( D_{G_1} \) are in one-to-one correspondence with the faces of \( D_{G_1}' \). For a face \( f_1 \) of \( D_{G_1} \), we denote the corresponding face by \( f_1' \), and similarly for a face \( f_2 \) of \( D_{G_2} \) and \( f_2' \).

Claim 2.4: The graphs \( G_1 \) and \( G_2 \) as above have planar drawings \( D_{G_1}' \) and \( D_{G_2}' \), respectively, that form a simultaneous embedding in which each edge of \( G_1 \) crosses each edge of \( G_2 \) at most \( C \) times, for a suitable constant \( C \); moreover,

---

\(^8\) Classifying inner holes according to the signature helps us to obtain a bound independent of the number of holes. Inner holes with the same signature are all treated in the same way, independent of their number.
$D'_G_1$ is directly equivalent to $D_G_1$; $D'_G_2$ is directly equivalent to $D_G_2$; all hole edges are drawn in the same way in $D'_G_1$ and $D'_G_2$; and whenever $(f_1, f_2)$ is a signature of an inner hole, the interior of the intersection $f'_1 \cap f'_2$ is nonempty.

We postpone the proof of Claim 2.4, and we first finish the proof of Theorem 1.1 assuming this claim.

For each inner hole $H$ with signature $(f_1, f_2)$, we introduce a closed disk $B_H$ in the interior of $f'_1 \cap f'_2$. We require that these disks are pairwise disjoint. In the sequel, we consider holes as subsets of $S^2$ homeomorphic to closed disks (in particular, a hole $H$ intersects $\mathcal{M}$ in $\partial H$).

**Claim 2.5:** There is an orientation-preserving automorphism $\varphi_1$ of $S^2$ transforming every inner hole $H$ to $B_H$ and $D_G_1$ to $D'_G_1$.

**Proof.** Using Lemma 2.1 again, there is an orientation-preserving automorphism $\psi_1$ transforming $D_G_1$ into $D'_G_1$ (since $D_G_1$ and $D'_G_1$ are directly equivalent).

Let $f_1$ be a face of $D_G_1$. The interior of $f'_1$ contains images $\psi_1(H)$ of all holes $H$ with signature $(f_1, \cdot)$, and it also contains the disks $B_H$ for these holes. Therefore, there is a boundary- and orientation-preserving automorphism of $f'_1$ that maps each $\psi_1(H)$ to $B_H$.

By composing these automorphisms on every $f'_1$ separately, we have an orientation-preserving automorphism $\psi_2$ fixing $D'_G_1$ and transforming each $\psi_1(H)$ to $B_H$. The required automorphism is $\varphi_1 = \psi_2 \psi_1$.

**Claim 2.6:** There is an orientation-preserving automorphism $\varphi_2$ of $S^2$ that fixes hole edges (of subdivided holes), fixes $B_H$ for every inner hole $H$, and transforms $\varphi_1(D_G_2)$ to $D'_G_2$.

**Proof.** By Lemma 2.1 there is an orientation-preserving automorphism $\psi_3$ of $S^2$ that fixes hole edges and transforms $\varphi_1(D_G_2)$ to $D'_G_2$.

If an inner hole $H$ has a signature $(\cdot, f_2)$, then both $\psi_3(B_H)$ and $B_H$ belong to the interior of $f'_2$. Therefore, as in the proof of the previous claim, there is an orientation-preserving homeomorphism $\psi_4$ that fixes $D'_G_2$ and transforms $\psi_3(B_H)$ to $B_H$. We can even require that $\psi_4 \psi_3$ is identical on $B_H$. We set $\varphi_2 := \psi_4 \psi_3$.$\blacksquare$

To finish the proof of Theorem 1.1, we set $\varphi = \varphi_1^{-1} \varphi_2 \varphi_1$. We need that $\varphi$ fixes the holes (inner or subdivided) and that $\alpha_1, \ldots, \alpha_m$ and $\varphi(\beta_1), \ldots, \varphi_1(\beta_m)$ have $O(mn)$ intersections. It is routine to check all the properties:
If \( H \) is a hole (inner or subdivided), then \( \varphi_2 \) fixes \( \partial \varphi_1(H) \). Therefore, \( \varphi \) also restricts to a \( \partial \)-automorphism of \( \mathcal{M} \).

The collections of curves \( \alpha_1, \ldots, \alpha_m \) and \( \varphi(\beta_1), \ldots, (\beta_m) \) have the same intersection properties as the collections

\[
\varphi_1(\alpha_1), \ldots, \varphi_1(\alpha_m) \quad \text{and} \quad \varphi_2(\varphi_1(\beta_1)), \ldots, \varphi_2(\varphi_1(\beta_m)).
\]

Since each \( \alpha_i \) and each \( \beta_j \) was subdivided at most three times in the construction, by Claims 2.4, 2.5 and 2.6, these collections have at most \( O(mn) \) intersections. The proof of the theorem is finished, except for Claim 2.4.

**Proof of Claim 2.4.** Given \( G_1 \) and \( G_2 \), we form auxiliary planar graphs \( \tilde{G}_1 \) and \( \tilde{G}_2 \) on a vertex set \( \tilde{V} \) by contracting all hole edges and removing the resulting loops and multiple edges. We note that a loop cannot arise from an edge that was a part of some \( \alpha_i \) or \( \beta_j \).

Then we consider planar drawings \( D_{\tilde{G}_1} \) and \( D_{\tilde{G}_2} \) forming a simultaneous embedding as in Theorem 2.3, with each edge of \( \tilde{G}_1 \) crossing each edge of \( \tilde{G}_2 \) at least once and at most a constant number of times.

Let \( v_H \in \tilde{V} \) be the vertex obtained by contracting the hole edges on the boundary of a hole \( H \). Since the drawings \( D_{\tilde{G}_1} \) and \( D_{\tilde{G}_2} \) are piecewise linear, in a sufficiently small neighborhood of \( v_H \) the edges are drawn as radial segments.

We would like to replace \( v_H \) by a small circle and thus turn the drawings \( D_{\tilde{G}_1}, D_{\tilde{G}_2} \) into the required drawings \( D'_{G_1}, D'_{G_2} \). But a potential problem is that the edges in \( D_{\tilde{G}_1}, D_{\tilde{G}_2} \) may enter \( v_H \) in a wrong cyclic order.

We claim that the edges in \( D_{\tilde{G}_1} \) entering \( v_H \) have the same cyclic ordering around \( v_H \) as the corresponding edges around the hole \( H \) in the drawing \( D_{G_1} \). Indeed, by contracting the hole edges in the drawing \( D_{G_1} \), we obtain a planar drawing \( D^*_{\tilde{G}_1} \) of \( \tilde{G}_1 \) in which the cyclic order around \( v_H \) is the same as the cyclic order around \( H \) in \( D_{G_1} \). Since \( \tilde{G}_1 \) was obtained by edge contractions from a maximal planar graph, it is maximal as well (since an edge contraction cannot create a non-triangular face), and its drawing is unique up to an automorphism of \( S^2 \) (Lemma 2.1). Hence the cyclic ordering of edges around \( v_H \) in \( D_{\tilde{G}_1} \) and in \( D^*_{\tilde{G}_1} \) is either the same (if \( D_{\tilde{G}_1} \) and \( D^*_{\tilde{G}_1} \) are directly equivalent), or reverse (if \( D_{\tilde{G}_1} \) and \( D^*_{\tilde{G}_1} \) are mirror-equivalent). However, Theorem 2.3 allows us to choose the drawing \( D_{\tilde{G}_1} \) so that it is directly equivalent to \( D^*_{\tilde{G}_1} \), and then the cyclic orderings coincide. A similar consideration applies for the other graph \( G_2 \).
The edges of $D\tilde{\var{}G}_1$ may still be placed to wrong positions among the edges in $D\tilde{\var{}G}_2$, but this can be rectified at the price of at most one extra crossing for every pair of edges entering $v_H$, as the following picture indicates (the numbering specifies the cyclic order of the edges around $H$ in $D_{G_1} \cup D_{G_2}$):

It remains to draw the edges of $G_1$ and $G_2$ that became loops or multiple edges after the contraction of the hole edges. Loops can be drawn along the circumference of the hole, and multiple edges are drawn very close to the corresponding single edge.

In this way, every edge of $G_1$ still has at most a constant number of intersections with every edge of $G_2$, and every two such edges intersect at least once unless at least one of them became a loop after the contraction. Consequently, whenever $(f_1, f_2)$ is a signature of an inner hole, the corresponding faces $f'_1$ and $f'_2$ intersect. This finishes the proof.

3. Reducing the genus to $O(m + n)$

In this section we prove Proposition 1.3(i) as well as Proposition 1.4. We begin with several definitions.

3.1. Cutting Along Curves. Let $\mathcal{M}$ be an (orientable or nonorientable) surface with boundary. By $h(\mathcal{M})$ we denote the number of holes in $\mathcal{M}$ and by $g(\mathcal{M})$ we denote the (orientable or nonorientable) genus of $\mathcal{M}$.

Now let $\delta$ be a properly embedded curve in $\mathcal{M}$ (i.e., either a simple closed curve that avoids the boundary $\partial \mathcal{M}$, or a simple arc whose endpoints lie on $\partial \mathcal{M}$). The curve $\delta$ is called separating if $\mathcal{M}\setminus \delta$ has two components. Otherwise, $\delta$ is non-separating.

We denote by $\mathcal{M}_\langle \delta \rangle$ the (possibly disconnected) surface obtained by cutting $\mathcal{M}$ along $\delta$. If $\delta$ is non-separating, then $\mathcal{M}_\langle \delta \rangle$ is connected. Otherwise, $\mathcal{M}_\langle \delta \rangle$ has two components, which we denote by $\mathcal{M}_1\langle \delta \rangle$ and $\mathcal{M}_2\langle \delta \rangle$. 
Now we recall basic properties of the **Euler characteristic** of a surface. Given a triangulated surface $\mathcal{M}$, the Euler characteristic $\chi(\mathcal{M})$ is defined as the number of vertices plus number of triangles minus the number of edges in the triangulation. It is well known that the Euler characteristic is a topological invariant and equals $2 - 2g(\mathcal{M}) - h(\mathcal{M})$ if $\mathcal{M}$ is orientable, and $2 - g(\mathcal{M}) - h(\mathcal{M})$ if $\mathcal{M}$ is nonorientable.

To work simultaneously with orientable and nonorientable surfaces, it is also convenient to define the **Euler genus** of $\mathcal{M}$ as

$$g_e(\mathcal{M}) := 2 - \chi(\mathcal{M}) - h(\mathcal{M}).$$

That is, $g_e(\mathcal{M}) = g(\mathcal{M})$ if $\mathcal{M}$ is nonorientable, and $g_e(\mathcal{M}) = 2g(\mathcal{M})$ if $\mathcal{M}$ is orientable.

We have the following relations for the Euler characteristic:

<table>
<thead>
<tr>
<th>$\delta$ is a closed curve</th>
<th>$\delta$ is non-separating</th>
<th>$\delta$ is separating</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\chi(\mathcal{M})$</td>
<td>$\chi(\mathcal{M}_{(\delta)})$</td>
<td>$\chi(\mathcal{M}) = \chi(\mathcal{M}_{(\delta)}) - 1$</td>
</tr>
<tr>
<td>$\delta$ is an arc</td>
<td>$\chi(\mathcal{M}) = \chi(\mathcal{M}_{(\delta)}) - 1$</td>
<td>$\chi(\mathcal{M}) = \chi(\mathcal{M}_{(\delta)})$</td>
</tr>
</tbody>
</table>

The relations above also allow us to relate the genus of $\mathcal{M}$ and the genus of the surface(s) obtained after a cutting.

Let us call a closed curve $\delta$ in $\mathcal{M}$ **two-sided** if a small closed neighborhood of $\delta$ is homeomorphic to the annulus $S^1 \times [0, 1]$; otherwise, $\delta$ is **one-sided** (and a small closed neighborhood of $\delta$ is a Möbius band). Note that every orientable surface contains only two-sided closed curves.

**Lemma 3.1:** We have the following relations for genera:

(a) If $\mathcal{N}$ is orientable, then

$$g(\mathcal{M}) = \begin{cases} g(\mathcal{M}_{(\delta)}) + g(\mathcal{M}_{(\delta)}) & \text{if } \delta \text{ is separating;} \\ g(\mathcal{M}_{(\delta)}) & \text{if } \delta \text{ is a non-separating arc connecting} \\ \text{two different boundary components;} \\ g(\mathcal{M}_{(\delta)}) + 1 & \text{if } \delta \text{ is a non-separating closed curve, or} \\ \text{a non-separating arc with both endpoints} \\ \text{in a single boundary component.} \end{cases}$$
(b) If $N$ is orientable or nonorientable, then
\[
    g_e(M) = \begin{cases} 
        g_e(M_1^{\delta}) + g_e(M_2^{\delta}) & \text{if } \delta \text{ is separating;} \\
        g_e(M_{\delta}) & \text{if } \delta \text{ is a non-separating arc connecting} \\
        \text{two different boundary components;} \\
        g_e(M_{\delta}) + 1 & \text{if } \delta \text{ is a non-separating one-sided closed curve;} \\
        g_e(M_{\delta}) + 2 & \text{if } \delta \text{ is a non-separating arc with both} \\
        \text{endpoints in a single boundary component, or} \\
        \text{a non-separating two-sided closed curve.}
    \end{cases}
\]

Note that (b) implies (a). However, it is still convenient to state (a) separately.

Proof. A simple case analysis yields the following relations for the numbers of holes:
\[
    h(M) = \begin{cases} 
        h(M_1^{\delta}) + h(M_2^{\delta}) - 2 & \text{if } \delta \text{ is a separating closed curve;} \\
        h(M_{\delta}) - 2 & \text{if } \delta \text{ is a two-sided non-separating} \\
        \text{closed curve;} \\
        h(M_{\delta}) - 1 & \text{if } \delta \text{ is a one-sided non-separating} \\
        \text{closed curve;} \\
        h(M_1^{\delta}) + h(M_2^{\delta}) - 1 & \text{if } \delta \text{ is a separating arc;} \\
        h(M_{\delta}) + 1 & \text{if } \delta \text{ is a non-separating arc connecting} \\
        \text{two different boundary components;} \\
        h(M_{\delta}) - 1 & \text{if } \delta \text{ is a non-separating arc with both} \\
        \text{endpoints in a single boundary component.}
    \end{cases}
\]

The proof now follows by simple computation from the table above the lemma and the relations $\chi(M) = 2 - 2g(M) - h(M)$ if $M$ is orientable and $\chi(M) = 2 - g_e(M) - h(M)$ if $M$ is orientable or nonorientable.

3.2. Orientable Surfaces. Let $M$ be a surface, which may be orientable or nonorientable. A handle-enclosing curve is a separating closed curve $\lambda$ in $M$ that splits $M$ into two components $M^+_{(\lambda)}$ and $M^-_{(\lambda)}$ such that $M^-_{(\lambda)}$ is a torus with hole—that is, an orientable surface of genus 1 with one boundary hole; here are two ways of looking at it:
A system $L$ of handle-enclosing curves is **independent** if $\mathcal{M}_{\langle \kappa \rangle} \cap \mathcal{M}_{\langle \lambda \rangle} = \emptyset$ for every two closed curves $\kappa, \lambda \in L$.

First we focus on proving Proposition 1.3 (i). For the remainder of this subsection, all surfaces will be orientable.

For an orientable surface of genus $g$ with $h$ holes, we fix a **standard representation** of this surface, denoted by $\mathcal{M}_{g,h}$. It is obtained by removing interiors of $h$ pairwise disjoint disks $H_1, \ldots, H_h$ in the southern hemisphere of $S^2$ and by removing interiors of $g$ pairwise disjoint disks $D_1, \ldots, D_g$ in the northern hemisphere of $S^2$ and then attaching a torus with hole along the boundary of each $D_i$; see Fig. 2. Note that $\{\partial D_i\}_{i=1}^g$ is an independent system of handle-enclosing curves.

Figure 2. The standard representation $\mathcal{M}_{3,2}$.

One of the tools we need (Lemma 3.3) is that if we find handle-enclosing curves in some surface $\mathcal{M}$ (of genus $g$ with $h$ holes), then we can find a homeomorphism $\mathcal{M} \to \mathcal{M}_{g,h}$ mapping these curves to $\partial D_i$ extending some given homeomorphism of the boundaries. However, we have to require a technical condition on orientations, to be described next.
Let \( \gamma_1, \ldots, \gamma_h \) be a collection of the boundary curves of an orientable surface \( M \) (of arbitrary genus) with \( h \) holes. We assume that \( \gamma_1, \ldots, \gamma_h \) are also given with orientations. Since \( M \) is orientable, it makes sense to speak of whether the orientations of \( \gamma_1, \ldots, \gamma_h \) are mutually compatible or not: Choose and fix an orientation of \( M \). Then we can say for each boundary curve \( \gamma_i \) whether \( M \) lies is on the right-hand side of \( \gamma_i \) or on the left-hand side (with respect to the chosen orientation of \( M \) and the given orientation of \( \gamma_i \)).

**Lemma 3.2:** Let \( M \) be a planar surface with \( h \) holes. Let \( \gamma_1, \ldots, \gamma_h \) be the boundary curves of \( M \) given with compatible orientations. Let \( \zeta : \partial M \to \partial M_{0,h} \) be a homeomorphism such that the orientations (induced by \( \zeta \)) of the curves \( \zeta(\gamma_1), \ldots, \zeta(\gamma_h) \) are compatible. Then \( \zeta \) can be extended to a homeomorphism \( \tilde{\zeta} : M \to M_{0,h} \).

The lemma is generally known and the proof is quite straightforward. We give the proof here for completeness (and for lack of a reference). Similar remark applies to Lemma 3.3 below.

**Proof.** If \( h = 0 \), then the claim follows immediately from the classification of surfaces. For \( h = 1 \), an arbitrary homeomorphism \( \partial M \to \partial M_{0,h} \) can be extended to a homeomorphism \( M \to M_{0,h} \) (between disks) by ‘coning’.

For \( h > 1 \) we prove the lemma by induction in \( h \). We connect two (closed) boundary curves \( \gamma_1, \gamma_2 \) with an arc \( \delta \) inside \( M \) attached at some points \( a \) and \( b \) and we also connect \( \zeta(\gamma_1) \) and \( \zeta(\gamma_2) \) inside \( M_{0,h} \) with an arc \( \delta' \) attached at \( \zeta(a) \) and \( \zeta(b) \). We cut \( M \) and \( M_{0,h} \) along \( \delta \) and \( \delta' \), obtaining surfaces \( M^* \) and \( M^*_{0,h} \) with one hole less.

The holes \( \gamma_3, \ldots, \gamma_h \) are kept in \( M^* \), while the holes \( \gamma_1 \) and \( \gamma_2 \) and the arc \( \delta \) in \( M \) induce a boundary curve \( \gamma^* \) in \( M^* \) composed of four arcs \( \gamma_1^*, \delta_1^*, \gamma_2^* \) and \( \delta_2^* \). Since the orientations of \( \gamma_1, \ldots, \gamma_h \) are compatible, the arcs \( \gamma_1^* \) and \( \gamma_2^* \) are concurrently oriented as subarcs of \( \gamma^* \), and they induce an orientation of \( \gamma^* \) still compatible with \( \gamma_3, \ldots, \gamma_h \).

---

\(^9\) If \( M \) is smooth, for instance, and if we choose a point \( p_i \) in each \( \gamma_i \), then there are two distinguished unit vectors in the tangent plane of \( M \) at \( p_i \): the inner normal vector \( \nu_i \) of \( \gamma_i \) within \( M \) (which is independent of any orientation), and the tangent vector \( \tau_i \) of \( \gamma_i \) (which depends on the orientation of \( \gamma_i \)). The orientations of the boundary curves \( \gamma_1, \ldots, \gamma_h \) are compatible if and only if each pair \( (\nu_i, \tau_i) \) determines the same orientation of \( M \).
Similarly, we obtain an orientation on the new hole $\gamma'$ in $M^*_{0,h}$. We can also extend $\zeta$ so that $\zeta(\gamma^*) = \zeta(\gamma'^*)$ (running along $\delta^*_1$ and $\delta^*_2$ with same speed). By induction, there is a homeomorphism

$$\bar{\zeta}^*: \mathcal{M}^* \to M^*_{0,h},$$

and the resulting $\bar{\zeta}$ is obtained by gluing $\mathcal{M}^*$ and $M^*_{0,h}$ back to $\mathcal{M}$ and $M_{0,h}$. ■

**Lemma 3.3:** Let $(\lambda_1, \ldots, \lambda_s)$ be an independent system of handle-enclosing curves in a surface $\mathcal{M}$ of genus $g$ with $h$ holes, $s \leq g$. Let $\{\gamma_i\}_{i=1}^h$ be the system of the boundary curves of the holes in $\mathcal{M}$. Then there is a homeomorphism

$$\psi: \mathcal{M} \to \mathcal{M}_{g,h}$$

such that $\psi(\gamma_i) = \partial H_i$, $i = 1, 2, \ldots, h$, and $\psi(\lambda_i) = \partial D_i$, $i = 1, 2, \ldots, s$. Moreover, $\psi$ can be prescribed on the $\gamma_i$, assuming that it preserves compatible orientations.

**Proof.** First we remark that we can assume that $s = g$. If $s < g$, we can extend $(\lambda_1, \ldots, \lambda_s)$ to an independent system of handle-enclosing of size $g$: We cut away each torus with hole $\mathcal{M}_{(\lambda_i)}$, obtaining a surface of genus $g - s$ homeomorphic to $\mathcal{M}_{g-s,h+s}$. Then we can find an independent system of $g - s$ handle-enclosing loops in this surface. In the sequel, we assume $s = g$.

Let us cut $\mathcal{M}$ along the curves $\lambda_1, \ldots, \lambda_s$. It decomposes into a collection $T_1, \ldots, T_g$, where each $T_i$ is a torus with hole (with $\partial T_i = \lambda_i$), and one planar surface $P$ with $g + h$ holes (the boundary curves of $P$ are the $\lambda_i$ and the $\gamma_i$). In particular, $\mathcal{M}$ decomposes into the same collection of surfaces (up to a homeomorphism) as $\mathcal{M}_{g,h}$ when cut along $\partial D_i$. Let $P'$ be the planar surface in this decomposition of $\mathcal{M}_{g,h}$.

As we assume in the lemma, $\psi$ can be prescribed on some closed curves of $\partial P$ while preserving compatible orientations. It can also be extended so that it maps each $\lambda_i$ to $\partial D_i$, while preserving compatible orientations between $P$ and $P'$. Then we have, by Lemma 3.2, a homeomorphism between $P$ and $P'$ extending $\psi$.

Finally, this homeomorphism can also be extended to all the $T_i$, one by one. Note that preserving the orientations is not an issue in this case since the torus with a hole admits an automorphism reversing the orientation of the boundary curve. ■
We state the following corollary of Lemma 3.3, which will be useful in Section 5.

**Corollary 3.4:** Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be two orientable surfaces of genus $g$ with $h$ holes. Let $\zeta: \partial \mathcal{M}_1 \to \partial \mathcal{M}_2$ be a homeomorphism of the boundaries that preserves compatible orientations. Then $\zeta$ extends to a homeomorphism $\psi$ of $\mathcal{M}_1$ and $\mathcal{M}_2$.

**Proof.** We find an arbitrary homeomorphism $\zeta_1: \partial \mathcal{M}_1 \to \partial \mathcal{M}_{g,h}$ that preserves compatible orientations. Then the homeomorphism $\zeta_2: \partial \mathcal{M}_2 \to \partial \mathcal{M}_{g,h}$ defined as $\zeta_2 = \zeta_1 \zeta^{-1}$ preserves compatible orientations as well. Using Lemma 3.3 (with $s = 0$), we obtain extensions $\psi_1: \mathcal{M}_1 \to \mathcal{M}_{g,h}$ and $\psi_2: \mathcal{M}_2 \to \mathcal{M}_{g,h}$. Then $\psi := \psi_2^{-1} \psi_1$ is the required homeomorphism. ■

**Lemma 3.5:** Let $\mathcal{M}$ be a surface of genus $g$ with $h$ holes. Let $(\delta_1, \ldots, \delta_n)$ be an almost disjoint system of curves on $\mathcal{M}$. Then there is an independent system of $s \geq g - n$ handle-enclosing curves $\lambda_1, \ldots, \lambda_s$ such that each of the tori with hole $\mathcal{M} - (\lambda_j)$ is disjoint from $\bigcup_{i=1}^n \delta_i$.

**Proof.** Let us cut $\mathcal{M}$ along $\{\delta_i\}_{i=1}^n$ obtaining several components $\mathcal{M}_1, \ldots, \mathcal{M}_q$. If we cut along the curves one by one, we see that Lemma 3.1(a) implies

$$g(\mathcal{M}_1) + \cdots + g(\mathcal{M}_q) \geq g(\mathcal{M}) - n.$$  

In each $\mathcal{M}_k$ we find an independent system of $g(\mathcal{M}_k)$ handle-enclosing curves (this can be done by transforming $\mathcal{M}_k$ into the standard representation). The union of these independent systems yields a system as in the lemma. ■

**Proof of Proposition 1.3(i).** Let $\mathcal{M}$ be a surface of genus $g$ with $h$ holes. Let $A = (\alpha_1, \ldots, \alpha_m)$ and $B = (\beta_1, \ldots, \beta_n)$ be two almost disjoint systems of curves in $\mathcal{M}$.

Our task is to find a $\partial$-automorphism $\varphi$ of $\mathcal{M}$ such that the number of crossings between $\alpha_1, \ldots, \alpha_m$ and $\varphi(\beta_1), \ldots, \varphi(\beta_n)$ is at most $f_{g-s,h+s}(m, n)$, where $s := \min(g - m, g - n)$. (Let us recall that we assume that $g > m, n$, and therefore $s > 0$.)

By Lemma 3.5 there is an independent system of handle-enclosing curves $\lambda_{1,\alpha}, \ldots, \lambda_{s,\alpha}$ such that the corresponding tori with a hole are disjoint from the curves in $A$. Consequently, by Lemma 3.3, we have a homeomorphism $\psi_\alpha: \mathcal{M} \to \mathcal{M}_{g,h}$, extending a fixed homeomorphism $\psi': \partial \mathcal{M} \to \partial \mathcal{M}_{g,h}$, which
preserves compatible orientations and maps each $\lambda_{k,\alpha}$ to $\partial D_k$ (using the notation from the definition of a standard representation).

Similarly, we have an independent system of handle-enclosing curves

$$\lambda_1, \beta, \ldots, \lambda_s, \beta$$

with the corresponding tori with a hole disjoint from the curves in $B$. We also have a homeomorphism $\psi_\beta: M \to M_{g, h}$ extending $\psi'$ that maps the (closed) curves $\lambda_{k, \beta}$ to $\partial D_k$.

Now we have two systems

$$A' = (\psi_\alpha(\alpha_1), \ldots, \psi_\alpha(\alpha_m)) \quad \text{and} \quad B' = (\psi_\beta(\beta_1), \ldots, \psi_\beta(\beta_m))$$

of curves in $M_{g, h}$ avoiding the tori with a hole bounded by the $\partial D_i$. Let us remove these tori (only for $i \leq s$) obtaining a new surface $M^*$ of genus $g - s$ with $h + s$ holes. We find a $\partial$-automorphism $\varphi^*$ of $M^*$ such that the number of intersections between $A'$ and $\varphi^*$-images of the curves in $B'$ is at most $f_{g-s, h+s}(m, n)$. Since $\varphi^*$ fixes the boundary, it can be extended to a $\partial$-automorphism $\varphi_{g, h}$ of $M_{g, h}$ while introducing no new intersections. Finally, $\varphi := \psi_{\alpha}^{-1} \varphi_{g, h} \psi_\beta$ is the required $\partial$-automorphism of $M$.

3.3. NONORIENTABLE SURFACES. The proof of Proposition 1.4 is similar to the previous proof but simpler, since we need not worry about orientations.

**Lemma 3.6:** Let $N$ and $N'$ be two nonorientable surfaces with the same genus and number of holes. Let $\psi_0: \partial N \to \partial N'$ be a homeomorphism of the boundaries. Then $\psi_0$ extends to a homeomorphism $\psi: N \to N'$.

**Proof.** By the classification of surfaces, $N$ and $N'$ are homeomorphic. Given two boundary components, there is a self-homeomorphism of $N$ that exchanges these components. Therefore, we know that there is a homeomorphism $\psi_1: N \to N$ such that for each component $C$ of $\partial N$ the images $\psi_0(C)$ and $\psi_1(C)$ coincide (as sets). However, if we equip $C$ with an orientation, it might happen that $\psi_0(C)$ and $\psi_1(C)$ have opposite orientations. In such a case, we consider a self-homeomorphism $\psi_C$ of $N$ that reverts the orientation of $C$ and fixes all other boundary components. Here is an example of such a self-homeomorphism:
Up to a homeomorphism, we can consider $\mathcal{N}$ as a polygon with holes whose edges are identified according to the labels. By moving the middle hole along $\gamma$, we revert its orientation without affecting the other holes.

By gradually composing $\psi_1$ with the $\psi_C$ for those $C$ on which orientations disagree, we can get a self-homeomorphism of $\mathcal{N}$ such that $\psi_0(C)$ and $\psi_2(C)$ have compatible orientations for every $C$. Finally, by a local modification of $\psi_2$ at a small neighborhood of every $C$ we can get a self-homeomorphism $\psi$ of $\mathcal{N}$ that agrees with $\psi_0$ on $\partial \mathcal{N}$.

Similar to the orientable case, we will use a certain canonical representation $\mathcal{N}_{g,h}$ for a nonorientable surface of genus $g$ with $h$ holes. We recall that a cross-cap in a nonorientable surface $\mathcal{N}$ is a subset of $\mathcal{N}$ which is homeomorphic to a Möbius band. Note that the boundary of a cross-cap is a single closed curve.

A standard way of representing a nonorientable surface of genus $g$ with $h$ holes is to remove $h$ disjoint disks from the 2-sphere and replace the other $g$ disjoint disks with cross-caps. However, here it is more convenient to replace all but at most two of the cross-caps by handles: indeed, for $g \geq 3$, a pair of cross-caps can be replaced with a handle (this is sometimes called Dyck’s Theorem, see, e.g., [FW99, Lemma 3]; note that it is essential that at least one cross-cap remains present).

Thus, we can define a convenient representation (as opposed to the standard one mentioned above) $\mathcal{N}_{g,h}$ as follows. We again start with the sphere $S^2$, and we remove $h$ pairwise disjoint disks $H_1, \ldots, H_h$. Then we remove $\lfloor (g-1)/2 \rfloor$ more disjoint disks $D_1, \ldots, D_{\lfloor (g-1)/2 \rfloor}$ and attach a torus with a hole along the boundary of each $D_i$. Finally, we remove one (for $g$ odd) or two (for $g$ even) extra disks and we attach Möbius bands along these disks. Here is the convenient representation of $\mathcal{N}_{6,2}$:
Lemma 3.7: Let \((\lambda_1, \ldots, \lambda_s)\) be an independent system of handle-enclosing curves in a nonorientable surface \(N\) of genus \(g\) with \(h\) holes, \(s \leq \lfloor (g - 1)/2 \rfloor\). Let \(\{\gamma_i\}_{i=1}^h\) be the system of the boundary curves of the holes in \(N\). Then there is a homeomorphism \(\psi: N \to \mathcal{N}_{g,h}\) such that \(\psi(\gamma_i) = \partial H_i, i = 1, 2, \ldots, h,\) and \(\psi(\lambda_i) = \partial D_i, i = 1, 2, \ldots, s\). Moreover, \(\psi\) can be prescribed on the \(\gamma_i\).

Proof. The proof is analogous to that of Lemma 3.3. Let us cut \(N\) along the curves \(\lambda_1, \ldots, \lambda_g\). It decomposes into a collection \(T_1, \ldots, T_s\), where each \(T_i\) is a torus with a hole (with \(\partial T_i = \lambda_i\)), and one nonorientable surface \(\hat{N}\) of genus \(g - 2s\) with \(h + s\) holes (the boundary curves of \(N\) are the \(\lambda_i\) and the \(\gamma_i\)). In particular, \(N\) decomposes into the same collection of surfaces (up to a homeomorphism) as \(\mathcal{N}_{g,h}\) when cut along the \(\partial D_i\). Let \(N'\) be the nonorientable surface in the decomposition of \(\mathcal{N}_{g,h}\).

By Lemma 3.6, we have a homeomorphism between \(\hat{N}\) and \(N'\) extending a given homeomorphism of the boundary curves. This homeomorphism can be also extended to all \(T_i\), one by one. \(\blacksquare\)

Lemma 3.8: Let \(N\) be a nonorientable surface of genus \(g\) with \(h\) holes. Let \((\delta_1, \ldots, \delta_n)\) be an almost disjoint system of curves on \(M\). Then there is an independent system of \(s \geq g/2 - 2n - 1\) handle-enclosing curves \(\lambda_1, \ldots, \lambda_s\) such that each of the tori with hole \(\mathcal{M}_{\langle \lambda_j \rangle}\) is disjoint from \(\bigcup_{i=1}^n \delta_i\).

Proof. Let us cut \(N\) along \(\{\delta_i\}_{i=1}^n\), obtaining several components \(\mathcal{M}_1, \ldots, \mathcal{M}_q\), \(q \leq n + 1\) (some of them may be orientable and some nonorientable). Cutting along the curves one by one, we see that Lemma 3.1(b) implies

\[ g_e(\mathcal{M}_1) + \cdots + g_e(\mathcal{M}_q) \geq g_e(\mathcal{M}) - 2n. \]
In each $M_k$ we find an independent system of at least $(g_e(M_k) - 2)/2$ handle-enclosing curves. Indeed, if $M_k$ is orientable, then we can find even $g_e(M_k)/2$ such curves by transforming $M_k$ to the standard representation. If $M_k$ is nonorientable, then we find at least $(g_e(M_k) - 2)/2$ such curves by transforming $M_k$ to the convenient representation.

The union of these independent systems yields a system as in the lemma (using $g = g_e(M)$ and $q \leq n + 1$).

Proof of Proposition 1.4. The proof is now almost the same as for Proposition 1.3(i).

Let $N$ be a nonorientable surface of genus $g$ with $h$ holes. Let

$$A = (\alpha_1, \ldots, \alpha_m) \quad \text{and} \quad B = (\beta_1, \ldots, \beta_n)$$

be two almost disjoint systems of curves in $N$.

Our task is to find a $\partial$-automorphism $\varphi$ of $N$ such that the number of crossings between $\alpha_1, \ldots, \alpha_m$ and $\varphi(\beta_1), \ldots, \varphi(\beta_n)$ is at most $\hat{f}_{g-2s,h+s}(m,n)$, where $s := \min(\lceil g/2 \rceil - 2m - 1, \lceil g/2 \rceil - 2n - 1)$. Note that $g - 2s = 4L + 2 - (g \mod 2)$ and $h + s = h + \lceil g/2 \rceil - 2L - 1$ as required ($L = \max(m, n)$). (Let us also recall that we assume that $g > 4L + 2$, and so $s > 0$.)

By Lemma 3.8 there is an independent system of handle-enclosing curves $\lambda_{1,\alpha}, \ldots, \lambda_{s,\alpha}$ such that the corresponding tori with a hole are disjoint from the curves in $A$. Consequently, by Lemma 3.7, we have a homeomorphism $\psi_\alpha : N \to N_{g,h}$, extending a fixed homeomorphism $\psi' : \partial N \to \partial N_{g,h}$, which maps each $\lambda_{k,\alpha}$ to $\partial D_k$.

Similarly, we have an independent system of handle-enclosing curves $\lambda_{1,\beta}, \ldots, \lambda_{s,\beta}$ with the corresponding tori with a hole disjoint from the curves in $B$. We also have a homeomorphism $\psi_\beta : N \to N_{g,h}$ extending $\psi'$ that maps each $\lambda_{k,\beta}$ to $\partial D_k$.

Now we have two systems

$$A' = (\psi_\alpha(\alpha_1), \ldots, \psi_\alpha(\alpha_m)) \quad \text{and} \quad B' = (\psi_\beta(\beta_1), \ldots, \psi_\beta(\beta_m))$$

of curves in $N_{g,h}$ avoiding the tori with a hole bounded by the $\partial D_i$. Let us remove these tori (only for $i \leq s$) obtaining a new surface $N^*$ of genus $g - 2s$ with $h + s$ holes. We find a $\partial$-automorphism $\varphi^*$ of $N^*$ such that the number of intersections between $A'$ and $\varphi^*$-images of the curves in $B'$ is at most $f_{g-s,h+s}(m, n)$.
Since \( \varphi^* \) fixes the boundary, it can be extended to a \( \partial \)-automorphism \( \varphi_{g,h} \) of \( \mathcal{N}_{g,h} \) while introducing no new intersections. Finally, \( \varphi := \psi^{-1}_\alpha \varphi_{g,h} \psi_\beta \) is the required \( \partial \)-automorphism of \( \mathcal{N} \).

4. Reducing the orientable genus to 0

Here we prove Proposition 1.3(ii). We start with some preliminaries.

Let \( g \geq 1 \) and let \( M_g \) be a \( 4g \)-gon with edges consecutively labeled \( a_1^+, b_1^+, a_1^-, b_1^-, a_2^+, b_2^+, a_2^-, b_2^-, \ldots, b_g^- \). The edges are oriented: the \( a_i^+ \) and \( b_i^+ \) clockwise, and the \( a_i^- \) and \( b_i^- \) counter-clockwise. By identifying the edges \( a_i^+ \) and \( a_i^- \), as well as \( b_i^+ \) and \( b_i^- \), according to their orientations, we obtain an orientable surface \( M_g \) of genus \( g \). The polygon \( M_g \) is a canonical polygonal schema for \( M_g \).

Removing the interior of \( M_g \), we obtain a system of \( 2g \) loops (closed curves with distinguished endpoints), all having the same endpoint. This system of loops is a canonical system of loops for \( M_g \). The loop in \( M_g \) obtained by identifying \( a_i^+ \) and \( a_i^- \) is denoted by \( a_i \). Similarly, we have the loops \( b_i \). In the sequel, we assume that an orientable surface \( \mathcal{M} \) is given and we look for a canonical system of loops induced by some canonical polygonal schema; here is an example with the double-torus:

![Double-Torus](image)

Given a surface \( \mathcal{M} \) with boundary, we can extend the definition of canonical system of loops for \( \mathcal{M} \) in the following way. We contract each boundary hole of \( \mathcal{M} \) obtaining a surface \( \tilde{\mathcal{M}} \) without boundary. A system of loops \( (a_1, b_1, a_2, \ldots, b_g) \) in \( \mathcal{M} \) is a canonical system of loops for \( \mathcal{M} \) if no loop intersects the boundary of \( \mathcal{M} \) and the resulting system \( (\tilde{a}_1, \tilde{b}_1, \tilde{a}_2, \ldots, \tilde{b}_g) \) after the contractions is a canonical system of loops for \( \tilde{\mathcal{M}} \).

**Lemma 4.1:** Let \( L = (a_1, b_1, \ldots, b_g) \) and \( L' = (a'_1, b'_1, \ldots, b'_g) \) be two canonical systems of loops for a given orientable surface \( \mathcal{M} \) with or without boundary.
Then, there is a $\partial$-automorphism $\psi$ of $\mathcal{M}$ transforming $L$ to $L'$ (it may not keep$^{10}$ the labels; that is, $a_1$ need not be transformed to $a'_1$, etc.).

Proof. If $\mathcal{M}$ has no boundary, then the lemma immediately follows from the definitions; $a_i$ is mapped to $a'_i$ and $b_i$ to $b'_i$.

If $\mathcal{M}$ has a boundary, we first contract each of the holes, obtaining a surface $\tilde{\mathcal{M}}$. In particular, each hole $H_i$ becomes a point $h_i$. Let $\tilde{L}$ and $\tilde{L}'$ be the resulting canonical systems on $\mathcal{M}$. We find an automorphism $\tilde{\psi}_1$ of $\tilde{\mathcal{M}}$ transforming $\tilde{L}$ to $\tilde{L}'$.

The automorphism $\tilde{\psi}_1$ may or may not be orientation-preserving. If $\tilde{\psi}_1$ preserves the orientation of $\tilde{\mathcal{M}}$, we set $\tilde{\psi}_2 := \tilde{\psi}_1$. If $\tilde{\psi}_1$ reverts the orientation, we set $\tilde{\psi}_2 := \tilde{\psi}_1 \tilde{\psi}$ where $\tilde{\psi}$ is an orientation-reversing automorphism of $\tilde{\mathcal{M}}$ transforming $\tilde{L}$ to $\tilde{L}'$; see Fig. 3. In any case, $\tilde{\psi}_2$ preserves the orientation and maps $L$ to $L'$.

We adjust $\tilde{\psi}_2$ to fix each $h_i$ (this is possible since $\tilde{\mathcal{M}}$ remains connected after cutting along $\tilde{L}'$ and also since the points $h_i$ are disjoint from the loops of $\tilde{L}$). Then we decontract the points $h_i$ back to holes, obtaining $\mathcal{M}$. After this $\tilde{\psi}_2$ induces the required $\partial$-automorphism $\psi$ of $\mathcal{M}$. (The obvious automorphism of $\mathcal{M}$ obtained by decontraction of the holes need not fix boundary; however, it can easily be modified to fix the boundary since $\psi_2$ preserves the orientation.)

$^{10}$ It can be seen from the proof that the labels are either kept or $\psi$ transforms $(a_1, b_1, \ldots, b_g)$ to $(b'_g, a'_g, \ldots, a'_1)$.
We need a theorem of Lazarus et al. [LPVV01] in the following version.

**Theorem 4.2** (cf. [LPVV01, Theorem 1]): Let \( \mathcal{M} \) be a triangulated surface without boundary with total of \( n \) vertices, edges and triangles. Then there is a canonical system of loops for \( \mathcal{M} \) avoiding the vertices of \( \mathcal{M} \) and meeting edges of \( \mathcal{M} \) at a finite number of points such that each loop of the system has at most \( O(n) \) intersections with the edges of the triangulation.

As we already mentioned in the introduction, the result is essentially due to Vegter and Yap [VY90]. Lazarus et al. provide more details ([VY90] is only an extended abstract), and they have a slightly different representation for the canonical system of loops, which is more convenient for our purposes.

From Theorem 4.2 we easily derive the following extension.

**Proposition 4.3:** Let \( \mathcal{M} \) be an orientable surface of genus \( g \) with or without boundary. Let \( D = (\delta_1, \ldots, \delta_n) \) be an almost disjoint system of curves on \( \mathcal{M} \). Then there is a canonical system of loops \( L = (a_1, b_1, \ldots, b_g) \) such that \( D \) and \( L \) have \( O(gn + g^2) \) crossings.

For the proof, we need the following lemma, which may very well be folklore, but which we haven’t managed to find in the literature.

**Lemma 4.4:** Let \( G \) be a nonempty graph with at most \( n \) vertices and edges, possibly with loops and/or multiple edges, embedded in an orientable surface \( \mathcal{M} \) of genus \( g \) without boundary. Then there is a graph \( G' \) without loops or multiple edges and with \( O(g + n) \) vertices and edges that contains a subdivision of \( G \) and triangulates \( \mathcal{M} \).

In the proof below we did not attempt to optimize the constant in the \( O \)-notation. We thank Robin Thomas for a suggestion that helped us to simplify the proof.

**Proof.** We can assume that every vertex is connected to at least one edge; if not, we add loops.

Let us cut \( \mathcal{M} \) along the edges of \( G \). We obtain several components

\[ \mathcal{M}_1, \ldots, \mathcal{M}_q. \]

By Lemma 3.1 we know that

\[ g(\mathcal{M}_1) + \cdots + g(\mathcal{M}_q) \leq g. \]
First, whenever $g(M_i) > 0$ for some $i$, we introduce a canonical system of loops inside $g(M_i)$. For this we need one vertex and $2g(M_i)$ edges, which gives at most $3g$ new vertices and edges in total. In this way we obtain a graph $G^1$ (containing $G$).

We cut $M$ along the edges of $G^1$; the resulting components are all planar. Inside each component $M^1_i$ we introduce a new vertex $v$ and connect it to all vertices on the boundary of $M^1_i$; $v$ can be connected to some boundary vertex $u$ by multiple edges if $u$ occurs on the boundary of $M^1_i$ in multiple copies. This is easily achievable if we consider, up to a homeomorphism, $M^1_i$ as a polygon, possibly with tiny holes inside; see the left picture:

Since we have added at most $\deg u$ edges per vertex $u$ of $G^1$, we obtain a graph $G^2$, still with $O(g + n)$ vertices and edges.

We cut $M$ along the edges of $G^2$. The resulting components $M^2_i$ are all planar and, in addition, they have a single boundary curve. We subdivide each edge of $G^2$ twice, we introduce a new vertex $w$ in each $M^2_i$, and we connect $w$ to all vertices on the boundary of $M^2_i$ (including the vertices obtained from the subdivision). If $w$ is connected to a vertex $u$ of $G^2$ on the boundary of $M^2_i$, we further subdivide the edge $uw$ and we connect the newly introduced vertex to the two neighbors of $u$ along the boundary of $M^2_i$; this is illustrated in the right picture above.

This yields the required graph $G'$. Indeed, we have subdivided all loops and multiple edges in $G^2$, and we do not introduce any new loops or multiple edges (because of the subdivision of $uw$ edges). Each face of $G'$ is triangular; therefore, we have a triangulation. The size of $G'$ is bounded by $O(g + n)$. ■
Proof of Proposition 4.3. If $\mathcal{M}$ contains holes, we contract them, find the canonical system on the contracted surface, and decontract the holes (without affecting the number of crossings). Thus, we can assume that $\mathcal{M}$ has no boundary.

Now we form a graph $G$ embedded in $\mathcal{M}$ in the following way. The vertex set of $G$ contains all endpoints of arcs in $D$. For a closed curve in $D$, we pick a vertex on the curve. Each arc of $D$ induces an edge in $G$. Each closed curve of $D$ induces a loop in $G$. This finishes the construction of $G$.

The graph $G$ has $O(n)$ vertices and edges. Let $G'$ be the graph from Lemma 4.4 containing a subdivision of $G$.

Now we can use Theorem 4.2 for the triangulation given by $G'$ to obtain the required canonical system of loops.

Proof of Proposition 1.3(ii). Let $\mathcal{M}$ be a surface of genus $g$ with $h$ holes. Let $A = (\alpha_1, \ldots, \alpha_m)$ and $B = (\beta_1, \ldots, \beta_n)$ be two almost-disjoint systems of curves. Our task is to find a $\partial$-automorphism $\varphi$ of $\mathcal{M}$ such that $\alpha_1, \ldots, \alpha_m$ and $\varphi(\beta_1), \ldots, \varphi(\beta_m)$ have at most $f_{0,h+1}(m', n')$ intersections, where

$$m' \leq cg(m + g) \quad \text{and} \quad n' \leq cg(n + g)$$

for some constant $c$. Proposition 1.3(ii) then follows from the monotonicity of $f_{g,h}(m, n)$ in $m$ and $n$.

Let $L_\alpha$ be a canonical system of loops as in Proposition 4.3 used with $(\alpha_1, \ldots, \alpha_m)$, and let $L_\beta$ be a canonical system of loops as in Proposition 4.3 used with $(\beta_1, \ldots, \beta_n)$.

According to Lemma 4.1, there is a $\partial$-automorphism $\psi$ of $\mathcal{M}$ transforming $L_\beta$ to $L_\alpha$. This homeomorphism induces a new system of curves

$$B_\psi := (\psi(\beta_1), \ldots, \psi(\beta_n)).$$

We cut $\mathcal{M}$ along $L_\alpha$, obtaining a new, planar surface $\mathcal{M}'$ with $h + 1$ holes (one new hole appears along the cut). According to the choice of $L_\alpha$ and $L_\beta$, the systems $A$ and $L_\alpha$ have at most $O(gm + g^2)$ intersections. Similarly, $B_\psi$ and $L_\alpha$ have at most $O(gn + g^2)$ intersections. Thus, $A$ induces a system $A'$ of $m' \leq cg(m + g)$ new curves on $\mathcal{M}'$, and $B_\psi$ induces a system $B'$ of $n' \leq cg(n + g)$ new curves on $\mathcal{M}'$. From the definition of $f$, we find a $\partial$-automorphism $\varphi'$ of $\mathcal{M}'$ such that $A'$ has at most $f_{0,h+1}(m', n')$ intersections with $\varphi'(B')$. Then we glue $\mathcal{M}'$ back to $\mathcal{M}$, inducing the required $\partial$-automorphism $\varphi$ of $\mathcal{M}$. ■
In this section, we prove Proposition 1.5.

Let $N$ be a nonorientable surface with $h \geq 0$ holes and nonorientable genus $g \geq 1$.

Our approach to prove Proposition 1.5 is similar in spirit to the proof of Proposition 1.3 (ii). The difference is that instead of cutting an orientable surface along a canonical system of loops to get a planar one, we cut the nonorientable surface $N$ along one distinguished closed curve so as to obtain an orientable surface.

We recall that, given a closed curve $\lambda$ on a surface $N$, the surface obtained by cutting $N$ along $\lambda$ is denoted by $N_{(\lambda)}$.

Formally, an orientation-enabling curve in a nonorientable surface $N$ is a properly embedded closed curve $\lambda$ such that $N_{(\lambda)}$ is orientable. It follows that an orientation-enabling curve is non-separating, since attaching two orientable components along a closed curve yields an orientable surface.

It is not hard to see that any nonorientable surface admits an orientation-enabling curve; it can be explicitly found in the convenient representation of the surface introduced in Section 3.3. For technical reasons, however, we will need to find an orientation-enabling curve $\lambda$ that also satisfies two additional properties: $\lambda$ should be compatible with orientations of the boundary curves of the holes in the surface (in a sense to be made precise below), and it should also be compatible with a given system $D$ of curves on $N$, in the sense that we can bound the number of intersections between $\lambda$ and $D$.

The first ingredient for the proof of Proposition 1.5 is an analogue of Lemma 4.1. A perfect analogue would be to show that any two orientation-enabling curves of $N$ can be transformed into one another by a $\partial$-automorphism of $N$. However, it turns out that for nonorientable surfaces with holes this is not true in general; see Example 5.4 below. For this reason, we need the requirement of compatible orientations in the following lemma.

**Lemma 5.1:** Let $N$ be a nonorientable surface with boundary curves $\gamma_1, \ldots, \gamma_h$ and let $\lambda$ and $\kappa$ be two orientation-enabling curves in $N$. Suppose that we have chosen orientations each of the curves $\gamma_1, \ldots, \gamma_h$ and for $\lambda$ and $\kappa$.

Supposed furthermore that the induced orientations of the boundary curves of $N_{(\lambda)}$ are mutually compatible, in the sense explained before Lemma 3.2, and that the same holds for the boundary curves of $N_{(\kappa)}$ (we stress that the
compatibility condition also applies to the boundary curves originating from λ and κ, respectively).

Then there is a ∂-automorphism ψ of $N$ transforming λ to κ.

The second ingredient for the proof of Proposition 1.5 is the following existence result, analogous to Proposition 4.3.

**Proposition 5.2:** Let $N$ be a nonorientable surface of genus $g$ with or without boundary. Let $γ_1, \ldots, γ_h$ be the boundary curves of $N$ given with some orientations. Let $D = (δ_1, \ldots, δ_n)$ be an almost disjoint system of curves on $N$. Then there is an orientation-enabling curve $λ$ such that $D$ and $λ$ have $O(g + n)$ crossings and such that $λ$ can be equipped with an orientation such that the induced orientations of the boundary curves on $N(λ)$ are mutually compatible.

Finally, we will need the following simple lemma that relates the genus and number of holes of $N$ to the corresponding quantities for $N(λ)$.

**Lemma 5.3:** Let $N$ be a nonorientable surface of genus $g$ with $h$ holes and let $λ$ be an orientation-enabling curve. Let $g_λ$ be the (orientable) genus of $N(λ)$ and $h_λ$ be the number of holes of $N(λ)$.

(a) If $g$ is odd, then $λ$ is one-sided, $g_λ = (g - 1)/2$, and $h_λ = h + 1$.

(b) If $g$ is even, then $λ$ is two-sided, $g_λ = (g - 2)/2$, and $h_λ = h + 2$.

**Proof.** Let us recall that we have the following relations for the Euler characteristic: $χ(N) = 2 - g - h$ since $N$ is nonorientable, and $χ(N(λ)) = 2 - 2g_λ - h_λ$ since $N(λ)$ is orientable. We also have

$$χ(N) = χ(N(λ))$$

since the Euler characteristic of the closed curve $λ$ is 0.

If $λ$ is one-sided, then $h_λ = h + 1$, implying $g_λ = (g - 1)/2$. In particular, $g$ must be odd. If $λ$ is two-sided, then $h_λ = h + 2$, implying $g_λ = (g - 2)/2$. In particular, $g$ must be even. This proves the lemma, since we have exhausted all possibilities. ■

Now we are ready to prove Proposition 1.5.

**Proof of Proposition 1.5.** Let $N$ be a nonorientable surface of (nonorientable) genus $g$ with $h$ holes. Let $A = (α_1, \ldots, α_m)$ and $B = (β_1, \ldots, β_n)$ be two almost-disjoint systems of curves. Our task is to find a ∂-automorphism $φ$ of
such that \( \alpha_1, \ldots, \alpha_m \) and \( \varphi(\beta_1), \ldots, \varphi(\beta_m) \) have at most
\[
 f_{g', h'}(c(g + m), c(g + n))
\]
intersections, with
\[
g' = \left\lfloor \frac{(g - 1)}{2} \right\rfloor \quad \text{and} \quad h' = h + 1 + (g \mod 2).
\]

Let us fix orientations of the boundary curves of \( N \) arbitrarily. Let \( \lambda_\alpha \) be an orientation-enabling curve obtained from Proposition 5.2 applied to \( N \) and the system \( A = (\alpha_1, \ldots, \alpha_m) \), and let \( \lambda_\beta \) be an orientation-enabling curve obtained from Proposition 5.2 used for \( N \) and the system \( B = (\beta_1, \ldots, \beta_n) \).

According to Lemma 5.1, there is a \( \partial \)-automorphism \( \psi \) of \( N \) transforming \( \lambda_\beta \) to \( \lambda_\alpha \). This homeomorphism induces a new system of curves \( B_\psi := (\psi(\beta_1), \ldots, \psi(\beta_n)) \).

We cut \( N \) along \( \lambda_\alpha \), obtaining a new, orientable surface \( M \). By Lemma 5.3, \( M \) has genus \( g' \) and \( h' \) holes. By the choice of \( \lambda_\alpha \), the system \( A \) and the (closed) curves \( \lambda_\alpha \) have at most \( O(m + g) \) intersections. Similarly, by our choices of \( \lambda_\beta \) and of \( \psi \), the system \( B_\psi \) and \( \lambda_\alpha = \psi(\lambda_\beta) \) have at most \( O(n + g) \) intersections. Thus, \( A \) induces a system \( A' \) of \( m' \leq c(m + g) \) new curves on \( M \), and \( B_\psi \) induces a system \( B' \) of \( n' \leq c(n + g) \) new curves on \( M \). By the definition of \( f \) and monotonicity, we find a \( \partial \)-automorphism \( \varphi' \) of \( M \) such that \( A' \) has at most
\[
 f_{g', h'}(c(g + m), c(g + n))
\]
intersections with \( \varphi'(B') \).

By the construction, \( \varphi' \) is compatible with the operation of undoing the cutting of \( N \) along \( \lambda_\alpha \), i.e., \( \varphi' \) induces a \( \partial \)-automorphism \( \varphi \) of \( N \), and this \( \varphi \) yields the desired bound on the entanglement number of \( A \) and \( B \).

5.1. Uniqueness of Orientation-Enabling Curves. In this section, we prove Lemma 5.1 (which is fairly easy, using the classification of surfaces). First, however, we briefly digress to describe the promised example that explains why the compatibility assumptions in the lemma are necessary. (The reader may skip this example since it is not used in any of the proofs.)

Example 5.4: Let us consider a fixed nonorientable surface \( N \); for concreteness, let us take the projective plane with 4 holes. We assume that \( N \) is obtained by identifying antipodal points on the boundary of the disk with holes. Let us consider orientation-enabling curves \( \kappa \) and \( \lambda \) as below:
We want to show that there is no $\partial$-automorphism of $N$ transforming $\lambda$ to $\kappa$.

We see that the holes $h_2$ are (locally) on the same side of $\kappa$ whereas they are on different sides of $\lambda$. Let $N'$ be the surface obtained by gluing $h_2$ and $h_3$ according to the indicated orientations. If there is a $\partial$-automorphism transforming $\lambda$ to $\kappa$, then the surfaces $N'_{(\lambda)}$ and $N'_{(\kappa)}$ must be homeomorphic. However, $N'_{(\kappa)}$ is obtained from $N_{(\kappa)}$ by introducing a cross-handle (i.e., two cross-caps) since the orientations of $h_2$ and $h_3$ are compatible on $N_{(\kappa)}$, and thus $N'_{(\kappa)}$ is a nonorientable surface. On the other hand, $N'_{(\lambda)}$ is obtained by introducing a handle (think of moving $h_3$ as the arrow in the picture above indicates). Therefore, $N'_{(\lambda)}$ is orientable. We conclude that there is no $\partial$-automorphism of $N$ transforming $\lambda$ to $\kappa$.

By this approach, if we have $h$ holes, we can construct $2^{h-1}$ different orientation-enabling curves with respect to $\partial$-homeomorphisms. (By an approach similar to the proof of Lemma 5.1, one can actually see that there are exactly $2^{h-1}$ different orientation-enabling curves, but we will not need this in what follows.)

We now proceed to provide the details for the proof of Lemma 5.1.

Proof of Lemma 5.1. Both $N_{(\lambda)}$ and $N_{(\kappa)}$ have the same number of holes and same genus according to Lemma 5.3, and so they are homeomorphic. The idea is that a homeomorphism $\psi'$ of $N_{(\lambda)}$ and $N_{(\kappa)}$ induces the required $\partial$-automorphism $\psi$ of $N$ simply by undoing the operations of cutting $N$ along $\lambda$ and $\kappa$, respectively. We need to be a little careful, however, and to check that $\psi'$ preserves the boundary and is compatible with the gluing.

Let $B_\lambda$ be the part of the boundary of $N_{(\lambda)}$ obtained from $\lambda$ when cutting $N$. According to Lemma 5.3, $B_\lambda$ consists of one or two closed curves, depending of the parity of $g$. We define $B_\kappa$ analogously. We have an involution $i_\lambda$ on $B_\lambda$ such that the identification of all pairs $x$ and $i_\lambda(x)$ yields $N$. We have an
analogous involution $i_\kappa$ on $B_\kappa$. We need a homeomorphism $\psi': \mathcal{N}(\lambda) \to \mathcal{N}(\kappa)$ that is compatible with these involutions (that is, $\psi'i_\lambda = i_\kappa\psi'$ on $B_\lambda$), so that gluing back induces an automorphism of $\mathcal{N}$. We also need that $\psi'$ fixes the other holes so that we obtain a $\partial$-automorphism.

We can define $\psi'$ first on $\partial\mathcal{N}(\lambda)$ so that the requirements above are satisfied. Due to our compatibility assumptions, we can use Corollary 3.4 to get $\psi'$ on the whole $\mathcal{N}(\lambda)$. As we have already mentioned, we obtain the required $\psi$ by gluing back $\mathcal{N}(\lambda)$ and $\mathcal{N}(\kappa)$ to $\mathcal{N}$.

\section*{5.2. Existence of Orientation-Enabling Curves} In this section, we prove Proposition 5.2.

The proof will be subdivided into several steps. As in the proof of Proposition 1.3 (ii), we will replace the given system $D$ of curves by a suitable triangulation of the surface and show that there exists an orientation-enabling curve $\lambda$ in $\mathcal{N}$ that intersects the edges of the triangulation in a controlled way. We will look for $\lambda$ by choosing local orientations of the triangles of (a suitable refinement of) the given triangulation of $\mathcal{N}$. Then $\lambda$ will appear as the “ceasefire line” where the local orientations disagree. This will automatically guarantee that the surface $\mathcal{N}(\lambda)$ obtained after cutting along $\lambda$ is orientable. However, we still have to argue that we can choose the local orientations so that $\lambda$ is a single closed curve, and so that it does not intersect the original triangulation too often. Below we provide the details.

\textit{Local orientations.} Let us assume that $\mathcal{N}$ is a triangulated surface. We equip each triangle with a local orientation (which can be given by a choice of a cyclic order on the vertices of triangle). We say that the orientations of two neighboring triangles are \textbf{coherent} if they are locally both clockwise or both counter-clockwise.\textsuperscript{11}

\textsuperscript{11} Note that we cannot speak of a clockwise or counter-clockwise direction in a global sense on the whole of $\mathcal{N}$ since we expect to work with nonorientable surfaces. However, we still can do this locally on a rather trivial orientable surface consisting of the two triangles.
Given a choice $\omega$ of local orientations on all triangles of $\mathcal{N}$ we create a graph $G_\omega$ embedded in $\mathcal{N}$ consisting of all edges of the triangulation for which the two neighboring triangles are not coherent. (Using the terminology of [MT01], this corresponds to the edges, in the dual graph, of signature $-1$.)

By $\mathcal{N}(\omega)$ we denote the (possibly disconnected) surface obtained from $\mathcal{N}$ by cutting along $G_\omega$. The surface $\mathcal{N}(\omega)$ is orientable by the choice of the cut edges. Therefore, in particular, if $G_\omega$ consists of a single closed curve, then this is an orientation-enabling curve.

Given these preliminaries we can prove the following auxiliary proposition resembling Proposition 5.2 for surfaces without boundary.

**Proposition 5.5:** Let $\mathcal{N}$ be a nonorientable surface without boundary with a fixed triangulation with a total of $n$ vertices and edges. Then there is an orientation-enabling curve avoiding the vertices of $\mathcal{N}$ and meeting the edges of $\mathcal{N}$ in at most $2n$ intersections.

**Proof.** First we create a certain collection of closed curves on $\mathcal{N}$. Let $\omega$ be a choice of local orientations. For every vertex $u$ we pair edges of $G_\omega$ incident to $u$ so that the two edges in every pair are neighbors in the cyclic order. This is possible since each edge corresponds to a change of local orientations and when we travel around $u$ we have to observe an even number of changes. We shorten each edge $\varepsilon$ of $G_\omega$ and shift it a little, obtaining a new edge $\hat{\varepsilon}$ that avoids the edges of the triangulation of $\mathcal{N}$. We connect these shortened edges according to the chosen pairs:
In this way, we obtain a system of closed curves $(\gamma_1, \ldots, \gamma_t)$ (understood as curves in $\mathcal{N}$). Moreover, we can consider this system of curves as $G_\eta$ where $\eta$ is a choice of local orientations of some suitable refinement of the original triangulation of $\mathcal{N}$.

Further, we observe that $G_\eta$ meets each edge of $\mathcal{N}$ at most twice (once next to each vertex of $G_\omega$; we emphasize that by an edge of $\mathcal{N}$ we mean an edge of the original triangulation of $\mathcal{N}$).

If we are lucky and $t = 1$, that is, $G_\eta$ consists of a single closed curve, then we deduce that this curve is the curve we seek and we are done.

If $t > 1$ we still have to modify the local orientations in order to obtain a single closed curve. In this case we will find a further refinement of the triangulation of $\mathcal{N}$ and a choice of local orientations $\vartheta$ such that $G_\vartheta$ consists of $t - 1$ closed curves and $G_\vartheta$ still meets each edge of the original triangulation of $\mathcal{N}$ at most twice. After repeating this step $(t - 1)$ times we obtain the required closed curve.

Let $G^*$ be the graph dual to the triangulation of $\mathcal{N}$. That is, the vertices of $G^*$ are the triangles of $\mathcal{N}$ and the edges of $G^*$ are the pairs of triangles sharing an edge. Let $\tau_1$ and $\tau_2$ be two triangles closest in $G^*$ such that $\tau_1$ contains a part of some curve $\gamma_i$ and $\tau_2$ contains a part of some curve $\gamma_j$ with $i \neq j$ (possibly $\tau_1 = \tau_2$).

We want to connect $\gamma_i$ and $\gamma_j$ with an arc $\delta$ that is minimal in the following sense. First of all we assume that $\delta$ belongs only to triangles of some preselected shortest path between $\tau_1$ and $\tau_2$ in $G^*$. We also assume that it intersects each edge of $\mathcal{N}$ at most once. Finally, we can also assume that $\delta$ intersects $G_\eta$ only in endpoints of $\delta$, for otherwise, we could shorten $\delta$ (this might require changing the indices $i$ or $j$ if $\tau_1 = \tau_2$ and this triangle contains other curve(s) $\gamma_k$). We observe that all the inner triangles on the preselected shortest path between $\tau_1$
and \( \tau_2 \) are disjoint from \( G_\eta \) due to our choice of \( \tau_1 \) and \( \tau_2 \). It follows that if \( \delta \) intersects an edge of \( N \), then this edge is not intersected by \( G_\eta \).

Now we consider two arcs \( \delta_1 \) and \( \delta_2 \) parallel to \( \delta \) (both of them join \( \gamma_i \) and \( \gamma_j \)). We join \( \gamma_i \) and \( \gamma_j \) into a single closed curve \( \gamma' \) along \( \delta_1 \) and \( \delta_2 \):

![Diagram](image1)

After a suitable refinement of the triangulation we change the orientation of the narrow region between \( \delta_1, \delta_2 \) and of the two tiny segments of \( \gamma_i \) and \( \gamma_j \):

![Diagram](image2)

This way we obtain the required new choice of local orientations \( \vartheta \). The corresponding graph \( G_\vartheta \) consist of \( \gamma' \) and all closed curves (cycles) of \( G_\eta \) except \( \gamma_i \) and \( \gamma_j \), that is, it has \( t - 1 \) closed curves as required. In addition, it intersects each edge of \( N \) at most twice due to the choice of \( \delta \). This finishes the proof.

Now we are ready to prove Proposition 5.2.

**Proof of Proposition 5.2.** First we contract all boundary holes \( \gamma_i \) to points \( \hat{\gamma}_i \); in this way, we obtain a surface \( \hat{N} \). We remember the orientation of \( \gamma_i \) as one of two possible directions of how to travel around \( \hat{\gamma}_i \) in some neighborhood of \( \hat{\gamma}_i \) (it does not make sense to consider whether this direction is clockwise or counter-clockwise, since \( N \) is not orientable). We also let \( \hat{D} = (\hat{\delta}_1, \ldots, \hat{\delta}_n) \) be the system of curves on \( \hat{N} \) corresponding to \( D \) on \( N \).
Now we form a graph $G$ embedded in $\hat{N}$ in the following way. The vertex set of $G$ consists of all endpoints of arcs in $\hat{D}$. For a closed curve in $\hat{D}$, we pick a vertex on this curve. Each arc in $\hat{D}$ induces an edge in $G$. Each closed curve in $\hat{D}$ induces a loop in $G$. This finishes the construction of $G$. Note that the $\gamma_i$ are situated either in the vertices of $G$ or in the faces, but not in the interiors of the edges. Also note that no two holes are contracted to the same vertex.

The graph $G$ has $O(n)$ vertices and edges. Let $G'$ be the graph from Lemma 4.4 containing some subdivision of $G$ and having $O(g + n)$ vertices and edges. By possibly perturbing $G'$, we can assume that the $\gamma_i$ are not in the interiors of edges of $G'$.

Using Proposition 5.5 we find an orientation-enabling curve $\lambda_0$ that intersects each edge of $G'$ at most twice. We would like to decontract the holes transforming $\lambda_0$ to $\gamma_0$ on $N$ getting the required curve. However, the problem is that the orientations of curves on $N(\lambda_0)$ may not be compatible as we require. We still have to modify $\lambda_0$. We use an approach similar the to proof of the previous proposition.

Let $G^*$ be the dual graph to $G'$. Let us also equip $\lambda_0$ with some orientation. Note that $\lambda_0$ can be one-sided or two-sided in $N$. In the second case, it is important to observe that the two closed curves originating from $\lambda_0$ on $N(\lambda_0)$ have compatible orientations. (Otherwise, gluing along them would mean introducing a handle, contradicting the non-orientability of $N$.)

Let $\gamma_i$ be a hole such that the orientation of $\gamma_i$ is not compatible with $\lambda_0$ on $N(\lambda_0)$. Let $\tau_1$ be a triangle containing $\gamma_i$ (if $\gamma_i$ is a vertex, it may be contained in several triangles). Let $\tau_2$ be a triangle containing a part of $\lambda_0$ closest to $\tau_1$ in $G^*$. We connect $\lambda_0$ with $\gamma_i$ by an arc $\delta$ minimal in the following sense. We assume that $\delta$ uses triangles of some prescribed shortest path between $\tau_1$ and $\tau_2$. It intersects each edge on this path at most once. It also has no other intersection with $\lambda_0$, for otherwise, it could be shortened.

We ‘pull a finger’ along $\delta$ obtaining a new curve $\lambda_1$: 

\begin{center}
\begin{tikzpicture}
\node[anchor=east] at (-3,0) {$\lambda_0$};
\node[anchor=east] at (-3,-3) {$\gamma_i$};
\node at (0,0) {$\delta$};
\end{tikzpicture}
\quad
\begin{tikzpicture}
\node[anchor=east] at (-3,0) {$\lambda_1$};
\node[anchor=east] at (-3,-3) {$\gamma_i$};
\end{tikzpicture}
\end{center}
After decontractions, we obtain that the resulting $\lambda_1$ and $\gamma_i$ are compatible on $\mathcal{N}_{(\lambda_1)}$. The compatibility of $\lambda_1$ with respect to other boundary curves is not affected.

The curve $\hat{\lambda}_1$ can have more intersections with the edges of $G'$. However, the new intersections appear either on edges that were not intersected previously (at most twice), or, if $\hat{\gamma}_i$ is a vertex, on the edges incident to it.

We can apply this procedure repeatedly, obtaining $\hat{\lambda}_2$, $\hat{\lambda}_3$, etc. After a finite number of steps we obtain a curve $\hat{\lambda}_k$ such that the corresponding $\lambda_k$ is already compatible with all holes on $\mathcal{N}_{(\lambda_k)}$. This curve is our desired curve $\lambda$, since during the procedure we have introduced at most $2|E(G')| + \sum \deg v$ new intersections, where the sum is over all vertices $v$ of $G'$. Thus we are still within the $O(g + n)$ bound.

Acknowledgement. We would like to thank the authors of [GHR13] for making a draft of their paper available to us, and, in particular, T. Huynh for an e-mail correspondence. We also thank an anonymous referee for many valuable comments and in particular for a suggestion of using local orientations in the proof of Proposition 5.2 which replaced our original (longer) homology-based proof.

References


Shellability of the higher pinched Veronese posets

Martin Tancer

Abstract The pinched Veronese poset $V^*_{n}$ is the poset with ground set consisting of all nonnegative integer vectors of length $n$ such that the sum of their coordinates is divisible by $n$ with exception of the vector $(1, \ldots, 1)$. For two vectors $a$ and $b$ in $V^*_{n}$, we have $a \preceq b$ if and only if $b - a$ belongs to the ground set of $V^*_{n}$. We show that every interval in $V^*_{n}$ is shellable for $n \geq 4$. In order to obtain the result, we develop a new method for showing that a poset is shellable. This method differs from classical lexicographic shellability. Shellability of intervals in $V^*_{n}$ has consequences in commutative algebra. As a corollary, we obtain a combinatorial proof of the fact that the pinched Veronese ring is Koszul for $n \geq 4$. (This also follows from a result by Conca, Herzog, Trung, and Valla.)

Keywords Shellable · Pinched Veronese poset · Cohen-Macaulay · Koszul

1 Introduction

In this paper, we focus on the following question: Is every interval in the pinched Veronese poset shellable? (Cohen-Macaulay?) Let us explain this question and its background in detail.

By the $m$-th Veronese poset on $n$ generators, denoted as $(V_{m,n}, \preceq)$, we mean the following poset. Its ground set consists of nonnegative integer vectors of length $n$ such that the sum of their coordinates is divisible by $m$. The partial order on $V_{m,n}$ is given so that $a \preceq b$ if and only if $a$ is less or equal to $b$ in each coordinate. It is not hard to see that every interval in $V_{m,n}$ is shellable and, therefore, Cohen-Macaulay.

If we set $m = n$, we just speak of the $n$-th Veronese poset $V_{n} := V_{n,n}$. We can pinch this poset in the following way. We remove the distinguished vector $j$ which
contains 1 in each coordinate. We also remove order relations between vectors that differ exactly by \( j \) (making them incomparable). In this way, we thus obtain the \( n \)-th "pinched Veronese poset" \( (V_n^\bullet, \preceq) \); see Fig. 1. (More details on this poset are discussed in Sect. 4.) It is very interesting that removing this single element \( j \) (and the corresponding order relations) strongly influences understanding the properties of the poset.

On the algebraic side, it follows that the \( n \)-th pinched Veronese ring is Koszul for \( n \geq 4 \) from a result by Conca et al. [9] (we will discuss this in more detail below). This is equivalent to stating that every interval in \( V_n^\bullet \) is Cohen-Macaulay; see [16, Corollary 2.2]. Later on, Caviglia [7] showed that the third pinched Veronese ring is Koszul. The methods used in [7] are based on computer calculations. Recently, a more general result was found by Caviglia and Conca [8] without the use of computer.

Our task is to focus on the combinatorial side of this question. That is, we focus on shellability of intervals in the pinched Veronese poset remarking that shellability implies Cohen-Macaulayness. We also remark that Cohen-Macaulayness of a poset implies several deep intrinsic properties of the poset, for example certain enumerative properties. The reader is referred, for example, to [1] for more details on Cohen-Macaulayness.

We develop a new method for showing that a certain poset is shellable. Using this method, we are able to prove the following theorem.

**Theorem 1.1** Let \( n \geq 4 \). For any \( z \in V_n^\bullet \), the interval \( [0, z] \) in \( V_n^\bullet \) is a shellable poset, where \( 0 \) is the zero vector of length \( n \).

Note that we do not lose anything by considering intervals \( [0, z] \) only, since an interval \( [a, b] \) is isomorphic to \( [0, b - a] \).

Our motivation for proving Theorem 1.1 can be seen from two sides. On one hand, the pinched Veronese poset is an interesting poset from a combinatorial point of view and it is interesting to understand its combinatorial properties, especially, if its combinatorial properties have further consequences in commutative algebra (see the text at the end of this section).

On the other hand, Theorem 1.1 can be seen as a testing example for a new method for showing that a certain poset is shellable. We establish inductive criteria showing that a certain poset \( P \) is shellable assuming that several subposets of \( P \) are shellable.
and that \( P \) satisfies few other properties. Let us remark that, in general, our method differs from a very standard tool which is lexicographic shellability.

A small drawback of our method is that it requires quite technical case analysis checking that all inductive criteria are satisfied. In this part, the main message for the reader is that the analysis can be done (still, it is fully included in the paper).

1.1 The third pinched Veronese poset

The reader might wonder what is the importance of our assumption \( n \geq 4 \) in Theorem 1.1. The case \( n = 1 \) does not make sense. The case \( n = 2 \) makes the most sense (in relation to the algebraic side of the question) if the elements \((\alpha_1, \alpha_2)\) are further removed from the poset whenever \( \alpha_1 \) and \( \alpha_2 \) are odd. However, in this case, \( V^*_2 \) is isomorphic to \( V_{1,2} \).

The only real issue occurs when \( n = 3 \). In this case, our method, as stated in Sect. 2, does not suffice to prove shellability of \( V^*_3 \). In fact, it is possible to show that some intervals in \( V^*_3 \) are not lexicographically shellable. It turns out that the reason why some intervals in \( V^*_3 \) are not lexicographically shellable also implies limitations for our method. Maybe a further improvement of our method might yield a solution for \( n = 3 \).

1.2 More detailed relation to commutative algebra

Let us fix an integer \( n \) and consider a subset \( A \) of \( \mathbb{N}^n_0 \). For simplicity we assume that the sum of the coordinates of all vectors in \( A \) equals a fixed integer \( m \). Given a commutative field \( \mathbb{k} \), we consider the ring \( \mathbb{k}[A] \) as a subring of \( \mathbb{k}[x_1, \ldots, x_n] \) generated by all monomials \( x^a \) for \( a \in A \) where \( x_1^{a_1} \cdots x_n^{a_n} \) if \( a = (a_1, \ldots, a_n) \).

We can also associate a poset \( P(A) \) to \( A \) in the following way. We let \( \Lambda \) consist of those vectors in \( \mathbb{N}^n_0 \) that are nonnegative integer combinations of vectors from \( A \) (including zero). Then, we set \( P(A) = (\Lambda, \leq_A) \) where \( a \leq_A b \) if and only if \( b - a \in A \).

Cohen-Macaulayness of intervals in \( P(A) \) is related to the Koszul property of \( \mathbb{k}[A] \) in the following way.

**Proposition 1.2** ([16, Corollary 2.2]) The ring \( \mathbb{k}[A] \) is Koszul if and only if every interval in \( P(A) \) is Cohen-Macaulay over \( \mathbb{k} \).

The reader is referred, for example, to [11] for more information about the importance of the Koszul property.

If we set \( A_{m,n} \) to consist of all vectors in \( \mathbb{N}^n_0 \) whose coordinates sum to \( m \), we get \( P(A_{m,n}) = V_{m,n} \). Similarly, if we set \( A^*_n \) to \( A_{n,n} \setminus \{j\} \), we get \( P(A^*_n) = V^*_n \). Thus, we have the following corollary of Theorem 1.1 and Proposition 1.2.

**Corollary 1.3** \( \mathbb{k}[A^*_n] \) is Koszul for any \( n \geq 4 \).

As we mentioned above, Corollary 1.3 also follows from the result of Conca et al. [9], and thus our contribution for the algebraic side is a combinatorial proof of this corollary.
For completeness, we explain how to derive Corollary 1.3 from Corollary 6.10 (2) in [9]. We set \( I \) to be the ideal \((x_1^2, \ldots, x_n^2)\) in \( k[x_1, \ldots, x_n] \). It is generated by a regular sequence since \( x_i^2 \) is a nonzero divisor in \( k[x_1, \ldots, x_n]/(x_1^2, \ldots, x_i^2) \). Setting \( d = 2, e = 1, c = n - 2 \), and \( r = n \) in Corollary 6.10 (2) from [9], we get that \( k[I_n] \) is Koszul where \( k[I_n] \) is generated by all monomials of degree \( n \) belonging to \( I \); that is, \( k[I_n] = k[A_n] \).

Very recently, Vu [17] proved a general result that for \( m, n \geq 2 \) and \( x \in A_m, n \), the ring \( k[A_m, n \{x\}] \) is Koszul unless \( m \geq 3 \) and \( x \) is \((0, \ldots, 0, 2, m - 2)\) or one of its permutations (this result also includes Corollary 1.3).

### 1.3 Further related work

Here, we very briefly mention further related work. We keep several terms undefined in this paragraph. The reader is welcome to consult the cited sources for more details. Eisenbud et al. [10] showed that the \( m \)-th Veronese subrings of \( k[z_1, \ldots, z_t] / I \) are Koszul where \( I \) is a homogeneous ideal and \( m \) is large enough (more precisely when \( m \geq \text{reg}(I)/2 \) where \( \text{reg}(I) \) is Castelnuovo-Mumford regularity of \( I \)). Further investigation of Koszulness of \( k[z_1, \ldots, z_t] / I \) can be found in [6, 12, 13, 16] in the context where the generators \( z_i \) correspond to monomials \( x^a \) as above and \( I \) records the syzygies between the monomials (and then \( k[z_1, \ldots, z_t] / I \simeq k[A] \)).

### 1.4 Structure

In Sect. 2, we explain our new method for showing shellability. In Sect. 3, we prove the correctness of the method. Section 4 serves as a preliminary section on properties of the (pinched) Veronese posets. In Sect. 5, we prove Theorem 1.1. Finally, in Sect. 6, we compare the strength of our shellability method (mainly) with standard chain-lexicographic shellability of Björner and Wachs [4]. If the reader is more interested in the shellability criteria rather than Theorem 1.1, we highly recommend to read Sect. 6 right after Sect. 2. Here, we offer the graph of the dependency of the sections:

![Graph dependency of sections]

The dashed arrow between Sects. 3 and 5 means that Sect. 3 is not necessary for understanding Sect. 5; however, the correctness of the proof in Sect. 5 is based on Sect. 3.

### 2 Method for showing shellability

In this section, we describe our main tools for the proof of Theorem 1.1. We need to set up some preliminaries first.
2.1 Poset preliminaries

Let $P = (P, \leq)$ be a graded poset with rank function $\text{rk}$. By $\hat{0}$ we mean the unique minimal element of $P$ (if it exists), and similarly, by $\hat{1}$ we mean the unique maximal element (if it exists). For $a, b \in P$, we say that $a$ covers $b$, $a \succ b$, if $a > b$ and there is no $c$ with $a > c > b$. Equivalently, $a > b$ and $\text{rk}(a) = \text{rk}(b) + 1$. Pairs of elements $a, b$ with $a \succ b$ are also known as edges in the Hasse diagram of $P$. Atoms are elements that cover $\hat{0}$. That is, atoms are elements of rank 1 in a poset that contains $\hat{0}$.

From now on, let us assume that $P$ contains a unique minimal element. Let $A$ be a set of some atoms in $P$. By $P \langle A \rangle = (P \langle A \rangle, \leq)$ we mean the induced subposet of $P$ with the ground set

$$P \langle A \rangle = \{\hat{0}\} \cup \{b \in P : b \geq a \text{ for some } a \in A\}.$$ 

2.2 Shellability

Now, we assume that $P$ contains both a unique minimal and a unique maximal element. Let $C(P)$ be the set of maximal chains of $P$. A shelling order is an order of chains from $C(P)$ satisfying the following condition.

(Sh) If $c'$ and $c$ are two chains from $C(P)$ such that $c'$ appears before $c$, then there is a chain $c^*$ from $C(P)$ appearing before $c$ such that $c \cap c^* \supseteq c \cap c'$ and also $c$ and $c^*$ differ in one level only (that is, $|c \Delta c^*| = 2$ where $\Delta$ denotes the symmetric difference).

A poset $P$ is shellable if it admits a shelling order. This is equivalent with saying that the order complex of $P$ is shellable (as a simplicial complex).

2.3 A-shellability

Now, let us assume that $A = (A, \leq^o)$ is a partially ordered set of some atoms in $P$. We say that $P \langle A \rangle$ is A-shellable if $P \langle A \rangle$ is shellable with a shelling order respecting the order on $A$. That is, if $c$ and $c'$ are two maximal chains on $P \langle A \rangle$ and the unique atom of $c'$ appears before the unique atom of $c$ in the $\leq^o$ order, then $c'$ appears before $c$ in the shelling. \footnote{For purposes of Theorem 1.1, it would be fully sufficient to consider $\leq^o$ as a linear order (a.k.a. total order). However, we use partial orders, because nothing new has to be done to obtain more general criteria with partial orders, and we believe that for some further applications partial orders might be important.}

2.4 Using A-shellability

Let $P$ be a poset for which we aim to show that $P$ is shellable (in our application $P = V_n^*$). Let us order all the atoms of $P$ into a sequence $a_1, \ldots, a_t$. For $k \in [t]$, let us set $A_k := \{a_1, \ldots, a_k\}$ and consider $A_k$ as a partially ordered set with the order $a_1 \leq^o a_2 \leq^o \cdots \leq^o a_k$. We would like to prove inductively that $P \langle A_k \rangle$ is...
Let us assume that we are able to perform the first induction step, that is, to show $A_1$-shellability of $P\langle A_1 \rangle$, and let us focus on the second induction step. We will provide two criteria, Theorems 2.1 and 2.2 below, how to prove $A_{k+1}$-shellability of $P\langle A_{k+1} \rangle$ assuming $A_k$-shellability of $P\langle A_k \rangle$.

This technique is quite similar to the technique using recursive atom orderings defined by Björner and Wachs [5] and a comparison of these two techniques is discussed in Sect. 6. In particular, the second criterion (Theorem 2.2) is set up in such a way that it covers the case of recursive atom orderings. However, the technique presented here allows more freedom. In particular, it allows to combine different criteria to achieve the task.

One technical issue is the following. In our application for the pinched Veronese posets, it is not enough to consider the induction steps along a single ordering $a_1 \leq \cdots \leq a_t$ of the atoms of $P$. If we aimed on a single ordering only, we would not have strong enough induction assumption to achieve the task. Thus, we will rather focus on many orderings of the atoms. For considering more orderings simultaneously, it pays off to set up a third criterion, Theorem 2.3, which allows to “restrict” an $A$-shelling of $P\langle A \rangle$ to an $A'$-shelling of $P\langle A' \rangle$ where $A'$ is a subset of $A$.

2.5 Necessity of the criteria

In our approach, the first criterion, Theorem 2.1, seems to be the most important. The remaining two theorems could, perhaps, be circumvented; however, they will simplify our analysis.

2.6 Setting up the criteria

To set up conditions in the criteria, we need some additional notation. We fix some partially ordered set $A = (A, \leq^0)$ of atoms of $P$ and a further atom $a^+$ which is not in $A$. Think of $A = A_k$ and $a^+ = a_{k+1}$ when comparing with the sketch above (it is more convenient to use a notation independent of the index $k$).

We set $A^+ := A \cup \{a^+\}$ and $Q := P\langle A^+ \rangle \setminus P\langle A \rangle$. The partial order on $A^+$, which we again denote by $\leq^0$, extends $\leq^0$ on $A$ so that $a^+ \geq^0 a$ for any $a \in A$. We also consider $Q = (Q, \leq)$ as a subposet of $P$ with the unique minimal element $a^+$ (it does not need to have a unique maximal element).

For $q \in Q$, we set $I(q)$ to be the interval $[q, \hat{1}]$. Elements of $P$ that cover $q$ are atoms of $I(q)$. By $A(q)$ we denote the set of (all) atoms of $I(q)$ which simultaneously belong to $P\langle A \rangle$. By $A^{\text{all}}(q)$ we denote the set of all atoms of $I(q)$. In particular, note that the poset $I(q)\langle A(q) \rangle$ is well defined (we will need this poset later on).

2.7 Edge falling property

Let $q \in Q$. We say that $q$ has the edge falling property if for every $p \in P\langle A \rangle$ with $p > q$ and every $q' \in Q \cup \{\hat{0}\}$ with $q > q'$ there is $p' \in P\langle A \rangle$ such that $p > p' > q'$.

See Fig. 2.
The edge falling property. The $P \langle A \rangle - Q$ edge $pq$ falls by one level to $p' q'$.

Important subposets appearing in the conditions of Theorem 2.1. We also recall the edge falling property by a little diamond between $P \langle A \rangle$ and $Q$.

2.8 Shellability criteria

Now, we can state our first criterion; see also Fig. 3.

**Theorem 2.1** (Criterion I) The poset $P \langle A^+ \rangle$ is $A^+$-shellable if the following conditions are satisfied.

(i) $P \langle A \rangle$ is $A$-shellable;
(ii) for every $q \in Q$ the interval $[a^+, q]$ is shellable;
(iii) every $q \in Q$ has the edge falling property; and
(iv) for every $q \in Q$ the poset $I(q) \langle A(q) \rangle$ is shellable.

The second criterion is similar to the first one; however, it focuses more on the structure of the interval $I(a^+)$ rather than on the structure of $Q$. See also Fig. 4.

**Theorem 2.2** (Criterion II) The poset $P \langle A^+ \rangle$ is $A^+$-shellable if the following conditions are satisfied.

(i) $P \langle A \rangle$ is $A$-shellable;
(ii) there is a linear order on $A^{all}(a^+)$ such that the elements of $A(a^+)$ appear before other elements in this order and such that $I(a^+) = I(A^{all}(a^+))$ is $A^{all}(a^+)$-shellable (with respect to this order); and
(iii) for every $q \in Q$ and for every $p \in P \langle A \rangle$ if $p \succ q$, then $p \in I(a^+) \langle A(a^+) \rangle$.

The third criterion that we provide below differs from the previous two. In this case, we rather reduce $A$ to $A'$ instead of enlarging it.
Theorem 2.3 (Criterion III) Let $A'$ be a subset of $A$, linearly ordered with the order inherited from $A$. The poset $P(A')$ is $A'$-shellable if the following conditions are satisfied.

(i) $P(A)$ is $A$-shellable; and
(ii) for every $b \in A \setminus A'$ and for every $p \in P(A')$ with $p > b$, there is $b' \in A'$ appearing before $b$ in $A$ such that $b' \in A'$ and $p > b'$ (see Fig. 5).

The proofs of all three criteria are given in Sect. 3.

We conclude this section by remarks about the differences in the criteria above and their comparison to lexicographic shellability.

2.9 Relation between Criterion I and Criterion II

A reader might check that Theorem 2.1 “almost” follows from Theorem 2.2. More precisely, it is not hard to see that conditions (i) and (iii) of Theorem 2.2 easily follow from the assumptions of Theorem 2.1. The main difference is that condition (ii) of Theorem 2.2 does not immediately follow from the assumptions of Theorem 2.1. (Assuming that the conditions of Theorem 2.1 are satisfied, we can immediately deduce that $I(a^+)\langle A(a^+) \rangle$ is shellable by setting $q = a^+$ in condition (iv) of Theorem 2.1; however, we do not have shelling of whole $I(a^+)$ yet).

Actually, the essence of the proof of Theorem 2.1 can be seen as verifying condition (ii) of Theorem 2.2 from conditions (ii), (iii) and (iv) of Theorem 2.1, which is solely a property of a certain decomposition of the interval $I(a^+)$. The interested
reader is welcome to formulate the criteria on extension of a shelling of $I(a^+)\langle A(a^+)\rangle$ to a shelling of whole $I(a^+)$ separately, following the proof of Theorem 2.1.

### 2.10 Relation of lexicographic shellability and $A$-shellability

A very standard notion for showing that a certain poset is shellable is the so-called (chain lexicographic) CL-shellability, introduced by Björner and Wachs [4], or even a further generalization, the so-called (chain compatible) CC-shellability introduced by Kozlov [14], still induced by a lexicographic order on chains. It is natural to ask what is the relation between $A$-shellability introduced here and lexicographic shellability (we will focus on CL-shellability only; some ideas can be carried for CC-shellability as well). We discuss this relation in more detail separately in Sect. 6, and the reader interested in these details is encouraged to read Sect. 6 immediately (perhaps after finishing this section). Questions addressed in Sect. 6 have arisen in discussions with Anders Björner and Afshin Goodarzi. Here, we briefly survey these questions.

It is not hard to see that every lexicographically shellable poset is $A$-shellable where $A$ is the set of all atoms equipped with an appropriate linear order. On the other hand, it is not hard to find an $A$-shellable poset (again with $A$ consisting of all atoms) which is not lexicographically shellable.

We can also ask more subtle questions about the relative power of Theorems 2.1 and 2.2 compared with lexicographic shellability. (We skip Theorem 2.3 since it is of a different spirit.)

The conditions of Theorem 2.2 are analogous to the conditions on recursive atom orderings from [5], and in particular, Theorem 2.2 preserves lexicographic shellability (if the “shellable” assumptions are changed into “lexicographically shellable”) as well as lexicographically shellable posets satisfy the conditions of Theorem 2.2. The added value of Theorem 2.2 appears when we use it with nonlexicographic assumptions.

Regarding Theorem 2.1, let us (again) consider the following two questions: whether a lexicographically shellable poset satisfies the criteria of Theorem 2.1; and whether lexicographic shellability is kept by the criteria of Theorem 2.1 (for linearly ordered $A$).

The answer to the first question is no. The answer to the second question is not known to the author. We just remark that the proof of Theorem 2.1 might produce nonlexicographic shelling even if all posets in the conditions of Theorem 2.1 are assumed to be lexicographically shellable (not even a CC-shelling). We again refer to Sect. 6 for more details.

The above-mentioned remarks suggest that $A$-shellability using Theorem 2.1 and lexicographic shellability are perhaps in “generic position” regarding applicability in various situations.

### 3 Proofs of shellability criteria

Here, we prove Theorems 2.1, 2.2, and 2.3. We keep the notation introduced in the previous section.
Fig. 6 Three cases when $c'$ appears before $c$

Below, we also set up an additional notation common to proofs of Theorems 2.1 and 2.2. Let $C := C(P\langle A\rangle)$ and $C^+ := C(P\langle A^+\rangle)$ be the sets of maximal chains in $P\langle A\rangle$ and $P\langle A^+\rangle$. We know that $P\langle A\rangle$ is shellable; therefore, there is some shelling order $c_1, c_2, \ldots, c_t$ of all chains from $C$ (note that $P\langle A\rangle$ contains both $\hat{0}$ and $\hat{1}$). We are going to describe a shelling order on $C^+$. In both cases, we start with $c_1, \ldots, c_t$ and then we continue with chains containing $a^+$. This way, if we show that we have a shelling order, it will immediately be an $A^+$-shelling.

3.1 Proof of Theorem 2.1

We choose some order $q_1, \ldots, q_u$ of elements of $Q$ such that $i \leq j$ if $\text{rk}(q_i) \leq \text{rk}(q_j)$. In particular, $q_1 = a^+$. For every $q_i \in Q$, we have an order of maximal chains in the interval $[a^+, q_i]$ inducing a shelling of this interval, by condition (ii).

Now, we describe a shelling order of all maximal chains from $C^+ \setminus C$. (We already have an order on $C$.) Let $c$ be a chain from $C^+ \setminus C$, and the index $i(c)$ is denoted in such a way that $q_{i(c)}$ is the element of $c \cap Q$ with the largest rank. Note that if $r \in c$, $r \neq \hat{0}$, and $\text{rk}(r) < \text{rk}(q_{i(c)})$, then $r \in Q$.

Now, let $c$ and $c'$ be two different chains from $C^+ \setminus C$ and we want to describe when $c'$ is before $c$.

The first criterion is whether $i(c') < i(c)$. That is, if $i(c') < i(c)$, then $c'$ is sooner in the order than $c$ (and symmetrically $c'$ is later if $i(c') > i(c)$); see Fig. 6, on the left.

If $i(c) = i(c')$, then we have the following second criterion. Let $q = q_{i(c)} = q_{i(c')}$. We look at the two maximal chains $\tilde{c} = c \cap [a^+, q]$ and $\tilde{c}' = c' \cap [a^+, q]$ in the interval $[a^+, q]$. As we sooner realized, if $\tilde{c} \neq \tilde{c}'$, then there is order of these chains inducing a shelling on $[a^+, q]$. This induces the order of $c$ and $c'$; see Fig. 6, in the middle. If $\tilde{c} = \tilde{c}'$, we need a third criterion.

Now, we assume that $i(c) = i(c')$ and $\tilde{c} = \tilde{c}'$. The element $q$ is defined as above. We set $\bar{c} = c \cap I(q)$ and $\bar{c}' = c' \cap I(q)$ recalling that $I(q)$ is the interval $[q, \hat{1}]$. Both chains $\bar{c}$ and $\bar{c}'$ are maximal chains in $I(q)\langle A(q)\rangle$ due to the choice of $q = q_{i(c)} = q_{i(c')}$. The condition (iv) in the statement of the theorem implies that $I(q)\langle A(q)\rangle$ is shellable.
We set that \( c' \) appears before \( c \) in our shelling if and only if \( \overline{c'} \) appears before \( \overline{c} \) in the shelling of \( I(q)\langle A(q) \rangle \); see Fig. 6, on the right.

We have described an order of chains in \( C^+ \). Now, we have to prove that it is indeed a shelling order. That is, we have to prove condition (Sh). In the sequel, we, therefore, assume that \( c \) and \( c' \) are given, as in (Sh), and we seek for \( c^* \).

If \( c \in C \), then we first required \( c^* \) immediately from shellability of \( P \langle A \rangle \). In the sequel, we assume \( c \in C^+ \setminus C \) and we set \( q = q_{i(c)} \). We distinguish several cases.

1. \( q \not\in c' \).

   In this case we use the edge falling property. Let \( q' \) be the element of \( c \) such that \( q > q' \) and \( p \) be the element of \( c \) such that \( p > q \). The edge falling property implies that there is \( p' \in P \langle A \rangle \) such that \( p > p' > q' \). We set up \( c^* = (c \cup \{ p' \}) \setminus \{ q \} \). Obviously, \( c^* \) satisfies the required properties.

2. \( q \in c' \), and \( 
\overline{c} = c \cap [a^+, q] \) and \( \overline{c'} = c' \cap [a^+, q] \).

   By their definition, \( c' \) appears before \( c \), and thus due to the first criterion we have that \( i(c') \leq i(c) \). Now, since \( q \in c' \), it follows that \( i(c') = i(c) \) and, therefore, \( q = q_{i(c')} \) (that is, \( q \) is the element of \( c' \cap Q \) of the highest rank). In addition, due to the second criterion, we know that \( c' \) appears before \( c \) in the shelling of \( [a^+, q] \). Therefore, there is a maximal chain \( c^* \) in \( [a^+, q] \) appearing before \( c \) which coincides with \( \overline{c} \) with exception of one level and such that \( \overline{c} \cap c^* \supseteq \overline{c} \cap \overline{c'} \). We set \( c^* \) so that it coincides with \( \overline{c^*} \) on \( Q \) and with \( c \) on \( P \langle A \rangle \).

3. \( q \in c' \), and \( \overline{c} = \overline{c'} \).

   We again have \( q = q_{i(c')} \). Hence, the third criterion on comparison of \( c \) and \( c' \) applies. That is, \( \overline{c} \) appears before \( \overline{c} \) in the shelling of \( I(q)\langle A(q) \rangle \). Similarly, as in the previous case, there is, therefore, a maximal chain \( \overline{c^*} \) in \( I(q)\langle A(q) \rangle \) appearing before \( \overline{c} \) which coincides with \( \overline{c} \) with exception of one level and such that \( \overline{c} \cap \overline{c^*} \supseteq \overline{c} \cap \overline{c'} \) (recall that \( \overline{c} = c \cap I(q) \) and \( \overline{c'} = c' \cap I(q) \)). We set \( c^* \) so that it coincides with \( c \) on \( Q \) and with \( \overline{c^*} \) on \( P \langle A \rangle \).

We have verified condition (Sh) in all cases. This concludes the proof of Theorem 2.1.

3.2 Proof of Theorem 2.2

In this case, it is easier to set up the order of shelling \( C^+ \setminus C \). (Let us recall that the order on \( C \) is already set up and that the chains from \( C^+ \setminus C \) will follow after the chains from \( C \).)

   Every chain \( c \in C^+ \setminus C \) contains \( a^+ \). Let \( \overline{c} \) be in this case \( c \cap I(a^+) \). We set that \( c' \) precedes \( c \) if and only if \( \overline{c'} \) precedes \( \overline{c} \) in the shelling from condition (ii) of the statement of the theorem.

   Now, we need to verify condition (Sh) to be sure that we have indeed a shelling order. Similarly, as in the proof of previous theorem, we assume that \( c \) and \( c' \) are given, as in (Sh), and we seek for \( c^* \). We distinguish several cases.

1. \( c \in C \).

   In this case, we know that \( c' \) appears before \( c \) and thus \( c' \in C \). Therefore, we can find suitable \( c^* \) from the shellability of \( P \langle A \rangle \).
2. $c \in C^+ \setminus C$ and $c' \in C^+ \setminus C$.

In this case, $\tilde{c}'$ appears before $\tilde{c}$; therefore, there is $\tilde{c}^*$ from shelling of $I(a^+)$ such that $\tilde{c}$ and $\tilde{c}^*$ differ in one level only and that $\tilde{c}^* \cap \tilde{c} \supseteq \tilde{c}' \cap \tilde{c}$. We set $c^* = \tilde{c}^* \cup \{\hat{0}\}$. This choice of $c^*$ obviously satisfies the required properties.

3. $c \in C^+ \setminus C$, $c' \in C$, and $c \cap A(a^+) \neq \emptyset$.

Let $b \in c \cap A(a^+)$. Then, there is $a \in A$ such that $b > a$ due to the definition of $A(a^+)$. Let us set $c^* := (c \{a^+\}) \cup \{a\}$. Then, $c^* \cap c \supseteq c' \cap c$ since $c'$ misses $a^+$. See Fig. 7, on the left.

4. $c \in C^+ \setminus C$, $c' \in C$, and $c \cap A(a^+) = \emptyset$.

As usual, let $q$ be the largest element of $c \cap Q$. Let $p$ be the element of $c \cap P\langle A \rangle$ such that $p > q$. See Fig. 7, on the right. Condition (iii) in the statement of the theorem implies that there is a maximal chain $c'_2$ in the interval $[a^+, p]$ such that $c'_2 \cap A(a^+) \neq \emptyset$. Let $\tilde{c}'_2$ be the maximal chain in $I(a^+)$ which agrees with $c'_2$ on $[a^+, p]$ and which agrees with $c$ on $[p, \hat{1}]$. Note that $\tilde{c}'_2$ precedes $\tilde{c}$ in the shelling of $I(a^+)$ since $\tilde{c}'_2 \cap A(a^+) \neq \emptyset$ whereas $c \cap A(a^+) = \emptyset$. Therefore, by (Sh), there is a chain $\tilde{c}^*$ in $I(a^+)$ which agrees with $\tilde{c}$ in all levels but one and which satisfies $\tilde{c}^* \cap \tilde{c} \supseteq \tilde{c}'_2 \cap c$. In particular, $\tilde{c}^*$ agrees with $\tilde{c}$ on $p$ and all elements above $p$. Now, we set $c^* := \tilde{c}^* \cup \{\hat{0}\}$. We have that $c^* \cap c \supseteq c' \cap c$ since $c' \cap c \subseteq P\langle A \rangle$.

This finishes the proof of Theorem 2.2.

3.3 Proof of Theorem 2.3

Let $C = C(P\langle A \rangle)$ and $C' = C(P\langle A' \rangle)$ be the sets of maximal chains of $P\langle A \rangle$ and $P\langle A' \rangle$. We have that $C' \subseteq C$. Since $P\langle A \rangle$ is $A$-shellable, we have a shelling order on $C$ respecting $A$. We simply inherit this order on $C'$. It respects $A'$; however, we have to show that it is indeed a shelling order.

Let $c$ and $c'$ be chains in $P\langle A' \rangle$ as in condition (Sh). We look for a suitable $c^*$ from (Sh).

Chains $c$ and $c'$ also belong to $P\langle A \rangle$. Since we started with a shelling on $C$, there is $c^{**} \in C$ such that $c^{**} \cap c \supseteq c' \cap c$ and $c^{**}$ differs from $c$ in one level. If $c^{**}$ belongs to $C'$, we set $c^* := c^{**}$ and we are done.
Now, let us assume that $c^{**} \not\in C'$. Let $b$ and $p$ be the elements of $c^{**}$ of rank 1 and 2, respectively, in particular $p > b$. Since $c^{**} \not\in C'$, it follows from the definition of $C'$ that $b \in A \setminus A'$. Moreover, $c$ and $c^{**}$ differ in only one level. Therefore, they differ in level 1 and $p \in c$. This implies that $p \in P\langle A'\rangle$. By applying now assumption (ii) of the theorem for elements $b$ and $p$, we conclude that there is $b' \in A'$ appearing before $b$ in $A$ such that $p > b'$. Let us set $c^* := (c^{**}\setminus\{b\}) \cup\{b'\}$. Then, $c^*$ appears before $c^{**}$ in the shelling of $C$ and hence also before $c$. In addition, $c^*$ and $c$ have to differ in level 1 (only) by definition of $c^*$. Thus, we obtain $c^* \cap c = c^{**} \cap c \supseteq c' \cap c$ as required.

This finishes the proof of Theorem 2.3.

4 Preliminaries on the (pinched) Veronese poset

The $n$-th Veronese poset $(\mathcal{V}_n, \leq)$ is given by

$$\mathcal{V}_n = \{(\alpha_1, \ldots, \alpha_n) \in \mathbb{N}_0^n : \alpha_1 + \cdots + \alpha_n \equiv 0 \pmod{n}\}$$

and $a \leq b$ for $a = (\alpha_1, \ldots, \alpha_n)$, $b = (\beta_1, \ldots, \beta_n)$ if and only if $\alpha_i \leq \beta_i$ for $i \in [n]$.

In the sequel, we often write $a = \alpha_1\alpha_2\alpha_3$ instead of $a = (\alpha_1, \alpha_2, \alpha_3)$ and so on for higher $n$. We can also use brackets to separate coordinates in expressions such as $(\alpha_1 + 1)01\alpha_4$ instead of $(\alpha_1 + 1, 0, 1, \alpha_4)$.

In slightly more general setting, for positive integers $m$ and $n$ we also define

$$\mathcal{V}_{m,n} = \{(\alpha_1, \ldots, \alpha_n) \in \mathbb{N}_0^n : \alpha_1 + \cdots + \alpha_n \equiv 0 \pmod{m}\}.$$  

We again have that $a \leq b$ if $a$ is less than or equal to $b$ in every coordinate. In particular, we have $\mathcal{V}_n = \mathcal{V}_{n,n}$.

The $n$-th pinched Veronese poset $(\mathcal{V}_n^*, \leq)$ is a (noninduced) subposet of $\mathcal{V}_n$ given by the following data.

$$\mathcal{V}_n^* = \{a \in \mathcal{V}_n : a \neq j\}.$$  

Here, $j = 1 \cdots 1$. The partial order on $\mathcal{V}_n^*$ is given by $a \leq b$ if $a \leq b$ and $b - a \neq j$.

We also define $0 = 0 \cdots 0$ to be the minimal element of $\mathcal{V}_n^*$.

4.1 Arithmetic operations on $\mathcal{V}_n$ and $\mathcal{V}_n^*$

We consider elements of $\mathcal{V}_n$ and $\mathcal{V}_n^*$ as vectors in $\mathbb{Z}^n$. We can then sum and subtract these vectors. For a set $X \subseteq \mathbb{Z}^n$ and vector $v \in \mathbb{Z}^n$, we let $X \oplus v$ to be the set $\{x + v : x \in X\}$. Similarly, $X \ominus v := \{x - v : x \in X\}$. Let $[0, z]$ be an interval in $\mathcal{V}_n^*$ and $x \in [0, z]$. In our considerations, we will often use the fact that $[x, z]$ and $[0, z - x]$ are isomorphic; more precisely, $[0, z - x] = [x, z] \ominus x$.

4.2 Shellability of intervals in $\mathcal{V}_{m,n}$

It is not hard to observe, using known results, that every interval in $\mathcal{V}_{m,n}$ is shellable. We will actually need this for considering the pinched version, and thus we provide full details.
Proposition 4.1 Let $m$ and $n$ be positive integers. For any $z \in V_{m,n}$, the interval $[0, z]$ in $V_{m,n}$ is a shellable poset.

Proof We have that $V_{m,n}$ is a subposet of $V_{1,n}$. We first observe that $[0, z]$ is shellable as an interval in $V_{1,n}$ and then we deduce that $[0, z]$ is shellable as an interval in $V_{m,n}$ as well.

It is not hard to observe that $[0, z]$ as an interval in $V_{1,n}$ is a graded modular lattice: By modular we mean that $\text{rk}(a) + \text{rk}(b) = \text{rk}(a \lor b) + \text{rk}(a \land b)$.

If $a = \alpha_1 \cdots \alpha_n$ and $b = \beta_1 \cdots \beta_n$, then

$$a \lor b = \max(\alpha_1, \beta_1) \cdots \max(\alpha_n, \beta_n)$$

and

$$a \land b = \min(\alpha_1, \beta_1) \cdots \min(\alpha_n, \beta_n).$$

These relations easily imply modularity of $V_{1,n}$. Therefore, $V_{1,n}$ is shellable by [2, Theorem 3.7] (semimodular would be sufficient).

The fact that $V_{m,n}$ is shellable follows from the fact that the shellability is preserved by rank selections. Indeed, if we start with $[0, z]$ as an interval in $V_{1,n}$, we remove elements exactly in levels not divisible by $m$ in order to turn it into an interval in $V_{m,n}$. This means that we remove the same number of elements from every maximal chain. Therefore, $V_{m,n}$ is shellable by [3, Theorem 11.13]. \hfill \square

5 Proof of Theorem 1.1

The task of this section is to prove Theorem 1.1. Throughout this section, we assume that $n \geq 4$ is fixed.

5.1 The induction mechanism

Let $A^{\text{all}}$ be the set of all atoms of $V_n^{\bullet}$. We will consider several linear orders on $A^{\text{all}}$ and some of its subsets. Let $x = \xi_1 \cdots \xi_n \in \mathbb{Z}^n$. For $\ell \in [n]$ we set $x^{(\ell)} = \xi_\ell \cdots \xi_n$. We also set $A^{(\ell)}$ to be the subset of $A^{\text{all}}$ made of all $x \in A^{\text{all}}$ such that $x^{(\ell)} \neq 0 \cdots 0$.

We consider two linear orders, $<^L$ and $<^S$ on $A^{\text{all}}$.

The first order is the lexicographic order given in the following way. Let $s = \sigma_1 \cdots \sigma_n$ and $t = \tau_1 \cdots \tau_n$. We set $s <^L t$ if and only if there is $j \in [n]$ such that $\sigma_i = \tau_i$ for $i < j$ and $\sigma_j < \tau_j$.

The second order is a specific order which we describe now. We set $A^S := A^{(n)} \setminus \{1 \cdots 102\}$. The smallest elements in $<^S$ order are the elements of $A^S$ sorted

---

2 It can be computed that $|A^{\text{all}}| = \binom{2n-1}{n} - 1$; however, we will not need to know this value explicitly.
The elements of $A_{\text{all}}\setminus A^{(n)}$ follow sorted again by the $<_L$ order. The reader is referred to Table 1 for more concrete comparison of these orders (for $n = 4$).

We will need to work with the following ordered sets. Let $a^L_i$ be the $i$th smallest element of $A_{\text{all}}$ in the $<_L$ order and similarly $a^S_i$ be the $i$th smallest element in the $<_S$ order. We then set $A^L_k := \{a^L_1, \ldots, a^L_k\}$ and $A^S_k := \{a^S_1, \ldots, a^S_k\}$. We also set $A_k^{(\ell)}$ to be the set of the first $k$ elements of $A^{(\ell)}$ in the $<_L$ order (this time, we omit the superscript $L$ for simpler notation).

Now, let $I = [0, z]$ be any interval in $V^*_n$. Our task is to show that $I$ is shellable. In order to explain our next step, let us use the following simplification of notation. Let $A$ be some set of atoms of $I$ equipped with the $<_L$ order (resp. with the $<_S$ order). Instead of saying that $I \langle A \rangle$ is $A$-shellable we say that $I \langle A \rangle$ is $(L)$-shellable (resp. $I \langle A \rangle$ is $(S)$-shellable). This simplifies the notation when our typical $A$ will be of form $A_k^{(\ell)} \cap I$. In addition, it also explicitly emphasizes whether $A$ is equipped with the $<_L$ order or the $<_S$ order.

Our task will be to prove the assertions below. The first two assertions depend on $k \leq |A_{\text{all}}|$. The third assertion depends on $\ell \in [n-1]$ and $k \leq |A^{(\ell+1)}|$. The poset $I \langle A^L_k \cap I \rangle$ is (L)-shellable (if nonempty).

$$(A^L_k) \text{ The poset } I \langle A^L_k \cap I \rangle \text{ is } (L)\text{-shellable if nonempty.}$$

$$(A^S_k) \text{ The poset } I \langle A^S_k \cap I \rangle \text{ is } (S)\text{-shellable if nonempty.}$$

$$(A_k^{(\ell+1)}) \text{ The poset } I \langle A_k^{(\ell+1)} \cap I \rangle \text{ is } (L)\text{-shellable.}$$

**Proposition 5.1** Let $I = [0, z]$ be any interval in $V^*_n$. Then, assertions $(A^L_k)$ and $(A^S_k)$ are valid for any positive integer $k \leq |A_{\text{all}}|$ and assertion $(A_k^{(\ell+1)})$ is valid for any $\ell \in [n-1]$ and any positive integer $k \leq |A^{(\ell+1)}|$.
Theorem 1.1 follows from the proposition by setting $k = |A^{\text{all}}|$ in $(A^{L}_k)$ (or $A^{S}_k$)).

The task is to prove Proposition 5.1 by a double induction. The first (inner) induction is slightly unusual—we first prove $(A^{L}_k)$ by induction in $k$ (see Lemmas 5.2, 5.3, and 5.4 below), then we prove $(A^{S}_k)$ by induction in $k$ (see Lemmas 5.2, 5.3, and 5.5 below), and finally, we prove $(A^{L}_k)$ already assuming $(A^{L}_{k'})$ directly with no induction (see Lemma 5.6 below). The fact that we use the induction is also the reason why we need to prove all assertions $(A^{L}_k)$, $(A^{S}_k)$, and $(A^{L}_{k'})$, although only $(A^{L}_k)$ is sufficient for deducing Theorem 1.1. We need the induction assumption strong enough so that the induction works well.

We also remark that $I$ does not need to contain all atoms from $A^{\text{all}}$ (for example, if the first coordinate of $z$ is zero). This is why we need to consider, for example, $(L)$-shellability of $I\langle A^{L}_k \cap I \rangle$ instead of (possibly expected) $(L)$-shellability of $I\langle A^{L}_k \rangle$.

For improved readability, we decompose the induction step into several lemmas, with different approaches on how to prove them. From now on, we assume that $z$ and $I = [0, z]$ are fixed.

**Lemma 5.2** Let us assume that Proposition 5.1 is valid for every interval $[0, y]$ with $\text{rk}(y) < \text{rk}(z)$. Then, $I\langle A^{L}_k \cap I \rangle$ is $(L)$-shellable and $I\langle A^{S}_k \cap I \rangle$ is $(S)$-shellable (if they are nonempty), that is, $(A^{L}_k)$ and $(A^{S}_k)$ are valid.

**Lemma 5.3** Let us assume that Proposition 5.1 is valid for every interval $[0, y]$ with $\text{rk}(y) < \text{rk}(z)$. Then, $I\langle A^{L}_k \cap I \rangle$ is $(L)$-shellable and $I\langle A^{S}_k \cap I \rangle$ is $(S)$-shellable (if they are nonempty), that is, $(A^{L}_k)$ and $(A^{S}_k)$ are valid.

**Lemma 5.4** Let $k \in \{3, \ldots, |A^{\text{all}}|\}$. Let us assume that Proposition 5.1 is valid for every interval $[0, y]$ with $\text{rk}(y) < \text{rk}(z)$. Let us also assume that $(A^{L}_k)$ is valid for the interval $I = [0, z]$ and for $k' < k$. Then, $I\langle A^{L}_k \cap I \rangle$ is $(L)$-shellable (if nonempty), that is, $(A^{L}_k)$ is valid.

**Lemma 5.5** Let $k \in \{3, \ldots, |A^{\text{all}}|\}$. Let us assume that Proposition 5.1 is valid for every interval $[0, y]$ with $\text{rk}(y) < \text{rk}(z)$. Let us also assume that $(A^{S}_k)$ is valid for the interval $[0, z]$ and for $k' < k$. Then, $I\langle A^{S}_k \cap I \rangle$ is $(S)$-shellable (if nonempty), that is, $(A^{S}_k)$ is valid.

**Lemma 5.6** Let $\ell \in [n - 1]$ and $k \in \{1, \ldots, |A^{(\ell+1)}|\}$. Let us assume that $(A^{L}_k)$ is valid for the interval $I = [0, z]$ and for $k' = |A^{\text{all}}|$. Then, $I\langle A^{L}_k \cap I \rangle$ is $(L)$-shellable (if nonempty), that is, $(A^{L}_k)$ is valid.

We remark that Lemma 5.3 implies Lemma 5.2. Similarly, Lemmas 5.4 and 5.5 together imply Lemma 5.3. The reason why we state Lemmas 5.2 and 5.3 separately is that Lemma 5.2 is used in the proof of Lemma 5.3, and this one is used in the proofs of Lemmas 5.4 and 5.5.

Assuming the validity of the lemmas, we immediately obtain a proof of Proposition 5.1 as described just below the statement of the proposition. Therefore, it is sufficient to prove the lemmas.
5.2 Proofs of Lemmas 5.2-5.6

Proof of Lemma 5.2 Let $A := A_1^L = A_1^S = \{ a^+ \}$ where $a^+ = 0 \cdots 0 n$. We also assume that $a^+ \in I$ otherwise we encounter the “empty” case. Thus, $(L)$-shellability of $I(A_1^L \cap I)$ and $(S)$-shellability of $I(A_1^S \cap I)$ coincide with the usual shellability of $I(A)$ (since $A$ contains a single atom). We easily observe that the interval $[a^+, z]$ is shellable, since it is isomorphic to $[0, z - a^+]$, and $[0, z - a^+]$ is shellable by our assumption. It follows that $I(A)$ is shellable by extending every maximal chain of $[a^+, z]$ by $\{ 0 \}$ and considering the same order of maximal chains as for shelling $[a^+, z]$.

For the proof of a next lemma, the following claim will be useful.

Claim 5.7 Let $u = \omega_1 \cdots \omega_n \cdot 0$ be a nonzero element of $\mathcal{V}_n^\bullet$ with the last coordinate 0, or $u = 1 \cdots 102$. Then, there is $v \in A^S$ such that $v < u + 1 \cdots 102$.

Proof If $u = 1 \cdots 102$, we can set $v := 1 \cdots 1003$, for example.

Further, we assume $u \neq 1 \cdots 102$. Let $i$ be such that $\omega_i \geq 1$ while we prefer $i \neq n - 1$ if possible; and furthermore, if we meet the first preference, we prefer $\omega_i \neq 2$ if possible.

If we meet both preferences, we set $v := 1 \cdots 121 \cdots 101$ where the “2” occurs in the $i$th position. In particular, $v \in A^S$. We also have $u + 1 \cdots 102 - v = \omega_1 \cdots \omega_i - 1(\omega_i + 1) \omega_{i+1} \cdots \omega_{n-1} 1$, which is different from $j$ since $\omega_i \neq 2$. That is, $u + 1 \cdots 102 > v$.

If we meet the first preference only, then we still set $v := 1 \cdots 121 \cdots 101$ where the “2” occurs on the $i$th position. This time we conclude $u + 1 \cdots 102 - v \neq j$ by realizing that there is $j \neq i, n - 1$ such that $\omega_j \neq 1$ (here, we use $n \geq 4$).

Finally, if we meet no preference, then $u = 0 \cdots 0(r \cdot n) 0$ for some integer $r$. In this case, we set $v = 1 \cdots 1021$ and we have $u + 1 \cdots 102 - v = 0 \cdots 01(rn - 2) 1 \neq j$.

Proof of Lemma 5.3 We have $a_1^L = a_1^S = 0 \cdots 0 n$ and $a_2^L = a_2^S = 0 \cdots 01(n - 1)$. We set $A := \{ a_1^L \}$ and $A^+ := \{ a_1^L, a_2^S \}$. With this setting, our only task is to show that $I(A^+ \cap I)$ is $(L)$-shellable (which coincides with $(S)$-shellability). We can assume that $a_2^L \in I$, otherwise $A^+ \cap I$ coincides with $A \cap I$ and we conclude by Lemma 5.2. We can also assume that $a_1^L \in I$; otherwise $A^+ \cap I$ contains a single atom only and we obtain $(L)$-shellability of $I(A^+ \cap I)$ in the same way as in the proof of Lemma 5.2.

Altogether, we assume $a_1^L, a_2^L \in I$ and, therefore, our task simplifies to showing $(L)$-shellability of $I(A^+)$. We are going to use Theorem 2.2 for this task. For consistent notation, we set $Q := I(A^+) \setminus I(A)$ and $a^+ = a_2^L$ (we prefer using bold $a^+$ rather than $a^+$ in Theorem 2.2 emphasizing that $a^+ \in \mathcal{V}_n^\bullet$). We also recall that $I(a^+) = [a^+, z]$ and $A^\text{all}(a^+)$ is the set of all atoms of $I(a^+)$ whereas $A(a^+)$ is the set of only those atoms of $I(a^+)$ which belong to $I(A)$ as well. We need to check the conditions of Theorem 2.2.

The first condition, $A$-shellability of $I(A)$, just follows from Lemma 5.2.

For checking the remaining two conditions, we need more intrinsic description of $Q$. Note that in our notation $(q - a^+)^{(n)}$ denotes the last coordinate of $q - a^+$. Consult Fig 8 while following the proof of the next claim and the rest of the proof of the lemma.
Claim 5.8 We have the following description of $Q$.

$$Q = \{ q \in I \langle a^+ \rangle : q \succeq a^+, (q - a^+)^{(n)} = 0 \text{ or } q - a^+ = 1 \cdots 102 \}. $$

Proof If $q \in Q$, then it must satisfy $q \succeq a^+$ and our task is to determine whether $q \in Q$.

Let us first consider the case $(q - a^+)^{(n)} = 0$. Then, $q^{(n)} = (a^+)^{(n)} = n - 1$, and, therefore, $q \not\succeq a^L = 0 \cdots 0n$. We conclude $q \in Q$ since $q \not\in I \langle A \rangle$.

Now, let us consider the case $(q - a^+)^{(n)} \geq 1$. Then, $q \geq a^L$. We deduce $q \geq a^L$ unless $q = a^L + j$. That is, $q \not\in Q$ unless $q = (0 \cdots 0n) + (1 \cdots 1) = 1 \cdots 1(n + 1)$. In this case, $q - a^+ = 1 \cdots 102$. \hfill \Box

Using Claim 5.8, it is easy to check the second condition in Theorem 2.2.

We first observe that Claim 5.8 implies the following description of $A(a^+)$:

$$A(a^+) = \{ p' \in I : p' - a^+ \in A^S \}. \quad (1)$$

Indeed, $A(a^+)$ consists of those elements in $I$ covering $a^+$ which do not belong to $Q$. By Claim 5.8 and the definition of $A_0$, we obtain that $A(a^+)$ consists of those elements $p' \in I$ covering $a^+$ such that $p' - a^+ \in A^S$. This immediately yields the required description (1) since if $p' - a^+ \in A^S$, then $p' \succ a^+$.

Now, by the assumptions of the lemma, the interval $[0, z - a^+]$ is $(S)$-shellable. This interval is isomorphic to $I \langle a^+ \rangle$ by adding $a^+$. Therefore, using (1), this isomorphism induces a shelling of $I \langle a^+ \rangle$ required by condition (ii) of Theorem 2.2.

Finally, we want to check condition (iii) of Theorem 2.2. Therefore, we are given $q \in Q$ and $p \in I \langle A \rangle$ such that $p$ covers $q$. Our task is to show that $p \in I \langle a^+ \rangle \langle A(a^+) \rangle$. Recalling (1), our task is to show that there is $p' \in I$ such that $p' - a^+ \in A^S$ and $p \succeq p'$. Note that the condition $p' \in I$ follows from $p \succeq p'$, and thus we do not need to check it in the following verification separately.

A natural candidate for $p'$ is the element $p'_{\text{cand}} := a^+ + (p - q)$. We have $p'_{\text{cand}} \leq p$ since $p - p'_{\text{cand}} = q - a^+$ and $q \succeq a^+$. Furthermore, $p'_{\text{cand}} - a^+ = p - q$; therefore, we are done if $p - q \in A^S$. See Fig. 9, on the left.
Fig. 9 Verifying condition (iii) of Theorem 2.2. If \( p - q \) does not belong to \( A^S \) (on the left), then we need to distinguish two further cases (in the middle and on the right). Label of an edge (a path) \( st \) is given by \( t - s \).

It remains to consider \( p - q \not\in A^S \). In this case, we have to choose \( p' \) different from \( p'_{\text{cand}} \). We further distinguish two cases whether \( q - a^+ = 1 \cdots 102 \) or \( (q - a^+)_{(n)} = 0 \) (which is sufficient due to Claim 5.8 using \( q \in Q \)) while we keep in mind that \( p - q \not\in A^S \). See Fig. 9, in the middle and on the right.

1. First, let us assume that \( q - a^+ = 1 \cdots 102 \).
   
   We let \( u := (p - q) \). In particular, either \( u = 1 \cdots 102 \), or \( u_{(n)} = 0 \) since \( p - q \not\in A^S \). By Claim 5.7, there is \( v \in A^S \) such that \( v < u + 1 \cdots 102 \). Let \( p' := v + a^+ \). Then, \( p' - a^+ \in A^S \) and also \( p' < p \) since \( p - p' = (p - q) + (q - a^+) - v = u + 1 \cdots 102 - v \) and \( u + 1 \cdots 102 \succ v \).

2. Now, we assume \( (q - a^+)_{(n)} = 0 \). Since \( p \not\in Q \), Claim 5.8 implies that \( (p - a^+)_{(n)} > 0 \) (and \( p - a^+ \not\in 1 \cdots 102 \)). Therefore, \( (p - q)_{(n)} > 0 \). Since \( p - q \not\in A^S \), we conclude \( p - q = 1 \cdots 102 \). (This also implies that \( q \not\in a^+ \)).

Now, let \( u := q - a^+ \). By Claim 5.7, there is \( v \in A^S \) such that \( u + 1 \cdots 102 \succ v \).

We set \( p' := v + a^+ \). Then, \( p' - a^+ \in A^S \) and also \( p > p' \) since \( p - p' = (p - q) + (q - a^+) + v = 1 \cdots 102 + u - v \succ 0 \).

We have checked all conditions of Theorem 2.2. This concludes the proof of the lemma.

The following claim will be useful for the proof of the next lemma. Item (ii) of the claim is trivial; however, it will be useful to refer to it as stated in the claim.

**Claim 5.9**  
(i) Let \( a \in A^{\text{all}} \) such that \( a \not\in a^L_1 \). Then, there is \( a' \in A^{\text{all}} \) such that \( a' \lessdot^L a \) and \( a' \lessdot a + j \). In addition, we can require \( a' \not= 1 \cdots 102 \).

(ii) Let \( a := a^L_1 \). Then, \( a' \lessdot a + j \) for \( a' = a^L_2 \).

**Proof** Let us start with item (i). Let \( a = 0 \cdots 0\alpha_\ell \cdots \alpha_n \) where \( \alpha_\ell \neq 0 \). That is, we require \( a' \lessdot 1 \cdots 1(\alpha_\ell + 1) \cdots (\alpha_n + 1) \). We have \( \ell \leq n - 1 \) since \( a \not\in a^L_1 = 0 \cdots 0n \).

Let \( b := 0 \cdots 0(\alpha_\ell - 1)\alpha_{\ell+1} \cdots \alpha_{n-1}(\alpha_n + 1) \). If \( b \not= j, 1 \cdots 102 \), then \( b <^L a \), and thus we can set \( a' := b \). (Note that \( b \leq a + j \) and \( b + j \neq a + j \) implying \( b < a + j \).)

If \( b = j \), then \( a = 21 \cdots 10 \) and we can, for example, set \( a' = 1 \cdots 120 \).

If \( b = 1 \cdots 102 \), then \( a = 21 \cdots 101 \) and we can set \( a' = 1 \cdots 1201 \).
Item (ii) is trivial just since by definition of $a_1^L$ and $a_2^L$ we have $a_1^L = 0 \cdots 0n$ and $a_2^L = 0 \cdots 01(n - 1)$.

**Proof of Lemma 5.4** We set $A := A_{k-1}^L \cap I$ and $A^+ := A_k^L \cap I$; we also set $a^+ = a_k^L$. Our task is to show that if $A^+$ is nonempty, then $I \langle A^+ \rangle$ is $(L)$-shellable.

We can assume that $a^+ \in I$ otherwise $(L)$-shellability of $I \langle A^+ \rangle$ coincides with $(L)$-shellability of $I \langle A \rangle$ which we conclude from the assumptions of the lemma (if $A^+ \neq \emptyset$).

We can also assume that $A \neq \emptyset$, otherwise $I \langle A^+ \rangle$ has a single atom only and we derive the lemma analogously as Lemma 5.2. In particular, from the assumptions of the lemma, we know that assertion $(A_k^L)$ is valid, and, therefore, we have that $I \langle A \rangle$ is $(L)$-shellable.

Our task is to use Theorem 2.1 for verifying $(L)$-shellability of $I \langle A^+ \rangle$. We set $Q := I \langle A^+ \rangle \setminus I \langle A \rangle$. We need to verify assumptions of Theorem 2.1.

We have already observed that item (i) of Theorem 2.1 is satisfied; that is, that $I \langle A \rangle$ is $(L)$-shellable.

For verifying other items, we need more intrinsic definition of $Q$. We will assume that $a^+ = 0 \cdots 0 \alpha_\ell \alpha_{\ell+1} \cdots \alpha_n$ where $\ell$ is the smallest integer such that $\alpha_\ell > 0$. Note that $\ell \leq n - 1$ since $k \geq 3$.

**Claim 5.10** We have the following description of $Q$.

(i) $Q = \{q \in I \langle A^+ \rangle : q \geq a^+, (q - a^+)^{(\ell+1)} = 0 \cdots 0\} \quad \text{if } a^+ \neq 201 \cdots 1$; and

(ii) $Q = \{q \in I \langle A^+ \rangle : q \geq a^+, (q - a^+)^{(2)} \in \{0 \cdots 0, 10 \cdots 0\}\} \quad \text{if } a^+ = 201 \cdots 1$.

**Proof** First, we assume that $a^+ \neq 201 \cdots 1$ and we want to prove item (i). Let $q \geq a^+$. Our task is to determine whether $q \in Q$. We also let $q - a^+ = \kappa_1 \cdots \kappa_n$.

We need to show two inclusions.

- For the first one, we assume that $(q - a^+)^{\ell+1} \neq 0 \cdots 0$, and we want to show that $q \notin Q$. That is, we want to find an atom from $A$ which is below $q$. In this case, we have $i \in \{\ell + 1, \ldots, n\}$ such that $\kappa_i \neq 0$. Let

  $$a := 0 \cdots 0(\alpha_\ell - 1)\alpha_{\ell+1} \cdots \alpha_{i-1}(\alpha_i + 1)\alpha_{i+1} \cdots \alpha_n.$$ 

  We have $0 < a < q$.

  If $a \neq j$ and $q - a \neq j$, then $0 < a < q$, and thus $a$ is the required atom of $A$ since $a$ precedes $a^+$ in the $<_L$ order.

  If $a = j$, then $a^+ = 21 \cdots 101 \cdots 1$ where the “0” appears in the $i$th position ($i \geq 3$ since $a^+ \neq 201 \cdots 1$). In particular, if $a' = 201 \cdots 1$, then $q \geq a'$ (since $\kappa_i \geq 1$) and $a'$ precedes $a^+$ in the $<_L$ order. Therefore, $a'$ is the required atom of $A$ unless $q = a' + j = 312 \cdots 2$. In this case, we can use $1 \cdots 102$ for example.

  If $q - a = j$ and $a \neq j$, we consider $a' = a + j = q$ obtained from Claim 5.9. We also have $a' <_L a^+$. This follows from Claim 5.9 (i) by $a' <_L a <_L a^+$ if $a \neq a_1^L$. It follows from Claim 5.9 (ii) if $a = a_1^L$ since $a' <_L a_2^L \leq a$.
• For the second inclusion, we assume that \((q - a^+)_{\ell+1} = 0 \cdots 0\) and we need to show that \(q \in Q\); that is, we need to show that \(a \nleq q\) for any \(a \in A\).

Let \(a = \alpha'_1 \cdots \alpha'_n \in A\). Since \(a <^L a^+\), we have that \(\alpha'_1 = \cdots = \alpha'_{\ell-1} = 0\) and \(\alpha'_\ell \leq \alpha_\ell\). This implies that there is \(i \in \{\ell + 1, \ldots, n\}\) such that \(\alpha'_i > \alpha_i\) (note that \(\alpha_1 + \cdots + \alpha_n = n = \alpha'_1 + \cdots + \alpha'_n\) since both \(a^+\) and \(a\) are atoms; note also that we get a strict inequality since \(a^+ \neq a\)). This implies \(q \nleq a\) since \(q\) and \(a^+\) agree in the \(i\)th position.

Now, we want to prove item (ii). That is, we assume that \(a^+ = 201 \cdots 1\). Similarly, as before, let \(q \succeq a^+\). Our task is to determine whether \(q \in Q\). We also let \(q - a^+ = \kappa_1 \cdots \kappa_n\). We again need to show two inclusions.

• For the first one, we assume that \((q - a^+)^{(2)} \notin \{0 \cdots 0, 10 \cdots 0\}\), and we want to show that \(q \nleq Q\).

If we apply the reasoning from item (i), we obtain that \(q \nleq Q\) if \(\kappa_i > 0\) for some \(i \geq 3\).

It remains to consider the case \((q - a^+)^{(2)} = \kappa_2 0 \cdots 0\) where \(\kappa_2 \geq 2\). In this case, we set \(a = 021 \cdots 1\). Thus, \(q > a\). In addition, \(q \nleq a + j\) since \((q - a^+)(2) = \kappa_2 0 \cdots 0\). Thus, \(q > a\). We also have \(a <^L a^+\), and, therefore, \(q \nleq Q\).

• For the second inclusion, we assume that \((q - a^+)^{(2)} \in \{0 \cdots 0, 10 \cdots 0\}\) and we need to show that \(a \nleq q\) for any \(a \in A\).

Let \(a = \alpha'_1 \cdots \alpha'_n \in A\). Since \(a <^L a^+\), we have that \(\alpha'_1 \leq 2\). This implies that either is \(i \in \{3, \ldots, n\}\) such that \(\alpha'_i > \alpha_i = 1\), or \(\alpha'_2 > \alpha_2 = 0\) and \(\alpha'_i = \alpha_i = 1\) for \(i \geq 3\).

In the first case, we have \(q \nleq a\) since \(q + a^+\) agree in the \(i\)th position. In the second case, we have \(\alpha_2 \geq 2\) since \(a \neq j\). Therefore, again \(q\) does not agree in the second position at most by 1.

Now, we verify condition (ii) of Theorem 2.1. Let \(J = [a^+, q]\) be an interval where \(q \in Q\). We recall that \([a^+, q]\) is isomorphic to \([0, q - a^+]\).

If \(a^+ \neq 201 \cdots 1\), then by Claim 5.10, \(J\) is isomorphic to an interval in \(\mathcal{V}_{n, \ell}\) (by forgetting last \(n - \ell\) coordinates of \(J \ominus a^+\)). Therefore, \(J\) is shellable by Proposition 4.1.

If \(a^+ = 201 \cdots 1\), then \(Q\) has a very simple structure by Claim 5.10; see Fig. 10. We could check that every interval in \(Q\) in this case is a modular lattice and deduce shellability of \(Q\) in the same way as in Proposition 4.1, using [2, Theorem 3.7]. However, this is perhaps just an overkill in this case and the shelling order of every interval can be easily found explicitly.

We continue with the verification of condition (iii) of Theorem 2.1; that is, we verify the edge falling property. Let \(q \in Q\), \(q' \in Q \cup \{0\}\) and \(p \in I(A)\) be such that \(p > q\) and \(q > q'\). Our task is to find \(p' \in I(A)\) such that \(p > p' > q'\).

Natural candidate for \(p'\) is \(p'_{\text{cand}} := q' + (p - q)\). We have \(p > p'_{\text{cand}} > q'\). If \(a^+ \neq 201 \cdots 1\), we immediately obtain that \(p'_{\text{cand}} \in I(A)\) from Claim 5.10 (i) as follows. We know that \((q - q')^{(\ell+1)} = 0 \cdots 0\) by Claim 5.10 (i) since \(q - q' = (q - a^+) - (q' - a^+)\).
Therefore, $p^{(\ell+1)} = (p_c^\ell)^{(\ell+1)}$, and it follows by Claim 5.10 (i) that $p_c^\ell$ indeed belongs to $I(A)$. Therefore, we can set $p' := p_c^\ell$.

If $a^+ = 201 \cdots 1$, we need to be more careful. We have $p^{(2)} - (p_c^{\ell})^{(2)} = q^{(2)} - (q')^{(2)}$. Therefore, if $q^{(2)} = (q')^{(2)}$, then we obtain $p_c^{\ell} \in I(A)$ by Claim 5.10 (ii) and we can set $p' := p_c^{\ell}$. However, it might also occur that $(q - a^+(2)) = 10 \cdots 0$ and $(q' - a^+(2)) = 0 \cdots 0$ by Claim 5.10 (ii). In this case, we focus on $(p - q)^{(2)}$. Claim 5.10 (ii) implies that $(p - q)^{(2)} \neq 0 \cdots 0$. If $(p - q)^{(2)} \neq 10 \cdots 0$, then $p_c^{\ell} \in I(A)$ again by Claim 5.10 (ii) and we can again set $p' := p_c^{\ell}$.

Finally, it remains to consider the case $(p - q)^{(2)} = 10 \cdots 0$. In this case, $p_c^{\ell} \in Q$ and we have to choose $p'$ differently. We actually obtain $p - q = (n - 1)10 \cdots 0$ since $p > q$. Similarly, we obtain $q - q' = (n - 1)10 \cdots 0$. We can then choose $p' := q' + (n - 2)20 \cdots 0$. Then, $p > p' > q'$ and $p' \in I(A)$ by Claim 5.10. See Fig. 11.

We conclude by verifying condition (iv) of Theorem 2.1. Let $q \in Q$; we need to show that the poset $I(q)(A(q))$ is shellable where $A(q)$ is defined as in the statement of the theorem. We observe that this poset is isomorphic with $I(q)(A(q)) \ominus q$, that is, with $[0, z - q]\langle A(q) \ominus q \rangle$. Note that $rk(z - q) < rk(z)$. Here, we plan to use our assumption that Proposition 5.1 is valid for intervals $[0, y]$ with $rk(y) < rk(z)$, in particular, for the interval $[0, z - q]$. Therefore, we want to determine $A(q) \ominus q$.

Let $a \in A^\text{all}$; we want to determine whether $a \in A(q) \ominus q$. This is equivalent with determining whether $q + a \in A(q)$ and using the definition of $A(q)$ with determining whether $q + a \in I(A)$ (assuming that $q + a \in I(q)$, otherwise $a \notin A(q) \ominus q$).
If \(a^+ \neq 201 \cdots 1\), we get that \(q + a \in I(A)\) if and only if \(a \in A(\ell + 1)\) and \(q + a \in I(A)\) by Claim 5.10 (i). Therefore, we obtain the required shellability of \([0, z - \mathbf{q}])\langle A(q) \ominus q\rangle\) from assertion \((A_k^{(\ell+1)})\) (with \(k = |A(\ell+1)|\)) for the interval \([0, z - \mathbf{q}])\).

If \(a^+ = 201 \cdots 1\) and \((q - a^+)^{(2)} = 1 \cdots 0\), then \(q + a^+ \in I(A)\) if and only if \(a \in A(2)\) and \(q + a \in I(A)\) by Claim 5.10 (ii). Therefore, we obtain the required shellability of \([0, z - q])\langle A(q) \ominus q\rangle\) from assertion \((A_k^{(2)})\) (with \(k = |A(2)|\)) for the interval \([0, z - q])\).

If \(a^+ = 201 \cdots 1\) and \((q - a^+)^{(2)} = 0 \cdots 0\), then \(q + a^+ \in I(A)\) if and only if \(a \in A(\ell + 1)\) and \(q + a \in I(A)\) by Claim 5.10 (ii). Therefore, we obtain the required shellability of \([0, z - q])\langle A(q) \ominus q\rangle\) from assertion \((A_k^{(\ell)})\) (with \(k = |A(\ell)|\)) for the interval \([0, z - q])\).

This covers all cases when \(a^+ = 201 \cdots 1\) by Claim 5.10 (ii). Thus, we have verified condition (iv) of Theorem 2.1 which concludes the proof of the lemma.

For the proof of the next lemma, we need the following extension of Claim 5.9.

**Claim 5.11** Let \(\ell \in [n - 1]\). Let \(a \in A^\text{all}\) such that \(a \neq a^L\). Then, there is \(a' \in A(\ell+1)\) such that \(a' <^L a\) and \(a' < a + j\). In addition, we can assume \(a' \neq 1 \cdots 102\).

**Proof** By Claim 5.9 (i) we have \(b'' \in A^\text{all}\) (playing the role of \(a'\) in Claim 5.9) such that \(b'' <^L a\) and \(b'' < a + j\) and \(b'' \neq 1 \cdots 102\). If \(b'' \in A(\ell+1)\), then we set \(a' := b''\) and we are done.

If \(b'' \neq A(\ell+1)\), then \(b'' := \beta_1 \cdots \beta_{n-1} 0\) for some \(\beta_1, \ldots, \beta_{n-1} \geq 0\). Let \(i \in [n - 1]\) be such that \(\beta_i \neq 0\) and \(\beta_i\) is as small as possible. We set \(a' := \beta_1 \cdots \beta_{i-1} (\beta_i - 1) \beta_{i+1} \cdots \beta_{n-1}\). We have that \(a' \neq j\) due to our choice that \(\beta_i\) is as small as possible. Thus, \(a' <^L b'' <^L a\). In addition, \(a' < a + j\) since \(a' \leq a + j\) (\(a'\) is dominated by \(b''\) in the first \(n - 1\) coordinates and dominated by \(j\) in the last coordinate) and \(a' \neq a\).

Finally, \(a' \in A(\ell+1)\) and \(a' \neq 1 \cdots 102\) since its last coordinate is 1.

**Proof of Lemma 5.5** The proof is similar to the proof of Lemma 5.4. It is only slightly more technical, since the \(<^S\) order is more complicated than the \(<^L\) order.

We set \(A := A^S_{k-1} \cap I\) and \(A^+ := A^S_k \cap I\); we also set \(a^+ = a^S_k\). Our task is to show that if \(A^+\) is nonempty then \(I(A^+)\) is \((S)\)-shellable.

Similarly, as in the proof of Lemma 5.4, we derive that we can assume \(a^+ \in I\), \(A \cap I \neq \emptyset\) and, therefore, \(I(A)\) is \((S)\)-shellable from the assumptions of this lemma.

Our task is to use Theorem 2.1 for verifying \((S)\)-shellability of \(I(A^+)\). We set \(Q := I(A^+) \setminus I(A)\). We need to verify assumptions of Theorem 2.1.

We have already observed that item (i) of Theorem 2.1 is satisfied; that is, that \(I(A)\) is \((S)\)-shellable.

For verifying other items, we need more intrinsic definition of \(Q\). We will assume that \(a^+ = 0 \cdots 0 \alpha_{\ell+1} \cdots \alpha_n\) where \(\ell\) is the smallest integer such that \(\alpha_{\ell} > 0\). Note that \(\ell \leq n - 1\) since \(k \geq 3\).

**Claim 5.12** We have the following description of \(Q\).

(i) \(Q = \{q \in I(A^+) : q \geq a^+, (q - a^+)(\ell+1) = 0 \cdots 0\}\) if \(a^+ \neq 1 \cdots 102, 201 \cdots 1, 201 \cdots 202\);
\(Q = \{q \in I^+(A): q \geq a^+, (q - a^+)(2)^n\} = \{0 \cdots 0, 10 \cdots 0\}\) if \(a^+ = 201 \cdots 1\) or \(a^+ = 201 \cdots 102\); and

(iii) \(Q = \{a^+\}\) if \(a^+ = 1 \cdots 102\).

Note that we crucially use that \(n \geq 4\) in order that this claim makes sense; that is, we use that \(201 \cdots 102\) belongs to \(\mathcal{V}_n^*\).

**Proof** The proof is similar to the proof of Claim 5.10; however, in this proof, there are more cases to consider. Keeping in mind the number of cases we want to consider, we use slightly different approach how to treat them, compared to Claim 5.10.

We assume that we are given \(q\) such that \(q \geq a^+\) (this is a necessary condition for \(q \in Q\)). We let \(q - a^+ = k_1 \cdots k_n\). If \(a^+ \notin \{1 \cdots 102, 201 \cdots 1, 201 \cdots 102\}\), we want to verify that \(q \in Q\) if and only if \(k_{\ell+1} = \cdots = k_n = 0\). If \(a^+ \in \{1 \cdots 102, 201 \cdots 102\}\), we want to verify that \(q \in Q\) if and only if \(k_2 \in \{0, 1\}\) and \(k_3 = \cdots = k_n = 0\). If \(a^+ = 1 \cdots 102\), we want to verify that \(q \in Q\) if and only if \(q = a^+\).

First, we distinguish cases according to whether \(k_{\ell+1} \cdots k_n = 0 \cdots 0\) (note that we also cover \(a^+ \in \{1 \cdots 102, 201 \cdots 1, 201 \cdots 102\}\) by setting \(\ell = 1\) in these cases).

1. \(k_{\ell+1} \cdots k_n \neq 0 \cdots 0\). In this case, we have \(i \in \{\ell + 1, \cdots, n\}\) such that \(k_i > 0\).
   
   We prefer \(i \neq 2\), if possible. We set
   \[
   a := 0 \cdots 0(\alpha_\ell - 1)\alpha_{\ell+1} \cdots \alpha_{i-1}(\alpha_i + 1)\alpha_{i+1} \cdots \alpha_n.
   \]

   Note that if \(a \neq j\), then \(a\) precedes \(a^+\) in the \(\prec^L\) order. (In fact, \(a\) precedes \(a^+\) in the lexicographic order in any case, but we do not define the \(\prec^L\) order for \(j\).) Note also that \(a^+ \prec q\). In some cases, we will manage to show that \(a \neq j\), \(a \prec^S a^+\), and \(a + j \neq q\). This will imply that \(a \in A\) and \(a \prec q\) and, therefore, \(q \notin Q\). In some other cases, we will replace \(a\) with another \(a'\) satisfying the above-mentioned conditions still deriving \(q \notin Q\). However, this will be impossible if \(a^+ \in \{1 \cdots 102, 201 \cdots 102\}\), \(i = 2\), and \(k_2 = 1\) when we will actually derive that \(q \notin Q\).

Now, we distinguish several subcases according to \(a^+\).

(a) \(a^+ \in A^n = A^S \cup \{1 \cdots 102\}\).

Before we start, we remark that all considerations are also valid if \(a^+ = 1 \cdots 102\). The atom 1 \cdots 102 is the last atom of \(A^n\) in the \(\prec^S\) order. This will reflect in such a way that in some cases we check for \(a^+ = 1 \cdots 102\) more than we need (which is not a big price for a coherent case analysis).

We have that \(a\) precedes \(a^+\) in the \(\prec^S\) order unless \(a \in \{j, 1 \cdots 102\}\). Therefore, for the beginning, we assume that \(a \notin \{j, 1 \cdots 102\}\). If, in addition, \(a + j \neq q\), then we have the required properties of \(a\) deriving \(q \notin Q\). However, if \(a + j = q\), then we obtain \(a' \neq 1 \cdots 102\) of required properties by Claim 5.11 (or by Claim 5.9 (ii) if \(a = 0 \cdots on\)).

If \(a = j\), then \(a^+ = 21 \cdots 101 \cdots 1\) where the “0” appears in the \(i^\text{th}\) position. We distinguish subsubcases according to \(i\).
i. $i \geq 3$.
   In this situation, we set $a' = 201 \cdots 1$. Then, $q > a'$ (since $\kappa_i \geq 1$ and $q > a^+$) and $a'$ precedes $a^+$ in the $<L$ order and, therefore, in $<S$ order as well. Therefore, $a'$ has the required properties unless $q = a' + j = 312 \cdots 2$. In this case, we can use $1 \cdots 1201 < q$, for example.

ii. $i = 2$.
   In this situation, $a^+ = 201 \cdots 1$. We also have $\kappa_3 = \cdots = \kappa_n = 0$ since we wanted $i \neq 2$ if possible.
   If $\kappa_2 \geq 2$, implying $q \geq 221 \cdots 1$, we still can set $a' = 021 \cdots 1$ deriving $q \notin Q$ (note that $q \neq a + j$ since $\kappa_n = 0$).
   If $\kappa_2 = 1$, we actually want to derive $q \in Q$ according to our description.
   In this case, it is easiest to refer to Claim 5.10 (ii) (since we have already done this analysis). The claim implies that there is no $a \in A^{all}$ such that $a < L a^+$ and $a \not\leq q$. In particular, there is no such $a \in A^{(n)}$. Since $a < L a^+$ is equivalent with $a < S a^+$ in this case, we deduce $q \notin Q$.

If $a = 1 \cdots 102$, then we can perform the same analysis as if $a = j$ just replacing the suffix $111$ with $102$. (The only major difference is that we cannot use the shortcut referring to Claim 5.10.) Here, the analysis follows in detail.

We have $a^+ = 21 \cdots 101 \cdots 102$ where the first “0” appears in the $i$th position or $a^+ = 21 \cdots 101$ if $i = n$. (In particular, $i \neq n - 1$.)

We distinguish subsubcases according to $i$.

i. $i \geq 3$.
   In this situation, we set $a' = 201 \cdots 102$. Then, $q > a'$ (since $\kappa_i \geq 1$ and $q > a^+$) and $a'$ precedes $a^+$ in the $<L$ order (hence in $<S$ order as well). Therefore, $a'$ has the required properties unless $q = a' + j = 312 \cdots 213$.
   In this case, we can use $1 \cdots 1201 < q$, for example.

ii. $i = 2$.
   In this situation, $a^+ = 201 \cdots 102$. We also have $\kappa_3 = \cdots = \kappa_n = 0$ since we wanted $i \neq 2$ if possible.
   If $\kappa_2 \geq 2$, implying $q \geq 221 \cdots 102$, we still can set $a' = 021 \cdots 102$ deriving $q \notin Q$ (note that $q \neq a + j$ since $\kappa_n = 0$).
   If $\kappa_2 = 1$, we actually want to derive $q \in Q$ according to our description.
   In this case, $q = (r \cdot n + 1) \cdots 102$ for some positive integer $r$. We want to show that there is no $a \in A^S$ such that $a < S a^+$ and $a \not\leq q$. For contradiction, there is such $a = a'_1 \cdots a'_n$. Condition $a < S a^+$ implies $a'_1 \leq 2$. Since the sum of the last $(n - 1)$ coordinates of $q$ equals $n - 1$, we derive either that $a'_1 = 1$ and $a$ agrees with $q$ on all remaining $n - 1$ coordinates or that $a'_1 = 2$ and $a$ agrees with $q$ on all remaining $n - 1$ coordinates except one coordinate, where it is one less. The first case is excluded since $1 \cdots 102 \not\in S a^+$. The second case is also excluded, since in such a case $a \not\leq L a^+$, implying $a \not\in S a^+$, a contradiction. We conclude that $q \notin Q$ if $\kappa_2 = 1$.

(b) $a^+ \in A^{all} \setminus A^{(n)}$.
   In this case, $\alpha_n = 0$. We also emphasize that $a$ precedes $a^+$ in the $<S$ order if $a \neq j$. This is simply because $a$ precedes $a^+$ in the $<L$ order and $a^+ \not\in A^{(n)}$ in this case. Therefore, we derive $q \notin Q$ if $a \neq j$ and $a + j \neq q$. 

$\square$ Springer
Now, let us consider the case \( a \neq j, \) but \( a + j = q \). Then, there is \( a' \) from Claim 5.9 (i) or (ii) such that \( a' <^L a^+, \ a < q \). We derive \( a' <^S a^+ \), and, therefore, \( q \not\in Q \).

Finally, we consider the case \( a = j \). We derive \( a^+ = 21 \cdots 10 \) and \( i = n \) (since \( \alpha_n = 0 \)). We set \( a' := 21 \cdots 101 \) or \( a' := 21 \cdots 1011 \) so that \( a' + j \neq q \). We derive \( a' < q, \ a' <^S a^+ \), and, therefore, \( q \not\in Q \) as desired.

2. \( \kappa_{l+1} \cdots \kappa_n = 0 \cdots 0 \).

In this case, we want to derive \( q \in Q \) for all possible choices of \( a^+ \) except \( a^+ = 1 \cdots 102 \) and \( q > a^+ \).

We distinguish subcases according to \( a^+ \).

(a) \( a^+ \in A^S \).

In this case, we refer to Claim 5.10 which implies that there is no \( a \in A^{all} \) such that \( a <^L a^+ \) and \( a < q \). Therefore, in particular, there is no \( a \in A^S \) with \( a <^S a^+ \) and \( a < q \) which is what we need.

(b) \( a^+ = 1 \cdots 102 \).

If \( q = a^+ \), then \( q \in Q \) as desired.

If \( q > a^+ \), then \( q = (rn + 1) \cdots 102 \) for some integer \( n \). Setting \( a = 21 \cdots 101 \), we get \( a \in A^{(n)} \) implying \( a <^S a^+ \) and also \( a < q \). Thus, \( q \not\in Q \) as required.

(c) \( a^+ \in A^{all} \setminus A^{(n)} \).

By Claim 5.10, there is no \( a \in A^{all} \) such that \( a <^L a^+ \) and \( a < q \). Therefore, in particular, there is no \( a \in A^{all} \setminus A^{(n)} \) with \( a <^S a^+ \) and \( a < q \).

On the other hand, there is no \( a \in A^{(n)} \) with \( a <^S a^+ \) and \( a < q \) either, because \( \alpha_n = \kappa_n = 0 \) implying that the last coordinate of \( q \) is 0 whereas \( a \) from \( A^{(n)} \) has the last coordinate nonzero.

Altogether, there is no \( a \in A^{all} \) with \( a <^S a^+ \) and \( a < q \) implying \( q \in Q \). \( \square \)

This finishes the proof of the claim.

Now, we verify condition (ii) of Theorem 2.1. However, the verification is almost the same as in the case of Lemma 5.4 using Claim 5.12 instead of Claim 5.10. This is because of the described structure of \( Q \). (Compare with the text below the proof of Claim 5.10.)

If \( a \notin \{1 \cdots 102, 201 \cdots 1, 201 \cdots 102 \} \), then we just use Proposition 4.1. If \( a \in \{201 \cdots 1, 201 \cdots 102 \} \), then we obtain shellability of \( Q \) referring to Fig. 10. Finally, if \( a = 1 \cdots 102 \) then the verification is trivial, since a poset with single element is shellable.

We continue with the verification of condition (iii) of Theorem 2.1; that is, we verify the edge falling property. If \( a^+ \neq 1 \cdots 102 \), then again this verification can be taken in verbatim from the analogous verification in the proof of Lemma 5.4 using Claim 5.12 instead of Claim 5.10, considering cases according to structure of \( Q \). We, therefore, do not repeat the relevant text again.

If \( a^+ = 1 \cdots 102 \), then the verification of the edge falling property is somewhat trivial. In this case, \( Q = \{a^+\} \) by Claim 5.12. Therefore, we are supposed to verify that if \( q = a^+, \ q' = 0, \) and \( p \in I(A) \) is such that \( p > q, \) then there is \( p' \in A \) covering \( 0 \) and covered by \( p \). But, this just immediately follows from \( p \in I(A) \) since \( \text{rk}(p) = 2 \).
We conclude by verifying condition (iv) of Theorem 2.1. We again refer that if $a^+ \neq 1 \cdots 102$, then this verification is already done in the proof of Lemma 5.4. It again solely depends on the structure of $Q$.

If $a^+ = 1 \cdots 102$, then we are just supposed to check that the interval $[a^+, z]$ is shellable. This follows from the assumptions of this lemma, since it is isomorphic to $[0, z - a^+]$.

\begin{proof}
First, we observe that it is sufficient to prove the lemma for case $k = |A^all|$ since an $(L)$-shelling of $I(A^{(\ell+1)} \cap I)$ restricts to an $(L)$-shelling of $I(A_j^{(\ell+1)} \cap I)$. Therefore, in case $k = |A^all|$, we just aim to show that $I(A^all \cap I)$ is $(L)$-shellable.

We plan to use Theorem 2.3 for the proof of this lemma where we set $A := A^all \cap I$ and $A' := A^{(\ell+1)} \cap I$.

Condition (i) of Theorem 2.3 follows from the assumptions of the lemma.

For checking condition (ii), we have $b \in I \cap (A^all \setminus A^{(\ell+1)})$ and $p \in I(A^{(\ell+1)} \cap I)$ covering $b$. We need to find $b' \in A^{(\ell+1)} \cap I$ such that $p \succ b'$ and $b' < L b$. Actually, we will only check $b' \in A^{(\ell+1)}$, $p \succ b'$, and $b' < L b$ since $p \succ b'$ implies $b' \in I$.

We have that $b = \beta_1 \cdots \beta_\ell 0 \cdots 0$ since $b \notin A^{(\ell+1)}$. On the other hand, if we let $p = \pi_1 \cdots \pi_n$, then there is $j \in \{\ell + 1, \ldots, n\}$ such that $\pi_j > 0$ since $p \in I(A^{(\ell+1)} \cap I)$. Let also $i \in [\ell]$ be such an index that $\beta_i > 0$ and $\beta_i$ is as small as possible. We set the following candidate for $b'$.

$$b'_{\text{cand}} := \beta_1 \cdots \beta_{i-1}(\beta_i - 1)\beta_{i+1} \cdots \beta_\ell 0 \cdots 010 \cdots 0,$$

where the “1” appears on the $j$th position. We have $b'_{\text{cand}} \leq p$. We also have $b'_{\text{cand}} \neq j$; this is obvious if $\ell \neq n - 1$, and it follows from our choice of $\beta_i$ if $\ell = n - 1$. In particular, $b'_{\text{cand}} < L b$ and $b'_{\text{cand}} \in A^{(\ell+1)}$. If $b'_{\text{cand}} + j \neq p$, then $b'_{\text{cand}} < p$ and consequently $p \succ b'_{\text{cand}}$ (by comparing ranks). Thus, we can simply set $b' := b'_{\text{cand}}$ in this case.

If $b_{\text{cand}} + j = p$, we think of $b'_{\text{cand}}$ as $a$ from Claim 5.11. We obtain the corresponding $a'$, and we just set $b' := a'$.

\end{proof}

6 Relation of lexicographic shellability and $A$-shellability

6.1 Lexicographic shellability

Here, we briefly recall the definition of lexicographic shellability. The reader interested in more details (including examples) is referred to sources such as \cite{4, 5, 14, 15}. The reader familiar with lexicographic shellability can skip this subsection.

As usual, we let $(P, \preceq)$ be a graded poset (with a unique minimal and maximal element), using the notation from Sect. 2. Given a maximal chain $c \in C(P)$, we label all of its edges with elements of some poset $A$ (typically, $A = \mathbb{Z}$). In this way, we label edges of all maximal chains in $C(P)$ (that is, a label of an edge might differ if we start with two different chains). We obtain a chain-edge labeling of $P$ if the following condition is satisfied. Whenever $c, c' \in C(P)$ are two chains sharing first $k$ edges (for some $k$), then the labels of these first $k$ edges have to coincide. Let us assume that $P$ is equipped with a fixed chain-edge labeling.
A rooted interval \([x, y]_r\) is an interval in \(P\) where the root \(r\) of this interval is a maximal chain in the interval \([0, x]\). Given a maximal chain \(c_0\) in \(C([x, y])\), we obtain (with respect to \(r\)) a labeling of edges of \(c_0\) induced from the labeling of a maximal chain \(c' \in C[x, y]\) obtained by composing \(r, c_0\), and an arbitrary maximal chain in interval \([y, \hat{1}]\). This labeling does not depend on the choice of the chain in \([y, \hat{1}]\) due to the definition of chain-edge labeling. In the sequel, we consider the labeling of \(c_0\) as a sequence of \(rk(y) - rk(x)\) elements of \(A\). In particular, we can say that \(c_0\) is increasing (in \([x, y]_r\)) if its labeling is increasing and \(c_0\) is lexicographically smaller than another maximal chain \(c_1\) in \(C([x, y])\) if the labeling of \(c_0\) is lexicographically smaller than the labeling of \(c_1\).

We say that a chain-edge labeling is a \(CL\)-labeling (chain-lexicographic labeling) if for every rooted interval \([x, y]_r\) in \(P\) the following two conditions are satisfied.

(i) There is a unique maximal increasing chain \(c_0\) in \([x, y]_r\); and
(ii) \(c_0\) is lexicographically smaller than any other maximal chain in \([x, y]_r\).

The poset \(P\) is \((chain-)\text{lexicographically shellable}\), abbreviated as CL-shellable, if it admits a CL-labeling.

It follows from [4] that every CL-shellable poset is indeed shellable. Actually, the order of shelling is given by the lexicographic order of chains in \(C(P)\) (with respect to given CL-labeling). The converse is not true—there are posets which are shellable but not lexicographically shellable; see [18, 19].

6.2 Lexicographic shellability versus \(A\)-shellability

In this subsection, we want to compare \(A\)-shellability and lexicographic shellability. This comparison makes sense if \(A = A^{all}\) is the set of all atoms. In addition, we also assume that \(A^{all}\) is linearly ordered. (If we allow arbitrary partial order on \(A^{all}\), then, for example, we can allow all elements incomparable; then, \(A^{all}\)-shellability just coincides with shellability.)

6.3 Lexicographic shelling is an \(A^{all}\)-shelling

Let \(P\) be a CL-shellable poset, and let us fix a CL-labeling of it. Given an atom \(a\) of \(P\), we observe that the edge \(e_a = \hat{0}a\) is labeled the same way in all maximal chains containing \(e_a\) (by the definition of chain-edge labeling). Thus, we can denote by \(A(e_a)\) this label of \(e_a\). By condition (ii) of the definition of CL-labeling, we have that \(A(e_a)\) and \(A(e_{a'})\) differ for two different atoms \(a\) and \(a'\), and in addition, they are comparable with \(A\). Thus, these labels induce a linear ordering \(\leq_A\) on \(A^{all}\). In this setting, the CL-shelling of \(P\) is also an \(A^{all}\)-shelling of \(P\) (where \(A^{all}\) is equipped with \(\leq_A\)).

6.4 \(A^{all}\)-shelling which is not lexicographic shelling

It is not hard to come up with an example of an \(A^{all}\)-shelling which is not a CL-shelling. Let \(P'\) be a poset which is shellable but not CL-shellable. Let us consider \(k\) copies
\(\hat{0}_1, \ldots, \hat{0}_k\) of the minimal element in \(P'\). The poset \(P\) is obtained by replacing the minimal element of \(P'\) by these \(k\) copies and then adding a new minimal element \(\hat{0}_{new}\) smaller than everything else. Note that \(A^{all} = \{\hat{0}_1, \ldots, \hat{0}_k\}\).

It is not hard to check that \(P\) is \(A^{all}\)-shellable where \(A^{all}\) is equipped with an arbitrary linear order (either by elementary means or using Theorem 2.1). On the other hand, \(P\) is not CL-shellable since \(P\) contains an interval isomorphic to \(P'\) and all intervals in a CL-shellable poset are CL-shellable as well.

6.5 Recursive atom orderings

Björner and Wachs [5] gave an equivalent reformulation of CL-shellability using recursive atom orderings. It is useful to compare \(A\)-shellability and recursive atom orderings. We first repeat their definition.

A poset \(P\) (graded, with a unique minimum and maximum) admits a recursive atom ordering if it has length 1 or if the length of \(P\) is greater than 1 and there is an ordering \(a_1, \ldots, a_t\) of all the atoms of \(P\) that satisfies:

(R1) For all \(k \in [t]\), the interval \([a_k, \hat{1}]\) admits a recursive atom ordering in which the atoms of \([a_k, \hat{1}]\) that come first in the ordering are those that cover some \(a_i\) where \(i < k\).

(R2) For all \(i < k\), if \(a_i, a_k < y\), then there is \(j < k\) and an element \(z\) such that \(a_j, a_k < z \leq y\).

Björner and Wachs [5] proved that a poset is CL-shellable if and only if it admits a recursive atom ordering.

In our notation, a recursive atom ordering induces an ordering of \(A^{all}\). From this point of view, recursive atom orderings are very strongly related to our second criterion, Theorem 2.2. Let us assume that condition (i) of Theorem 2.2 is satisfied in a slightly stronger form, that is, we assume that \(P\langle A\rangle\) admits a recursive atom ordering (which induces an \(A\)-shelling). Similarly, let us assume that we can replace \(A^{all}(a^+)\)-shellability of \(I(a^+)\) with a recursive atom ordering on \(I(a^+)\) inducing \(A^{all}(a^+)\)-shellability. Then, we can deduce that \(P\langle A^+\rangle\) admits a recursive atom ordering:

Indeed, condition (R1) translates to condition (ii) of Theorem 2.2 (it is sufficient to check (R1) only for \(a_k = a^+\) since we already assume that \(P\langle A\rangle\) admits a recursive atom ordering). Similarly, we will check that condition (R2) translates to condition (iii) of Theorem 2.2. Given \(a_i, a_k,\) and \(y\) from (R2), we can again assume that \(a_k = a^+\). We choose a maximal chain \(c\) in \([a^+, y]\) and set \(p\) to be the smallest element of \(c\) belonging to \(P\langle A\rangle\) (note that \(y \in P\langle A\rangle\) since \(a_i < y\); see Fig. 12). Then, we can set \(q\) to be the element of \(c\) one rank below \(p\). Then, by assuming (iii) of Theorem 2.2, \(p\) is above some \(z \in A(a^+)\). This is the required \(z\) since \(z \in A(a^+)\) implies that \(z\) covers some atom \(a_j\) preceding \(a_k\).

Altogether, we see that the method using \(A\)-shellability includes the recursive atom ordering method. On the other hand, it is not hard to see, that if we were allowed to use only Theorem 2.2, we would not get more than recursive atom orderings. However, Theorem 2.2 is still more flexible since, for example, it does not need to assume that \(P\langle A\rangle\) comes with a recursive atom ordering. This is useful, when it is combined with Theorem 2.1.
6.6 Lexicographic shellability versus Theorem 2.1

Now, we compare our first criterion, Theorem 2.1, to lexicographic shellability (in this case, it is more natural to choose lexicographic shellability rather than recursive atom orderings). In this case, Theorem 2.1 seems to be in more “generic” position in relation with lexicographic shellability.

6.7 CL-shellable poset which does not satisfy assumptions of Theorem 2.1

First, we provide an example of a poset that is CL-shellable, but which does not satisfy assumptions of Theorem 2, with respect to a given CL-shelling. This example arose in discussions with Afshin Goodarzi.

Let $P$ be the poset from Fig. 13. It is lexicographically shellable: we first label edges as on picture; and then we label chains according to labels of edges. The reader is welcome to check that we indeed obtain a CL-labeling. (Actually, we obtain a so-called EL-labeling where, in addition, the label of an edge does not depend on the considered chain.) Note also that chains containing $a$ appear before chains containing $123$. 

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{fig13}
\caption{Lexicographically shellable poset which does not satisfy assumptions of Theorem 2.1}
\end{figure}
Fig. 14 Theorem 2.1 does not produce a lexicographic shelling of this poset

\[ a^+ \] in the corresponding lexicographic shelling. In particular, \( P \) is \( A^+ \)-shellable where \( A^+ := \{a, a^+\} \) and \( a \) appears before \( a^+ \).

On the other hand, if we intend to use Theorem 2.1 for showing \( A^+ \)-shellability of \( P \), we will not succeed. The condition (iii) (edge falling property) is not satisfied for the edge \( q \hat{1} \).

6.8 Theorem 2.1 does not provide a CL-shelling

Let us imagine that we replace our shellability assumptions in Theorem 2.1 by CL-assumptions. That is, for condition (i), we would assume that \( P \langle A \rangle \) is CL-shellable (and the corresponding CL-shelling is \( A \)-shelling as well); and for condition (iv), we would assume that \( I(q) \langle A(q) \rangle \) is CL-shellable. Does it follow that \( P \langle A^+ \rangle \) is CL-shellable?

The author does not know the answer to this question; but, it seems that the more probable answer is “no.” If the answer is indeed “no,” then this would mean further difference in applicability of Theorem 2.1 and CL-shellability (or even more general CC-shellability of Kozlov [14] as remarked below). However, even if the answer is “yes,” Theorem 2.1 still provides particular conditions that might possibly be checked in an easier way than establishing CL-labeling (or establishing recursive atom ordering).

Here, we can at least provide a simple example showing that the current proof of Theorem 2.1 does not provide CL-shelling even if we ask CL-shelling assumptions. Let \( P \) be the poset on Fig. 14. If we set \( a^+ \) as in the picture, we can easily check that all assumptions of Theorem 2.1 are satisfied even with lexicographic assumptions. We label elements of \( Q \) as \( q_1, \ldots, q_5 \) consistently with the proof of Theorem 2.1. Then, the proof provides shelling such that the chains \( \hat{0}q_1q_2q_4\hat{1}, \hat{0}q_1q_3q_4\hat{1}, \hat{0}q_1q_2q_5\hat{1}, \) and \( \hat{0}q_1q_3q_5\hat{1} \) appear in this order; consult also Fig. 6. This cannot be a CL-shelling due to the alternation of edges \( q_1q_2 \) and \( q_1q_3 \). (The reader familiar with Kozlov’s CC-shellability [14] is welcome to check that this is not even a CC-shelling.)

Acknowledgments I am very thankful to Anders Björner, Aldo Conca, and Afshin Goodarzi for valuable discussions. In addition, I further thank Anders Björner for suggesting this topic for research and reading preliminary versions of this paper, Aldo Conca for letting me know that Corollary 1.3 follows from results in [9], and Volkmar Welker for pointing out the reference [16]. Finally, I thank two anonymous referees for many valuable remarks. Supported by the Göran Gustafsson postdoctoral fellowship.
References

Recognition of Collapsible Complexes is NP-Complete

Martin Tancer\(^1\),\(^2\)

Received: 29 April 2013 / Accepted: 12 October 2015 / Published online: 30 November 2015
© Springer Science+Business Media New York 2015

Abstract We prove that it is NP-complete to decide whether a given (3-dimensional) simplicial complex is collapsible. This work extends a result of Malgouyres and Francés showing that it is NP-complete to decide whether a given simplicial complex collapses to a 1-complex.

Keywords Collapsibility · Simplicial complex · NP-hardness · Bing’s house

1 Introduction

A classical question often considered in algebraic topology is whether some topological space is contractible. When we consider this question as an algorithmic question, that is, we consider the topological space as an input for an algorithm (say as a finite simplicial complex; Many topological spaces cannot be represented as finite simplicial complexes. However, we only consider those that can be represented this way. Already for such spaces, this question is undecidable), then it turns out that this question is algorithmically undecidable by a result of Novikov; see [16, §10] (see also Appendix).

An important, algorithmically recognizable, subclass of contractible complexes is the class of collapsible complexes, introduced by Whitehead. Roughly speaking, a simplicial complex is collapsible if it can be shrunk to a point by a sequence of face collapses, which preserve the homotopy type. (The precise definition is given in the following section.)
We focus on the computational complexity of the collapsibility problem, considered as an algorithmic question. We show that this question is NP-complete. More precisely, we obtain NP-completeness even if we restrict the input to 3-dimensional complexes.

**Theorem 1** It is NP-complete to decide whether a given 3-dimensional simplicial complex is collapsible.

It is easy to see that this problem belongs to NP (it is just sufficient to guess a right sequence of elementary collapses), thus the core of our paper relies on showing the NP-hardness.

By attaching a $d$-simplex to the complexes used in the proof of Theorem 1 it is easy to observe that Theorem 1 is also valid if we replace ‘3-dimensional’ with ‘$d$-dimensional’ for any $d \geq 4$. We provide more details in the conclusion (Sect. 8).

**Previous work.** Eğecioğlu and Gonzalez [4] have shown that it is (strongly) NP-complete to decide whether a given 2-dimensional complex can be collapsed to a point by removing at most $k$ triangles where $k$ is a part of the input. This problem, however, becomes polynomial-time solvable when $k$ is fixed as pointed out by Joswig and Pfetsch [9] or Malgouyres and Francés [10]. In particular, deciding whether a 2-dimensional complex collapses to a point is polynomial-time solvable. The same approach yields to the fact that deciding whether a $d$-complex collapses to a $(d-1)$-complex is polynomial time solvable. Since the author is not aware of a reference, we include a simple proof here; see Proposition 5.

Malgouyres and Francés [10] have shown that it is NP-complete to decide whether a given 3-dimensional complex collapses to some 1-complex. Naturally, they asked about the complexity of the problem of deciding whether a given 3-complex is collapsible. Theorem 1 answers this question in terms of NP-completeness.

Our approach relies, in a significant part, on the work of Malgouyres and Francés. We sketch their proof as well as point out the differences in Sect. 3. One of the important differences is that we need to replace a very simple ‘clause gadget’ of Malgouyres and Francés with something more suitable for our setting. For this we need to introduce Bing’s house with three rooms, which is done in Sect. 4. Then we construct all gadgets needed for our reduction in Sect. 5 and we finish the proof of Theorem 1 in Sect. 6.

**Links to discrete Morse theory.** Using the result of Eğecioğlu and Gonzalez [4], Joswig and Pfetsch [9] proved that it is (strongly) NP-complete to decide whether there exists a Morse matching with at most $c$ critical cells where $c$ is a part of the input. Our main result can be reformulated in terms of Morse matchings in the following way.

**Theorem 2** It is NP-complete to decide whether a given 3-dimensional simplicial complex admits a perfect Morse matching.

In the short proof below we keep a few notions undefined. We refer to Joswig and Pfetsch [9] for these notions as well as for further details on Morse matchings in the computational complexity context.

**Proof** We use the fact that a simplicial complex $K$ collapses to a point $x$ if and only if it admits a Morse function such that $K \setminus \{x, \emptyset\}$ contains no critical cells (see Forman [5]). Therefore, $K$ is collapsible if and only if it admits a perfect Morse matching.

(see the exposition by Joswig and Pfetsch [9]). Consequently, Theorem 2 is equivalent to Theorem 1.

Efficient search for collapsing sequences also plays an important role for searching a CW-complex homotopy equivalent to a given simplicial complex with number of critical cells as small as possible. This has further impact on efficient homology computations. See [3].

Links to shape-reconstruction. The task of shape reconstruction is to reconstruct a shape from a set of points that sample it. An important subtask is to reconstruct the homotopy type or the homeomorphism type of the shape. In a recent work of Attali and Lieutier [1], the aim is to collapse the Rips complex or the Čech complex of the sampling set to a complex homeomorphic with the shape. In this context, Theorem 1 means certain limitations for results one can expect. In particular, specific treatment in [1], using the properties of the Rips and Čech complexes, seems important.

Another notion of collapsibility. In the context of Helly-type theorems in discrete geometry, Wegner [17] introduced a notion of \( d \)-collapsibility. This notion shares some properties with collapsibility, but for example, it does not preserve the homotopy type. The author has shown in [14] that the recognition of \( d \)-collapsible complexes is NP-hard for \( d \geq 4 \). We remark that the approach in that case is different and the result from [14] should not be confused with the result presented here on classical (Whitehead’s) collapsibility.

2 Preliminaries

Simplicial complexes. We work with finite (abstract) simplicial complexes, that is, with set systems \( K \subseteq 2^V \) such that \( V \) is a finite set and if \( \alpha \in K \) and \( \beta \subset \alpha \), then \( \beta \in K \). We recall few basic definitions; however, we also assume that the reader is familiar with some basic properties of simplicial complexes. Otherwise we refer to any of the books [7,11,12]. In particular, we assume that the reader is familiar with the correspondence of abstract simplicial complexes and geometric simplicial complexes since it will be very convenient in the further text to define some simplicial complexes by pictures.

Elements of a simplicial complex \( K \) are faces (or simplices). A \( k \)-face is a face of dimension \( k \), that is, a face in \( K \) of size \( k + 1 \). 0-Dimensional, 1-dimensional, and 2-dimensional faces are vertices, edges, and triangles, respectively.

When we consider a simplicial complex as an input for an algorithm, it is given by a list of all faces.

Collapsibility. Let \( \sigma \) be a non-empty non-maximal face of \( K \). We say that \( \sigma \) is free if it is contained in only one maximal face \( \tau \) of \( K \). Let \( K' \) be the simplicial complex obtained from \( K \) by removing \( \sigma \) and all faces above \( \sigma \), that is,

\[
K' := K \setminus \{ \varnothing \in K : \sigma \subseteq \varnothing \}.
\]

We say that \( K' \) arises from \( K \) by an elementary collapse (induced by \( \sigma \) and \( \tau \)). We say that a complex \( K \) collapses to a complex \( L \) if there exists a sequence of complexes \((K_1 = K, K_2, \ldots, K_{m-1}, K_m = L)\), called a sequence of elementary
collapses (from $K$ to $L$), such that $K_{i+1}$ arises from $K_i$ by an elementary collapse for any $i \in \{1, \ldots, m-1\}$. A simplicial complex $K$ is collapsible if it collapses to a point.

Let $(K_1 = K, K_2, \ldots, K_{m-1}, K_m = L)$ be a sequence of elementary collapses. Then for every $\eta \in K \setminus L$ there is a unique complex $K_i$ such that $\eta \in K_i$ and $\eta \notin K_{i+1}$. Then we say that $\eta$ collapses in this step. In particular, we will often use phrases such as ‘$\eta_1$ collapses before $\eta_2$’.

**Collapsibility with constrains.** In our constructions, we will often encounter the following situation: We will be given a complex $L$ glued to some other complexes forming a complex $M$. We will know some collapsing sequence of $L$ and we will want to use this collapsing sequence for $M$. This might or might not be possible. We will set up a sufficient condition.

**Definition 3** Let $M$ be a simplicial complex and $L$ be a subcomplex of $M$. We define the constrain complex of pair $(M, L)$ as

$$
\Gamma = \Gamma(M, L) := \{ \varnothing \in L : \varnothing \subseteq \eta \text{ for some } \eta \in M \setminus L \}.
$$

The constrain complex is obviously a subcomplex of $L$. Now we can present an elementary condition when collapsing of $L$ induces collapsing of $M$.

**Lemma 4** Let $M$ be a complex, $L$ subcomplex of $M$ and $\Gamma$ be the constrain complex of $(M, L)$. We also assume that $L$ collapses to $L'$ containing $\Gamma$. Then $M$ collapses to

$$
M' := L' \cup (M \setminus L).
$$

**Proof** Let $(L_1 = L, L_2, \ldots, L_{m-1}, L_m = L')$ be a sequence of elementary collapses. Let $\sigma_i$ be the face of $L_i$ which is collapsed in order to obtain $L_{i+1}$ and let $\tau_i$ be the unique maximal face in $L_i$ containing $\sigma_i$. We also set $M_i = L_i \cup (M \setminus L)$. The assumption ensures us that all superfaces of $\sigma_i$ in $M_i$ belong to $L_i$. Therefore $(M_1 = M, M_2, \ldots, M_{m-1}, M_m = M')$ is a sequence of elementary collapses still induced by $\sigma_i$ and $\tau_i$. \hfill \Box

**Collapsibility in codimension 1.** Here we show that collapsibility in codimension 1 is polynomial-time solvable. This result is not needed for the proof of Theorem 1; it only serves as a complementary result. We need the proposition below. The proposition implies that we can collapse an input $d$-complex $K$ greedily, and with this greedy algorithm, we obtain a $(d-1)$-complex if and only if $K$ collapses to a $(d-1)$-complex $L$.

**Proposition 5** Let $K$ be a $d$-complex which collapses to a $(d-1)$-complex $L$ and to some $d$-complex $M$. Then $M$ collapses to a $(d-1)$-complex.

**Proof** Let $(K_1 = K, K_2, \ldots, K_{m-1}, K_m = L)$ be a sequence of elementary collapses where the collapse from $K_i$ to $K_{i+1}$ is induced by faces $\sigma_i$ and $\tau_i$. Note that every $d$-dimensional face of $K$ is $\tau_i$ for some $i$.

Let $j$ be the smallest index such that $\tau_j$ belongs to $M$ and let $\eta_j$ be a $(d-1)$-face with $\sigma_j \subseteq \eta_j \subseteq \tau_j$. Note also that no $\tau_i$ with $i < j$ belongs to $M$ due to our choice of $j$. However, since $\sigma_j$ is free in $K_j$ and therefore $\eta_j$ is free as well, the only $d$-faces of
K containing η_j might be the faces τ_i with i ≤ j. Altogether, τ_j is the unique d-face of M containing η_j and we can collapse η_j (removing τ_j). If M is still d-dimensional, we repeat our procedure. After finitely many steps we obtain a (d − 1)-complex. □

3 The Approach by Malgouyres and Francés

In this section we describe the approach of Malgouyres and Francés [10]. In some steps we follow their approach almost exactly; however, there are also steps that have to be significantly modified in order to obtain our result. We will emphasize the steps where our approach differs.

The reduction is done, as usual, from 3-satisfiability problem which is well known to be NP-complete. We assume that the reader is familiar with the related terminology. Given a 3-CNF formula \( \Phi \), Malgouyres and Francés construct a 3-dimensional complex \( C(\Phi) \) such that \( C(\Phi) \) collapses to a 1-complex if and only if \( \Phi \) is satisfiable. They compose \( C(\Phi) \) of several smaller complexes that we will call gadgets. For every literal \( \ell \) in the formula they introduce a literal gadget \( C(\ell) \). (This includes introducing \( C(\bar{\ell}) \) where \( \bar{\ell} \) is the negation of \( \ell \).) The gadgets \( C(\ell) \) and \( C(\bar{\ell}) \) are glued along an edge so that a major part of only \( C(\ell) \) or only \( C(\bar{\ell}) \) can be collapsed in the first phase of collapsing. Another gadget is a conjunction gadget \( C_{\text{and}} \) glued to literal gadgets via clause gadgets so that \( C_{\text{and}} \) can be collapsed at some step if and only if every clause contains a literal \( \ell \) such that the major part of \( C(\ell) \) was already collapsed, that is, if and only if \( \Phi \) is satisfiable. As soon as \( C_{\text{and}} \) is collapsed, it makes few other faces of the literal gadgets free which enables to collapse the whole complex to a 1-dimensional complex. As it follows from the construction of Malgouyres and Francés, if the formula is satisfiable, the resulting 1-complex contains many cycles and therefore it cannot be further collapsed to a point.

Our idea relies on filling the cycles of the resulting 1-complex so that we can further proceed with collapsings. However, we cannot fill the cycles completely naively, since we do not know in advance which 1-complex we obtain. In addition filling these cycles naively could possibly introduce new collapsing sequences starting with edges on the boundaries of the filled cycles which could possibly yield to collapsing the complex even if the formula were not satisfiable. Therefore, we have to be very careful with our construction (which unfortunately means introducing few more technical steps).

We are going to construct a simplicial complex \( K(\Phi) \) such that \( K(\Phi) \) collapses to a point if and only if \( \Phi \) is satisfiable. In fact, our complex \( K(\Phi) \) will always be contractible, independently of \( \Phi \) (although we do not need this fact in our reduction; and therefore we do not prove it). We reuse literal and conjunction gadgets of Malgouyres and Francés (only with minor modifications regarding distinguished subgraphs). Unfortunately, we need to replace the very simple clause gadget of Malgouyres and Francés (it consists of two triangles sharing an edge or is even simpler, depending on the clause). For this we need to introduce Bing’s house with three rooms and three thick walls. We also need disk gadgets which fill the cycles in the resulting 1-complex. We remark that the disk gadgets will not be topological disks but only some contractible

---

1 That is, a formula in conjunctive normal form.
complexes. However, we keep the name disk gadgets because of the idea of filling the cycles.

4 Bing’s Rooms and Bing’s House with Three Rooms

In our reduction we will need several auxiliary constructions that we suitably glue together. We present them in this section.

Bing’s rooms. We will consider Bing’s house as a simplicial complex obtained by gluing two smaller simplicial complexes called Bing’s rooms. Later we will use these rooms for building more complicated Bing’s house with three rooms. Bing’s room with a thin wall is a complex depicted in Fig. 1 on the left and Bing’s room with a thick wall is in the middle. The room with a thin wall contains only 2-dimensional faces whereas the room with a thick wall contains one 3-dimensional block obtained by thickening one of the walls. Both rooms contain two holes in the ground floor and one hole in the roof. If, starting from Bing’s house with thick wall, we collapse away the thick wall, we obtain a complex that we call Bing’s room with collapsed thick wall, shown on the right. (Note that the left bottom edge of the collapsed thick wall is still present although it is not contained in any 2-dimensional face.)

Bing’s house with one thin and one thick wall. If we rotate the ground floor of one of the rooms and we glue the two rooms together along the ground floor, we obtain Bing’s house with one thin and one thick wall as introduced by Malgouyres and Francés [10]. See Fig. 2. Similarly, we can obtain Bing’s house with two thin walls or Bing’s house with two thick walls.

![Fig. 1 Bing’s room with a thin wall (on the left), Bing’s room with a thick wall (in the middle), and Bing’s room with collapsed thick wall (on the right)](image1)

![Fig. 2 Bing’s house with one thin and one thick wall](image2)
Fig. 3 The base floor of Bing’s house with three rooms and with three thick walls

**Bing’s house with three rooms and three thick walls.** As an auxiliary construction, we also need to introduce Bing’s house with three rooms. First we consider the base floor depicted in Fig. 3. It consists of three quadrilaterals with holes, glued together. For simplicity of explanations, we will assume that all these three quadrilaterals are squares. Also, we assume that the holes are squares. Now we consider three Bing’s rooms with thick walls labeled 1, 2 and 3. The Bing room with label $i$ is glued to the two squares with label $i$ so that the grey part of one of the squares with label $i$ is the place where the thick wall of the room is glued to the base floor. Here, it is important that we do not have to distinguish whether the rooms are glued to the base floor from below or from above, since we could not place them in such a way (in 3D) simultaneously. The resulting complex we call *Bing’s house with three rooms (and three thick walls)*. We remark that Bing’s house with three rooms is contractible which can be shown in a similar way as contractibility of classical Bing’s house. (If glue cuboid bricks to the base floor instead of Bing’s rooms, we obtain a complex $L$ which is obviously contractible. Bing’s house with three rooms is obtained by ‘digging holes’ into $L$.) Later on, we will need a specific collapsing sequence of Bing’s house with three rooms. The existence of such a sequence implies contractibility as well.

**Bing’s house with three collapsed walls.** For further purposes it will be convenient to work with Bing’s house with three rooms where the thick walls are collapsed. We let each of the thick walls collapse to the edge on the base floor. This way we obtain *Bing’s house with three collapsed walls*. We provide the reader with a drawing with two rooms only (this can still be done in three dimensions); see Fig. 4. We also distinguished edges $x_1$, $x_2$ and $x_3$ such that $x_i$ is the only remaining edge of the thick wall in room $i$ after collapsing the wall. Note that $x_1$, $x_2$ and $x_3$ are the only free faces of Bing’s house with three collapsed walls. See also Fig. 9 for the base floor.

**Triangulations.** In order to obtain simplicial complexes we need to triangulate our gadgets that we obtain from Bing’s rooms, Bing’s houses, etc. It will not be important for us how do we precisely triangulate pieces in the construction of dimensions 2 or less. For example, the middle level of Bing’s house with one thick and one thin wall can be triangulated as suggested in Fig. 5 keeping in mind that the triangulations of

---

2 It can be shown that the resulting complex does not topologically embed into $\mathbb{R}^3$ since it contains a Möbius band and an annulus glued to the central cycle of the band along one of the boundary components of the annulus. (The annulus is one third of the base floor and the band is formed by parts of the outer walls of the three rooms.) However, this non-embeddability fact is far beyond the needs of this paper.
Fig. 4 Two blocks of Bing’s house with three collapsed walls. The edges marked with $e$ are glued together.

Fig. 5 A suitable triangulation of the middle level of Bing’s house with one thick and one thin wall.

Fig. 6 Collapsings of the thick wall. (Usually, we use gray only for 3-cells or attachments of 3-cells. In this case, the objects in the middle and on the right are two dimensional while grey emphasize which 2-cells are still in the object.)

particular 2-cells have to be compatible on the intersections. In some cases, we will need gadgets with many prescribed edges in some part of the triangulation where the number of these edges depends on the size of the 3-CNF formula we will work with (see Figs. 8 or 10). In such cases we require that the size of the triangulation is polynomial in the number of prescribed edges.

The only 3-cells appearing in our construction are thick walls of Bing’s rooms (houses). For these thick walls we use particular triangulations of Malgouyres and Francés [10]. The thick wall is subdivided into four prisms 012389, 014589, 236789 and 456789. See Fig. 6, left. Each prism is further subdivided into two simplices (which are not shown on the picture). This triangulation allows collapsing the thick wall into two smaller complexes from Fig. 6, middle and right (in the middle picture, the edge 01 is contained in no 2-cell; the right picture is drawn from behind and the edge 89 is contained in no 2-cell). In particular, the collapsing from the middle picture is used when obtaining Bing’s room with a collapsed thick wall from Bing’s room with a thick wall.

5 Construction of the Reduction Gadgets

Here we start filling in details of the construction sketched in Sect. 3.
Given a 3-CNF formula $\Phi$ we construct a three-dimensional simplicial complex $K(\Phi)$ such that $K(\Phi)$ is collapsible if and only if $\Phi$ is satisfiable. We assume that every clause of $\Phi$ contains exactly three literals and also that no clause contains a literal and its negation.

The complex $K(\Phi)$ will consist of several gadgets described below. For each of the gadgets we also need to find some suitable collapsing sequence. We usually postpone the proofs that such sequences exist to Sect. 7 so that the main idea can be explained while the technical details are left to the end.

**Literal gadget.** First we establish the literal gadget $K(\ell, \bar{\ell})$ for every pair of literals $\ell$ and $\bar{\ell}$. This gadget is by Malgouyres and Francés [10], we only glue it to other gadgets in a different way. It consists of two smaller gadgets $X(\ell)$ and $X(\bar{\ell})$ suitably glued together.

We set $X(\ell)$ to be Bing’s house with two thick walls as in Fig. 7. It contains two distinguished edges $e(\ell)$ and $f(\ell)$. Furthermore, it contains a distinguished path $p(\ell)$ joining the common vertex of $e(\ell)$ and $f(\ell)$ with the upper thick wall (this path contains neither $e(\ell)$ nor $f(\ell)$). Let us emphasize that in this case, we use particular triangulation by Malgouyres and Francés [10] that subdivides the upper thick wall into four prisms which are further triangulated. The path $p(\ell)$ enters the upper wall in vertex 0 of this triangulation and it continues to vertex 8. For $\bar{\ell}$ we construct $X(\bar{\ell})$ analogously.

The complex $K(\ell, \bar{\ell})$ is composed of $X(\ell)$ and $X(\bar{\ell})$ glued together along edge 89. The common vertex 8 will be important for further constructions; and therefore we rename it to $u_{\ell, \bar{\ell}}$ emphasizing dependency on $\ell$ and $\bar{\ell}$. The following lemma describes a particular sequence of collapsing the literal gadget that we will use later. It also says that at least one of the edges $f(\ell)$, $f(\bar{\ell})$ has to be collapsed before collapsing the literal gadget to a 2-complex.

**Lemma 6.** 1. $K(\ell, \bar{\ell})$ collapses to a complex that contains only path $p(\ell)$ and edges $e(\ell)$ and $f(\ell)$ from $X(\ell)$ while it contains almost all $X(\bar{\ell})$ with exception that the upper thick wall of $X(\ell)$ was collapsed to a thin wall keeping only rectangles 0462, 0451, 4576, 2673, and 0132. (Consult Fig. 6 right, if you remove the edge 89.) The role of $\ell$ and $\bar{\ell}$ can be interchanged.

---

3 For simplicity of notation, we keep the same numbers of vertices either for upper thick wall of $X(\ell)$ or of $X(\bar{\ell})$. However, we once more emphasize that these two walls share the edge 89 only.
2. Let $L(\ell, \bar{\ell})$ be the complex resulting in item 1 without the edge $e(\ell)$. This complex further collapses to the union of the paths $p(\ell), p(\bar{\ell})$ and the edge $e(\bar{\ell})$.

3. Let $T(\ell)$ be any of the two triangles containing $e(\ell)$ and $T(\bar{\ell})$ be any triangle containing $e(\bar{\ell})$. Before collapsing both $T(\ell)$ and $T(\bar{\ell})$, at least one of the edges $f(\ell), f(\bar{\ell})$ must be collapsed.

**Proof** We postpone the proof of items 1 and 2 on precise collapsing sequences to Sect. 7. Item 3 is already proved by Malgouyres and Francés; see Remark 1, Example 3 and the proof of Theorem 4 in [10]. We sketch here that if neither $f(\ell)$ nor $f(\bar{\ell})$ is collapsed, then the only one of the two upper thick walls, one of $X(\ell)$ and one of $X(\bar{\ell})$, can be collapsed so that its edge 01 becomes free. Hence, only one of the triangles $T(\ell)$ and $T(\bar{\ell})$ might become free before collapsing $f(\ell)$ or $f(\bar{\ell})$.

**Conjunction gadget.** Next we define the conjunction gadget $K_{\text{and}}$. It is Bing’s house with one collapsed thick wall and one thin wall. See Fig. 8 on the left. We also distinguish several edges and vertices of the gadget.

As an auxiliary construction, for every pair $\ell, \bar{\ell}$ of literals, we create an anchor-shaped tree $A(\ell, \bar{\ell})$ formed of $u(\ell, \bar{\ell}), p(\ell), p(\bar{\ell}), f(\ell)$ and $f(\bar{\ell})$ from $K(\ell, \bar{\ell})$ and furthermore of newly introduced edge $a(\ell, \bar{\ell})$ and vertex $v_{\text{and}}$. See Fig. 8 on the right. We glue all trees $A(\ell, \bar{\ell})$ in vertex $v_{\text{and}}$ obtaining a tree $A$.

Finally, we let $e_{\text{and}}$ to denote the only free edge of $K_{\text{and}}$ and we glue $A$ to the lower left wall of $K_{\text{and}}$ as in Fig. 8 on the left. Note that, in particular, every literal gadget is glued to the conjunction gadget.

After we introduce the remaining gadgets, we will see that $K_{\text{and}}$ is glued to other gadgets only along $A$ and $e_{\text{end}}$. The following lemma states that if we want to collapse $K_{\text{and}}$ at some phase of collapsing, we have to make $e_{\text{end}}$ free first in whole $K(\Phi)$ and only then we can continue with collapsing $K_{\text{and}}$. On the other hand, as soon as we make $e_{\text{and}}$ free, we can collapse the complex to $A$.

**Lemma 7** 1. $K_{\text{and}}$ collapses to $A$.
2. Before collapsing any triangle containing one of the edges $f(\ell, \bar{\ell})$, the edge $e_{\text{and}}$ has to be collapsed.

**Proof** We prove item 1 in Sect. 7. Item 2 of the lemma is explained in [10, Remark1].
Clause gadget. We proceed with introducing the clause gadget. For a clause $c = (\ell_1 \lor \ell_2 \lor \ell_3)$ we set $K(c)$ to be Bing’s house with three collapsed walls as described in Sect. 4 and with several distinguished edges and paths; see Fig. 9. Namely, the only three free edges of $K(c)$ are labeled $(\ell_i, c)$. We also distinguish three paths $p(\ell_i, c)$ connecting the center of the base floor with $(\ell_i, c)$ (we assume that $(\ell_i, c)$ is not contained in the path). We also distinguish one other edge emanating from the center inside the base floor and we label it by $e_{\text{and}}$. This last edge $e_{\text{and}}$ is glued together with the edge of the conjunction gadget labelled $e_{\text{and}}$ so that the central vertex of the base floor becomes the vertex $v_{\text{and}}$ of the conjunction gadget.

Lemma 8 1. $K(c)$ collapses to a complex composed of $e_{\text{and}}$, three paths $p(\ell_i, c)$ and two of the three edges $(\ell_i, c)$.
2. Any collapsing of $K(c)$ starts with one of the edges $(\ell_i, c)$.

Proof We again postpone the proof of item 1 to Sect. 7. Item 2 is obvious as soon as we realize that the only free faces of $K(c)$ are the three edges $(\ell_i, c)$.

Disk gadgets. Finally, for every pair of literals $\ell, \bar{\ell}$ we construct the disk gadget $D(\ell, \bar{\ell})$ filling empty cycles in the construction of Malgouyres and Francés [10]. As we mentioned before, these gadgets will not be topological disks. However, they are contractible and play a similar role as disks.

We start with Bing’s house with one collapsed thick wall and one thin wall; see Fig. 10. We label the only free face of this complex with $e(\ell)$ and glue it to the edge $e(\ell)$ of $K(\ell, \bar{\ell})$. We pick a vertex on the edge connecting the left and the bottom wall and label it $v_{\text{and}}$. We also glue this vertex to $v_{\text{and}}$ vertex of the conjunction gadget. The edge connecting $v_{\text{and}}$ and one of the vertices of $e(\ell)$ is labelled by $b(\ell)$. Next, for every clause $c_j$ containing the literal $\ell$ we make a copy of path $p(\ell, c_j)$ and edge $(\ell, c_j)$ (where the template comes from $K(c_j)$) starting in vertex $v_{\text{and}}$. In particular, $B(\ell)$ is glued to the complexes $K(c_j)$ along these paths and edges. The resulting complex is denoted by $B(\ell)$. We perform an analogous construction for $B(\bar{\ell})$.

This complex can be collapsed (inside whole $K(\Phi)$) as soon as the edge $e(\ell)$ is free. Then it collapses to a complex composed of the distinguished edges and paths, as the following lemma summarizes.

Lemma 9 1. $B(\ell)$ collapses to the 1-complex composed of $b(\ell)$, paths $p(\ell, c_j)$ and edges $(\ell, c_j)$.
2. Any collapsing of $B(\ell)$ starts with the edge labelled $e(\ell)$. 
Proof As usual, item 1 is proved in Sect. 7. Item 2 is true since $e(\ell)$ is the only free edge of $B(\ell)$. □

Now we can finally construct $D(\ell, \bar{\ell})$; see Fig. 11. We fill two cycles with a disk. The first cycle is formed by $b(\ell), p(\ell)$ and $a(\ell, \bar{\ell})$, the second cycle by $b(\bar{\ell}), p(\bar{\ell})$ and $a(\ell, \bar{\ell})$. This finishes the construction of $D(\ell, \bar{\ell})$ and since we have already described all gluings, it also finishes the construction of $K(\Phi_1)$.

6 Collapsibility of $K(\Phi)$

In this section we prove that $K(\Phi)$ is collapsible if and only if $\Phi$ is satisfiable. Thereby we prove Theorem 1.

Satisfiable formulas. Let us first assume that $\Phi$ is satisfiable and fix one satisfying assignment of $\Phi$. We construct a collapsing sequence for $K(\Phi)$. We proceed in several steps (each step will still consist of many elementary collapses). By $K^{(i)}(\Phi)$ we denote the complex obtained after performing $i$th step of collapsing. We use similar notation for gadgets, for example, $K^{(i)}_{and}$ is the remaining part of $K_{and}$ after $i$th step, that is, $K^{(i)}(\Phi) \cap K_{and}$.

Step 1. For every literal $\ell$ we start with “partial” collapsing of $K(\ell, \bar{\ell})$ such as in Lemma 6 (1). Note that the constrain complex of the pair $(K(\Phi), K(\ell, \bar{\ell}))$ consists of $p(\ell), p(\bar{\ell}), e(\ell), e(\bar{\ell}), f(\ell)$ and $f(\bar{\ell})$; therefore Lemma 4 induces collapsing on whole $K(\Phi)$. If $\ell$ has positive occurrence in the assignment, we let $X(\ell)$ collapse to...
p(ℓ), e(ℓ) and f(ℓ) while in X(¯ℓ) only the upper thick wall of X(¯ℓ) collapses to a thin wall. This collapsing makes the edge e(ℓ) free.

We gradually perform this collapsing for all literals with positive occurrence. Note that by considering positive occurrences only, we do not “miss” negative ones since for every variable u exactly one literal among u and ¬u has positive occurrence.

**Step 2.** We continue with collapsing B(ℓ) as stated in Lemma 9 (1). Observe that at this stage, the constrain complex of the pair \((K^{(1)}(Φ), B^{(1)}(ℓ)) = (K^{(1)}(Φ), B(ℓ))\) contains only \(b(ℓ)\), paths \(p(ℓ, cj)\) and edges \((ℓ, cj)\) (in particular the edge \(e(ℓ)\) is not in it as well as the vertex of \(e(ℓ)\) which is not adjacent to \(b(ℓ)\)), therefore collapsing from Lemma 9 induces collapsing of \(K^{(1)}(Φ)\) by Lemma 4.

In further text we will use Lemma 4 many times in a similar fashion without mentioning it explicitly. (We will describe the constrain complex only.)

**Step 3.** Now, since the assignment is satisfying for every clause \(c\) at least one of the edges \((ℓ_i, c)\) became free. Therefore, every clause gadget collapses to a 1-complex described in Lemma 8 (1). The constrain complex for the pair \((K^{(2)}(Φ), K^{(2)}(c))\) is a subcomplex of the complex formed by paths \(p(ℓ_j, c)\) and edges \((ℓ_j, c)\) with \(j ≠ i\).

**Step 4.** Now we focus on the edge \(e_{and}\). At the beginning, it was contained in triangles in clause gadgets and in a single triangle of the conjunction gadget. All triangles of clause gadgets were collapsed, therefore \(e_{and}\) is free now. According to Lemma 7 (1), we can collapse the conjunction gadget \(K_{and}\) to \(A\) now (checking that the constrain complex for \((K^{(3)}(Φ), K^{(3)}_{and})\) is \(A\)).

**Step 5.** In this step, we will collapse the literal and the disk gadgets. The important fact is that the edges \(f(ℓ)\) and \(f(¯ℓ)\) are already free. Therefore, we can proceed with collapsing \(K^{(4)}(ℓ, ¯ℓ)\) according to Lemma 6 (2). This leaves \(e(ℓ)\) free as well as all remaining edges of paths \(p(ℓ)\) and \(p(¯ℓ)\). Now we can easily collapse \(B(¯ℓ)\) according to Lemma 9 (1) and consequently also the \(D(ℓ, ¯ℓ)\) (having all boundary edges free).

**Step 6.** Now we have a collection of paths emanating from \(v_{and}\) (which are remainders of clause gadgets). This collection can be easily collapsed to a point, say \(v_{and}\).

**Non-satisfiable formulas.** Now we show that \(K(Φ)\) is not collapsible for non-satisfiable formulas. More precisely, we assume that \(K(Φ)\) is collapsible and we deduce that \(Φ\) is satisfiable.

If \(K(Φ)\) is collapsible, then in particular some triangle of \(K_{and}\) has to be collapsed. We investigate what had to be collapsed before collapsing a first triangle of \(K_{and}\). According to Lemma 7 (2) the edge \(e_{and}\) has to be made free before this step. Lemma 8 (2) implies that for every clause \(c\) there is a literal \(ℓ(c)\) in this clause such that the edge \((ℓ(c), c)\) was made free prior this step. This means by Lemma 9 (2) that the edge \(e(ℓ(c))\) had to be made free previously. Now we recall that no triangle of \(K_{and}\) was collapsed yet (including triangles of \(K_{and}\) attached to \(f(ℓ)\) and \(f(¯ℓ)\)). Therefore, Lemma 6 (3) implies that only one of the edges \(e(ℓ)\) and \(e(¯ℓ)\) can be collapsed at this
stage. This gives a satisfying assignment to $\Phi$ by setting a variable $u$ to be TRUE if $e(u)$ was collapsed (before collapsing a triangle from $K_{and}$) and FALSE otherwise. The existence of $\ell(c)$ implies that every clause $c$ is indeed satisfied.

7 Collapsing Sequences

Here we prove technical lemmas used previously in the text. It is convenient to change the order of the proofs.

Proof of Lemma 7

(1) We recall that our task is to collapse the conjunction gadget from Fig. 8 to the tree $A$. We start with collapsing the wall below the edge $e_{and}$ and then the lowest floor (except edges belonging to $A$). We continue with collapsing all walls that used to be attached to the lowest floor. At this step we have the complex depicted in Fig. 12. This complex is already a 2-sphere with a hole and with $A$ attached to it. It is easy to collapse it to $A$ in the directions of arrows.

Proof of Lemma 8

(1) We recall that our task is to collapse the clause gadget from Fig. 9. We will provide a collapsing to the union of paths $p(\ell_1, c)$ and edges $(\ell_2, c)$ and $(\ell_3, c)$. Other cases are analogous. For the picture, we will assume that Bing’s room number 1 is above the base floor and Bing’s room number 2 is below the base floor as in Fig. 13 on the left.

Since $(\ell_1, c)$ is allowed to be collapsed, we can collapse the left wall of room 2 and then the bottom wall. In next step, we can collapse all walls of room 2 perpendicular to the base floor. We obtain complex as in Fig. 13 on the right. Next we collapse the interior of the 23 square so that the edges left of $(\ell_2, c)$ become free. This means that the room 3 can be collapsed in a similar fashion as we collapsed room 2 (note that after this step only $(\ell_2, c)$, $p(\ell_2, c)$ and part of $p(\ell_1, c)$ remain of the 23 square). Finally we can collapse room 1 in a similar fashion taking care that the edge $e_{and}$ remains uncollapsed.

Proof of Lemma 6

(1) First we collapse the thick wall of $X(\ell)$ in the way in Fig. 6, on the right. This makes the common edge 89 of $X(\ell)$ and $X(\ell)$ free. Then the thick wall of $X(\ell)$ can be collapsed so that the upper Bing’s room of $X(\ell)$ becomes Bing’s room with collapsed thick wall. Then the rest of $X(\ell)$ can be collapsed in very same way as in the proof of Lemma 7 (1) while keeping $p(\ell)$ and $f(\ell)$.

Proof of Lemma 6

(2) As soon as we are allowed to collapse $f(\ell)$, the lower thick wall of $X(\ell)$ can be collapsed obtaining Bing’s room with collapsed thick wall from
1
(1, c)
(2, c)
(3, c)
e and 1
(1, c)
(2, c)
(3, c)…

Fig. 13 Rooms 1 and 2 of the clause gadget while collapsing it

the lower Bing’s room. Now $X(\bar{\ell})$ can be collapsed in analogous way as was presented in the proof of Lemma 7 (1) while keeping the required subcomplex (the role of the lower and upper room are interchanged). \hfill \Box

Proof of Lemma 9 (1) We use almost the same collapsing procedure as in the proof of Lemma 7 (1). We just remark, that the wall below $e(\ell)$, split by $b(\ell)$ is collapsed in two stages. First the half containing $e(\ell)$ is collapsed; then the lowest floor of $B(\ell)$ is collapsed and finally, the second half of this wall is collapsed. \hfill \Box

8 Conclusion

We have shown that it is NP-hard to decide whether a 3-dimensional complex collapses to a point. Here we mention few (simple) corollaries of our construction as well as several related questions.

Collapsing $d$-complexes to $k$-complexes. Motivated by a question of Malgouyres and Francés [10] about higher dimensions, we set up question $(d, k)$-COLLAPSIBILITY asking whether a given $d$-dimensional complex collapses to some $k$-dimensional complex where $d > k \geq 0$ are fixed parameters.

Our result shows that $(3, 0)$-COLLAPSIBILITY is NP-complete; however, it is not difficult to observe that our result can be extended to showing that $(d, k)$-COLLAPSIBILITY is NP-complete for any $d \geq 3$ and $k \in \{0, 1\}$. For this it is sufficient to attach a $d$-simplex to $v_{\text{and}}$, say, and remark that if $\Phi$ is not satisfiable, then any collapsing of $K(\Phi)$ yields a complex of dimension 2 or more. (We also remark that the case $d \geq 3$, $k = 1$ can be already obtained from the construction of Malgouyres and Francés.)

As we mentioned in the introduction, it is not hard to see that $(d, k)$-COLLAPSIBILITY is polynomial time solvable whenever $d \leq 2$, and also in the codimension 1 case (see Proposition 5).
In the remaining cases, \( d \geq k + 2 \geq 2 \), it is reasonable to believe that an \( NP \)-hardness reduction can be obtained with higher dimensional analogues of the gadgets in our construction. However, it does not follow from our construction immediately, therefore we pose this case as a question.

**Question 10** What is the complexity status of \( (d, k) \)-Collapsibility for \( d \geq k + 2 \geq 2 \)?

**Collapsing to a fixed 1-complex.** In fact, our construction also shows that it is \( NP \)-complete to decide whether a 3-dimensional complex collapses to a fixed 1-complex. For this, it is sufficient to attach the fixed 1-complex to \( v \) and \((and eventually a \( d \)-simplex for \( d \geq 3 \) again if we want to reach higher dimension).

**Collapsing of complexes from a specific class.** In general we can consider two collections of simplicial complexes, the initial collection \( I \) and the target collection \( T \). The \((I, T)\)-Collapsibility question asks whether the given input complex from \( I \) collapses to some complex from \( T \). It would be interesting to know whether this question is polynomial time solvable for some natural choices of \( T \) and \( I \). One natural choice, in the author’s opinion, is when \( I \) is a collection of triangulated \( d \)-balls for some \( d \geq 3 \) and \( T \) is simply a point. Note that even in this setting the question is non-trivial since there exist non-collapsible \( d \)-balls; see, e.g., [2, Corollary 4.25]. However, even in this case we suspect \( NP \)-hardness.

**Acknowledgments** I would like to thank Dominique Attali, Bruno Benedetti, Anders Björner, Alexander Engström, André Lieutier and Rémy Malgouyres for helpful answers to my questions and/or providing useful references. I also thank the anonymous referee for a list of very helpful comments. Supported by the Göran Gustafsson postdoctoral fellowship. Partially supported by the project CE-ITI (GA ČR P202/12/G061) of the Czech Science Foundation.

**Appendix: Unrecognizability of Contractible Complexes**

Here we briefly discuss the current state of art regarding the recognition of contractible complexes. We first focus on the case of complexes of dimension at least 5.

**Theorem 11** (Novikov) For every \( d \geq 5 \), it is algorithmically undecidable whether a given simplicial complex of dimension at most \( d \) is contractible.

The proof of Theorem 11 easily follows from [16, §10]. We sketch a proof here.

**Sketch of a proof of Theorem 11** Novikov [16, §10] shows the existence of efficiently constructible sequence \( M_j \) of \( d \)-manifolds such that \( M_j \) is a ball if and only if \( \pi(M_j) \) is trivial and it is algorithmically undecidable whether \( \pi(M_j) \) is trivial. In particular, \( M_j \) is contractible if and only if \( \pi(M_j) \) is trivial. In order to finish the proof, we need to know that \( M_j \) can be efficiently constructed as a simplicial complex. This can be indeed done by inspecting the proof in [16] with not too much effort. \( \square \)

An alternative proof can be obtained from a more complete exposition by Nabutovsky; see the appendix of [13]. This is done in an earlier version of this paper [15]. The proof there is in full detail. It is easier to get a triangulation, because the analogues
of $M_j$ are zero sets of some polynomials with rational coefficients. On the other hand, the overall proof is slightly more complicated since it is necessary to transform spheres into balls.

The dimension 5 in Theorem 11 can be dropped to 4, if we greedily collapse 5-dimensional simplices of $\Theta_i$ via some of their 4-dimensional faces. Note that we cannot get stuck on a 5-dimensional complex, since $\Theta_i$ is connected. (The author learnt this idea, in a different context, from Bruno Benedetti.)

On the other hand, the contractibility question for complexes of dimension at most 1 is trivially polynomial-time solvable since it is equivalent with recognition of trees (as graphs).

Regarding complexes of dimension at most 2 or 3, the decidability of the contractibility question is open in these two cases to the best knowledge of the author. In particular, in dimension 2, this question is equivalent to the triviality of finite balanced representations of groups; see the exercise above in [8, Sect I.1.4].

We also remark that it is well-known that the triviality of the fundamental group is algorithmically undecidable already for complexes of dimension 2; see, e.g., [6]. This is equivalent with contractibility of each loop in the complex. However, this question should not be confused with the contractibility of the complex. On the level of group presentations, the triviality of the fundamental group corresponds to the triviality of any finite presentations of groups, not necessarily balanced.

References